Biomedical Applications of Belief Networks

Daniel Jonathan Cunliffe

PhD ~ The University of Edinburgh ~ 1996

Biomedical Applications of Belief Networks

Daniel Jonathan Cunliffe

PhD ~ The University of Edinburgh ~ 1996

Abstract

Biomedicine is an area in which computers have long been expected to play a significant role. Although many of the early claims have proved unrealistic, computers are gradually becoming accepted in the biomedical, clinical and research environment. Within these application areas, expert systems appear to have met with the most resistance, especially when applied to image interpretation.

In order to improve the acceptance of computerised decision support systems it is necessary to provide the information needed to make rational judgements concerning the inferences the system has made. This entails an explanation of what inferences were made, how the inferences were made and how the results of the inference are to be interpreted. Furthermore there must be a consistent approach to the combining of information from low level computational processes through to high level expert analyses.

Until recently *ad hoc* formalisms were seen as the only tractable approach to reasoning under uncertainty. A review of some of these formalisms suggests that they are less than ideal for the purposes of decision making. Belief networks provide a tractable way of utilising probability theory as an inference formalism by combining the theoretical consistency of probability for inference and decision making, with the ability to use the knowledge of domain experts.

The potential of belief networks in biomedical applications has already been recognised and there has been substantial research into the use of belief networks for medical diagnosis and methods for handling large, interconnected networks. In this thesis the use of belief networks is extended to include detailed image model matching to show how, in principle, feature measurement can be undertaken in a fully probabilistic way. The belief networks employed are usually cyclic and have strong influences between adjacent nodes, so new techniques for probabilistic updating based on a model of the matching process have been developed.

An object-orientated inference shell called FLAPNet has been implemented and used to apply the belief network formalism to two application domains. The first application is model-based matching in fetal ultrasound images. The imaging modality and biological variation in the subject make model matching a highly uncertain process. A dynamic, deformable model, similar to active contour models, is used. A belief network combines constraints derived from local evidence in the image, with global constraints derived from trained models, to control the iterative refinement of an initial model cue.

In the second application a belief network is used for the incremental aggregation of evidence occurring during the classification of objects on a cervical smear slide as part of an automated pre-screening system. A belief network provides both an explicit domain model and a mechanism for the incremental aggregation of evidence, two attributes important in pre-screening systems.

Overall it is argued that belief networks combine the necessary quantitative features required of a decision support system with desirable qualitative features that will lead to improved acceptability of expert systems in the biomedical domain.

Declaration

This thesis describes my own research and I am solely responsible for the composition of this thesis.

Daniel Jonathan Cunliffe
Edinburgh
September 1, 1996

Acknowledgements

This research was undertaken at the MRC Human Genetics Unit, Edinburgh under the supervision of Dr Richard Baldock, and the Department of Artificial Intelligence under the supervision of Dr John Hallam.

Thanks are due to all my colleagues from the former Pattern Recognition and Automation Section at the Human Genetics Unit, especially Dr Ian Poole, Patricia Eason, Mark McKie and Derek Charleston for their help with the cytology data, Stuart Wyles, John Ireland and Elizabeth Guest. Sheila Mould, the Unit Librarian deserves special thanks too.

This research was partially funded by MRC Research Studentship G78/0019.

Notes

POPLOG is a trademark of the University of Sussex. The X Window System is a trademark of the Massachusetts Institute of Technology. UNIX is a registered trademark of AT&T. Sun and SPARC are trademarks of Sun Microsystems Incorporated. QMR is a registered trademark of the University of Pittsburgh.

Other trademarks may also exist on the products mentioned in this thesis. The omission of trademark information should not be taken as proof of the absence of trademark registration.

The examples used in this thesis should not be considered to necessarily represent true medical opinion, they are presented for illustrative purposes only.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this thesis we consider the application of computers in the biomedical domain. In particular we focus on issues concerning the use of expert systems technology in the area of computer assisted decision making. One of the most important issues is the ability to present the results of uncertain computerised inference in a way that allows a rational decision maker to make choices based on, and regarding the validity of, those inference results. We review some expert systems topics that are relevant to the presentation of inference results, particularly the nature of the inference itself. A number of *ad hoc* uncertain inference formalisms are examined in order to demonstrate some of the pitfalls of *ad hoc* approaches. As an alternative to these approaches the belief network formalism is outlined and discussed. This formalism is referred to by a variety of names in the technical literature including, Bayesian networks, probabilistic inference diagrams, causal networks, causal probabilistic networks, Bayesian belief networks and, in a more general sense, influence diagrams. Belief networks have the advantage of a statistically sound axiomatic foundation, namely probability theory. We argue that belief networks are capable of supporting many of the additional functions of expert systems, such as the generation of explanations, the selection of action plans, and so on. Belief networks are already being applied to a variety of biomedical applications, we present a gazetteer of some of these as an appendix.

Having presented a case for the use of belief networks and provided a review of a number of biomedical belief network systems, we apply belief networks to two tasks. The intention is to demonstrate the use of probability theory to drive inference in image processing, from the lowest levels of data manipulation, to the highest level of image interpretation, diagnosis. Diagnosis has been investigated by other researchers, so we have concentrated principally on low-level data assessment, but we also consider intermediate level diagnosis.

The first of these tasks is examined in the context of fetal ultrasound image interpretation, via a trainable model matching system that utilises both local image data and global shape constraints to refine an initial image cue. This problem was selected as an example of the type of low-level task undertaken in medical image interpretation that must be integrated into the inference system.

The second task uses the domain of cervical screening as a typical diagnostic problem. Generally there is a wide variety of different information sources that should be considered during diagnosis. A belief network provides a mechanism for the principled incremental aggregation of this information, whilst simultaneously possessing an easily understandable and explicit inference structure.

Both applications are implemented in FLAPNet, a general-purpose network-based inference shell. FLAPNet was developed by the author in order to investigate belief network applications and the use of other network based inference methods.

In Chapter 2 we provide an overview of the field loosely described as computing in medicine (CIM[1]). The potential of CIM is assessed and contrasted with the apparent lack of success of certain CIM applications, particularly those using expert systems technology. Some of the reasons for the relatively poor acceptance of such systems are discussed.

Chapter 3 examines more closely the expert systems field, with particular emphasis on mechanisms for handling uncertainty. Various expert systems themes are explored with a view to the requirements of CIM applications.

Chapter 4 briefly introduces a particular formalism for uncertain reasoning, namely belief networks. Trends in the development of belief networks are outlined, suggesting how the belief network approach is capable of supporting the facilities that will be required of future expert systems if such technology is to become common place in the biomedical domain.

[1]This acronym is also used to denote Computers In Manufacturing.

An overview of the two medical applications used is presented in Chapter 5, along with a brief analysis of the concept of the belief network solutions.

Chapter 6 describes the FLAPNet tool in some detail, illustrating the assumptions and principles underlying its design. The implementation of particular types of inference strategy is also discussed.

In Chapter 7 an application in the area of fetal ultrasound imaging is described in some detail with particular reference to the characteristics of ultrasound images that makes image processing so difficult.

The application in cervical screening is presented in Chapter 8.

Chapter 9 draws together the various threads that run through this thesis, analysing the failures of CIM and discussing whether or not belief networks have anything to offer above and beyond that offered by more traditional expert system technologies.

The gazetteer of biomedical applications of belief networks is presented in Appendix A. This is intended to illustrate the breadth of potential applications as well as indicating the amount of research being conducted in this field. A more comprehensive list of references to work on belief networks and uncertain inference in general, can be found in the bibliography which includes many not directly cited in the text.

Chapter 2

Computers in Medicine

People's expectations of the role of computers in medicine (CIM) range from mundane clerical tools, such as patient databases, through to fully autonomous robot surgeons. Thirty years of piecemeal research has resulted in a variety of successes and failures. Computers have definitely established themselves in the medical domain, and few can doubt that they will become increasingly important as the technology develops.

In this chapter we examine the need for CIM, its potential, and reasons for its variable success. Some factors that could lead to improved acceptability in the future are identified.

2.1 Potential

Advocates of CIM can see a role for computers in almost every aspect of medical practice, whilst even the cynics acknowledge some role. Possibly because of this, there has been little formal justification for the continuing research into CIM. Informally, many different justifications are possible:

- To improve the accuracy, consistency and reproducibility of clinical diagnoses through the perceived superiority of computerised decision making methods over human decision makers [Shortliffe *et al* 84, Hand 87].
- To improve the cost effectiveness of tests and therapy plans [Shortliffe *et al* 84].
- To provide tools that enable physicians to cope with the complexity of modern medicine and the rapid rate of change [Anderson & Kettel 82, Hand 87].

- To alleviate the scarcity of able, qualified physicians and specialists [Schwartz 70, Hand 87] and reduce local variation of health standards due to increased availability of electronic medical information [Hafner *et al* 89, Laxminarayan & Kristol 92].

- To increase our understanding of medical decision making in order to improve it and to improve the design of medical computer systems [Shortliffe *et al* 84].

- To improve medical record keeping and enable more effective and novel methods for generating and using the information contained therein.

- To provide new types of medical information (perhaps previously considered impractical or impossible) and to identify gaps in current knowledge [Shortliffe *et al* 84, Holmes 94].

- To provide more effective medical education through simulation technology, such as virtual reality, and other teaching aids.

This variety of interrelated and overlapping aims has led to a fragmentation of the CIM field, with applications addressing different needs and different opportunities within the medical domain. These aims are linked by one common purpose, to improve the clinical care a patient receives, whilst efficiently using the resources available.

2.2 Application Types

The success of CIM applications is typically dependent on the role the computer is intended to fulfil and the degree to which the underlying computer technology is perceived to be proven technology.

A popular application is the provision of on-line literature databases such as MEDLINE. The vast increase in medical knowledge has been identified as one of the two dominant characteristics of current medical practice [Barnett 90]. Writing in 1990, Barnett says,

> If the most conscientious physician were to attempt to keep up with the [biomedical] literature [published each year] by reading two articles per day, in 1 year even this individual would be more than 800 years behind.

The importance of a physician keeping abreast of the literature should not be underestimated, either from the patients' viewpoint or the physicians'. An assessment of the liability issues of CIM [Hafner *et al* 89], suggests that failure to properly conduct a literature search, even if such technology is not widespread, may be considered negligence and may constitute malpractice.

Hospital information systems have become an almost indispensable management tool for many modern hospitals [Hasman 87, Dasta *et al* 92]. These systems are used for a variety of administrative and management tasks, such as patient admissions, accounting and stock control.

Pharmacological systems are used to check drug dosages and to test for adverse drug interactions [Engelbrecht *et al* 87].

Some authors have suggested that image processing is probably the most successful application of computers in medical care [Batson 84]. Computers have not only been used to enhance existing imaging modalities, but they have also led to the development of new modalities, giving the physician access to information about the patient that was only previously available using invasive methods, if at all.

Patient monitoring systems have also been introduced successfully into medical care. Such systems are used to provide a constant analysis of certain patient data, possibly triggering alarms when the data deviate from predicted norms [Osborn 82].

Computers have the potential to enhance medical education in a number of ways [Hand 87, Henry 90]. Unlike many educational media, the computer is interactive, which makes it an ideal tool for simulation. As simulation technology becomes more advanced, with the development of virtual reality tools, so the level of simulation will increase, allowing, for instance, virtual surgery. Even with current technologies, such as hypermedia, it is possible to combine disparate teaching resources into a single interactive package. Other benefits include the accessibility of information through bulletin boards, newsgroups, the World Wide Web, and so on.

Probably the least successful and most controversial CIM application is that of decision support systems and expert systems [Kulikowski 84, Shortliffe *et al* 84]. These systems are often considered to be a distinct research area, usually termed artificial intelligence in medicine (AIM[1]). AIM systems are designed to emulate certain diagnostic and decision making tasks usually carried out by the physician. Such systems could contribute greatly to fulfilling some of the aims of CIM, but despite decades of research they have probably made the least medical impact of all CIM applications.

2.3 Failure of AIM

Many reasons have been suggested for the limited success of AIM systems. Early AIM systems in the 1960s were held back by technical problems with the hardware and poor project management, resulting in inadequate systems [Young 84]. The field has now matured, hardware is more reliable, more powerful and far less costly, yet AIM systems are still failing to make an impact [Heathfield & Wyatt 93, Marquardt, Jr 93].

As several authors have noted [Teach & Shortliffe 81, Shortliffe *et al* 84, Young 84], the principal obstacle to improved acceptance of AIM systems is the continued failure to focus on the needs of the physician. Surveys of physician attitudes towards computers [Teach & Shortliffe 81, Knapp *et al* 87, Al-Hajjaj & Bamgboye 92] report overall positive attitudes to the introduction of computers into the medical field, patients also appear to be generally in favour [Cruickshank 84]. Whilst a number of physicians and students in surveys [Jones *et al* 91, Al-Hajjaj & Bamgboye 92] have indicated a lack of familiarity with computers, it is unlikely that this is a major factor in the slow introduction of AIM systems.

It would appear that whilst AIM is providing interesting research problems for the AI community, there has been relatively little emphasis on providing usable tools for the medical community. As one article [Cooper & Musen 90] puts it, "Although it is clear the AIM community has contributed substantively to AI, the contributions to medicine are far less palpable." The situation is schematically illustrated in figure 2.1, where medical acceptance and human control are shown to increase as the system moves from knowledge based reasoning to data processing, as the AI research interest diminishes.

The true picture is, of course, more complex than this and it is likely that as more knowledge based reasoning systems are adopted in non-medical domains so they will gradually become more accepted within the medical domain. There are several ways in

[1]In certain contexts this acronym is used for Advanced Informatics in Medicine.

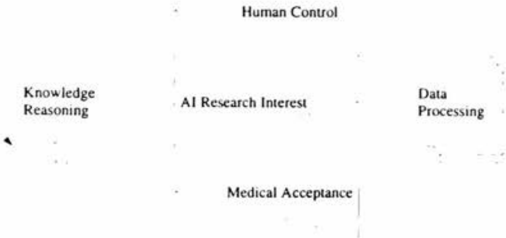


Figure 2.1: Reasoning with knowledge versus data processing

which this process of assimilation could be assisted.

2.3.1 Perspective

One of the problems with AIM systems is that they are designed and 'marketed' in the wrong way. The evidence suggests that physicians are well disposed towards systems that they perceive of as clinical tools or assistants [Teach & Shortliffe 81, Young 84]. On the other hand they are, perhaps understandably, not well disposed towards systems claiming to be autonomous physicians, what Suermondt and Cooper [Suermondt & Cooper 93, page 242] describe as "the doc-in-a-box". From an AI perspective this distinction may seem unimportant or even false. This may be partially due to the different viewpoint AI researchers and physicians take of the patient/physician relationship. It has been suggested [Collste 92] that there are two ways to view this relationship:

1. The patient visits the physician with a complaint, the physician assesses the symptoms, forms a diagnosis and determines a therapy. This is an asymmetrical relationship with the physician assuming the role of repairing a broken object.
2. In addition to the assessment of symptoms, diagnosis and therapy, the medical consulation is an encounter between two human beings, who attempt to arrive at mutual understanding through full communication, including a shared set of values, empathy, and an understanding of the patient as a whole.

Much of the AIM research seems to have followed the first model of the relationship, viewing medical diagnosis are merely a subset of a wider field of diagnostic tasks. This is perfectly adequate for research purposes, but is liable to prove inappropriate for real-world systems.

Under the second model the physician's humanity is recognised as an important factor in the role they play. It suggests that there are skills beyond those strictly defined as medical expertise, which form part of the complete diagnostic process. Some people might use this observation to prove that the concept of useful diagnostic computers is fatally flawed, a more profitable approach being to define more clearly the relationship between the computer and physician. Rather than placing the computer in the role of physician, the computer is assigned a role more akin to that of a fellow consultant. It offers advice and opinions, with explanations, leaving the physician free to make their own decisions about the management of their patient. The central role of the physician is recognised and preserved, and the physician/patient relationship remains undisturbed.

2.3.2 Communication

If an AIM system is to act in a similar way to a human consultant then its means of communication must be familiar to the physician. This is not to suggest that the medium need be the same (typically it will not), rather that the form of communication is appropriate. In order to achieve this the computer must be able to adapt its communication to the clinical situation and the needs and knowledge of the physician [Shortliffe & Clancey 84, Preece 90]. It must be able to provide appropriate justification of its decisions, constructing a diagnostic argument rather than simply stating its diagnosis. In some situations the ability to produce a reasoned argument will be as important as arriving at an appropriate diagnosis. This argument should be based on reasoning from first principles only when appropriate. A simple restatement of the diagnostic steps will not usually be sufficient as the reasoning mechanism will often be unfamiliar to the physician [Young 84]. Interactive dialogue should be supported whenever possible, allowing the physician to suggest alternative diagnoses, counter-supportive evidence, and so on with the computer arguing for or revising its initial diagnosis. The system could also be expanded to suggest medication regimes and appropriate test programmes, again with the ability to provide interactive justifications.

Ideally the knowledge base will also be accessible to the physician in order that they can examine the underlying clinical model the system is using. This requires that the knowledge base either is easily understood by a non-technical user or an appropriate interface is provided. The knowledge base should be easily maintained, possibly including a facility to learn, either from experience or from tuition. In conjunction with this, knowledge acquisition should be improved, and checks for consistency and completeness included [Shortliffe & Clancey 84].

2.3.3 Assimilation

Physicians usually practice according to a set clinical routine. Any system that is intended to be incorporated into such a routine must be designed with that routine in mind. If a system requires a large deviation from the accepted routine or increases the time a physician spends on a task without a significant and demonstrable benefit, then the system is unlikely to be used [Teach & Shortliffe 81, Preece 90].

2.3.4 Evaluation

It is important not only to be able to demonstrate to a physician that a systems works, but also to quantify its accuracy and, if possible, define the limits of its expertise. Part of this process will be to measure the accuracy of the physicians themselves. In many cases this will in itself be problematic. It will also be useful to evaluate the success of the system in a real-world environment for the purposes of improving system design.

There are several different aspects in the evaluation of an AIM system [Nykanen *et al* 91], those presented here are largely from the reviews of Miller [Miller 86] and Shortliffe and Clancey [Shortliffe & Clancey 84], a more AI orientated perspective is presented by Cohen and Howe [Cohen & Howe 88]. The terms of reference for the evaluation of an AIM system are:

- Is the knowledge-base sufficiently complete and consistent for operational use? In general, the more restricted the problem domain is, the easier it is to define, but proving completeness and consistency is difficult for any reasonably sized knowledge base. Does the system perform reasonably at the limits of its expertise? There is a well-known phenomenon called the *plateau effect*, which describes the way in which systems tend to work consistently whilst on their knowledge-base plateau, but suddenly fail when they reach the edge of their knowledge. The behaviour of the system at the edge of the plateau is therefore important.

- Is the performance acceptable, what is the 'gold standard' against which it should be measured? In many cases there has been little study to determine the accuracy of physicians, making comparisons even more difficult. There is also a 'superhuman' bias that says that a system must outperform the physician before it is considered to be of use. What should count as a correct response? If a system produces, for instance, a list of ranked alternative diagnoses, is membership of that list sufficient, or must the correct diagnosis be ranked highest? In some cases the only measure of correctness is agreement with the physician, but if we wish to create a system that outperforms the physician, then at some point in the evaluation process it will be necessary to accept the system diagnosis in preference to the physician diagnosis and monitor the condition of the patient. Clearly this kind of experimentation raises important ethical issues.

- Is there a demonstrable need for the system and is the system usable in a clinical environment, with respect to response times, disruption of clinical routine and so on? Do physicians alter their behaviour based on a system's advice and do they continue to use and maintain the system once it has been introduced?

- Is cost/benefit analysis possible? If so what quantities should be measured? How should qualitative factors be measured?

2.3.5 Ethics and Liability

Ethical considerations have always been important within the medical domain. The introduction of computers has raised new ethical problems for physicians as well as focusing attention on some traditional ones, for example confidentiality [Sieghart 84, de Dombal 87].

As computer systems gradually assume a greater decision taking role, so it becomes more difficult to establish the locus of responsibility within the complete clinical process [Hollnagel 90]. In order for the physician to assume responsibility the physician's autonomy must be maintained and the physician must be in possession of enough knowledge to foresee the consequences of their action [Collste 92]. This requires that the physician is able to make informed judgements about the advice offered by an AIM system so that they may accept or reject that advice. To facilitate such judgements the system must be able to communicate, its limitations and effectiveness must be quantified. In addition to this the legal implications of a physician either not consulting, or ignoring the advice of a decision support system must be clarified [de Dombal 87].

Similarly the liability issues of a system that misdiagnoses or that advocates an inappropriate treatment require investigation. It is important to determine whether an AIM system constitutes a tool or a qualified colleague, as physicians are generally protected from liability in the case of a negligent colleague, but not in the case of a defective tool or negligent subordinate [Hafner *et al* 89]. In cases where the system makes an inappropriate diagnosis and is held responsible, assigning liability will still prove problematical as the final system will typically be the product of several different people each providing expertise in a different field. The very definition of an inappropriate response from a system that is not claiming 100% accuracy is also problematic.

Until the ethical and legal implications of AIM systems are fully understood it is unlikely that physicians will be enthusiastic about using them and AI researchers may be uncomfortable about placing products in the real world.

2.4 Summary

Computers are already playing a significant role in medical care. The extent of this role in the future critically depends on the recognition of the importance of physician acceptance. In the biomedical domain, probably more than any other, it is not sufficient to ask the user to accept a system either on trust or on solely on the basis of its output. The concerns that physicians have can only be allayed by presenting them with systems in which the process by which a result is achieved can be explained as clearly as the result itself.

AIM systems must consider the physician at all stages of the design, from the knowledge-base, through the reasoning mechanism and interface, to deployment in a clinical environment. As part of this process, additional psychological studies should be conducted, both into the diagnostic and communicative behaviour of physicians

[Shortliffe & Clancey 84], and effective mechanisms for human/computer interaction. If AIM systems fully embrace the concept of physician orientated support systems and the design challenges raised by such systems, then we can expect an increased acceptance on the part of physicians.

Chapter 3

Expert Systems

The archetypal medical task is diagnosis, the determination of the disease state of the patient, given a set of symptoms, background medical and patient knowledge, and a set of tests that can be performed. The purpose of diagnosis is to facilitate therapy planning. Therapy planning is based on a patient diagnosis, knowledge of the available therapies and their effects, background medical and patient knowledge, and a set of therapeutic goals. Typically both diagnosis and therapy planning are resource bounded.

There are a number of questions a decision maker may want to ask of a diagnostic result before taking any action based on that result:

- What diagnoses are possible?
- What are the relative probabilities of those diagnoses?
- How has the available evidence influenced the diagnoses?
- How has the background knowledge influenced the diagnoses?
- How could the expenditure of resources influence the diagnoses?

Therapy planning adds a further level of questions:

- What therapies are possible?
- What are the relative utilities of the therapies?
- How has the available evidence influenced the therapy plan?

- How has the background knowledge influenced the therapy plan?

- How have the diagnoses influenced the therapy plan?

- How could the expenditure of resources influence the therapy plan?

We have suggested that the ability of a medical expert system to answer these types of questions will have an important role to play in improving acceptance. Answers to questions such as these will, almost without exception, involve some degree of uncertainty. Part of the task of understanding these answers lies in being able to interpret the terms used to quantify uncertainty. Expert systems have used a variety of different terms to quantify uncertainty and a corresponding variety of mechanisms for drawing inferences from uncertain information. In this chapter we examine both the potential sources of uncertainty and several mechanisms for handling it.

We will also examine the problems of control and resource allocation which depend upon being able to answer questions similar to those above.

The final area we will examine is that of interface design and communication. We have already stressed that the central task for many medical expert systems will be to explain and justify the decisions it has reached. Indeed the entire concept of the medical decision support system rests on the assumption that the physician will act on, or at least consider, the recommendations of the system.

3.1 Sources of Uncertainty

It has been recognised that in almost any complex domain there will be some element of uncertainty. Areas which can give rise to uncertainty are the real world, the experts' testimony regarding the domain, the instrumentation used, and the design of the expert system itself. Each of these areas must be examined to determine the possible effects of the uncertainty and the most appropriate ways of representing and handling it.

3.1.1 Domain Experts

Whilst human expertise is often the primary or sole knowledge source consulted when creating the knowledge base, it is important to realise that it potentially contains inaccuracies, errors and omissions. This is particularly true in cases where only a single expert source is consulted. It has been suggested [Cheeseman 84] that where possible an expert system should be based on empirical data rather than expert testimony. In many domains this is not possible or is simply impractical, though expert systems could be designed to modify initial expert knowledge on the basis of experience in the domain.

Perhaps the most obvious source of uncertainty in an expert testimony is the estimation of probabilities. There are many parts to the problem, including the assignment of a precise numerical probability, the use of non-numerical probability terms, the violation of statistical techniques, and inter-expert variation.

It appears that the assessment of probabilities, even by trained people, is often prone to severe and systematic errors. In an investigation by Tversky and Kahneman into the mechanisms used by humans, a number of heuristic techniques and biases were identified [Tversky & Kahneman 90a]. These fall into three main categories, *representativeness*, *availability*, and *adjustment and anchoring*.

Representativeness occurs when people are asked to answer questions of the type, what is the probability that object A belongs to class B, what is the probability that event A originates from process B, what is the probability that process B will generate event A, and so on. According to Tversky and Kahneman, people employ a *representativeness heuristic* by which the probability is assessed on the degree of similarity between A and B. The effects of using this heuristic include insensitivity to prior probabilities, insensitivity to sample size, misconceptions of chance (people expect long term probabilities to be manifested in small samples), insensitivity to predictability, illusions of validity and misconceptions about regression.

The *availability heuristic* refers to situations in which people assess the probability of an event or the frequency of a class based on the ease with which an instance can be brought to mind. The effects of this include biases due to the retrievability of instances, biases due to the effectiveness of the search set, biases of imaginability, and illusionary correlation.

Adjustment and anchoring errors are observed when people are asked to make estimates by starting from an initial value and adjusting it to find the final answer. Errors include insufficient adjustment, biases in the evaluation of conjunctive and disjunctive events, and the incorrect construction of probability distributions.

When eliciting measures of probability from experts it is often the case that such measures are expressed in linguistic rather than numeric terms. There are two approaches to handling such linguistic terms in expert systems. The first is simply to convert the linguistic term into a numeric probability. The second is to design an expert system that uses linguistic measures. Both approaches are fraught with difficulties as there is great variability in the way linguistic uncertainty terms are used. Different people will use the same term to refer to different levels of probability and people also rank uncertainty terms differently [Buxton 89]. It is also not clear that a person uses linguistic terms in a consistent manner, hence if an expert says that A is "likely", and then in a different context says that B is "likely", it is not necessarily true that A and B are equally probable.

Fuzzy terms represent a special kind of uncertainty in which non-probabilistic terms convey concepts which are not well defined, *e.g.* small, soft, fast, and so on [Dutta 85]. Such terms are typically both relative and subjective, therefore, when attempting to define a fuzzy term, it is important to consider the context in which the term is being used.

Another source of linguistic uncertainty is *lexical imprecision* [Henkind 88]. All problem domains have a special-purpose vocabulary, called an *explananda* [Pylyshyn 86], which describes concepts within the domain. The explananda therefore is a valuable expression of the entities and situations that should be considered when modelling the domain. Lexical imprecision refers to the situation where a particular term in the explananda has multiple, though similar, definitions (hence it is different from a fuzzy term where the definitions themselves are imprecise). A good example of lexical imprecision, taken from the fetal ultrasound domain, is the definition of intrauterine growth retardation. As has been pointed out by Jeanty and Romero [Jeanty & Romero 83], "Some authors use the 3rd percentile, while others the 10th percentile, or 2 standard deviations below the mean, of the birth weight." In fact other definitions are also in use, for instance the fifth percentile [Beischer *et al* 84]. In view of the fact that intrauterine growth retardation is outranked only by prematurity and major malformations as a cause of perinatal death [Beischer *et al* 84], the variation in the definition could be seen as cause for concern.

Whilst the most satisfactory solution to the problem of lexical imprecision would be the standardisation of definitions within an explananda, it is unlikely to happen, though the development of expert systems could prove to be a catalyst. Techniques for handling lexical imprecision include the direct input of data rather than the expert's interpretation and the incorporation of definitions into the system which can then be used to guide the user.

3.1.2 Real World

Uncertainty in the real world comes principally from two sources, randomness and variation. Randomness refers to genuinely random events, *e.g.* the results of spinning a fair coin. Variation refers to natural distributions, for instance, where some process produces artifacts of a nominal length, or biological variation, *e.g.* in height. Both randomness and variation may be expressible in terms of a distribution function. In both these cases a true description of the randomness or variation (and hence the uncertainty) can only be obtained through repeated observations. In practice this data may not be available and it may be necessary to use a subjective estimate of the underlying frequencies which may in itself introduce further uncertainty. For instance it is common to assume that distributions approximate some well understood and easily modelled distribution, such as the normal distribution, the worse the approximation the greater the uncertainty.

3.1.3 Instrumentation

It has long been acknowledged that certain types of instrument are prone to introduce errors, known as noise, into signals [Gonzalez & Wintz 87]. In some cases the noise component of the signal is relatively low and can effectively be ignored, in other cases noise can introduce uncertainty into later processes. In the majority of situations the noise will be randomly distributed. In some it will be highly structured noise which is often hard to distinguish from genuine signals [Kremkau & Taylor 86]. Where noise cannot be dealt with by simple procedures it may be necessary to allow for noise at a higher level within the system.

Other sources of instrument uncertainty involve features of the equipment, such as the resolution of a camera. Such features cannot generally be changed so their impact should be assessed with reference to subsequent processing. Typically the system will simply be designed to work within the limits of the available instrumentation.

3.1.4 Expert System

When analysing sources of uncertainty it is important to consider the design of the expert system itself. Whilst some of the uncertainty may be due to features of the system, others may be due to the way in which the system handles uncertainty.

If the mapping between terms in the explananda and the expert system's internal representation is poor, uncertainty will result. A poor mapping may mean that entities are assigned to one class when they actually belong to another, unmodelled class. Also numeric values may be truncated, or represented to unrealistic accuracy, both of which can lead to results with inaccuracies.

The use of so called 'magic numbers' within the system may also introduce uncertainty. Such numbers are usually embedded within procedures rather than being explicitly represented within the system.

If the control element of the expert system involves some degree of choice, then the system will benefit from knowledge of its own evidence gathering capabilities. This would give the system the potential for reducing overall uncertainty by directed evidence gathering. For instance, a system may have a number of processes for finding low level image features, such as lines or arcs, from image data. If it is known that process A is susceptible to noise but process B isn't, and that the image data is noisy, it is possible for the system to keep uncertainty low by selecting process B. Similarly, where time is an important constraint the system may be able to trade off a fast process which produces limited or unreliable (uncertain) information, against more time consuming processes which produce higher quality results. This again requires that the system have knowledge of its own procedures and that some measure of uncertainty can be associated with the results of a process based on the appropriateness of applying that process to the data under interpretation. We will discuss this further in section 3.3.

There are several problems in handling uncertain information that will also bear on the uncertainty in an expert system. In the propagation and aggregation of measures of uncertainty, whether numeric or linguistic, the context is always liable to play an important role. Perhaps the principal role is in determining what interdependencies exist between the various pieces of evidence. It is, for instance, important to make sure that a particular piece of evidence is counted only once and that the inference is not cyclic. Other factors should also be taken into account when assessing support from a body of evidence, including the degree of compatibility between the evidence and the hypothesis, the amount of evidence, the variety of the evidence and what to do in cases where the evidence is incomplete [Buxton 89]. It is therefore important to analyse the structure of inference within a domain and attempt to represent this structure within the expert system.

3.2 Formalisms

Many formalisms for modelling uncertainty have been developed. It has been suggested that the choice of formalism should be mediated by several factors in addition to the merits of the formalism itself. Among these factors are the difficulty of acquiring the initial uncertainty estimates, the computational complexity of inference, the semantics which guide the acquisition of the original estimates and guide the interpretation of computed results, and how the chosen representation is used in decision making [Lemmer & Kanal 88]. Indeed, it has been argued that the problem domain itself will, at least to some degree, influence the choice of formalism and problem solving strategy that should be used [Chandrasekaran & Tanner 86].

It has been shown, for example by Cox [Cox 61, Cox 90], that from a small set of intuitive properties a measure of belief should possess, a set of axioms can be defined. This set of axioms provides a normative basis for theories of uncertain reasoning. These axioms are precisely those of probability theory, which has led some authors to express strong opinions about probability theory [Lindley 87, page 17],

> Our thesis is simply stated: *the only satisfactory description of uncertainty is probability*. By this is meant that every uncertainty statement must be in the form of a probability; that several uncertainties must be combined using the rules of probability; and that the calculus of probabilities is adequate to handle *all* situations involving uncertainty. In particular, alternative descriptions of uncertainty are unnecessary ... We speak of "the inevitability of

probability."

Given this *inevitability of probability* it is perhaps surprising that so many alternative formalisms have been proposed. There appear to be three main motivations for the development of alternative formalisms, firstly the belief that complete probabilistic models are infeasible and inference on such models is intractable, secondly that there are particular types of uncertainty that are either poorly represented by probabilities or cannot be represented by probabilities at all, thirdly that probabilities are not cognitively valid. Good overviews of the objections to probability theory with counter arguments are presented by Cheeseman [Cheeseman 85] and Henrion [Henrion 87].

In the next chapter we will discuss a particular formalism, belief networks, which answer some, if not all of the objections to probability theory, and offer a tractable mechanism for normative probabilistic inference in complex domains. In the remainder of this section we give a brief overview of several alternative formalisms that have been proposed.

3.2.1 Certainty Factors

MYCIN was one of the ground-breaking expert systems, developed in the 1970s to assist in the diagnosis of bacteremia, and later applied to meningitis [Shortliffe & Fagan 82, Shortliffe & Buchanan 90]. A domain independent expert system shell EMYCIN (Essential MYCIN), based on MYCIN, was used to develop several medical applications, including PUFF [Aikins *et al* 84], an expert system for interpreting pulmonary (lung) function data.

MYCIN is a goal-driven, rule-based system which uses certainty factor (CF) calculus to measure support for hypotheses. The CF approach is interesting as it was specifically developed within the context of expert systems, rather than being introduced from another field. It was also designed to reflect the way in which physicians in the domain tended to approach the diagnostic task.

The general format of a MYCIN rule is:

```
IF <premise assertions are true>
THEN <consequent assertions are true> <with confidence weight W>
```

The assertions can be Boolean combinations of clauses, each of which consists of a predicate statement about an <attribute - object - value> triple. The triple represents medical facts and hypotheses about the patient and related objects or contexts, such as infections, cultures or organisms. For example, a paraphrased rule (after [Shortliffe & Buchanan 90, page 261]):

```
IF (1) the stain of the organism is gram positive,
AND (2) the morphology of the organism is coccus,
AND (3) the growth conformation of the organism is chains
THEN there is suggestive evidence (0.7) that the identity
    of the organism is streptococcus
```

The quantification of the suggestive evidence is a domain expert rating of the confidence in the rule (normalised to lie in the range [0, 1]). These values cannot be strictly treated as probabilities as they do not satisfy the additivity axiom. This reflects the physician's domain model in which evidence that only partially confirms an hypothesis is not also considered to partially disconfirm that hypothesis[1]. In the rule above, the complementary inference not-streptococcus is not confirmed to degree 0.3 by the negation of the antecedents of the rule. It was observed that the physicians tended to gather confirming evidence and disconfirming evidence independently. The formalism reflects this by maintaining a separate measure of belief (MB) and disbelief (MD). MYCIN treats both rules and data as uncertain, so both rules and hypotheses have either an MB or MD associated with it, depending on whether the truth of the antecedents provided confirmation or disconfirmation of the consequents.

Given $P(h)$ is the prior probability of hypothesis h, and $P(h \mid e)$ is the posterior probability given evidence e, MD, MB and CF are defined by the following relationships [Clark 90]:

[1] See [Cheeseman 85, page 1006] for an interesting perspective on this.

If e favours h then MB increases, $MB = (P(h \mid e) - P(h))/1 - P(h)$

If e counts against h then MD increases, $MD = (P(h) - P(h \mid e))/P(h)$

The CF of an hypothesis determined by, $CF = MB - MD$

In EMYCIN the CF definition was changed in order to prevent a single strong piece of disconfirmatory evidence outweighing the impact of several weaker pieces of confirmatory evidence. The EMYCIN CF definition is:

$CF = (MB - MD)/(1 - min(MB, MD))$

Thus the value of a CF lies in the range [-1, 1], with -1 representing certain falsehood and 1 representing certain truth. The two definitions of CF differ only when both MB and MD are non-zero.

Uncertainty in the input data, as indicated by the physician, is combined with the uncertain rules by use of three functions [Clark 90]:

1. Determine a pooled CF for the set of antecedents of a rule. For a conjunctive set of premises the pooled MB is the individual minimum MB and the pooled MD is the individual maximum MD. For a disjunctive set the pooled MB is the maximum of the MBs and the pooled MD is the minimum of the MDs.

2. Combine the pooled CF of a set on antecedents with the MB or MD of a rule to propagate an MB or MD to the consequents of a rule when it fires. When a rule is fired the MBs or MDs of the consequents are the product of the MDs or MBs of the rule and the pooled CF of the premises.

3. Pool evidence from different rules to produce an overall CF for each proposition. New evidence is pooled with existing evidence in proportion to the outstanding uncertainty:

   if (both $X, Y > 0$) then combined CF(X,Y) $=X + Y(1 - X)$

   if (either $X, Y < 0$) then combined CF(X,Y) $= (X + Y)/(1 - min(\mid X \mid, \mid Y \mid))$

   if (both $X, Y < 0$) then combined CF(X,Y) $= X + Y(1 + X)$

MYCIN uses a very simple control strategy [Kulikowski 84], goal-directed backwards chaining. This process starts with the rule containing the highest level goal — select treatments for all the infections of the patient. As the infections are usually unknown this will generate subgoals directed at the identification of the infections, which in turn will generate further subgoals. This process of goal refinement eventually results in an assertion that can only be confirmed by information from the user. The system can then begin to work backwards satisfying subgoals. A hierarchical context tree (patient - infections - cultures - organisms) is used to constrain the order in which rules are invoked. The final goal, selection of a therapy, is carried out by a specialised algorithm using the deduced knowledge of the patient's infections and the causative organisms, and the ranking of drugs by sensitivity and effectiveness.

The CF formalism has several useful attributes:

- [Clark 90] The separation of MD and MB makes it possible to distinguish between situations of ignorance MB = MD = 0 and equivocation MB = MD $\neq$ 0, though neither MYCIN nor EMYCIN have made use of this.
- [Clark 90] The CF formalism appears to provide a method for both formalising heuristic reasoning as rules whilst simultaneously allowing uncertainty to be quantified and combined with a formal calculus.
- MYCIN's formalism was developed on the basis of observed physician behaviour, and can therefore be considered to be based, at least in part, on natural reasoning strategies.

On the other hand, and in spite of its good diagnostic performance, MYCIN has been criticised on a number of important points:

- [Bonissone 87, Clark 90] The CF formalism attempts to provide quantification based on relative changes in belief rather than absolute probability.
- [Heckerman & Horvitz 88, Clark 90] The CF formalism has syntactic modularity but lacks semantic modularity. This implies that the strength of association between the antecedents and consequents in a non-categorical rule (MD or MB $\neq$ 0) will change when other rules are added to or deleted from the knowledge base.
- [Clark 90] The MDs and MBs attached to rules were used to represent utility considerations as well as probabilities. This was done by assigning higher CFs to

rules with serious consequences. This is a criticism of the application rather than the formalism itself.

- [Buxton 89] The use of MB and MD suggests that the quantities $P(h)$ and $P(h \mid e)$ contained in their definitions cannot be ordinary probabilities. If this is correct then it is not clear how particular numerical values should be interpreted. On the other hand if they are treated as probabilities, then MYCIN appears to be equivalent (or similar [Henkind & Harrison 88]) to a system of Bayesian updating with highly restrictive independence assumptions.

- [Henrion 87, Buxton 89, Langlotz 89] MYCIN does not explicitly represent prior probabilities, effectively assuming that the priors are equal. Any background information is presumed to be represented by MBs and MDs. In the application domain this is not entirely unreasonable as the organisms considered all had small, similar prior probabilities. Any errors could be corrected by adjusting the number and character of rules that concluded a particular organism or by including some prevalence information in the CFs associated with the rules.

- [Buxton 89] Any interrelationships or dependencies among hypotheses seem to be ignored both in the assessment of MB and MD and also in the rules for calculating the support for a conjunction or disjunction.

- [Buxton 89] MYCIN fails to take into proper account the interrelationships among evidence, which can lead to unrealistic assessments of the support provided by a new piece of evidence.

- [Bonissone 87, Buxton 89] The CFs calculated do not appear to have a simple interpretation. In particular they cannot be interpreted in the same way as the measure of belief and disbelief that are assessed by the experts.

- [Henkind & Harrison 88] The nature of the combining functions causes the MBs and MDs to converge quickly to 1, whilst the CFs stay near 0. Therefore the CF calculus is not suited to situations where there are large quantities of evidence to combine.

- [Shortliffe & Fagan 82] MYCIN-like rules need substantial modification in order to analyse temporal trends and rapid parameter changes.

- [Saffiotti 87] The hypothesis with the highest CF is not guaranteed to be the most probable one.

- [Saffiotti 87, Heckerman 90a] The unexplained insensitivity of MYCIN to a change in the CFs of its rules. This may in part be explained by the small number of rules that are typically applied to reach a conclusion.

3.2.2 Dempster-Shafer Theory

Dempster-Shafer (DS) theory, or the theory of belief functions, was developed in the 1960s by A. P. Dempster and extended in the 1970s by G. Shafer [Gordon & Shortliffe 85, Shafer & Srivastava 90]. Unlike MYCIN, DS theory is known more for the formalism than its applications, which include, by way of examples, GERTIS, a prototype system for diagnosing rheumatoid arthritis [Yen 89], and a system for knowledge-based computer vision [Wesley 86].

Two of DS theory's more distinctive features are the use of an interval representation of belief as opposed to point values, and the ability to assign belief to sets of hypotheses rather than an individual hypothesis.

The basics of DS theory are as follows [Clark 90, Lowrance *et al* 90]:

1. A set of mutually exclusive and exhaustive base elements forms the *frame of discernment*, $\{\Theta\}$. The impact of evidence is defined over the power set (set of all subsets) 2^{Θ}.

2. A mass probability function assigns a value [0, 1] to every disjunctive subset of hypotheses, so that the sum (or total probability mass) is 1 and the probability assigned to the empty set is 0, *i.e.* $m(\emptyset) = 0$. $m(\Theta)$ will be the probability mass that cannot be committed to any smaller subset of Θ, it represents ignorance.

3. The certainty of a particular proposition or hypothesis A is represented by the evidential interval $[Spt(A),\ Pls(A)]$ where (for subsets B of A)

   $Spt(A) = \sum_{B \subseteq A} m(B)$

$Pls(A) = 1 - Spt(\neg A)$

Spt is a measure of the belief in A given the evidence, while Pls represents the degree to which the evidence fails to refute A. The difference between the two values is a measure of ignorance, the belief that is committed to neither A nor $\neg A$. Some interpretations of interval values are offered by Wesley [Wesley 86]:

- Completely true proposition — [1, 1]
- Completely false proposition — [0, 0]
- Completely ignorant about proposition — [0, 1]
- Tends to support proposition — [Spt, 1] ($0 < Spt < 1$)
- Tends to refute the proposition [0, Pls] ($0 < Pls < 1$)
- Tends to support *and* refute the proposition [Spt, Pls] ($0 < Spt \leq Pls < 1$)

4. Evidence is combined using Dempster's rule of combination:

   $m(C) = \frac{\sum_{A \cap B = C} m_1(A)m_2(B)}{1 - \sum_{A \cap B = \emptyset} m_1(A)m_2(B)}$

   for two pieces of evidence 1 and 2 where A represents hypothesis subsets that are supported by 1 and B represents hypothesis subsets supported by 2. The new belief in subset C that is supported by both 1 and 2, is defined as the sum of the products of the masses assigned to subsets A and B whose intersection is C, divided by a normalisation factor equal to 1 minus the sum of the products of belief masses of subsets A and B whose intersection is the empty set $\{\emptyset\}$.

   Probability mass is assigned to the empty set whenever 1 and 2 assign mass to two disjoint sets. This would violate one of the axioms, so a normalisation factor is used to redistribute the unassigned mass.

The use of DS theory is strongly advocated by many researchers. Among its positive points are:

- [Gordon & Shortliffe 85] The ability to model the narrowing of the hypothesis set with accumulation of evidence, recognised as being a natural reasoning strategy characterising diagnostic reasoning in medicine and expert reasoning in general.

- [Bonissone 87, Stephanou & Sage 87, Clark 90] The representation of ignorance through the use of belief intervals rather than point values, the assignment of non-zero values to the base element set and the use of subsets.

- [Saffiotti 87] The ability to specify domains of discourse to suit the model information available.

There are, however, also problems associated with it:

- [Bonissone 87, Henkind & Harrison 88, Clark 90] One of the major criticisms is the computational complexity involved when large sets of hypotheses are considered, though the introduction of some restrictions can reduce this.

- [Clark 90] When a large probability mass would be assigned to the empty set, the normalisation procedure used to redistribute this unassigned mass can produce counter-intuitive results.

- [Henkind & Harrison 88] DS theory assumes independence of evidence.

- [Gordon & Shortliffe 85, Henkind & Harrison 88] The combination rule has no theoretical justification, it is based on intuitions about the pooling of evidence.

- [Thompson 85, Saffiotti 87] The design of the structure of domain of discourse is non-trivial as it must ensure inclusion of subsets that can serve as recipients of mass from each and every report that may be received during processing. The translation of a piece of information into a mass distribution over the domain of discourse may be a burden both from computational and design points of view.

- [Thompson 85, Wesley 86, Saffiotti 87, Clark 90] DS theory lacks a decision mechanism and the design of decision mechanisms for intervals is still a research problem.

- [Saffiotti 87, Black & Laskey 90] DS theory only defines probability updates due to evidence acquisition; propagation through local constraints must be defined in the application.

3.2.3 Fuzzy Set Theory

The concept of a fuzzy set was first presented in 1965 by L. Zadeh [Zadeh 65]. It was originally developed in response to certain types of paradox that arise in classical set theory [Henkind & Harrison 88], consider:

1. A heap containing one stone is small.
2. If you add one stone to a small heap it remains small.
3. Therefore (by induction), every heap is small.

When it was initially proposed, fuzzy set (FS) theory was considered highly controversial in the field of mathematical systems theory which Zadeh worked. Gaines and Shaw [Gaines & Shaw 85], present an interesting historical and philosophical view of FS theory which includes the following quotation from a paper by Kalman, published in 1974:

> His [Zadeh's] proposals could be severely, ferociously, even brutally criticised from a technical point of view. ... No doubt Professor Zadeh's enthusiasm for fuzzy sets has been reinforced by the prevailing political climate in the U.S.: one of unprecedented permissiveness.

It is not clear, given that the quotation is taken out of context, whether Kalman intended this to be tongue-in-cheek or not. More recently FS theory has prompted a great deal of mathematical work exploring the properties of fuzzy sets, along with the application of FS theory to expert systems.

FS theory is closely related to natural language expressions of 'fuzzy' concepts [Zadeh 86], for example

- Fuzzy predicates: small, large, young, safe, much larger than, soon ...
- Fuzzy quantifiers: most, many, few ...
- Fuzzy probabilities: likely, unlikely, not very likely ...
- Fuzzy truth values: very true, quite true, mostly true ...

The basics of FS theory are as follows [Henkind & Harrison 88]: in classical set theory the proposition of set membership is categorical. Let X denote a universe of objects, and x denote an individual element from that universe. Let A be a subset of X. μ_A is the *characteristic function* or *membership function* of A if:

$$\mu_A(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{otherwise} \end{cases}$$

FS theory calls the set $\{0, 1\}$ the *valuation set* and the value assigned by the function for a given x the *degree of membership*. From the above definition, in classical theory the degree of membership is either 0 or 1 (it is categorical).

In fuzzy set theory the valuation set is expanded to the interval [0,1], so the degree of membership can range between 0 and 1. Intuitively the larger the value of $\mu_A(x)$ the more $x \in A$, *i.e.* the better x satisfies the definition of A.

For example: let X be the set of all people and T be the set of tall people. Choose x to be Tom, who is 7 feet tall, then $\mu_T(x) = 1$, intuitively Tom is tall. Now choose x to be Dick who is 6 feet tall, then $\mu_T(x) = 0.5$, intuitively Dick is 'somewhat tall'. Finally choose Harry who is 5 foot tall then $\mu_T(x) = 0$, Harry is not tall.

In FS theory it is necessary to provide a characteristic function (or definition) for each set of interest, so a function for 'tall' based on a person's height h might be:

$$\mu_T(h) = \begin{cases} 0, & h < 5 \\ \text{(h-5)/2}, & 5 \leq h \leq 7 \\ 1, & h > 7 \end{cases}$$

The values assigned by the characteristic function are chosen by the person who defines the function and are therefore subjective (Henkind who provides the above example is 6'4"). The definition of a characteristic function may also be context dependent, consider the definitions of 'tall' for basketball players and pygmies, for example. It has been suggested [Cayrol *et al* 80] that it is the general shape of the characteristic function that is important rather than its precise determination. Fuzzy membership functions for natural language are prototypically characterised as single peaked, or monotonic with a maximum of 1 [Clark 90], for example in figure 3.1, 'tall' (note the different characteristic function) has a monotonic function while 'middle aged' is single peaked.

The concept of a fuzzy set can be broadened to that of a linguistic variable [Zadeh 75a, Zadeh 75b, Zadeh 75c]. A linguistic variable, such as '*age*' is composed of a set of fuzzy

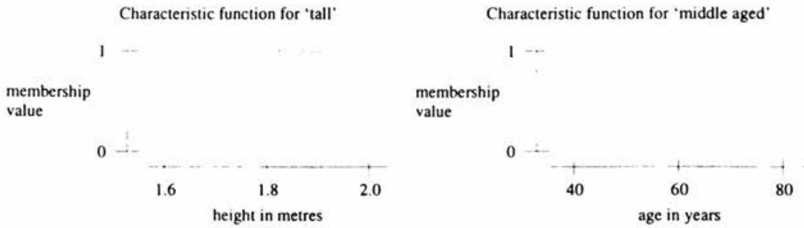


Figure 3.1: Monotonic and single peaked characteristic functions [Clark 90, page 125]

sets, such as {*'very young', 'young', ..., 'old', 'very old'*}, each of which is defined over the set of values of years lived since birth [Zimmermann 90].

Techniques for using FS theory in expert systems fall broadly into two types, fuzzy logics and possibility theory. The term 'fuzzy logic' has several interpretations in the literature, some treat it as meaning multivalued logics whilst others see it as a logic for manipulating fuzzy sets. Dubois and Prade [Dubois & Prade 90a] give an introduction to possibilistic and fuzzy logics. The fuzzy logic approach presented here as an example, is taken from Thompson [Thompson 85]:

Zadeh defines the axioms of fuzzy logic as:

- $0 \leq \mu(x) \leq 1$
- $\mu(\neg x) = 1 - \mu(x)$
- $\mu(x\ AND\ y) = min[\mu(x), \mu(y)]$
- $\mu(x\ OR\ y) = max[\mu(x), \mu(y)]$
- $\mu(x \Rightarrow y) = min[1, (1 - \mu(x) + \mu(y))]$
- $\mu(x \equiv y) = min[(1 - \mu(x) + \mu(y)), (1 + \mu(x) - \mu(y))]$

Suppose the following statements define the characteristic function of the fuzzy pred icate 'low reflectance' (LR):

$$\mu_{LR}(r) = \begin{cases} 1.0, & r = 0.0 \\ 0.6, & r = 1.0 \\ 0.1, & r = 2.0 \end{cases}$$

This can be represented as the fuzzy set, $LR = \{0 \mid 1.0,\ 1 \mid 0.6,\ 2 \mid 0.1\}$.

Suppose we wish to classify an object into one of n classes, $c_1, \ldots, c_n$, given some evidence, $E_1, \ldots, E_k$. Based on evidence E_1 we can develop characteristic membership functions $\mu_{11}, \ldots, \mu_{1n}$ to form a fuzzy set:

$$A_1 = \{c_1 \mid \mu_{11},\ c_2 \mid \mu_{12}, \ldots, c_n \mid \mu_{1n}\}$$

similarly for evidence E_2:

$$A_2 = \{c_1 \mid \mu_{21},\ c_2 \mid \mu_{22}, \ldots, c_n \mid \mu_{2n}\}$$

k sets of evidence can be combined to give:

$$B(k) = \{c_1 \mid \mu(k)_1,\ c_2 \mid \mu(k)_2, \ldots, c_n \mid \mu(k)_n\}$$

where $\mu(k)_1, \ldots, \mu(k)_n$ are integrated membership functions for each of the n classes, these are obtained using:

$$\mu(k)_j = D_{xxx}(\mu_{1j},\ \mu_{2j}, \ldots, \mu_{kj})$$

where D_{xxx} is one of several alternative fuzzy decision functions:

$$D_{int}(\mu_{1j}, \ldots, \mu_{kj}) = min(\mu_{1j}, \ldots, \mu_{nj})$$

$$D_{pro}(\mu_{1j}, \ldots, \mu_{nj}) = \prod_{i=1}^{k} \mu_{ij}$$

$$D_{con}(\mu_{ij}, \ldots, \mu_{nj}) = \sum_{i=1}^{k} a_{ij}\mu_{ij} \qquad (\sum_{i=1}^{k} a_{ij} = 1)$$

The use of D_{int} suggests that E_1 and E_2 interact in a more or less independent fashion and that the presence of a smaller μ should be preserved. The use of D_{pro} suggests E_1 and E_2 act like identical independent trials so that repetitive observations cause marked changes in relative values of membership. The use of D_{con} suggests that E_1 and E_2 act in a reinforcing fashion so that membership is intermediate between the two input values. There is no generally accepted criterion for selecting a particular D_{xxx}.

Fuzzy logic lacks a well defined decision model. One approach is that described by Thompson [Thompson 85], which follows from the explanation of fuzzy logic above. This approach combines a fuzzy set of goals and a fuzzy set of constraints to form a fuzzy set, known as the confluence set. The combination is achieved through one of the D_{xxx} methods above or some other suitably defined mechanism. This confluence set can then be used in the decision making process. There is no generally accepted method for using this set, though suggestions include:

1. Choose action having greatest degree of membership

2. Choose action that is a mixture of all actions weighted by their degree of membership

3. Choose an action that is an equal mixture of the two actions having the minimum and maximum degree of membership

Possibility theory is a theory of approximate reasoning based on the concept of a *possibility distribution.* A possibility distribution measures the degree to which something is feasible, for example, considering the definition of 'tall' given earlier, if Tom is a tall person then it is quite possible (0.75) that he is 6'6" high, it is less possible that he is 5'6" high (0.25). The possibility of an event is not the same as its probability, the probability that Tom is 6'6" is not 0.75. There is also no requirement that possibilities sum to one. A related notion is that of *necessity*, the degree to which something *must* be true. These values can then be used to represent a degree of certainty. Bonissone [Bonissone 83, Bonissone 87] proposes a method based on fuzzy intervals with necessity as the lower bound and possibility as the upper, similar to DS theory. Function evaluations based on triangular norms and co-norms are used in the weighting and aggregation of conclusions. A mapping between fuzzy intervals and linguistic labels is used to convert between internal representations and external, user-based expressions. Some work on using possibility and necessity in pattern-matching is reported by Cayrol *et al* [Cayrol *et al* 80].

One of the fundamental tools in approximate reasoning is the rule of compositional inference. The classical inference rule is *modus ponens*:

1. X is A implies Y is B

2. X is A

3. Conclude Y is B

The rule of compositional inference extends this definition to a *generalised modus ponens*:

1. X is A implies Y is B

2. X is A′

3. Conclude Y is B′

where A and A′ and B and B′ are 'similar' fuzzy sets. For example [Zimmermann 90]:

1. If a tomato is red then the tomato is ripe

2. This tomato is very red

3. Conclude this tomato is very ripe.

There are several different formulations of this inference rule, generally based on min and max compositions. A schematic illustration of the reasoning process taken from Dutta [Dutta 85] is shown in figure 3.2. Henkind and Harrison [Henkind & Harrison 88, page 707] explain the difference between fuzzy logics and approximate reasoning in the following way:

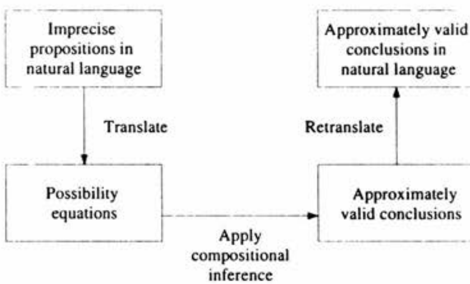


Figure 3.2: Approximate reasoning process [Dutta 85, page 22]

> ... fuzzy logics deal with propositions of the form "$x \in A$," where A is a fuzzy set and the truth of the proposition is given by $\mu_A(x)$. Approximate reasoning, on the other hand, deals with propositions of the form "x *is* A" where A is a fuzzy set. Thus fuzzy logics manipulate numerical values (acquired from a fuzzy membership function), whereas approximate reasoning manipulates fuzzy sets.

FS theory's main strength lies in the fact that it was developed in response to a particular class of problem. This means that it is generally strong on representing certain types of concept, and in relating to natural language:

- [Dutta 85] FS theory is intrinsically linked to natural language, which makes fuzzy sets a natural form of representation.
- [Henkind & Harrison 88] FS may provide a solution to some problems arising from lexical imprecision.
- [Henkind & Harrison 88] FS theory is flexible as many possible operator definitions are possible.
- [Zadeh 65] FS theory is a natural way to model imprecisely defined classes which play an important role in human thinking, particularly in the domains of pattern recognition, communication and abstraction.

However, whilst it has strengths in terms of representation, it is less good at reasoning with those representations:

- [Cayrol *et al* 80] A grade of possibility has only an indicative value because it is more or less subjectively assessed.
- [Dutta 85] Validity in fuzzy logic is approximate rather than exact.
- [Dutta 85] It is not clear that natural language is the most appropriate general knowledge representation technique.
- [Saffiotti 87, Stephanou & Sage 87] The definition of characteristic membership functions is both subjective and context dependent.
- [Clark 90] Empirical derivation of characteristic membership functions is hard.
- [Henkind & Harrison 88] It is not always clear how to construct reasonable membership functions, no completely general technique exists.
- [Henkind & Harrison 88] The choice of operator definitions can be problematical and different definitions may be needed in different situations. Some definitions

with nice mathematical properties perform poorly in the real world, whilst others which perform well are *ad hoc* and lack mathematical rigour. There is little guidance as to which methods should be used for a given problem.

- [Thompson 85] FS theory lacks a well defined decision mechanism.

3.2.4 Numerical and Non-numerical Formalisms

The above formalisms can loosely be categorised as numerical, they represent belief by a numerical value or range of values. The alternative approach, based on symbolic representations, argues that numerical techniques are inappropriate and inadequate for a variety of reasons. The four principal arguments for non-numerical representations are summarised by Buxton [Buxton 89]:

1. Most people find it difficult to express, or think about, uncertainty in numerical terms.
2. The use of an exclusively numerical approach may restrict the knowledge that a person uses in arriving at assessments of support or belief.
3. Numerical approaches to uncertainty are representationally inadequate — they fail to capture all aspects of uncertainty that are relevant to subsequent reasoning.
4. Numerical assessments of uncertainty may hide more specific knowledge which we could specify or collect if we took the trouble.

People recognise levels of belief associated with many of the rules they use, but these are not routinely expressed in numerical terms, nor are they used in any formal statistical manner [Shortliffe & Clancey 84]. When people discuss uncertain events they will typically quantify the uncertainty through the use of imprecise terms, such as 'likely', 'fairly unlikely' and so on. These natural terms lack any precise definition and tend to be highly context sensitive even within the usage of a particular individual. Whilst people are often able to express associations, *e.g.* causes, with confidence, the precise quantification of that association is often problematic [Pearl 88b, pages 20 and 79]. When people are asked to express uncertainty in numerical terms the reliability and meaning of the resulting numbers is questionable [Cohen 86]. As was noted earlier, people generally use *ad hoc* techniques in determining and using numerical representations of uncertainty [Tversky & Kahneman 90a].

It has been suggested [Buxton 89] that the use of numerical representations may change the cognitive structure of a problem and hence the expert system derived from that cognitive structure.

Whether or not a number is an adequate representation of uncertainty, or how many numbers are required, and so on, has long been a subject of debate. Cohen [Cohen 89] argues that it is necessary to represent uncertainty about different kinds of evidence in different ways. Numerical approaches are often criticised for using point values rather than ranges, requiring an unfeasible degree of accuracy which in turn leads to conclusions that are deceptively precise. The use of numerical measures when generating explanations has also been questioned [Cohen 86], as has their *psychological meaningfulness* [Pearl 88b, page 78].

It has further been argued that a numerical quantification is often used in cases where a more complex relationship is in fact true [Saffiotti 87]. The ability to abstract what is perhaps a poorly understood relationship into a single number is seductive.

Clark [Clark 90] summarises the difference between quantitative and symbolic ap proaches:

1. Symbolic techniques derive inspiration more from patterns of competent human reasoning.

2. Symbolic techniques make fewer and weaker assumptions about independence and exclusivity and are therefore more robust in some circumstances. However, by making stronger assumptions, quantitative approaches achieve greater precision in the combination of evidence.

3. Symbolic techniques are more amenable to implementation of metalevel control.

3.2.5 Endorsement Theory

Endorsement-based reasoning (ER) or endorsement theory is a method of symbolic reasoning about uncertainty developed by P. R. Cohen and co-workers. The emphasis of the theory is on actively managing uncertainty rather than simply measuring it [Cohen 86] Precise quantification of uncertainty is shunned and an approach that recognises different types of uncertainty and different methods of combining evidence is proposed This qualitative theory centres on the explicit recording of the justifications for a proposition. These justifications are tagged with endorsements which classify them according to the type of evidence and the possible actions required to solve uncertainty relating to that evidence [Bonissone 87]. Endorsements are attached to inference rules, program tasks, data, conclusions and so on [Saffiotti 87].

The meaning of particular endorsements is determined by the way in which they are used during the reasoning process, they can be said to have operational semantics [Cohen & Grinberg 83]. The mnemonic labelling of an endorsement, *i.e. corroboration* is chosen to reflect those operational semantics. The meaning of an endorsement is made up of three parts, the situations under which the endorsement can be applied to an interaction between evidence, how the endorsement affects the relative ranking of propositions that carry it, and how the endorsement interacts with other endorsements.

As an example, consider the following rule taken from SOLOMON, a system for giving investment advice [Cohen & Grinberg 83, Cohen 86]:

IF age > 65 THEN risk-tolerance = low

This rule could carry the endorsement *overgeneralisation*, indicating that for some individual the conclusion *could* be false even though the premise were true. ER suggests searching for another rule with the same conclusion but a different premise, *i.e.* a *corroborating* rule. If such a rule were found then the conclusion that risk-tolerance is low could be endorsed as *corroborated.*

The endorsements and a set of operational methods are used to reason over an inference net, which includes data nodes, intermediate nodes and conclusion nodes. Each domain may have its own characteristic set of endorsements and methods. Endorsements are propagated over the inference net in a manner that is sensitive to the context of the inference. Uncertain conclusions can be resolved in one of four ways [Cohen & Grinberg 83]:

1. A node's endorsements can be judged sufficient for the goal under consideration. This is not equivalent to saying that the value of the node is certain, only that it is not uncertain enough to warrant further action.

2. An endorsement of an earlier node, although sufficient for some previous goal, is judged to be insufficient for the current goal. The earlier value is retrieved and reconsidered in the current context and another endorsement is assigned. This will involve backtracking.

3. There is uncertainty about the value of the current node but it is discounted by (a) picking the value that has the highest endorsement or (b) generating a new value according to some heuristic method.

4. The uncertainty of the current node cannot be resolved in a way that preserves a minimum endorsement so the multiple values of the current node are propagated on to the next node.

ER concepts have been applied in several experimental systems in order to explore the applicability of the approach [Cohen 86]. The first of these was SOLOMON, which used endorsements in place of numeric degrees of belief. SOLOMON's inference rules used three main types of endorsement [Bhatnagar & Kanal 86]:

1. Model based when it is possible to provide a principled explanation of why the state in the condition leads one to believe the state in the conclusion.

2. Causal — when the state described in the condition of the rules causes the state described in the rule's action.

3. Correlational when the state in the condition is associated with the state in the conclusion but no definite causal link can be established.

Some of the rule endorsements used are [Clark 90]:

- Maybe too general — more cases satisfy the condition than merit the conclusion.
- Maybe-too-specific — fewer cases satisfy the condition than merit the conclusion.
- Exact — neither too general, nor too specific, nor a negation.

- Supportive — increases the confidence if true, but does not cast doubt on the conclusion if false.
- Necessary — conceptually the converse of supportive.
- Hard_not — not(X) must be adequately endorsed.
- Cwa_not — closed-world-assumption, not(X) appears in the database, or attempts to prove (X) fail.
- Ostrich_not — (X) does not *currently* appear in the database.
- Flexible — believable if a proposition is found that is approximately equal.
- Inflexible — values must be precisely met.

SOLOMON uses a goal driven control strategy directed by a task agenda [Clark 90]. When the goal driven strategy fails to produce a conclusion with sufficient endorsements to satisfy a task goal, a new task is created that will attempt to resolve or discount uncertainty. Uncertainty can be *discounted* by selecting a very general course of action that covers all possible outcomes. Uncertainty can be *resolved* by collecting further information, corroborating conclusions with weak endorsements, or by attempting to resolve conflicts by reducing or removing the endorsements of conflicting propositions. Tasks placed on the agenda also carry endorsements [Bhatnagar & Kanal 86]:

- P-corroborate — the conclusion of the task corroborates the conclusion of another task already on the agenda.
- P-conflict — the conclusion of the task conflicts with the conclusion of another task on the agenda.
- P-potential-conflict — the conclusion of the task may conflict with the conclusion of another task on the agenda.
- P-redundant — the conclusion of the task is identical to the conclusion of another task that was derived from the same rule and that has the same rule endorsement.

The conclusions resulting from these tasks carry endorsements that generally mirror their corresponding task endorsements, *e.g. corroborate*, *conflict*, etc..

The endorsements are combined by straight forward inheritance. For example, if the endorsements of P are E_P and $P \Rightarrow Q$ is endorsed by $E_{P \Rightarrow Q}$ then the endorsements of Q, are the set $\{E_P,\ E_{P \Rightarrow Q},\ E_O\}$ where E_O are the endorsements on other conclusions relating to Q. Clearly SOLOMON has the potential to construct large sets of endorsements after only relatively few inferences.

The second system, HMMM, was developed to explore the way in which reasons are adjusted in the context of new evidence [Sullivan & Cohen 90]. The domain concerned reasoning about simple devices that execute plans consisting of a sequence of steps:

plan 1: A B C

plan 2: B D E

The questions HMMM addressed were of the form "if the endorsement of the plan 1 interpretation of step A is *may-be-a-mistake*, what happens when more evidence in the form of subsequent plan steps, becomes available?" It was argued that answering questions of this type is analogous to finding a function for the combination of numerical measures of uncertainty. A set of combining schema was developed that "captured the flux of our reasons for uncertainty in the plan recognition problem" [Cohen 86, page 421]. For instance, if the first step in a sequence of operations has the endorsement MAY-BE-A-MISTAKE (perhaps some other sequence of operations was expected to be chosen), then evidence in the form of the second step in the sequence suggests that this endorsement is incorrect and that the first step was indeed intentional [Cohen 86, page 422]:

```
IF step i IS-UNIQUE-TO plan N
AND step j IS-UNIQUE-TO plan N
AND j FOLLOWS-IN-THE-PLAN step i
AND the plan N interpretation of step i is endorsed by MAY-BE-A-MISTAKE
THEN erase the endorsement
```

This provides a framework for subjectively combining endorsements, though a more realistic method might involve changing the weights of endorsements rather than deleting them [Sullivan & Cohen 90].

Cohen and his colleagues then moved on to consider the source of domain endorsements [Cohen 86]. They recognised that for a complex domain a large number of endorsements and combination methods would be required and that ideally these should be derived directly from other knowledge of the domain. They developed a theory of *path endorsements*, embodied in a system called GRANT. Path endorsements were used to describe typical patterns of association between evidence and conclusions, derived directly from inference rules. Consider the example in figure 3.3: there is only a single causal link to **Nap**, so given **Nap**, **Fatigue** can be credibly inferred. On the other

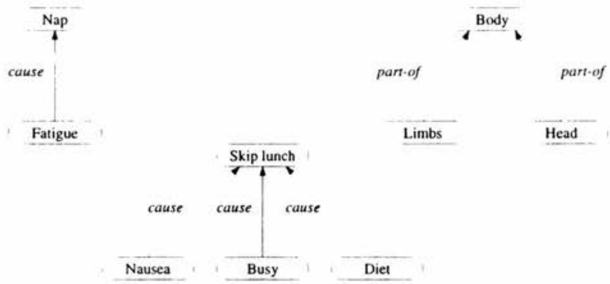


Figure 3.3: Inference rules from a semantic network [Cohen 86, page 427]

hand the inference of **Nausea** from **Skip lunch** is far less credible due to the number of potential alternate causes. Similarly the sibling relation between **Head** and **Limbs** is a poor inference path, as something that is true of an object is not necessarily true of that object's sibling. The endorsements, at least in some cases, can therefore be derived directly from the associations present in a semantic network.

The emphasis in MUM [Cohen 89] was on the use of uncertainty to constrain action in control problems. The argument was that problem solving under uncertainty is doubly constrained, actions must be selected both for their domain effects and for their effects on uncertainty. For example, if the treatments for disease 1 are A, B and C and for disease 2 are B, D, and E, and it is not possible to decide between diseases 1 and 2, then treatment B is the only course of action available. MUM was designed to create work-ups (a diagnostic sequence of questions, tests and treatments) for diseases that manifest themselves through chest and abdominal pain. Its goal is to create a work-up that conforms to that produced by a physician. It uses a large inference net with disease nodes at the top, data nodes at the bottom, and intermediate clusters of clinically significant groupings in between. Data provides support for, or detracts support from clusters which in turn support, or detract from disease hypotheses. MUM recognises seven levels of belief — *confirmed, strongly supported, supported, unknown, detracted, strongly detracted* and *disconfirmed.* The objects in the inference net are linked by endorsed paths which specify the role the evidence plays with respect to the conclusion. Endorsements include *potentially-confirming* and *potentially-detracting* which enable MUM to reason about the utility of nodes in confirming or disconfirming nodes above them in the net. Each node has a local evidence combining function, which makes explicit the dependence of the belief of a node on the levels of belief of its supporting and detracting nodes as illustrated in figure 3.4. These combining functions also support reasoning about evidence gathering, for instance by allowing MUM to select discriminatory clusters. MUM's control cycle

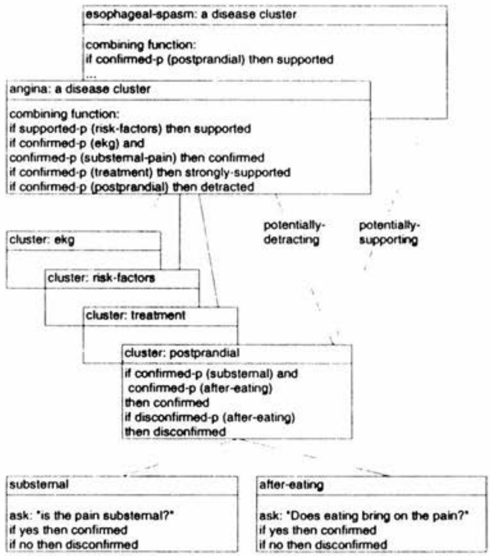


Figure 3.4: Internal structure of inference net clusters [Cohen 89, page 266]

followed the processing loop — establish the focus of attention, decide which question, test or treatment to invoke, given the current state of the net. Propagate the result through the net, updating the parameters on which control decisions depend in the next cycle.

MUM was later generalised to a domain independent tool MU which, among other features, allowed the user to define new control parameters in terms of those already defined, *e.g.* *critical* could be defined as *at-least-supported* and *dangerous*, which could then be associated with all nodes of a particular type, *e.g.* disease nodes. MU was able to answer five classes of question about the value of control parameters [Cohen 89]:

1. Questions about the current state — *e.g.* what is the monetary cost of a particular test?
2. Questions about how to change features — *e.g.* how can I increase the level of belief in angina?
3. Focusing questions — *e.g.* what set of diseases are both triggered and dangerous?
4. Affect questions — *e.g.* for what diseases does age affect the level of belief?
5. Questions about multiple feature changes — *e.g.* what discriminates between angina and esophageal-spasm?

It is difficult to identify specific positive features of ER which are not also features of other qualitative approaches:

- [Cohen & Grinberg 83] The propagation of inference is dependent on the local context.
- [Cohen & Grinberg 83] ER is able to represent domain specific heuristic approaches for handling uncertainty.
- [Bonissone 87] Endorsements provide a good mechanism for explanation as they explicitly maintain the entire history of justification and relevance of any proposition with respect to a given goal.
- [Clark 90] ER is useful as a declarative representation of uncertainty.
- [Sullivan & Cohen 90] ER does not preclude the use of numerical quantifications.
- ER is potentially a very natural reasoning mechanism.

Similarly many concerns about ER are concerns about non-numeric techniques generally. The major concerns about ER relate to its applicability to real-world applications:

- [Bhatnagar & Kanal 86, Saffiotti 87] No well defined mechanism for selecting between two competing hypotheses supported by different bodies of endorsements.
- [Saffiotti 87, Clark 90] Method of endorsement combination is inefficient.
- [Bonissone 87, Clark 90] Endorsement propagation methods are ill defined.
- An expert system based on ER may be highly expert dependent.

3.3 Control

Ideally the knowledge-base and inference mechanism are separate and distinct. In practice the inference mechanism is to a large extent constrained by the design of the knowledge base and in some cases the knowledge-base contains implicit control information. In this section the control of inference in an uncertain environment will be considered. The formalism selected for *representing* uncertainty should not influence the selection of strategies for *reasoning* under uncertainty (though in many situations it will act as a constraint in the same way as it may constrain, or be constrained by, the knowledge base design). A *control strategy* specifies three main behaviours [Cohen 90]:

1. Selection of focus of attention — the sub-part of the overall problem that should be considered next. If the focus is constantly shifting, no coherent diagnostic path will be followed. If the focus is too fixed, too many resources may be expended on an incorrect diagnosis.
2. The control of actions — the collection of evidence etc.. A balance between the cost of an action and the diagnostic importance of the action must be struck.
3. The control of inference — the selection of reasoning methods to apply to the focus of attention. If the propagation of inference is uncontrolled the system may become overburdened with the number of possibilities. If it is too restrictive, then important possibilities may be missed.

These mechanisms are illustrated in the following example [Cohen 90, page 178]:

> A patient walks into the doctor's office with a high temperature. The findings are consistent with dozens if not hundreds of disease and ailments. The

doctor must control the inferences that can be made from the finding of high temperature, or else be swamped by hypotheses about what is wrong with the patient. Control of inference implies that not all possible inferences are made. Of those that *are* made, the doctor will recognise one or two possibilities as especially likely, or serious, or worthy of attention. One will become the focus of attention, and the doctor will formulate a goal to confirm it or rule it out. Now the doctor must decide which actions will obtain more information at an acceptable cost — which questions, tests, or treatments will provide needed evidence. Once obtained, evidence will enable further inferences, and perhaps a revised focus of attention and other goals.

If the primary goal is to reduce uncertainty, then actions (such as performing a test) are the means by which uncertainty is reduced. The selection of an action must therefore be based on its domain effects and its effect on reducing uncertainty. This suggests that the representation of uncertainty will play an important role in determining the control of problem solving, as a means for selecting a focus, selecting an action and controlling inference. In order to achieve this, three *control features* are required [Cohen 90]:

1. A measure of the current degree of belief in a hypothesis.
2. The prior probability of the hypothesis.
3. The potential change of belief given some action.

Other domain level control features, such as *dangerousness* can also be used to guide control reasoning. Given control features such as these a variety of problem solving strategies are possible, *e.g.* select actions that potentially confirm the current most likely hypothesis. General strategies of this sort often have direct parallels in human problem solving methods, and may constitute natural strategies. Some general strategies, taken from Hearsay-II by Cohen [Cohen 90] are:

- Efficiency: reliable and inexpensive knowledge sources should be executed before less reliable or more expensive ones.
- Validity: knowledge sources operating on the most valid data should be executed first.

- Significance: some knowledge sources are defined *a priori* to be more significant than others.

- Goal satisfaction: knowledge sources that satisfy goals are preferred to those that do not.

- Competition: given a choice among several actions pick the 'best'.

There are many different approaches to control (see [Szolovits & Pauker 78] for some examples), two complementary approaches to control are described below. The first of these, exemplified by Protos, investigates computational models of control for agents acting under real-world constraints. The second, represented by ONYX and TA, performs analysis of a cognitive task in order to derive a computational model.

3.3.1 Computational Approaches

The Protos project [Horvitz 88, Horvitz 89, Horvitz *et al* 89a, Horvitz *et al* 89b, Horvitz 90, Horvitz & Rutledge 91] concerns the development of computerised agents that perform rationally under resource constraints. Typically every action incurs some cost, perhaps financial, or in terms of patient inconvenience, or time taken before the results are known. Because of this, the management of uncertainty through control involves a compromise between certainty and cost. In a domain where there are no cost considerations, control strategies are irrelevant. The tradeoff between cost and certainty provides a metric for judging the efficiency of problem solving, a more efficient solution maximises certainty with respect to cost.

The task of uncertain reasoning is composed of three elements [Horvitz 89]:

- Problem formulation — the task of modelling or constructing the reasoning problem, this often involves enumeration of relevant hypotheses and dependence among hypotheses.

- Belief entailment or inference — process of updating measures of truth assigned to to alternative hypotheses as new evidence becomes available.

- Decision making — process of selecting the best action to take, an irrevocable allocation of resources.

The implicit assumption of normative approaches to decision making (*i.e.* those consistent with the axioms of decision theory) is that sufficient resources exist to enable the determination of an optimal action. In many real world situations this assumption is false. This has led to an emphasis on non-normative, heuristic approaches to the control of uncertain reasoning. Horvitz and his co-workers are interested in normative reasoning under resource constraints and argue that this will lead to solutions that are more optimal.

Protos initially focussed on the notion of a partial or approximate result [Horvitz 88] under time constraints, concentrating on the question of how good a result can be achieved given the time available. A measure of *computational utility* was defined. This is a measure of the net value associated with the commitment to a particular computational strategy. It is composed of two parts, an object level utility which measures the benefit attributed to acquiring the result regardless of the computational cost, and an inference related utility which measures the cost of reasoning, *e.g.* in time used. The relationship between partial results and prototypical time constraints, *e.g.* deadline, urgency, etc., was investigated using these measures. It was shown that in situations of uncertain resource constraints, a strategy that is less optimal in the absence of constraints but which continuously refines its results may be preferred to one that only generates a complete solution after a set resource expenditure.

The project then moved on to consider the broader question of a rational control strategy with decision-theory as a normative basis [Horvitz 90], using the term *bounded optimality* to distinguish their approach. Such systems must reason about the solution methodologies available, the costs of reasoning resources and the expected challenges that will be faced in the environment during the problem solving process. All these considerations will involve elements of uncertainty. The task is not to determine the ideal result but to generate the best possible result given the resources.

Traditional normative applications of decision theory define rationality in terms of a model, the goal being then to select an action that has the greatest utility according to that model. The selection of an action generally assumes unconstrained resources though the creation of a model and inference within the model are resource intensive. The Protos project explores ways of extending normative rationality into domains of uncertain, varying and restricted resources. This has focussed on the use of decision theoretic models to reason about the utility of problem solving strategies as well as domain level utilities, and on partial results as mentioned above.

Strategies identified as being particularly relevant are those that provide partial results that have domain level value which increases monotonically with increasing resource expenditure, finally converging on the ideal result after consuming some quantity of resources. Several promising strategies for probabilistic inference have been identified [Horvitz 89]:

- Bounded calculation and propagation — the use of bounds on probability rather than point probabilities. These bounds can be refined as additional constraints are considered.
- Simulation — approximation strategies that report a probability distribution or partial characterisation of a distribution over probabilities of interest through a process of weighted random sampling which converge to true probabilities.
- Completeness modulation — simplification of model by deletion of classes or relationships based, for example, on a measure of importance.
- Abstraction modulation — it may be more useful to do complete normative reasoning on an abstracted model than to do approximate reasoning on a complete model. In many domains, models at higher levels of abstraction are more tractable.
- Local reformulation — use of local approximations in otherwise complete models for normative reasoning in cases where local complexity renders a full solution intractable.
- Default reasoning and compilation — under severe time pressure default beliefs and policies may be of more value than computed results. This may be particularly true when concerned with actions of great importance, high frequency or time-criticality.

A multilevel approach is proposed, whereby metareasoners use attributes of reasoning problems and of reasoning that can serve as indications about the value of future computations. These attributes serve to partially characterise the nature of reasoning at lower metalevels and and the domain level. The aim of metareasoning is to optimise the utility of a decision. In order to do this it may be necessary to expend a proportion of the reasoning resources to deliberate how best to optimise utility. The primary task of metareasoning about decision making is the determination of the expected value of computation for the available strategies. A simple, approximate, myopic measure, the *expected value of computation*, defined as the difference between the current value and the sum over expected future utility of states weighted by the probability of achieving that state under a particular strategy, is developed. The EVC does not consider all possible combinations of strategies in an attempt to find a global optimisation as this could introduce undesirable, resource intensive complexity into the metareasoning task. The applicability of metareasoning depends on the existence of approximation strategies that allow a trade off in quality of an ideal (precise) analysis against more tractable, less precise results to be made. Decision-analytic metaknowledge must include knowledge about model construction, inference, metareasoning and interactions among these three.

- Model construction metaknowledge captures attributes useful in reasoning about the value of continuing to employ strategies for generating and refining distinctions and relations in a decision model.

- Inference metaknowledge includes distinctions useful in estimating the value of future inference.

- Metareasoning metaknowledge is information about distinctions that are used to characterise the expected values of increasing the fraction of time dedicated to metareasoning or moving to a higher metalevel.

- Interaction metaknowledge captures knowledge about the interaction between model building, inference and metalevel deliberation, such as the relationship between models of higher quality and the growth of complexity of inference.

The research also suggests that pre-compiled knowledge, i.e. pre-computed complete or partial results, will have an important role to play in reasoning under resource constraints. Three general classes of compiled rules are identified [Horvitz 90]:

- Situation-action rules — observed situation linked directly to final action without deliberation.

- Platform rules — used in conjunction with deliberation to make deliberation more efficient and reduce computational burden. Includes cached partial results for gen eral classes of problem that can be refined with additional computation.

- Resource rules — situation-action rules and platform rules are both means of increasing efficiency of decision making. Resource rules are compiled behaviours that generate additional resources, for example by reducing the cost of a delay, or by extending a deadline. These are typically knowledge intensive and highly domain specific. For instance, if a patient's blood pressure is falling rapidly, a blood transfusion may be used as a temporary remedy, buying time to investigate the cause of the blood loss.

The question of higher levels of metareasoning, where such knowledge is available, was also considered. A balance needs to be made between the value associated with the flexibility gained from being able to reason about existing levels of analysis, and the costs incurred because of the additional complexity. Regardless of the level of metareasoning employed it was suggested that appropriate situation-action rules acting at the domain level should always be available and built, for example, into a compiled metareasoning policy — IF an object level situation-action rule is available THEN act ELSE deliberate for a time dictated by the metareasoning model.

Other interesting suggestions are the use of quiet-time during reasoning for expectation driven creation of partial solutions and the use of idle time for the refinement and learning of behaviours. This short term and long term learning are valuable compilation strategies.

3.3.2 Cognitive Approaches

In the past, much of the emphasis on control strategies was focussed on the achieving of a specific task, *e.g.* diagnosis of a particular disease. Recently there has been a move towards the development of control strategies for more general classes of task, such as diagnosis, therapy planning and patient monitoring. This approach uses the general control strategy in conjunction with a domain specific world model. It acknowledges that strategies for managing uncertainty are part of the domain expertise and should therefore be acquired from domain experts [Cohen 90]. Two examples of this approach, both concerned with therapy planning, are outlined briefly below. Therapy planning can be defined as [Quaglini *et al* 92, page 208] "...the task of selecting the best action to improve a patient's condition given the available clinical information."

ONYX is a prototype system designed to plan therapies for use in the treatment of cancer [Langlotz *et al* 87]. The basic therapy planning task involves a representation of the current state, a desired goal state and a set of operators (actions) which can be executed. The task is to select a sequenced subset of operators on the basis of the current state, that results in the optimal satisfaction of the goal state. Four major difficulties in therapy planning in this domain were identified:

1. Explicit guidelines for plan selection are not available.
2. The current state is not known with certainty.
3. The consequences of an action cannot be predicted with certainty.
4. The planning goals cannot be satisfied completely as they are inherently contradictory.

Clearly these concerns are equally true of other therapy planning domains and other classes of planning problem.

The ONYX control strategy was based on observations of domain experts, who exhibited a consistent problem solving approach:

1. Develop a set of possible therapy plans.
2. Envision the possible consequences of applying each plan.
3. Assess the predicted outcomes against the therapeutic goals.

ONYX translates these cognitive tasks into processes which closely reflect the intermediate solution steps listed above:

1. Plan generation — general treatment strategies are used to create a small set of reasonable plans by selecting combinations of appropriate treatment components given the current state of the patient.

2. Plan prediction — simulations based on the structure and behaviour of the human body are used to predict future states of the patient under each suggested therapy plan.

3. Plan evaluation — decision analysis is used to rank the plans according to how well they are predicted to satisfy the therapeutic goals for the patient.

The plan generation phase involves the traversal of a hierarchical therapy generation model under the control of heuristic strategies embodied in production rules. These rules fall into two main categories, control rules such as (paraphrased):

```
IF: a problem is encountered with a treatment
THEN: try to eliminate the least significant component of the
    treatment that might be causing the problem
```

or

```
IF: The patient is being treated according to a protocol
AND
The protocol requests drug treatment for this visit
OR
    The patient's tumour is sensitive to drug treatment
THEN: Consider drug treatment
```

and generation rules:

```
IF: a drug can exacerbate a toxicity already being experienced by the
patient
THEN: Propose reducing the dosage of that drug
```

The plan generation process starts at the most general node in the hierarchy, the control rules associated with a node are used to determine which descendant node should be examined next. When a terminal node is reached, the generation rules are used to create a set of plan steps. A complete plan is formed from a consistent set of proposed plan steps. A small set of plans forms the input to the next phase of the system.

The plan simulation phase predicts the behaviour of important patient variables under each of the proposed therapeutic plans. ONYX uses a qualitative simulation to embody the oncologist's knowledge which is itself often qualitative and uncertain. The simulation models physiological entities and behaviour in an object-orientated framework. The behaviour of an entity is described by three sets of rules:

1. Describes how object model parameters change with respect to incoming messages.
2. Contains knowledge including symbolic information on object state based on recent internal changes.
3. Determines how new object state will be transmitted to neighbouring objects.

The simulation is used to predict the outcomes of the proposed therapeutic plans. These predictions are then used in a decision theory based ranker to calculate a quality adjusted life-years utility value for each plan. This utility model considers the utilities associated with four treatment goals:

1. Decrease the risk of death.
2. Decrease need for supportive care and hospitalisation.
3. Decrease the discomfort for the patient.
4. Remain close to treatment guidelines.

In the second example, Therapy Advisor (TA), an epistemological model of the therapy planning task is developed, figure 3.5, that defines the relationship between the domain knowledge and task knowledge. According to this analysis, therapy planning is performed on the basis of a diagnosis. The diagnostic process is typically concerned with producing a qualitative description of the state of a patient. Therapy planning requires more exact quantitative information, so the first step may be to gather further information. Once all the necessary information is available, *abstraction* is used to derive a restricted set of critical attributes of the patient's state. This set of attributes can be interpreted immediately as a set of crucial therapy targets which compose therapeutic problems. Therapeutic problems fall into two classes — simple and complex.

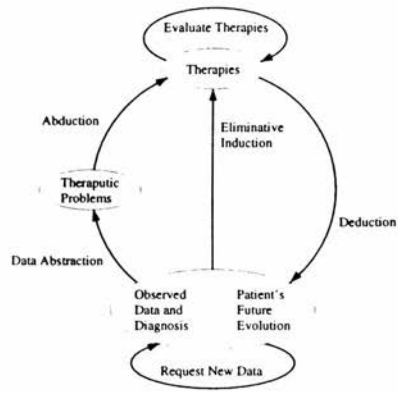


Figure 3.5: An epistemological model of therapeutic reasoning
[Quaglini *et al* 92, page 208]

Simple problems admit an immediate well-accepted solution without further deliberation. These are sometimes referred to as categorical decisions [Szolovits & Pauker 78]; such decisions depend on relatively few facts, their appropriateness is easy to judge and their result is unambiguous. There is a clear connection between the notion of a simple problem, a categorical decision, and the compiled behaviours and stimulus-action rules of Protos. Complex problems may involve several different therapeutic actions and their appropriateness is not easily established due to uncertainty regarding their outcomes and the decision maker's preferences. These problems require further deliberation.

The therapeutic problems serve as the basis for the *abduction* of a set of therapies. The set of therapies consists of those that are potentially relevant to the problems under consideration. Depending on the number and complexity of the therapeutic problems, a therapy can be adopted at this point, or will form a hypothesis. If a hypothesis is formed then therapy evaluation must be performed to determine which of the competing hypotheses is to be selected.

This task model underlies TA, a system concerned with the general task of therapy planning, though applied to the specific domain of anemia [Quaglini *et al* 92]. TA is part of NEOANEMIA [Quaglini *et al* 89], which provides the patient diagnosis that forms part of TA's input. TA's domain knowledge uses a mixture of representations including taxonomies of medical entities, production rules and hand-crafted anemia-specific influence diagrams. The control knowledge is in the form of metarules, such as:

```
IF abduction has been performed
THEN perform evaluate-therapies
```

where `abduction` and `evaluate-therapies` are operators that execute the appropriate reasoning steps. A taxonomy of subtasks that make up the complete epistemological model is used to guide the control process which is based on a blackboard system [Hayes-Roth 85, Nii 86a, Nii 86b]. Different inference steps are achieved by specific knowledge sources acting over the blackboard.

A possible criticism of these approaches is the difficulty of obtaining reliable models from human experts, this is touched on by Szolovits [Szolovits 79].

3.4 Communication

In the previous chapter the importance of the physician-system interface was identified. In many early expert systems the significance of the interface was not fully appreciated and in several cases it was added almost as an afterthought, once the knowledge base and inference mechanism were designed. Furthermore the interface design tended to be carried out without reference to the user, resulting in interfaces which were perfectly adequate from the point of view of the system's designer, whose understanding of the system was good, but less than adequate for the user, whose understanding and rationale for interaction were very different. As result of this these interfaces generally failed to meet user expectations and later work to use the knowledge base for a different mode of explanation often proved problematic [Clancey 83]. There is now a greater realisation of the importance of the interface and the fact that interface design is intrinsically linked to the design of the knowledge base and inference engine.

When designing an interface it is necessary to compromise between the need to provide the user with as natural and flexible an interface as possible, and the fact that computer cognition is impoverished, inflexible and qualitatively different from human cognition [Dodson 90]. The earliest interfaces relied heavily on being able to anticipate user queries by constraining the range of questions that could be asked; the answer to each query could simply be stored. This could be extended by generating text which included blanks which were filled depending on the context. This canned-text may be practical for a very limited type of expert system, where the domain of discourse is small and a single level of inflexible responses is adequate.

The second type of early interface relied on describing the processing steps of the inference engine. This type of interface may be acceptable for system designers who are essentially interested in how the system works, but they are generally not able to give explanations in terms familiar to the domain expert.

Both these approaches relied on tightly restricted user input and providing explanations that were closely related to the design of the expert system. More recently it has been recognised that what is actually required is interactive dialogue based on the domain explananda, with the user and the system sharing the locus of understanding. This is particularly true in cases where the patient is involved in the interaction as they are liable to need a greater degree of support and explanation than a physician who uses the system on a regular basis [Jimison 90, Jimison *et al* 92]. In order to be able to support this kind of dialogue it is necessary to have a clear separation between the knowledge base, defined in terms of the domain explananda, and the inference engine, defined in terms of domain independent problem solving strategies. By representing domain knowledge and inference strategies separately, it becomes easier to answer different types of question, for instance about causal relationships between domain entities or the definition of domain terms, in addition to describing the problem solving strategies employed.

It has been recognised [Clancey 83] that in some expert systems a certain amount of the domain knowledge is stored implicitly, either in the knowledge base or in the inference engine itself, or has been 'compiled out' [Swartout 81]. Such implicit knowledge may be necessary in an explanation, and must therefore be represented explicitly in the knowledge base. In order to do this it will often be necessary to have a very rich knowledge base, breaking down the high level associative relationships typically used for reasoning into a hierarchy of causal relationships that describe the underlying domain processes. This hierarchy of knowledge can then be used to generate explanations at different levels of detail.

There are many different knowledge representation schemes (for an overview see [Brachman & Levesque 85]), the one we discuss here is *structured object schemes.* The phrase *structured object schemes* is intended to define those knowledge representation schemes that rely fundamentally on a network or graph structure. The essence of such schemes is the representation of entities as nodes within a network, where the links in the network are indicative of some form of relationship between the entities they connect. Clearly this class of representations is large and includes traditional representations such as semantic networks [Woods 75] and frames [Kuipers 75, Minsky 80] as well as more recent ones, such as belief networks [Pearl 88b] and neural networks [Rumelhart *et al* 86].

In such representations the form that a node takes is as varied as the entities represented. In some cases, for example frames, a node can be a highly complex object, containing both declarative and procedural knowledge. In other cases, such as belief networks and neural networks, knowledge is limited to input/output relationships.

One of the strengths (and indeed weaknesses) of the network based approach is the variety of meanings that can be ascribed to a link. In some representations only a single type of link is used, for instance causal links in belief networks. Others, such as semantic networks, have many different types of link denoting particular types of relationship, see for example figures 3.3 and 3.4.

Networks that contain certain types of links have the advantage of being *cognitively meaningful.* By this we mean that certain network models form an analogue representation of the way in which people think about or communicate about the domain. Examples of this include belief networks and a number of hierarchical models, such as IS-A networks. Semantic networks and neural networks do not fall into this category as the meaning of their structure is generally opaque with respect to the domain.

Cognitively meaningful networks can have several desirable features, including:

- An explicit, meaningful structure that is easily modified and verified.
- The structure can be used as a basis for the generation of meaningful explanations.
- A structure that should be easy to elicit from domain experts.
- A structure that should be fairly stable between domain experts.

The use of a network representation typically does not place any constraint upon the types of inference that may be made, assuming the network contains all the necessary information.

Some authors [Ben-Bassat & Teeni 84, Horvitz *et al* 86] have argued that in certain cases the problem solving strategies employed by inference engines are themselves opaque to users. Natural reasoning strategies, *i.e.* those that humans perceive as simple and intuitive, are constrained by human cognitive ability and are often less than optimal, for instance in diagnosis physicians will often traverse a class hierarchy of possible causes rather than consider the entire set of possible causes. Humans often select a natural reasoning strategy designed to satisfy their own strategic goals, such as the collection of supportive information for the most plausible alternatives. In cases where the ability to explain the strategy is paramount, it may be desirable to adopt a less optimal, but more natural reasoning strategy. Wherever possible the inference engine should use strategies defined in terms, such as *find* or *avoid*, that can be appropriately and intuitively understood by the user [Swartout & Smoliar 88].

In addition to this it will be necessary to handle dialogue in an intelligent way which will involve a degree of user modelling and role behaviour on behalf of the system. This would allow the system to tailor its interaction to suit the needs and experience of the user, by selecting different levels of explanations and different modes of discourse (interrogative, tutorial, etc.). This technology is still very much a research topic. Though some work [Worden *et al* 87] has been done on defining appropriate behaviours for expert systems performing as assistants and some on dialogue heuristics is cited by Langlotz *et al* [Langlotz *et al* 88b], it is likely to be some time before user modelling and computer role filling with dialogue handling is developed enough to be in routine use.

As well as the changing perspectives on interface design, there is the development of new interface technologies to consider. Early interfaces were text based but the widening availability of low priced graphics hardware is leading to an increased use of graphics within interface design. Some interfaces use graphic facilities to replicate existing paper based interfaces [Lane *et al* 86, Marin *et al* 93]. These interfaces are already familiar to the user and are therefore readily accepted. Other researchers have suggested that new forms of mixed graphical/textual languages should be developed, allowing the user to interact with a graphical analogue representation of some aspects of the domain [Hayes 87, Dodson 90]. Beyond this lies Hypermedia technologies, designed to integrate text, still pictures, animated pictures and sound, and beyond that virtual reality (VR) where interactive 3-dimensional simulated worlds can be created, adding touch to the interface senses. Even though VR is cutting edge technology people are already experimenting with its use in medicine in applications such as virtual surgery [Rubenstein 94, Wright 94] and the simulation of brain cell function [Holmes 94]. Whether VR and graphical languages can be successfully combined to provide interfaces to applications that are essentially non-visual remains an intriguing question.

3.5 Conclusion

In this chapter we have examined some of the issues that are involved in expert systems design. The unifying thread has been the notion that the choice of uncertainty formalism fundamentally affects the control of the expert system, the ability of the system to make decisions, and the ease with with the decisions and results of the reasoning process can be explained to an external observer. As we have stated several times, it is precisely the reasoning process that must act as the context for communication between the system and the physician.

All of the formalisms discussed are modelled on largely *ad-hoc* cognitive theories of human problem solving, they are based on *cognitive axioms*. The developers of such systems typically claim that this makes them more amenable to communication and this is certainly true to some degree. On the other hand, the use of an *ad-hoc* cognitive formalism places the rational decision maker in a difficult situation. There are two main reasons for this. Firstly any uncertain reasoning formalism that does not conform to the normative axioms will, in certain circumstances, make invalid inferences. As a result of this the measure of belief assigned to a proposition may be incorrect. Secondly these *ad-hoc* formalisms typically have an *ad-hoc* decision mechanism, if they have one at all. This implies that not only must the assigned measures of belief be treated with suspicion but so must any utilities assigned to actions.

How can the assertion that the results of *ad-hoc* formalisms are unreliable be reconciled with the demonstrable success of some of these formalisms? It is important to distinguish between a successful application and a successful formalism. MYCIN, for example, is often cited as an example of a successful expert system, yet the CF formalism is *ad-hoc*. For instance, the implicit assumption of equal priors was largely true given the domain. It could also be justified on cognitive grounds, as human problem solvers are typically insensitive to prior probabilities. It is possible that even if the assumption of equal priors had been false, MYCIN could produce correct results as the domain experts might subconsciously include this information in their assessment of the CF values [Horvitz & Heckerman 86]. The implication of this is that the CF formalism may be successful for a particular class of problem, but outwith that class its results will be meaningless.

The difficulties in a formalism based on *cognitive axioms* are many, including:

- The precise definition of the axioms.
- The proof of validity of the axioms.
- Does a single axiomatic framework apply to all problems?
- Do the axioms guarantee rational results?
- Do the axioms provide a sound basis for decision making?
- On what basis are propositions assigned a measure of belief?
- Is it possible to empirically derive the measure of belief for a proposition?
- Can a tractable computational system be constructed on the basis of these axioms?

Unless a cognitive formalism can address these difficulties it is impossible to conduct a meaningful comparison between competing cognitive formalisms and between cognitive formalisms and formalisms based on the axioms of probability.

Chapter 4

Belief Networks

In the previous chapter we saw that whilst cognitively based formalisms may possess certain attractive features, they are unable to provide a general purpose mechanism for decision support. Probability theory, despite its axiomatic foundations, was until recently, largely unused in expert systems due to the perceived intractability of inference and the complexity of model creation. The major breakthrough as far as the *application* of probability theory in expert systems was concerned, has been the development of network representations that embody independence and causal relations. These belief networks provide a mechanism for tractable inference using probability theory, but also make use of expert knowledge in the construction of the network model. They facilitate the combination of the statistical power of probability theory with domain expertise. The expression of complex probabilistic relationships in the form of a directed graph and the development of an efficient inference algorithm have resulted in research into both the theory and applications of probability theory and decision theory, and raised the possibility of a normative formalism for handling uncertainty in expert systems.

Different authors have explained and introduced belief networks in a number of different ways. In this chapter we present an introduction starting from a consideration of the nature of causality and probabilistic independence. More rigorous mathematical introductions are given by Pearl [Pearl 88b] and Neapolitan [Neapolitan 90]. A less formal introduction may be found in Morawski [Morawski 89a, Morawski 89b], Charniak [Charniak 91] or Jensen [Jensen 93]. An introductory discussion of the more complex issues, such as decision making and learning, is given by Andreassen *et al* [Andreassen *et al* 91b]. This formalism is referred to by a variety of names in addition to belief networks, including, Bayesian networks, probabilistic inference diagrams, causal networks, causal probabilistic networks, Bayesian belief networks and, in a more general sense, influence diagrams.

4.1 Probability Theory

Probability theory provides a mechanism for reasoning about propositions, given uncertain evidence and noncategorical relationships between the evidence and the propositions. In probability theory the probability of a proposition A, is denoted by $P(A)$, and the probability of a proposition A given some evidence K, by $P(A \mid K)$.

The three rules of Bayesian probability specify that:

$$0 \leq P(A) \leq 1 \tag{4.1}$$

$$P(\text{certainty}) = 1 \tag{4.2}$$

$$P(A \text{ or } B) = P(A) + P(B) \text{ if } A \text{ and } B \text{ are mutually exclusive} \tag{4.3}$$

From the above rules it can be seen that:

$$P(A) + P(\neg A) = 1$$

The basic mechanism for updating the probability of a proposition, H, given new evidence, e, is Bayes' Rule, which states:

$$P(H \mid e) = \frac{P(e \mid H)P(H)}{P(e)} \tag{4.4}$$

We can ignore the denominator, $P(e)$, as it can be regarded as a normalising constant, determined from the condition $\sum_{H_i} P(H_i \mid e) = 1$. This results in equation 4.5, which states that the belief in H after receiving evidence e is the product of the prior probability of H (how probable we considered H before we received e) multiplied by the probability that e would have been encountered given that H is true (how indicative e is of H).

$$P(H \mid e) = P(e \mid H)P(H) \tag{4.5}$$

We can restate this relationship in terms of odds and likelihoods. The odds of a proposition are given by:

$$O(H) = \frac{P(H)}{P(\neg H)}$$

To define equation 4.5 in terms of odds we divide it by the complementary form for $P(\neg H \mid e)$, giving:

$$\frac{P(H \mid e)}{P(\neg H \mid e)} = \frac{P(e \mid H)}{P(e \mid \neg H)} \frac{P(H)}{P(\neg H)}$$

We can define the prior odds on H (the odds before e is considered) to be:

$$O(H) = \frac{P(H)}{P(\neg H)}$$

and the likelihood ratio of e given H as:

$$L(e \mid H) = \frac{P(e \mid H)}{P(e \mid \neg H)}$$

then the posterior odds (the odds on H after considering e) are given by:

$$O(H \mid e) = \frac{P(H \mid e)}{P(\neg H \mid e)} = L(e \mid H)\, O(H) \tag{4.6}$$

Formulated in this way, Bayes' Rule states that the posterior odds on hypothesis H, given evidence e, is the product of the prior odds on H and the likelihood ratio of e given H.

This is in essence the inference algorithm for probability theory, but in the traditional expression of probability theory the specification and calculation of inference in a probabilistic model is intractable. The reason for this lies with the definition of the probability of a proposition, $P(A)$, or more correctly $P(A \mid K)$, where K is the body of background knowledge against which $P(A)$ is being assessed. Conditional probabilities require that the probabilistic relations between propositions within the event space be quantified in some way. This can be done in terms of a joint distribution function, that is a function which assigns a probability to every possible combination of propositions in the event space. Whilst this approach may be applicable in cases where the event space is small and the joint distribution function is easily available, in any complex domain it becomes impractical.

Pearl's approach, now taken up by others [Henrion 89, Spiegelhalter & Lauritzen 90a, Andreassen *et al* 91b], is to reduce K, for a given proposition A, to only those propositions which impact directly on A. By representing independencies explicitly through the use of a network, the calculation of $P(A \mid K)$ can be expressed locally in terms of $P(A \mid L)$, where L is the set of propositions which impact directly on A and which will typically be smaller than K, thereby rendering the calculations tractable. These lower order relationships are implicitly represented within the joint distribution function but it can be difficult to identify them unless is possible beforehand to state L for a given A [Cheeseman 84]. Belief networks rely on experts to identify local causal relationships between propositions, thereby determining the relevant lower order relationships. It should be noted that the causal relationships identified by the expert are intended to be real causal relationships inherent in the domain itself as opposed to *ad-hoc* structures imposed on the domain for the sake of computational convenience.

4.2 Causality and Independence

The twin notions of causality and independence are central to belief network approaches to uncertainty management. Essentially probabilistic dependencies in the world, identified by the existence of direct causal relationships between propositions, are represented explicitly in the form of a network. This network serves both as a model of the domain and as a mechanism for reasoning within the domain. This mechanism is derived from the rules of probability theory and the independence relationships represented in the network.

The identification of independencies in the domain to be modelled is essential for the effective application of belief networks. By using these independencies it is possible to express the joint distribution function over all states of all propositions in the domain in terms of local conditional probabilities between neighbouring propositions in the network. This greatly reduces the number of probabilities that have to be directly specified and leads to a simple and elegant method for the propagation of evidence between propositions.

Pearl refers to the links in his belief network as being causal links. This terminology has not been universally accepted by others as they argue that the notion of causality is a vague one. This is largely an issue of semantics and the links can be interpreted in a variety of ways depending on the application [Lauritzen & Spiegelhalter 90, Neapolitan 90]. The underlying principle of independence remains the common factor. The notion of causality is often useful in the analysis of a problem when establishing an appropriate network model. A representation based on causation is attractive from the point of view of expert systems development as people often find it convenient and natural to express relationships between events in causal terms. They are able to say that event A caused event B or that event C was irrelevant to events B's happening, and so on, even though they may not understand the mechanism underlying the causal connection. The sophistication of the concept of causality is belied by the apparent ease with which we use it. The ability to abstract from the intricacies of a relationship to a simple notion of causality is a powerful cognitive tool. At least part of its power derived from the observation that causality allows one to to reason not only from cause to effect, but also from effect to cause. This is particularly important when a diagnosis is to be made solely on the basis of some observed evidence, for instance if we know that measles causes spots, and we observe spots then we can identify measles as a possible cause of those spots. Our belief that a cause is or was present can therefore be increased by the observation of its effect, even though we may be unable to directly prove that the cause is or was present.

Because the notion of causality is so fundamental to the way people experience the world, it is an ideal language to communicate knowledge both to and from an expert system. This communication is further enhanced if the internal representation of the expert system is itself expressed in terms of causal relationships. Nowhere is this more ably demonstrated than in the process of knowledge acquisition. It is generally acknowledged that people are poor estimators of probabilities [Tversky & Kahneman 90a], but they are typically able to identify independence between propositions and causal orderings among propositions.

4.3 Belief Networks

Belief networks use the notion of causality as the primary relationship in both the design of the knowledge base and in the implementation of its reasoning mechanisms, which makes them an ideal tool for expert systems.

A belief network represents propositions as nodes in a network. Each node, for example *measles*, consists of a number of states, for example $\{present, absent\}$. These states are exhaustive (there are no other states of *measles* that lie outwith the states encompassed by *present* and *absent*) and mutually exclusive. This is not to suggest that in the measles case we can only state that *measles* is *present* or *absent*, it merely means that the probability of *measles* must sum to one over the states *present* and *absent*, so if *measles* was suspected then $P(measles = present)$ might equal 0.7 in which case $P(measles = absent) = 0.3$. One of the criticisms levelled at probability theory is the relationship between the probability of a proposition and the negation of that proposition.

There is no requirement that a node have binary states, so we could redefine the states of *measles* to be $\{severe, moderate, mild, absent\}$, the probability at the node must still sum to one across the states. We can then introduce some evidence nodes — that is symptoms that could be caused by *measles*. We will consider *fever* and *spots*. In a belief network causality is represented by a directed link from cause to effect, as shown in figure 4.1. Even this very simple network can be used to illustrate some of

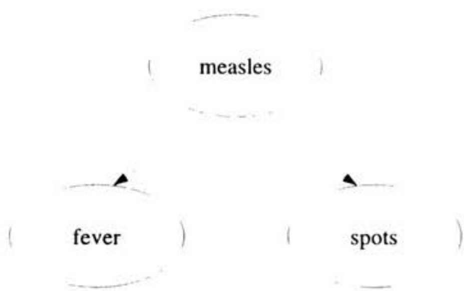


Figure 4.1: Diverging causal influences

the implications of causality. If we know nothing other than that *spots* are present this will increase our belief that *fever* will also be found, as the presence of *spots* is evidence for the presence of *measles* which in turn is known to cause *fever*. If, however we knew that *measles* was present, then the presence or absence of *spots* has no bearing upon our belief in *fever*. This relationship is termed *conditional independence*, *spots* and *fever* are conditionally independent given *measles*. This causal influence is described as *diverging*.

Suppose we introduce another disease into our model, *flu*, which is known to cause *fever*. This can be incorporated into our network, shown in figure 4.2. We now have competing potential causes: if *fever* is observed, it could be due to *flu* or *measles*. If we observe *fever* then our belief in both *flu* and *measles* will increase. However, if we observe both *fever* and *spots* then our belief in *measles* will increase, but our belief in

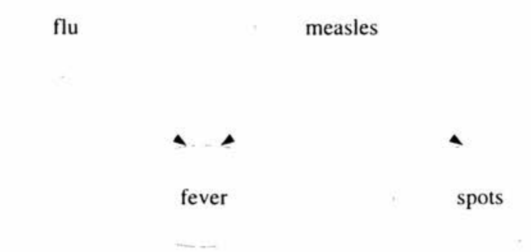


Figure 4.2: Diverging and converging causal influences

flu will decrease. *Measles* provides an explanation for the observed *fever* and *explains away* the observation, making *flu* less likely. When nothing is known about *fever*, *flu* and *measles* are *independent*, the occurrence of *spots* by itself should not change our belief in *flu*. When evidence about *fever* is available then the notion of explaining away becomes relevant. We say that *flu* and *measles* are *conditionally dependent* given *fever*. The influence is described as *converging*.

Consider the model shown in figure 4.3, where *fever* is a cause of *sweating*. If *flu* is

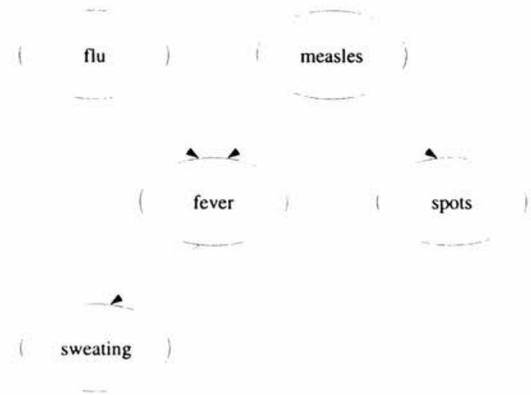


Figure 4.3: Diverging, converging and linear causal influences

known to be present then our belief in *sweating* will be increased as *fever* causes *sweating* and *flu* is a cause of *fever*. If the state of *fever* is known with certainty then *flu* and *sweating* are separated and our belief in *flu* cannot influence our belief in *sweating* and vice versa. *Flu* and *sweating* are conditionally independent given *fever*. The influence is described as linear. It should be noted that this holds true *only* if the state of *fever* is known with certainty. This is also the case with conditional independence between diverging causal influences. Conditional dependence in converging causal influences holds true in the absence of certain knowledge.

These relationships are the basic building blocks of belief networks and are formalised in the work of Kim and Pearl [Kim & Pearl 83]. The proof that a directed acyclic graph can properly embody independence relationships is complex and is not presented in full here. The interested reader is referred to Pearl [Pearl 88b, chapter 3].

Pearl identifies five axiomatic conditions that must be satisfied by the relation "**X** is independent of **Y** given **Z**", where **X**, **Y** and **Z** are three disjoint sets of variables taken from a finite set of discrete random variables **U**. These axioms are labelled *symmetry, decomposition, weak union, contraction* and *intersection* [1]. The intuitive interpretations given to these axioms by Pearl, illustrated in figure 4.4 (where unshaded sets are independent of each other if they are separated by a given shaded set), are as follows:

Symmetry: if **X** is independent of **Y** given **Z**, then **Y** is independent of **X** given **Z**.

Decomposition: if **X** is independent of **Y** and **W** together given **Z**, then **X** is independent of **Y** and **W** individually given **Z**.

Weak Union: if **X** is independent of **Y** given **Z**, and **X** is independent of **W** given **Z**, then **X** is independent of **Y** given **W** and **Z** together.

Contraction: if **X** is independent of **W** given **Y** and **Z** together, and **X** is independent of **Y** given **Z**, then **X** is independent of **W** given **Z**.

Intersection: if **X** is independent of **Y** when **W** is held constant, and **X** is independent of **W** when **Y** is held constant, then **X** is independent of **Y** and **W** individually and **Y** and **W** together.

From these axioms Pearl shows that directed acyclic graphs (DAGs) can embody independence relationships based on the criterion of *d-separation.* If **X**, **Y** and **Z** are three disjoint subsets of nodes from a DAG, then **X** is independent of **Y** given **Z** if **Z** *d-separates* **X** from **Y**. This is true if for every node **n** on every path from a node in **X** to a node in **Y** the following conditions hold (following the explanation given by Charniak

[1] The *intersection* axiom only holds under certain conditions.

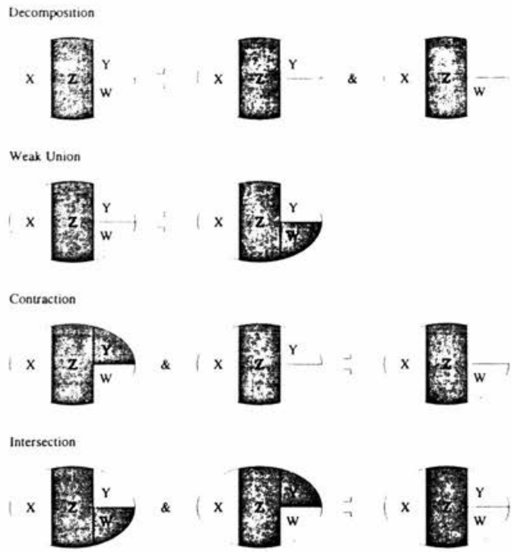


Figure 4.4: Graphical interpretation of axioms [Pearl 88b, page 86]

[Charniak 91]):

1. The relationship is linear or diverging and $\mathbf{n}$ is a member of $\mathbf{Z}$.

2. The relationship is converging and neither $\mathbf{n}$ nor any of its descendants are in $\mathbf{Z}$.

For example, in figure 4.5, where the paths indicate probabilistic dependence and the arrows the direction of causality, if $\mathbf{X} = \{2\}$ and $\mathbf{Y} = \{3\}$ then they are d-separated by $\mathbf{Z} = \{1\}$ as the path $2 \leftarrow 1 \rightarrow 3$ is blocked by $1 \in \mathbf{Z}$, and the path $2 \rightarrow 4 \leftarrow 3$ is blocked because 4 and all its descendants are outwith $\mathbf{Z}$. If $\mathbf{Z} = \{1, 5\}$ then path $2 \rightarrow 4 \leftarrow 3$ would become active.

A belief network uses a DAG as its underlying representation, with the links representing direct causal relationships. So far we have made no mention of the strength of a causal link, for instance *measles* may cause *fever* only rarely, whereas *flu* may often cause it. Given that our network model shows precisely which relationships must be considered, *e.g.* *flu* and *fever*, and those which can be ignored *e.g.* *flu* and *spots*, we can see that a considerable reduction in the number of probabilities that must be specified

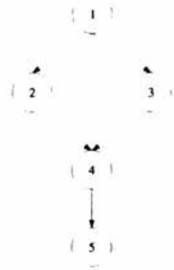


Figure 4.5: A DAG illustrating d-separation [Pearl 88b, page 118]

has been achieved.

The links are quantified with conditional probability matrices that specify the prob abilistic relationship between the states of the child given the states of the parent. Each node can calculate the belief across its states by combining the causal and diagnostic evidence from its direct neighbours and the matrices that specify the relationship. Root nodes of the network require prior probabilities in order to complete the probabilistic model.

Having specified both the qualitative network structure and the quantitative prob abilistic relationships, the network can now be used for inference. Consider again our simple network model shown in figure 4.3. Evidence of *sweating* will be propagated to *fever*, which must then propagate its updated belief to its direct neighbours. Clearly the message propagated to *sweating* should not result in any change in belief at *sweating* as no new evidence has been made available. The messages to *flu* and *measles* should similarly not reflect any evidence back to the nodes they came from. The local propagation from a node to its direct neighbours, whenever it receives a message from one of those neighbours (or a direct observation is made), can be performed asynchronously and will result in a consistent, correct probability distribution across the network in a time proportional to the length of the longest path. The only restriction is that there be no cyclic paths in the network.

4.4 Extensions

Thus far we have described a theoretically sound inference mechanism that can be applied to a restricted class of network model. Whilst the generation of sound probabilistic inferences is undoubtedly of value, we have shown in Chapter 2 that this is not in itself sufficient for a medical expert system. Belief networks are relatively new on the expert systems scene and so many of the important design issues are still to be resolved. In some areas, such as widening the class of network models to which the inference mechanisms can be applied, there has been a great deal of both interest and progress. In the following sections we examine trends in belief network development with reference to those facilities that will be required of a medical expert system.

4.5 Inference Algorithms

The inference algorithm described earlier placed severe restrictions on the structure of the underlying network. There are three important classes of belief network structure, illustrated in figure 4.6:

1. Singly connected — where there is at most one path between any two nodes in the network.

2. Multiply connected — where there is more than one path between a pair of nodes within the network, but no node has a *directed* path to itself.

3. Cyclic — where a node has a directed path to itself.

Cyclic networks are generally accepted as falling outwith the domain of application of belief networks [Pearl 88b, page 195], but at least one application has included cycles in the model, resolving them via heuristic methods [Long 89] (see page 218). Our work, presented in Chapter 7, also uses cyclic networks.

The algorithms for probabilistic inference so far described are only mathematically correct when applied to singly connected networks. As many real world domains cannot adequately be modelled within the singly connected restriction, algorithms are required for the more complex multiply connected networks.

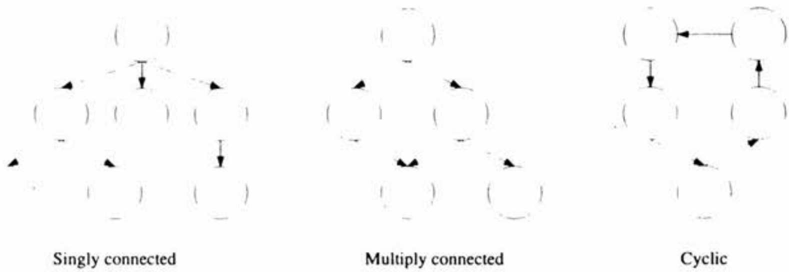


Figure 4.6: Three classes of network topology

However, probabilistic inference in arbitrary belief networks has been shown to be NP-hard (no general purpose, polynomial time algorithm exists) [Cooper 90]. This suggests that research efforts should be directed towards the development of approximate inference algorithms and more particularly, special-case algorithms, as the general task of approximating inference in a belief network has also been shown to be NP-hard [Dagum & Luby 93].

Due partly to the prevalence of domains requiring multiply connected networks, and partially due to efficiency considerations, almost all complex belief network applications rely on either a clique-tree or stochastic simulation method for inference. Stochastic simulation methods are also attractive from the point of view of resource bounded computation as they produce partial solutions.

4.5.1 Multiply Connected Networks.

A multiply connected network is one in which there are cycles in the underlying, undirected network, *i.e.* there is more than one path between a particular pair of nodes. These are distinct from cyclic networks in which a node can be one of its own causal predecessors, *i.e.* there is a directed causal path between the node and itself. Undirected cycles in multiply connected networks are referred to as loops. A network containing loops causes problems on two accounts:

Computational problem — in a communication scheme based on local computations it is possible that a message could circle indefinitely around a loop.

Consistency problem — the Bayesian foundation of belief network algorithms requires that each item of evidence be assimilated at a node only once. By permitting loops we are providing alternative routes for an item of data to reach a node on that loop (double counting). As a result of this, the calculations performed within the network can lead to incorrect probability assignments at the nodes.

In some cases it may be possible to design out the loops at the network specification level by simplifying the underlying model. Often this will not be possible without compromising the representational power of the model. It should be remembered that any loops in the causal network are reflecting complexity that is inherent in the domain, they are not a creation of causal network models *per se*.

Ideally any solution to this problem should attempt to retain the power of the existing algorithms, *i.e.* local computations and message passing mechanism based on network structure.

Three methods will be considered here; clustering; conditioning; and stochastic simulation. The first two of these methods break, or simulate the breaking of, loops within the network, rendering it singly connected and allowing the existing algorithm to be applied without modification. Clustering can also be used as a probabilistic inference mechanism using algorithms developed by Lauritzen and Spiegelhalter [Lauritzen & Spiegelhalter 90]. Stochastic simulation uses the network to generate possible world states, the frequency of occurrence of events can then be used to estimate probabilities of those events.

4.5.2 Clustering/Clique-tree Propagation.

Clustering is a method whereby nodes in the original belief network are combined together into clusters, forming a new network devoid of loops. For a given belief network there may be many possible sets of clusters that could be formed. The most popular method for generating clusters is based on junction trees [Olesen *et al* 89, Jensen *et al* 90c, Lauritzen & Spiegelhalter 90, Andreassen *et al* 91b].

A junction tree is formed by the following steps, as illustrated in figure 4.7:

1. Form the "moral" graph. This is done by inserting a link between all parents in the network with a common child, and dropping the directions from the links.

2. Triangulate the moral graph. All cycles of length greater than 3 are broken by the insertion of new links.

3. Identify cliques in triangulated graph (a clique is a maximal set of nodes all of which are pairwise linked)

4. Introduce links between the cliques such that a tree is formed with the property that for all pairs of cliques that contain a common set of nodes, each clique on the unique path between them also contains that set of nodes.

Once the junction tree is formed it can be used to propagate inference. There are several ways that this can be achieved, for example using Pearl's algorithms [Pearl 88b], Lauritzen and Spiegelhalter's [Lauritzen & Spiegelhalter 90], or those of Jensen et al [Andreassen *et al* 87, Jensen *et al* 90a, Jensen *et al* 90c].

As an example clique-tree propagation algorithm, consider that described by Andreassen *et al* [Andreassen *et al* 91b]. A simple, four node network shown on the left of figure 4.8 is converted into the junction tree shown on the right. The junction tree contains two cliques, C1 and C2, and a *separation set*, S that contains the nodes that are common to the cliques it separates. An initial belief table containing the joint probability distribution over the states of the nodes in each clique is calculated given the prior probabilities and the conditional probability tables, *i.e.* $BEL(C1) = P(Flu) \cdot P(Throat\ Infection) \cdot P(Fever \mid Flu,\ Throat\ Infection)$. The initial belief table for S can be calculated as the marginal distribution of either clique, *e.g.* $BEL(S) = \sum_{Fever} BEL(C1)$. The junction tree is now *consistent* and the probability of a node can be calculated by marginalising over any clique or separation set containing that node, *e.g.* $P(Fever) = \sum_{Flu,Throat\ Infection} BEL(C1)$.

If Fever is observed, the belief table of C1 is multiplied by the evidence vector and the resulting belief table is normalised. To propagate this evidence to C2 a new separation set belief table, S′ is calculated by marginalisation of C1 with respect to Fever. The belief table of C2 is updated by multiplying it by the ratio S′/S. At this point the network is again consistent. Full details of this particular mechanism are given by Jensen *et al*

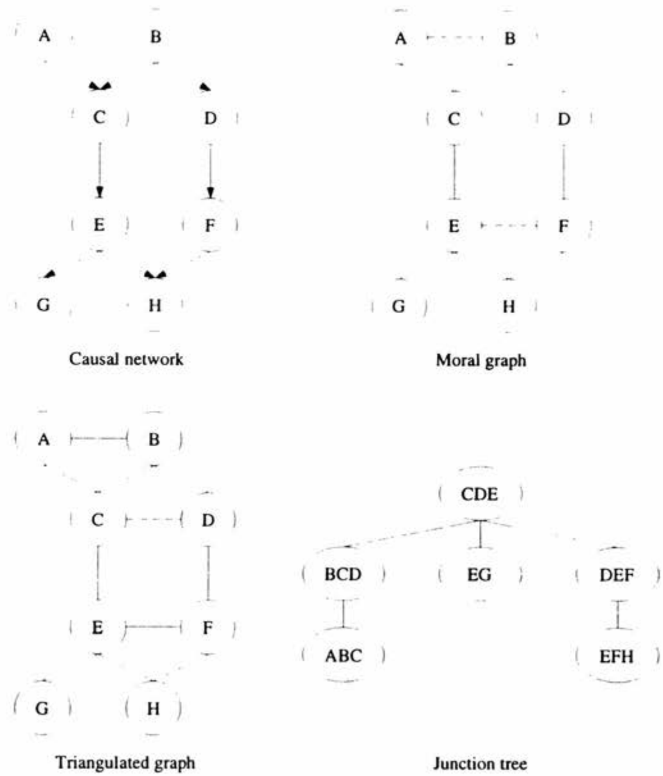


Figure 4.7: The creation of a junction tree [Olesen *et al* 89, page 393]

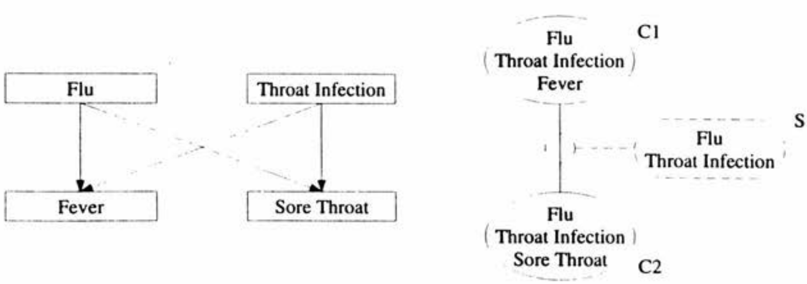


Figure 4.8: An example network and junction tree [Andreassen *et al* 91b]

[Jensen *et al* 90a, Jensen *et al* 90c].

The critical step in this method is the triangulation of the moral graph as the computational efficiency of the resultant junction tree depends on the quality of the triangulation. Different triangulations of the same graph can result in different numbers of cliques and in cliques of different sizes. Efficient computation requires a triangulation that results in the minimum overall table size. The table size for a clique is the product of the number of states in the nodes of the clique. Therefore the size of the table grows exponentially with the size of the clique [Olesen *et al* 89, page 310].

It has been proved that finding the triangulation with the minimum fill-in (that which requires the minimum number of new links to be added) is NP-complete. Olesen et al [Olesen *et al* 89] have conjectured that finding the triangulation with the minimum total table size is also NP-complete. Researchers have come up with a variety of heuristic techniques which result in minimal triangulations [Kjaerulff 90] and other methods for node aggregation [Chang & Fung 89].

4.5.3 Conditioning.

This method is based on conditioning the values of a set of nodes called the *loop cutset*. Conditioning a node simply involves fixing the value of that node such that one of its states is certain. The loop cutset is a set of nodes such that conditioning the nodes renders the network functionally singly connected. This is possible because of the blocking conditions that are inherent in the belief network model. The blocking conditions arise from the underlying formalism:

1. A fixed value node (*i.e.* a conditioned one) does not send information from its children to its parents or from its parents to its children.
2. A fixed value node does not send information from one child to another
3. A node whose value is not fixed and that does not have any fixed value descendants does not send information from one of its parents to any other parent.

If the value of a cutset node for a particular loop is fixed, then because of the blocking conditions, that loop will be rendered singly connected and the the problems outlined earlier will no longer apply. In order to be able to apply Pearl's algorithm to a multiply connected network we must satisfy the loop cutset condition, "Instantiate at least one node from every loop in the belief network such that this node is child to no more than one other node in the same loop" [Suermondt & Cooper 90].

Each unique set of conditioning values for the cutset nodes can be treated as a distinct singly connected network and beliefs can be calculated using standard propagation algorithms. The individual results must be weighted by the joint probability of the particular conditioning values used, and then summed to obtain the total belief. The joint probability of the conditioning values can be obtained during initialisation, and updated as further evidence is collected [Suermondt & Cooper 89].

Because each possible combination of conditioning values for nodes in the loop cutset must be considered, it is important to find a minimal loop cutset, defined as "the set of nodes satisfying the requirements of the method of conditioning such that the product of the number of values in these nodes is minimal" [Suermondt & Cooper 90]. Finding a minimal loop cutset is is NP-hard and heuristic methods are used to find near minimal cutsets [Stillman 91].

As each possible set of conditioning values must be processed, the task could be executed in parallel, with each set represented by a separate network. Conditioning is a method that is suitable for networks which are not highly connected and which therefore have a small loop cutset. With increasing cutset size the method rapidly becomes impractical as the computational time is exponential relative to the number of nodes in the cutset.

4.5.4 Stochastic Simulation.

Stochastic simulation (or Monte Carlo Methods) uses the belief network as a model from which possible scenarios can be generated. The probability of a particular set of events occurring can be determined from the fraction of randomly generated scenarios that contain that set of events.

Pearl [Pearl 86b, Pearl 88b], has developed a two phase stochastic simulation method in order to efficiently handle cases where the values of some nodes are known prior to the simulation.

Given a belief network:

1. Fix observed nodes to their known states.

2. Assign arbitrary initial states to unobserved variables.

3. Let each variable in succession choose another state in accordance with the conditional probability of that variable given the state of the other variables in its neighbourhood (defined below).

4. Once each non-fixed node has calculated a state for itself, a complete coherent scenario has been generated.

5. Repeat from 2 until sufficient scenarios have been generated.

The probability of a given value at a node can be calculated either as the fraction of times the value occurred during simulation, or by taking the average of the conditional probabilities computed for the occurrence of that value during the simulation. The second of these normally yields faster convergence.

Due to the blocking conditions mentioned earlier, the neighbourhood of nodes that need to be considered when performing step 3 (its *Markov blanket*), consists of those variables who, once their states were known, would render the node under examination independent of the rest of the network.

The Markov blanket of a node X consists of:

1. The parents of X.

2. The children of X.

3. The parents of those children of X that are neither parents nor children of X themselves.

The probabilities of a node conditioned on the states of its Markov blanket neighbours can be calculated from the conditional probability matrices in the belief network and the values of the instantiated neighbours. Typically many runs will be required before the probabilities generated from the simulation converge on the true probabilities. Each simulation run is composed of only N + L steps (where L is the number of links and

N the number of nodes) and computation time is determined mainly by the degree of accuracy required. This property makes stochastic simulation ideal in situations with highly connected networks where an estimate of the probabilities is sufficient. Simulation approaches must also ensure that the distributions generated correspond closely to the true distribution and that a large enough set of trials is conducted so that sampling errors are avoided [Chavez & Cooper 90].

4.5.5 Discussion of Inference Algorithms

Of the three methods for avoiding the problem of loops in belief networks there is none that stands out as the perfect solution. Both conditioning and clustering rely on heuristic algorithms which can result in inefficient solutions. Stochastic simulation avoids this problem but requires a large number of computations in order to converge on the true probabilities. In a domain in which new information is continually arriving, stochastic simulation will require an even larger number of simulations.

The choice of method clearly depends on the application domain which determines the resource constraints and the required accuracy of the results. In some domains a combination of methods may provide a better solution than any individual method. A combination of conditioning and clique-tree propagation has been used with success [Suermondt *et al* 91], due to the specific network structure. It may be that certain network structures will prove common across domains and that a set of standard approaches may be appropriate.

4.6 Network Modelling

The most significant development in network models has been the move to multiply connected networks, discussed in section 4.5. There have also been a number of relatively minor developments which influence current network modelling.

It was recognised early in the development of belief networks, that the specification of conditional probability matrices was potentially a major burden in network modelling. This led to the introduction of canonical models of multicausal interactions [Pearl 88b, page 184]. The most popular of these has proved to be the *noisy-OR gate*, which can be found in many applications. The *noisy-OR* is a probabilistic generalisation of the standard Boolean OR, each cause having an independent probability of being sufficient to cause the effect [Horvitz *et al* 88, page 275]. The use of the *noisy-AND gate* is reported less frequently (for example in BaRT [Booker *et al* 90]), perhaps suggesting that the *noisy OR* is more general, or the current application domains have a common underlying causal structure. Several variants of the *noisy-OR gate* can be found in the literature, probably the most sophisticated in modelling terms is the *generalised noisy-OR gate* [Henrion 89, Diez & Mira 94]. This model can include *leak probabilities* which allow for the probability of an effect occurring in the absence of any modelled cause. It can also include multiple states across the effect variable. The use of canonical models greatly reduces the number of probability assessments required when model building.

Another frequently reported approximation for the reduction of network model complexity, is the use of two state nodes. The question of the granularity of model representations is indicative of the general need to compromise between detail and tractability. Traditionally the granularity of the network was assumed to be static, but more recent work has focussed on the possibility of dynamically adjusting the granularity of nodes [Chang & Fung 91, Provan & Clarke 93]. It is suggested that the opportunistic refinement and coarsening of node states could improve both the accuracy and efficiency of the inference procedure. A role is also posited in the knowledge acquisition task.

One of the limiting factors in belief network models is the use of nodes with discrete states when modelling continuous variables. The most common approach to this problem is to divide the continuous range into discrete regions, typically grouping those parts of the range that are identical with respect to inference or decision making. A propagation scheme solely for continuous variables, under limiting assumptions, is presented by Pearl [Pearl 88b, pages 344–357] and a scheme for mixed networks, under similar limitations, is discussed by Olesen [Olesen 93].

There is little research into temporal reasoning with belief networks [Berzuini 90, page 16]. Most of the current approaches rely on the implicit inclusion of temporal relationships within the network structure. The most popular form of this is *network duplication*, in which the temporal axis of the problem is discretised into a number of time-slices or intervals. Each of these time-slices is represented by a distinct portion of the belief network. The results of inference at a particular time-slice Δ^t are influenced only by those of time-slice Δ^{t-1} and influence those of time-slice Δ^{t+1}. In some applications the time-slice networks are identical, for example in SWAN (see page 246) as illustrated in figure 4.9. In SWAN each time-slice contains a complete network that models glucose metabolism over a period of one hour. The results of this network are then used as inputs into an identical network representing the subsequent time-slice hour. Networks can be linked together in this way to provide a model covering a period of several hours. In other

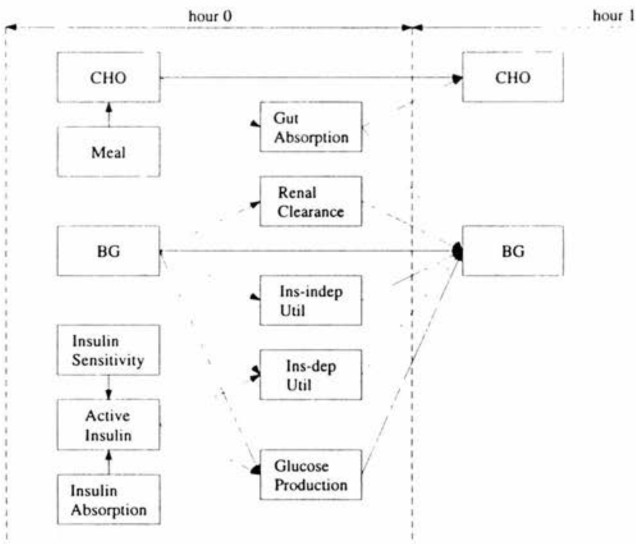


Figure 4.9: SWAN — time-slice model for glucose metabolism [Andreassen 94, page 106]

systems the discretisation is defined not with respect to the temporal axis, but according to the occurrence of events, as illustrated in figure 4.10. In these event-sliced networks, partial duplication is supplemented by nodes representing the events by which the event-slice is defined and events specific to that event-slice. Clearly these discrete network duplication models will not be applicable in all domains. Typically this approach assumes that the evidence is constant over the slice and that slices have definite transition points [Provan & Clarke 93]. An alternative to the static duplication of network structure is a dynamic approach. In the DYNASTY system [Provan & Clarke 93] influence diagrams are dynamically created, replaced and refined over time. This is potentially more efficient than wholesale duplication.

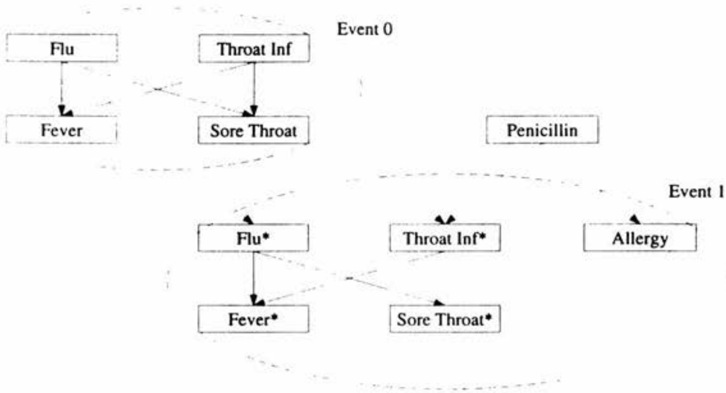


Figure 4.10: Quark — event-slice model for penicillin treatment
[Andreassen *et al* 91b, page 6]

Work on continuous-time representations is presented by Berzuini [Berzuini 90]. Two methods are discussed, the first includes nodes in the network that represent 'event-occurrence times', and nodes that represent the 'state' of the system at these times. The second method considers interactions between events occurring at certain 'dates'. Some relationships (causal, inhibitive and competitive) are defined between dates and can be coherently embedded within a belief network.

4.7 Network Creation

One of the criticisms often levelled against belief networks is their reliance on subjective probabilities. Regardless of concerns over the use of subjective probabilities, belief networks typically have to rely on them as the objective probabilities are unavailable or inaccessible. Several different approaches for improving the probability distributions used in network models have been investigated.

Some research into the generation of belief networks directly from databases of statistics has been conducted [Pearl & Dechter 89, Cooper & Herskovits 92], particularly in the areas of ranking belief network structures according to their posterior probability relative to a database of examples, and deriving numerical probabilities from a database, given the belief network structure. The ALARM system (see page 214) containing 46 arcs and 37 nodes of between two and four states has been used to test the K2 algorithm [Cooper & Herskovits 92]. The ALARM network was used to generate 10,000 example cases used by K2, along with a partial ordering among the nodes, to create a new network. The network created by K2 was identical to the ALARM network except for one missing and one additional link. Whilst such approaches are still in the developmental stage, the ability to generate belief networks directly from data would be a powerful one.

Testing and refinement of the network after creation is another option, typically relying on experts to provide the initial structural and probabilistic model and using subsequent data to refine the model [Sucar *et al* 93]. This relies on the availability of monitors to test the model during use. Some are presented by Cowell *et al* [Cowell *et al* 93b]. These monitors can then be used to refine the structural and/or probabilistic models [Spiegelhalter & Lauritzen 90a, Spiegelhalter & Lauritzen 90b, Spiegelhalter & Cowell 92, Cowell *et al* 93b, Sucar & Gillies 94].

Other research has concentrated on improving the way in which the models are obtained from the experts. Several tools for creating or eliciting decision structures have been developed, including work described by Leal and Pearl [Leal & Pearl 77], KNET [Chavez & Cooper 88], DAVID [Shachter 88], BaRT [Booker *et al* 90] and GAMEES [Bellazzi *et al* 91a]. A method of decomposing the task of constructing a belief network into the construction of small locally defined networks, based on the notion of *similarity networks* is described by Heckerman [Heckerman 90b].

4.8 Decision Making

Decision theory is a normative (though not descriptive) method for decision making under uncertainty [North 90]. It is based jointly on the axioms of probability theory and utility theory. It formalises the relationship between *preferences*, relative valuations for possible world states, and *decisions*, irrevocable allocations of resources. The axioms of utility theory are [Horvitz *et al* 88]:

Orderability — all outcomes can be compared, given outcomes x and y, a decision maker either prefers outcome x to y, y to x, or is indifferent.

Transitivity — if a decision maker prefers x to y and prefers y to z then they also prefer x to z.

Monotonicity — given two lotteries with the same outcomes but different probabilities, a decision maker chooses the lottery with the higher probability of the preferred outcome.

Decomposability — a decision maker is indifferent between two lotteries that have the same outcomes with the same probabilities, regardless of differences in the internal structure of the lotteries.

Substitutability — if a decision maker is indifferent between a lottery and some certain outcome, then substituting one for the other as an outcome in another more complex lottery will not affect the decision maker's preference for that lottery.

Continuity — if a decision maker prefers outcome x to y and y to z, then there is some probability p for which the decision maker is indifferent between getting y for certain and a lottery with a chance p of getting x, the most preferred outcome, and a $(1-p)$ chance of getting z, the least preferred.

The axioms define a scalar utility function $U(x, d)$ which assigns a number on a cardinal scale to each outcome x and decision d, indicating its relative desirability. When x is uncertain the preferred decisions are those that maximise the expected utility over the probability distribution for x. Decision theory essentially states that given a set of preference utilities, a set of probability distributions denoting beliefs, and a set of decision alternatives, the decision maker should choose the course of action that maximises the expected utility [Horvitz *et al* 88].

Belief networks and decision theory come together in the influence diagram approach to decision modeling. An influence diagram is a graphical, network based representation for decision making, as illustrated in figure 4.11.

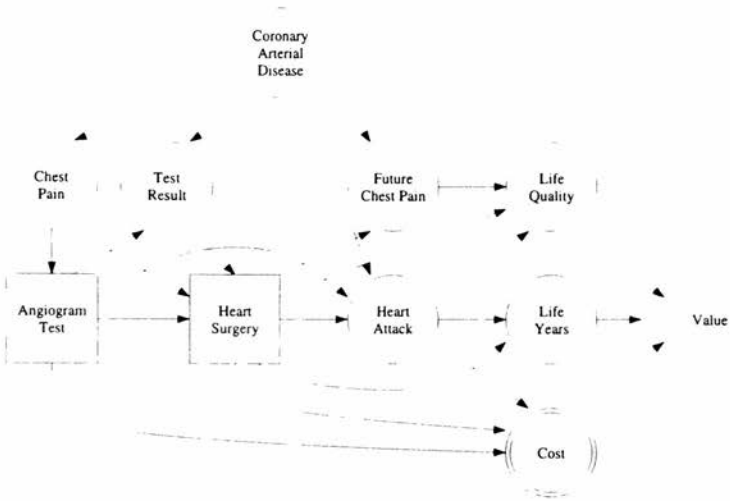


Figure 4.11: Influence diagram for a patient with heart disease
[Henrion *et al* 91, page 72]

The node types in an influence diagram are:

Decision nodes — rectangular nodes representing the actions available to the decision maker.

Chance nodes — circular or oval nodes representing stochastic chance (single line) and deterministic chance (double line). A stochastic node's outcome is a probabilistic function of its predecessors. A deterministic node's outcome is determined with certainty by its predecessors.

Value nodes — diamond nodes representing the decision maker's preferences.

A belief network is a special-case influence diagram containing only chance nodes. This relationship between influence diagrams and belief networks allows the use of decision theory as a decision making mechanism for belief networks. Many areas of research into influence diagrams are applicable to belief networks, particularly those that embody decision problems. Interesting topics include generating explanations of influence diagrams [Langlotz *et al* 88b, Langlotz & Shortliffe 90, Jimison 90, Jimison *et al* 92], tools for automatic and manual influence diagram construction [Sonnenberg *et al* 94] (see also section 4.7), and the use of belief networks as influence diagrams [Cooper 88].

4.9 Control

There are two main aspects of control in belief networks, the specification of the network model and the propagation of inference. The mathematical foundation of belief networks means that control decisions can be taken and implemented in a principled manner, in particular:

- A variety of exact and approximate inference mechanisms are available.
- Many approximate inference mechanisms produce partial results.
- Networks can be dynamically refined and coarsened.
- Networks can be created incrementally.
- Networks for classes of problems can be pregenerated and stored.
- The networks can contain influence diagrams.
- The results of network inference are sound for decision making purposes.
- Different types of inference are appropriate for different network models.

This gives belief networks a great deal of flexibility with respect to the control that can be exerted.

The Protos system (see page 47) has been adapted to perform metareasoning about the use of belief networks to solve problems under resource constraints [Horvitz 89, Horvitz *et al* 89a, Horvitz *et al* 89b, Horvitz 90]. Belief networks are ideal inference mechanisms for the Protos approach for the reasons given above.

The BaRT system [Musman *et al* 90] uses a hierarchical taxonomy of belief networks in order to prune the search space. The transition between network models within the taxonomy is made on the basis of the results of inference at the current level. Evidence gathering is prioritised on the basis of its effects on the belief of a dynamic target node.

The IDEAL system [Breese & Horvitz 91] considers the task of reformulating belief networks into clique-trees under different value functions, figure 4.12. Evaluation using

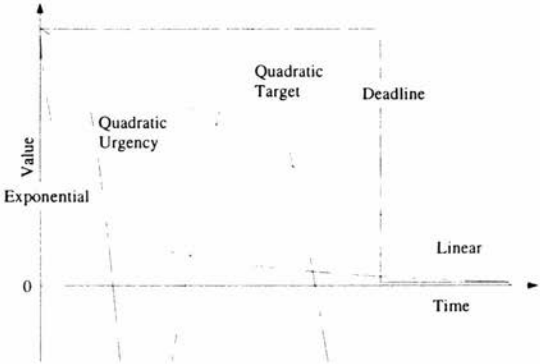


Figure 4.12: Value/time functions used in the evaluation of IDEAL
[Breese & Horvitz 91, page 141]

these different value functions, showed that under certain circumstances, time spent on forming alternative clique-trees was more valuable than simply opting to use the first clique-tree generated. The authors suggest that the kind of techniques they develop could lead to optimal control of the dynamic construction and solution of belief networks.

The ADRIES/SUCCESSOR system [Levitt 88, Binford *et al* 89, Levitt *et al* 90a, Levitt *et al* 90b] controls the dynamic creation of influence diagrams for efficient, opportunistic reasoning and decision making. A similar approach is adopted in the DYNASTY system [Provan & Clarke 93], which dynamically constructs pruned influence diagrams over time using sensitivity analysis. This sensitivity is defined with respect to decision equivalence, the degree to which the utilities of actions differ between models of adjacent time intervals. It also makes use of refinement and coarsening of nodes states, which is discussed in greater detail by Chang and Fung [Chang & Fung 91].

Network pruning can be applied in cases where it is possible to identify a subset of nodes of interest, the query nodes, Q. Pruning the network involves finding the smallest subgraphs that are *computationally equivalent* with respect to Q, given a particular set of evidence. The condition for computational equivalence is that the probabilities calculated at Q from the subgraphs are identical to those calculated from the complete network. A method based on d-separation (see page 69) and the recursive pruning of leaf nodes has been proposed [Baker & Boult 91] and is illustrated in figure 4.13. Pruning using d-separation involves finding the set of nodes d-separated from the query nodes by the evidence nodes, providing the evidence is known with certainty. Leaf nodes without evidence, so called *barren nodes*, can be recursively removed without altering the computation at the query nodes, regardless of whether or not they are d-separated from the query nodes. Propagation through these subgraphs is both necessary and sufficient

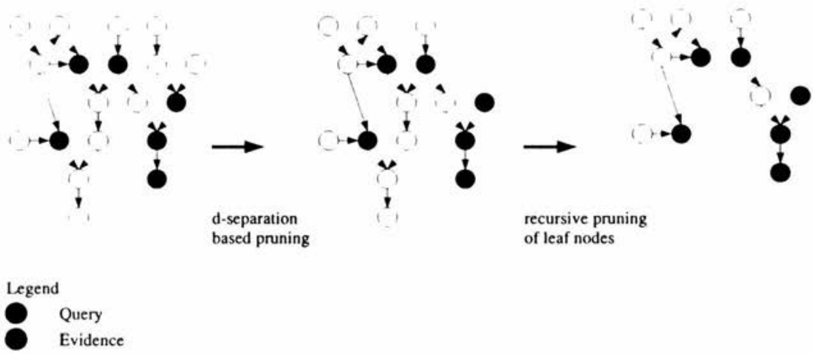


Figure 4.13: Pruning networks for efficient computation [Baker & Boult 91, page 227]

for the calculation of the query node probabilities, as the subgraphs are minimal with respect to topological criteria. This method will be less efficient when evidence arrives incrementally, unless efficient procedures for restoring pruned nodes to the subgraphs can be devised.

A method conceptually related to network pruning is relevance-based propagation [Pinto 86] in which the decision of whether or not to propagate a piece of evidence through the network is based on its effect with respect to reducing uncertainty at some pre-defined target node. This measure also allows priorities to be assigned to evidence nodes and decisions to be made regarding the halting of evidence collection.

Somewhat further removed from the notion of pruning is that of precomputation [Herskovits & Cooper 91a]. In situations where a small number of typical cases (sets of findings) account for a large proportion of the expected use of the system, efficiency can be improved by precomputing the results of these cases. When a query is made of the network, if its details match those of a precomputed case, the result of that case can be returned without further computation. Precomputation would be most applicable in situations where standard sets of evidence were used in a time-critical application, for example in patient monitoring. The selection of cases for precomputation can be made according to a number of criteria, such as expected evidence patterns, utility, or combinations of criteria. A similar approach based on the incremental construction of an instance of a prestored class of probabilistic model, specialised by parameters describing the current problem, has been proposed [Goldman & Charniak 91, Goldman & Charniak 93].

Many of the above methods assume that the network contains a small set of nodes in which the user is interested, and a larger number of nodes in which, aside from their influence on inference, the user is disinterested. This assumption suggests that such methods will be of particular use in medical applications where the number of diagnostic nodes is generally small relative to the evidence and intermediate nodes. However, it should also be noted that in the medical domain it is often vitally important to identify extremely rare cases and any form of pruning must ensure that rare events are not removed on the basis of their low prior probabilities.

4.10 Interfaces

Although belief networks are being adopted into mainstream expert systems, the majority of the work is still of a developmental nature. Perhaps as a result of this little consideration has been paid to the design of interfaces for belief networks. In cases where the belief network forms only a part of the expert system, the problem of interface design for belief networks will be subsumed by the general task of interface design. Those belief networks that have had some interface developed point towards a general approach that can be expected to underlie many future belief network interfaces. The general approach has two main guiding principles:

- The network structure is meaningful in domain terms and should therefore form the qualitative aspect of the interface.

- The network structure reflects probabilistic, causal relationships, therefore the quantitative information presented to the user will be based on the probabilistic and causal parameters in the network.

Assuming these principles hold true, and given the trends in graphical computing, belief network interfaces are likely to be graphics based, using a representation of the belief network as an interactive map of the domain.

MUNIN (see page 231) communicates with the user through what its authors describe as a spread-sheet like graphical interface. This interface uses the network model for its basic structure, with the nodes and links being represented graphically on the screen. Each node is represented by a box containing a list of its states. The belief at each node is represented as a bar chart across the states. When categorical evidence is entered into the network, either from the user or directly from an automatic system, it is represented by a broken bar, as illustrated in figure 4.14. The interface is also designed to provide

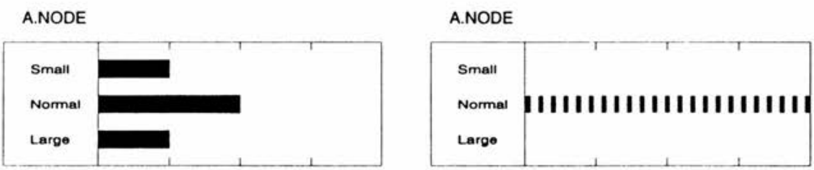


Figure 4.14: MUNIN — probabilistic and categorical node displays
[Jensen *et al* 87a, page 6]

some simple explanation facilities. One of these explanation facilities is illustrated in figure 4.15. It shows the evidence that impacts upon the focal node, MUP.CONCLUSION, from its direct neighbours. The evidence at these neighbours is expressed in terms of the states of the focal node. In this way the user can see the relative contributions of these neighbours. Another facility is illustrated in figure 4.16, in which the support that each state in MUP.CONCLUSION gets from each state of MU.STRUCTURE is shown. This allows the user to see the transformation of evidence between state sets. These

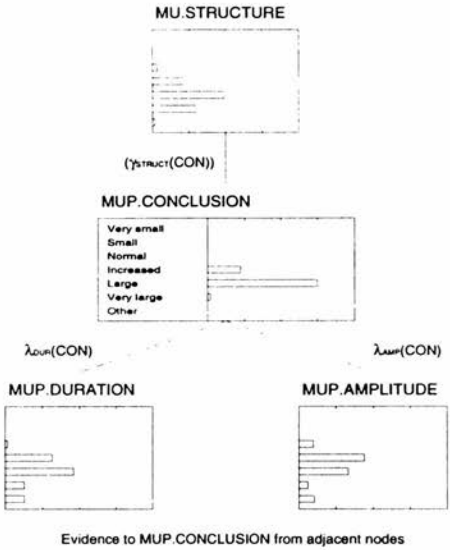


Figure 4.15: MUNIN — relative contributions of neighbouring nodes [Jensen *et al* 87a, page 8]

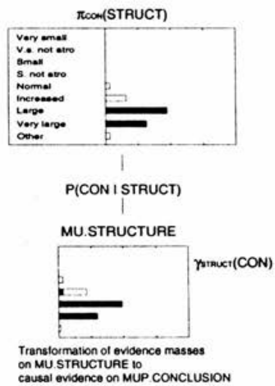


Figure 4.16: MUNIN — evidence transformation between state sets [Jensen *et al* 87a, page 8]

type of graphical tools allow the user to browse the network, examining the relationships between the nodes. The use of a standard presentation format allows the user to swiftly become familiar with the way that information is communicated by the interface. This graphical interface allows the user to interact with the network by incrementally entering and retracting evidence. In this way the user can examine different scenarios, using the network as a tool for exploration.

In the GAMEES system [Bellazzi *et al* 91a], the authors view the network as a probabilistic knowledge base and believe that constructing the network is a knowledge elicitation process which should ideally be executed jointly and interactively by the knowledge engineer and the domain expert. In order to facilitate this type of interaction, GAMEES incorporates a graphical interface which enables the user to interactively construct and execute belief networks and influence diagrams. This allows easy modification of the model, encouraging exploration of the model and incremental model development. The user is also able to experiment with different propagation methods. The graphical interface allows network creation through a combination of predefined node formats and menu and text-based format creation. Node types are arranged in a frame based system which supports inheritance of format slots. Different node types are displayed as different shapes allowing easy identification. The beliefs at the nodes are not usually shown, but the user can select target nodes' beliefs for display. The beliefs are displayed as both a histogram across the states and as a numerical value. Part of the interface is shown in figure 4.17.

The Angina Communication Tool [Jimison 90, Jimison *et al* 92] is a decision theoretic system with specific emphasis on the communication of information to the patient and physician. Among other ideas, they suggest the use of three metrics that are used to modify the graphical display:

- *Value of Contribution* in terms of the change in expected utility based on the patient specific model over the generic model.
- *Deviation* of patient specific variable values from the mean of the generic model distribution.

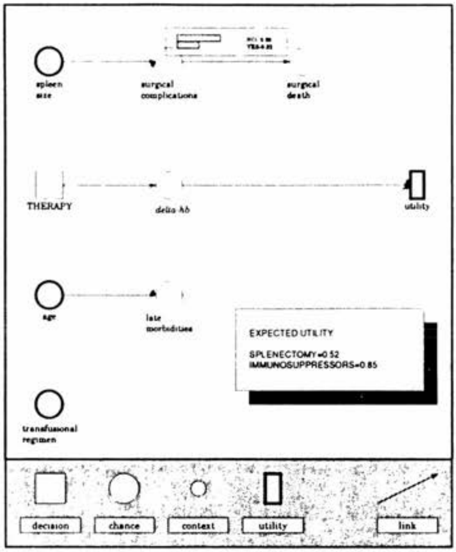


Figure 4.17: GAMEES — graphical interface [Bellazzi *et al* 91a, page 188]

- *Sensitivity* of a decision based on changes within the patient specific distribution for a variable.

On the bases of these metrics, the decision network can be collapsed and expanded to reflect important aspects of the patient specific decision model. The display of the network can be enhanced by including this information, and information about the strengths of probabilistic dependencies. This is illustrated schematically in figure 4.18.

4.11 Key Points

- Belief networks embody the normative theory of probability. Traditionally probability theory has not been used for uncertain reasoning in expert systems due to its apparent intractability. Belief networks rely heavily on the existence of conditional independences within the domain as established by a domain expert. This is exploited in order to reduce the number of probabilities that must be specified.

- Belief network inference is tractable because it is based on local information and asynchronous message passing.

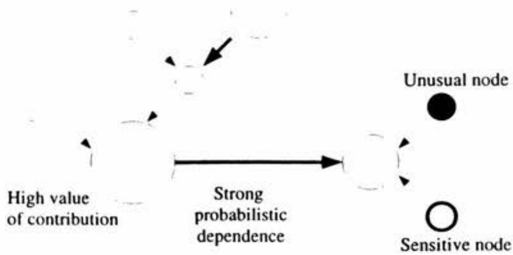


Figure 4.18: Angina Communication Tool — schematic network graphics [Jimison *et al* 92, page 200]

- Belief networks are able to perform both diagnostic (symptoms to causes) and causal (causes to effects) reasoning.

- Decision theory is a normative method for decision making based on utility theory and probability theory. It provides a powerful mechanism for decision making in belief networks.

- The use of probability theory facilitates the application of a wide range of statistical techniques that cannot be applied to *ad hoc* formalisms.

- For restricted classes of belief network model exact, efficient inference mechanisms exist. The problem of inference in an arbitrary belief network is NP hard. A variety of approximate inference techniques have been developed. The problem of approximate inference in an arbitrary belief network is also NP hard.

- The variety of exact and approximate inference algorithms, and the fact that many of the approximate algorithms produce partial results, makes belief networks ideal for use in resource bounded environments where complex metareasoning is appropriate.

- A belief network model's structure is often easily and intuitively expressed by a domain expert. The causal structure is an intuitively meaningful representation, making it an ideal communication tool.

- The quantitative probabilistic data is often difficult for a domain expert to assess. There are many reasons for doubting the accuracy of numerical estimates provided by experts. The derivation of the required numbers directly from data, or refinement of initial subjective estimates in the light of experience address this concern to a certain degree. Probabilities are not necessarily a meaningful tool for human communication.

- Tools for verifying the qualitative and quantitative components of a belief network model are being developed, lessening the potential effects of domain expert error.

Chapter 5

Applications — Overview

Many of the developments discussed in the previous chapter have been application driven. Belief networks are being applied to problems in a variety of domains, in Appendix A we present a gazetteer of 21 projects that apply belief networks to the biomedical domain. The most popular biomedical application is diagnosis, often combined with some decision making task. The size of the networks used varies from five nodes to over a thousand, indicating that the method scales to complex domains. In the less complex domains there are a number of applications that use Pearl's algorithm for inference in singly connected networks. In multiply connected networks some form of clique-tree propagation is typically employed.

Given the importance of imaging in the biomedical domain, a larger number of image-based applications might have been expected. Some image-based work has been performed in other domains, the ADRIES/SUCCESSOR system [Levitt 88, Binford *et al* 89, Levitt *et al* 90a, Levitt *et al* 90b] for instance, is concerned with the development of a system for model-based machine vision using influence diagrams. The influence diagram is used both to control the processing actions as well as propagating the evidence on which those actions are taken. Work along similar lines includes a system based on the HUGIN shell [Jensen *et al* 90b] and the TEA-1 system [Rimey & Brown 92].

The challenges of biomedical image processing are as difficult as those faced in any area of image interpretation. In the area of representation alone there are problems of fundamental importance for image interpretation as a whole. Biological objects exhibit a wide variety of forms both across classes of object (*e.g.* between teeth and hearts) and within classes (*e.g.* biological variation between hearts). They are often flexible and deformable and of different structure at different levels of resolution. Objects of interest may vary their shape with time, either due to function (*e.g.* the heart beating) or development (*e.g.* growth or decay). There can be exceptions to the way an object appears due to medical conditions and the same object may appear differently or may not appear at all depending on the imaging modality used.

The interpretation of a biomedical image is almost always an expert task, requiring knowledge of different domains to be used together in an appropriate way. In many cases the interpretation involves more than simply instantiating a model within an image, rather it involves making an assessment of how good the match is based on incomplete evidence from the image, mediated by uncertain background information which may determine possible alternative sub-models which account more satisfactorily for image features, circumstances under which model features will not be in the image and how likely this is, and so on. How knowledge of medical conditions, imaging modalities, patient history and so on is to be represented and used in conjunction with object representations to arrive at an interpretation for an image and to make judgements on the basis of that interpretation are questions of great importance.

In a diagnosis process dependent on expert image interpretation the rewards of automation are obvious. Due to the shortage of experts relative to the work load, the interpretation task will tend to become a bottleneck in the diagnostic process both in terms of time and money. An automated system could speed up the time between the images being taken and the results being available, it could perform measurements not normally performed due to time factors or the difficulty in taking the measurement accurately, and it could perform diagnostic tasks based on its own measurements. These types of benefits could enhance current diagnostic practice and make a wider range of tests economical. In practice though, it is likely to be many years before a computer-based diagnosis system is fast enough and accurate enough to be used in a clinical environment.

Given that the domain of biomedical image interpretation is in itself vast and that it is the general principles of the applicability of belief networks that is the main area of interest, two specific applications have been selected. The first of these is the task of locating the outline of the fetal head in a particular class of ultrasound scan that is routinely taken. This task emphasises the use of belief networks at a relatively low level of the image interpretation process. The second application is the classification of cervical smears as part of an automated pre-screening system. This task is concerned with incremental decision making at a high level in the image interpretation process. The biological background, clinical motivation and an outline of the concept behind the belief network approach to each of these tasks is presented below.

5.1 Fetal Ultrasound

The task of fetal ultrasound interpretation is interesting for a number of reasons. Firstly, it is very much an expert vision task, requiring the identification and accurate measurement of physiological structures. Secondly, if the process could be automated it would make scanning more widely available. Lack of trained personnel, time and equipment have been cited [Pearce & Campbell 84] as factors preventing all pregnant women from having detailed ultrasound scans. Thirdly, fetal ultrasound is an important clinical tool allowing a large number of fetal conditions to be detected. Current techniques often involve taking accurate measurements from ultrasound scans but the type and accuracy of the measurements is limited by time and repeatability constraints. Complex measurements, such as surface areas and volumes, could more easily and accurately be performed by computers, which could also provide three-dimensional visualisation facilities. Such developments could greatly enhance diagnosis based on ultrasound scans. Fourthly, it raises a number of very hard problems in terms of how a computer can represent the required uncertain expert knowledge and deploy it in a productive way. Finally the system could be tested directly against human experts and its performance properly evaluated, though it should be noted that clinical trials have not been conducted as part of this project.

5.1.1 Physics of Ultrasound Images

An ultrasound image is a representation of the strength of echoes produced when a beam of sound is directed into a medium [Campbell & Pearce 85]. The creation of an ultrasound image is possible as echoes are reflected back when the beam crosses interfaces between different tissues. These echoes are converted into electrical currents that can be displayed as an image. The strength of the echo depends on a number of factors, one of which is the difference in acoustic impedance (a measure of tissue density) between the tissues forming the interface. Bone and soft tissue, for instance, have very different acoustic impedances and such an interface will produce a strong echo, however as much of the sound wave is reflected back, visualisation of structures beyond this interface becomes more difficult. Across some interfaces, such as air/soft tissue, the reflection is almost total, making scanning through certain areas, *e.g.* the lungs, impossible. The size of the echo depends also on the angle at which the beam crosses the interface, a right angle producing a larger echo. However, as most interfaces will be irregular, the echoes from crossing the interface will tend to be scattered rather than producing a strong echo. Scattering of the beam also occurs within tissues as they are rarely homogeneous. This produces a phenomenon known as speckle. The ultrasound beam is reduced in strength (attenuated) as it is absorbed during its passage through a medium. Air is a particularly strong attenuator which is why a coupling medium, typically a gel or oil, is used between the transducer and the patient's skin.

The resolution in an ultrasound image is directly related to the wavelength of the ultrasound beam. Structures smaller than the wavelength cannot generally be resolved. Increasing the frequency will reduce the wavelength and thereby improve the resolution, but higher frequencies are absorbed more readily, resulting in lack of depth, so some compromise is necessary.

Ultrasound is a highly noisy imaging modality, the signal-to-noise ratio is of order two [Baldock & Towers 87]. There are two basic types of noise, speckle produced by scattering, and highly structural artifacts. Such artifacts occur in the image as data that are added, missing, or are of incorrect location, brightness, shape or size [Kremkau & Taylor 86]. Some of these artifacts are due to assumptions inherent in the ultrasound equipment, others arise from incorrect use of that equipment. It is possible that some artifacts and speckle are of diagnostic or interpretive use [Clarke 87].

There are three different types of ultrasound scan, A, B and M. A scans and M scans produce a one-dimensional image through the body, M scans being used to scan moving parts, such as the heart, by producing a series of scans along the same path. B scans produce a two-dimensional plane through the body by means of oscillating or rotating the transducer. It is these B scans which will be the focus of attention for this project.

5.1.2 Obstetrical Ultrasound

Ultrasound is an important tool for the diagnosis of the fetal condition for a variety of reasons. It is non-invasive, unlike fetoscopy or amniocentesis, it is relatively cheap, and there is no evidence to show that it is not safe under normal use[1]. It provides images of sufficient quality to enable a wide variety of diagnoses to be made, including the detection of structural abnormalities, the demonstration of life, the estimation of gestational age and the sexing of the fetus [Jeanty & Romero 83, Campbell & Pearce 85, Chudleigh & Pearce 86, McNay 87, Manning *et al* 89].

The nature of the diagnoses fall into two main groups, qualitative and quantitative. Qualitative assessments include the grading of the placenta as an aid to the determination of fetal maturity [Granum 79], the detection of multiple gestations and the assessment of fetal anatomy [Garrett 79, Pearce & Campbell 84]. Quantitative assessments are based on direct measurement of the ultrasound image using, for instance, map measurers or calipers. Length or width is a common parameter to measure as it is simple and highly reproducible. Area is also used though this is often estimated from lengths and some simple formula, resulting in an approximation of the true value. More complex and possibly more informative measures, such as volume or surface area are considered to be unreliable and too time consuming to calculate. In both the qualitative and quantitative assessments a high degree of skill is required to make a correct diagnosis or to take an accurate measurement.

There are two diagnostic uses of fetal ultrasound that deserve special mention both because of their clinical importance and their relation to the particular scans chosen as the test domain. These are the estimation of gestational age, and the detection of intrauterine growth retardation [DeVore & Hobbins 79, Sabbagha 79, Bowie & Andreotti 83, Deter *et al* 83, Beischer *et al* 84]. The accurate estimation of gestational age is important for the mother so that she can plan for the child's arrival. For the clinician it is critical to a whole range of disease detection and management decisions, where a balance must be struck between the potential damage caused by the disease if the fetus remains in utero and the risks associated with early delivery, primarily due to

[1] Following the completion of this section a number of reports have questioned the value of routine ultrasound scans. A summary of this research is presented by Saul [Saul 94].

immaturity of the infant's lungs and brain. It has been reported (Dewhurst, Beazley and Campbell in [Sabbagha 79]) that there is a four fold increase in perinatal mortality in cases where the gestational age is unknown. There are basically three methods for estimating gestational age, clinical, biochemical and sonographic.

The most common and best single clinical estimator is based on the last menstrual period (LMP or LNMP, last normal menstrual period), the estimated date of delivery being calculated from this using Naegele's rule. Unfortunately this estimation is unreliable under a number of circumstances [Chudleigh & Pearce 86]:

- When the date of the LMP is not accurately known.
- When the menstrual cycle is not 28 days long.
- When the menstrual cycle is irregular.
- When the patient only stopped taking the combined oral contraceptive (the pill) within the last three months.
- When the patient bled early in pregnancy.

Some 20-40% of patients are unable to give accurate dates, there being a reported bias (Zador, Hertz *et al* in [Bowie & Andreotti 83]) towards the first, fifth and fifteenth of the month! In addition to this 20% of patients with apparently reliable dates have discrepancies of one to six weeks [McNay 87]. For these reasons estimates based on Naegele's rule generally lack the precision necessary when taking important clinical decisions. Another popular clinical estimator is based on the physical examination of the uterine size which is considered to be accurate to within one week during the first trimester of pregnancy.

Biochemical estimators are typically based on the analysis of amniotic fluid as an indication of lung maturity. There appears to be some dispute as to how useful such estimators are [Bowie & Andreotti 83].

Sonographic estimations are based on the measurement of some well behaved parameter such as the width of the head (the biparietal diameter, BPD), for which a standard growth curve exists. The measurement taken can be cross referenced to the growth curve to obtain an estimation of gestational age. Such curves are available for a variety of parameters and have been in use long enough to verify and quantify the accuracy of estimations based on them and to determine which parameters should be used under different conditions. As fetal growth deviates little from the mean in the first 24 weeks of gestation and pathological growth retardation is uncommon during this period, estimates during this time will have a high degree of accuracy. Ultrasound has now gained wide acceptance as the preferred method for the estimation of gestational age.

Intrauterine growth retardation (IUGR) can be defined in lay terms as a condition in which the growth of a fetus is slower than would be expected. The clinical definition is based on measured parameters lying some distance (which varies) from the expected mean for the fetal age. The evaluation of a fetus as growth retarded obviously is heavily dependent on an accurate estimation of fetal age, a fetus which is average for 28 weeks may be small for 32 weeks. Retarded fetal growth is recognised as one of the three major contributors to perinatal deaths [Beischer *et al* 84] and has been associated with a variety of medical conditions, continued slow growth after birth, educational difficulties and possible neurological problems (Campbell reported in [Campbell & Pearce 85]). Many growth deficiencies are a result of fetal malnutrition associated with maternal diseases which interfere with the utero-placental blood flow. In the majority of cases however, the cause is unknown. As IUGR generally develops late in pregnancy it is possible to save many fetuses by early delivery. IUGR can be detected using the same parameters as used in the estimation of gestational age, providing the gestational age has already been determined. Some parameters are of less use depending on the type of retardation which may result in some area of the fetus, typically the head, being spared.

5.1.3 Fetal Head and Associated Parameters

The fetal head is an important region to image, not only can it reveal a number of cranial defects, but it can also provide a number of parameters useful for estimating gestational age. The most widely used of these is the biparietal diameter (BPD) mentioned earlier [Chudleigh & Pearce 86]. Others are the occipito-frontal diameter (OFD, the length of the head), the head circumference and the head area. All of these measurements must be taken in the correct plane through the fetal head, therefore the two main tasks are to assess whether an image is in the correct plane, and then take the measurements from the image. Where the image is not in the correct plane it may still be possible to approximate the measurement, though the confidence limits on the measurement will be wider. The correct plane for a measurement can be located by ensuring the plane images include certain brain structures. The recommended plane for the BPD has a short midline (an echo produced by the interhemispheric fissure) and the thalamus is visualised. If the head circumference and perimeter are also to be measured, the cavum septum pellucidum and the basal cisterns must also be imaged [Evans *et al* 89]. The skull itself should appear oval in shape.

In terms of ultrasound images, the midline appears as a straight, bright line running length ways, equidistant from the sides of the skull. The cavum septum pellucidum appears as two short lines parallel to and at either side of the midline, and at the front-most end of it. The basal cisterns appear as an almost semicircular arc towards the rear end of the midline, curving towards the cavum septum pellucidum, and bisected by the midline. The thalamus appears as a textured region straddling the midline, and running between the cavum septum pellucidum and the basal cisterns [Baldock & Towers 87]. The skull itself is probably the most obvious feature, appearing as a bright, rugby-ball shape. In this plane the BPD is the greatest width across the skull perpendicular to the midline and the OFD is the skull length along the midline. These features are illustrated diagrammatically in figure 5.1. If the required features are not imaged then

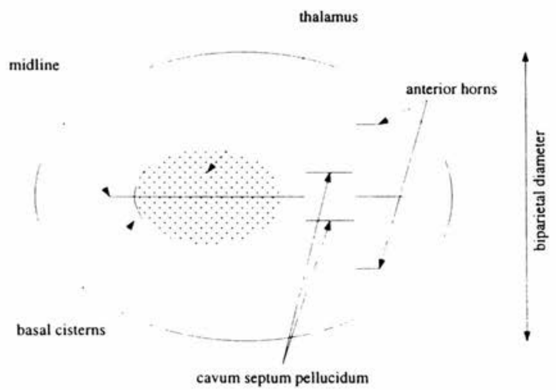


Figure 5.1: Diagrammatic representation of the fetal skull

there are three possible explanations: the feature is obscured by noise, the structure giving rise to the feature is missing due to some clinical abnormality, or the image is in the incorrect plane. The recognition of which of these explanations is correct (bearing in mind that all three could be true of an image) is crucial to the interpretation of an image. This recognition is complicated by the fact that each explanation cannot be evaluated independently, *e.g.* in deciding that a feature is obscured by noise we must check that other features confirm the correct image plane and that a disease state cannot explain the observed situation more completely. Obviously there is knowledge that can be used to constrain this process. For instance, if an image is believed to lie in an incorrect plane this can be checked by searching for features that confirm an alternative plane. If it is believed that a structure is absent then it should be possible to identify disease states that cause such an abnormality and maybe to confirm them by identifying other abnormalities associated with the condition. The interpretation of an image will therefore be an iterative process where partial interpretations are used to confirm or reject hypotheses about image features and their interpretations.

In terms of image processing, interpreting ultrasound scans of the fetal head involves most of the problems that will be encountered in ultrasound images, aside from motion. Some of these problems [Baldock & Towers 87] include:

- Ultrasound scans of the head often have strong shadowing where the beam is tangential to the skull so the boundary becomes broken. Also there are other bright curved lines in the image so it may be difficult to determine which correspond to the skull echo.
- Given that the bright regions corresponding to the skull outline have been found, how are the edges defined in terms of the fall-off of intensity, either side of the bone?
- How should gaps in the outline be traversed? One answer to this is to fit a model, *e.g.* a smooth curve, but there must be careful control of this procedure or else important diagnostic information may be lost.

5.1.4 Approaches to Automation

The potential for automated assistance in the fetal ultrasound image interpretation task has been clearly recognised by both researchers and clinicians [Zador & Sokol 92]. Despite this recognition relatively little research has actually been conducted and reliable commercial systems are still some way off.

SBS is a blackboard based expert system for model based interpretation of fetal head ultrasound images [Baldock *et al* 87, Baldock & Towers 88, Towers & Baldock 88]. SBS was designed to work with almost any data structures but in practice the most useful for image processing was found to be frames [Minsky 80]. The models in SBS are represented as a network of frames. Each frame is of one of three types; composite, relation or primitive. Composite frames are used to group frames together to represent a model or model sub-part. The combination of composite and primitive frames makes it possible to build complex models from a set of simple low-level components. The relation frame holds knowledge about the relationship that exists between frames. The work on SBS only progressed as far as simple models using basic primitives, such as lines and arcs, but it is potentially a very flexible representation. Two of the slots of a composite frame are worth noting, the confidence slot and the key features slot. The confidence slot, although never used, indicates the importance that measures of uncertainty would be likely to have in a full system. The key features slot provides a means of identifying which features (sub-frames) are necessary or important when matching the composite. The matching strategy employed in SBS is essentially the cue-hypothesis-test paradigm used in a number of image processing systems. Image primitives are used to cue hypotheses based on the key-features of a model. These hypotheses can then be tested by attempting to match other parts of the model.

A group at the Wayne State University in America have investigated the task of fetal image interpretation using standard image processing techniques [Salari *et al* 90, Zador *et al* 91]. They analyse images of the fetal head to extract the BPD, OFD and head circumference. The fetal skull is modelled using a simple five parameter ellipse model. The five parameters are the X and Y coordinate of the centre of the ellipse, the major and minor axis, and the orientation. These parameters are extracted using a four step process:

1. Local thresholding of image and rejection of small regions.
2. Location of edge points by application of a gradient operator and image thresholding.
3. Finding the centre of the ellipse using the Hough transform technique.
4. Determination of additional parameters by least-squares method.

Acceptable results are reported in 74 out of 75 test scans, when compared to a human operator. Care was taken to ensure that the scans used were free from obvious structural abnormality. The project hopes to develop a low cost, real-time system capable of taking a variety of measurements and of making decisions about the ultrasound imaging process. The appropriateness of the simple ellipse model is one of a number of factors still under investigation.

One aspect of image processing that may prove particularly influential in fetal image interpretation is three-dimensional visualization [Brinkley 87, Nelson & Pretorius 92]. By combining multiple two-dimensional ultrasound scans it is possible to form a three-dimensional model of the fetus, allowing the inspection of fetal surfaces, extremities and internal anatomy. In many cases, and especially those where the fetus is abnormal, the three-dimensional model would make it much easier to determine the spatial relationships between parts of the fetus. The model also has the potential to reveal defects that are difficult to identify in a two-dimensional image. Currently three-dimensional approaches rely on standard image processing techniques and therefore encounter a number of problems, such as determining surfaces in the presence of high noise levels, in addition to image capture problems, for instance those due to fetal movement.

Although there has been relatively little research concerned directly with fetal ultrasound images, there are many other image processing tasks that will pose similar problems and whose solutions will have a bearing on fetal ultrasound interpretation. Whilst lack of space precludes any in depth review of image processing as a whole, it is possible to single out *snakes*, or *active contour models*, as being of particular relevance to the belief network approach discussed later.

An active contour model is an energy minimizing spline which is guided by external constraints and influenced by image forces that pull it toward features such as lines and edges [Kass *et al* 87]. Internal forces on the contour may enforce continuity and curvature constraints. Local minima are alternative solutions. Typically an initial starting position is provided for the contour model which then adjusts itself iteratively to a minimum. These dynamic models are an example of a general approach to image interpretation based on deformable models as opposed to rigid geometrical models.

In addition to the detection of boundaries, active contour models have been applied to motion detection, surfaces and three dimensional shapes [Terzopoulos *et al* 87, Baumberg & Hogg 94, Byrne *et al* 94].

One of the recognised problems with the energy minimizing approach is that the contour can become trapped in a local minimum representing a false boundary. Global information, for instance in terms of a shape model, can be used to bias the contour towards the target shape, effectively restricting the space of allowed deformations [Baumberg & Hogg 93, Gunn & Nixon 94]. Ideally these global shape models will reflect the natural variation in the modelled shape and may best be derived from training examples [Cootes *et al* 95].

Other problems with active models include sensitivity to the initial location of the contour and undesirable attractions to irrelevant image features.

5.1.5 Belief Network Approach

The task of identifying the fetal skull in an ultrasound scan is an example of a generic low level model matching problem. If the model is described by a series of points along its length at some level of resolution, then the matching task can be expressed in terms of finding a maximally satisfying match between the points and the xy coordinates of the image. The definition of the degree of satisfaction of a particular match will depend on the information that is available. At a particular point there may be several different types of information that should be considered. Some of this will be local to the point, some will depend on the position of the other points and some may be due to higher level input. This information can be utilised to improve a given match, by using it to evaluate alternatives and to select the maximal match. Typically it is impractical to find the maximal match in a single step and an iterative approach is adopted instead. There are a number of different techniques that could be used to maximise the match.

Within this application, a belief network is used to perform the low level matching task, controlling and propagating evidence from the image, under constraints derived from pre-stored models based on a training set. This means that the low-level processing could be properly integrated into the diagnostic process within a theoretically sound statistical model and that errors at this level could propagate to uncertainties in the parameter estimation and diagnosis levels.

Each point in the model is represented by a node in the network. Only nodes that represent adjacent points in the model are directly linked in the network. Through these links the influence due to the position of neighbouring points can be modelled. The domain of potential matches for a point includes the entire image space, which is impractical. In order to restrict this domain, our starting point for the refinement is a user generated estimate of the boundary in the image, which provides an initial estimate of the match for each point. In our model we assume that the error associated with this cue is unknown[2], it is possible that the best match lies outwith an arbitrarily defined domain. To allow for this the domain itself is redefined iteratively with respect to the current match, so as a point ideally moves towards the maximal match, so the domain of possible matches is redefined to include the maximal match. The movement of a point towards the maximal match when that match lies outwith the current domain is possible due to the non-local information that is available. Similarly a locally maximal match that is not also part of a globally maximal match will be avoided. The available information can be viewed as constraints on the way a point can be moved with respect to the part of the image defined via the match domain.

In the ultrasound example we consider three main sources of information, the values of potential matches for the other points, the match to a local grey-level profile model and a parameterised shape model. At each point we consider two distinct match domains. The first domain is defined by an orthogonal to the model line that extends in both directions from a midpoint centred on the point. This domain is represented by the states at the point node. The belief values across these states are used as a measure of the match, a high belief representing a good match. The beliefs are propagated between nodes as determined by the network model, therefore nodes that are distant

[2]This is not normally the case and integrating a prior probability of this type is straightforward with respect to the probabilistic model.

with respect to their ordering along the model will have less influence on each other's matches than those that are direct neighbours. The grey-level profile defined by the orthogonal is used to provide local match information. By using a grey-level profile model created from training data, the likelihood of each possible match in the domain can be calculated from a measure of the match between the profile model and the image profile observed. Both the grey-level profile constraints and the constraints from the belief values of the other nodes in the network operate in the domain defined by the orthogonal. The parameterised shape model, which is also derived from training data, is theoretically unconstrained with respect to any domain. In practice, however, the current model points act as a global constraint on the shape model as the parameters are used to transform the current set of model points rather than generating a completely independent shape estimate. The result of the shape model estimate is a single maximal match for each model point. As the two match domains may have maximal matches that do not lie within the intersection of the domains, a mechanism for determining the combined best maximal match is required. This combined match value is used to determine the match for each point, this set of matches is then used as the initial set of points for the next iteration. The iterative process is repeated until the best match for all the points is identical (to within some degree of precision) to the current set of points, *i.e.* the match is *stable.*

5.2 Cervical Screening

Cervical cancer is the second most common form of cancer in women throughout the world [W.H.O. 88]. Currently primary prevention is not considered to be a viable option because, as the World Health Organisation notes [W.H.O. 88, page 1],

> ...the causes of cervical cancer are not yet fully understood ...and because the main factor thought to be associated with increased risk (sexual activity of both men and women) is not very amenable to regulation or control ...

The key to the treatment of cervical cancer is early detection, 80% of cases that are detected early can be cured as against 5 or 6% of late cases [N.E.F.J. 72]. Early detection is possible due to a recognizable pre-malignant phase, carcinoma *in situ*, which may last as long as 10 to 15 years before invasive cancer develops.

The detection of carcinoma *in situ* is based on the examination of cells exfoliated by the cervical surfaces. This technique of *exfoliative cytology* was described by Papanicolaou and Traut in the 1940s [Shingleton & Orr, Jr 87]. A spatula, such as an Ayre Type or a Cervix-brush, is used to collect a sample of cells which are then spread across a microscope slide. The slides are then sent to a diagnostic laboratory where they are stained using a Papanicolaou Stain which aids identification. The slides are then screened for cellular abnormalities under a light microscope by a trained cytotechnician and/or cytopathologist.

Screening to detect women in the pre malignant period has been shown to reduce both the morbidity and mortality of cervical cancer [Miller 89]. However the effects of different screening programmes are highly variable. Chamberlain [Chamberlain 89] suggests that the major reasons for the apparent failure of some screening programmes are:

- Failure to reach high risk groups.
- Inadequate follow-up of abnormal smears.
- Long intervals between smears.
- False-negative rate.

The identification of high risk groups is complicated by the interrelationships between those factors believed to be associated with high risk [Shingleton & Orr, Jr 87]:

- Early intercourse (less than 17 years old).
- Multiple sexual partners.
- Early pregnancy.
- Urban population.
- Low socio-economic status.
- Immunocompromised.

- Smoker.
- Previous abnormal smear.
- Failure to participate in screening.
- Nutritional defects.
- Infertility (fallopian tube damage).
- Use of contraceptives.
- High risk male partners (multiple sexual partners, previous partner with cervical cancer).

5.2.1 Approaches to Automation

Researchers since the 1950s have suggested that the process of cervical screening be automated in order to improve the throughput of specimens and the accuracy of diagnosis [Evans 70, Wied *et al* 76, Eason]. Much of the research has focussed on the concept of pre-screening. An automated pre-screening system is designed to identify those specimens that are definitely normal or abnormal, those which are unclear being referred to the cytologist. Pre-screening could potentially reduce the proportion of specimens that must be examined by the cytologist, it could also add quantitative as well as qualitative data to the analysis of specimens. Research in this area has advanced to the point where several mature systems exist, although a reliable automated system has yet to be produced [Banda-Gamboa *et al* 92, Linder 92, Eason].

The majority of the systems developed have taken a similar approach. Images of objects on the slide are captured, object features (such as diameter and optical density) are calculated and individual objects are classified on the basis of their set of feature values. The classification for the entire slide depends on the profile of classifications of individual objects in the sample. The particular features that are used and the way in which the individual features are combined to produce an object classification vary between systems, with at least one system using neural network techniques [Bartels & Weber 92, Linder 92].

There are, however, many uncertainties inherent in the prescreening process [Bartels & Weber 92]. All slides, normal or abnormal, will contain some objects that appear abnormal, and an abnormal slide may only contain a small number of abnormal cells. Cell abnormality itself is a continuum and cell classification is error prone. Each slide contains a large number of uninformative artifacts, some are easy to detect and ignore, but others appear cell like. It may only be possible to examine a sample of objects on the slide.

As a result of these uncertainties, the overall profile of objects on the slide is also uncertain due to misclassifications. Furthermore there are statistically significant overlaps between the object profiles of normal and abnormal slides [Bartels & Weber 92].

Investigations into improving automated classification through different staining methods, disaggregation of cells and automatic slide preparation have also been conducted [Banda-Gamboa *et al* 92, Eason].

5.2.2 Belief Network Approach

Automated prescreening involves scanning the microscope slide for suspicious cells. The scan produces a series of tens of thousands of "objects" (essentially dark image structures) as potential cell candidates, and at the lowest level of this process are the tasks of object detection, feature extraction and object classification. Whilst belief networks could be applied at this level, we have chosen to examine the classification of the specimen as a whole on the basis of the available information. Part of this information will be the classification of objects in the specimen, other information includes counts of the number of objects of different types and patient data that might indicate high risk, for instance. With this example we demonstrate how information can be incrementally introduced into a task bridging the gap between low-level processing and higher-level diagnosis.

The motivation for developing belief network models for cervical specimen prescreening was the concept of an explicit model. Many automated systems contain some form of 'black-box' statistical classifier that is typically highly tuned by a domain expert. These classifiers are totally opaque to the user who is then forced to accept or reject the system's classification without the benefit of understanding the reasoning behind the classification. On the other hand an *ad-hoc* classification based, for example on rules, will not have the rigorous statistical underpinning typically required of a medical system. Belief networks offer an ideal compromise, having a qualitative structure that is meaningful to the user, whilst preserving the statistical attributes of probability theory in its results. It also offers the potential for using a single inference mechanism to model the entire image processing task, from low level object classification to high level decision making.

Chapter 6

FLAPNet — Overview

In order to explore potential applications of belief networks, a flexible network propagation shell called FLAPNet (**FLA**VOURS **P**ropagation **Net**work) has been designed and implemented by the author. The shell supports a general network propagation mechanism that is not constrained by any particular inference algorithm. By allowing a variety of different node types it is possible to include special-purpose nodes into belief networks or to construct networks that do not follow the belief network paradigm at all.

The FLAPNet system has been developed using the POPLOG toolkit (version 14.1 with a couple of minor extensions from version 14.2). The code is written in POP-11 and makes use of the FLAVOURS package which adds object-orientated functionality to the POP-11 language. It is designed to interface via the X Windows System. It is currently being run under UNIX on a Sun Microsystems SPARC Station.

Although the FLAPNet core is application independent, when applying it to the problem domains it proved necessary to customise the code to some degree, particularly that concerned with interface management.

In addition to FLAPNet itself, the fetal ultrasound interpretation system, discussed in the following chapter, made use of image display and interaction software kindly provided by Dr Richard Baldock. As well as displaying stored boundaries and images and allowing the user to set parameter values and to interactively create boundaries, the software calculated the Mahalanobis distance which was then used as an input to FLAPNet.

Each application consists of approximately 15,000 lines of POP-11 code, all of which have been developed incrementally, with little concern for optimisation. In some areas ease of use has been sacrificed in exchange for flexibility. FLAPNet is not intended to be anything more than a research tool.

6.1 Design Overview

The guiding principle in the FLAPNet design has been to make the network propagation mechanism as general as possible and not follow any particular propagation paradigm. However, FLAPNet was originally a belief network specific tool and as a result of this influence the terminology used (*e.g. pi* and *lambda* vectors) is often that of belief networks and some of the assumptions incorporated arise from belief networks.

FLAPNet makes minimal assumptions about the physical network model:

- A node N may only communicate directly with nodes defined as either a *child* of N or a *parent* of N.
- If a node N is a *child* of another node **O**, then **O** is a *parent* of N and *vice versa.*
- A node cannot be defined as its own *parent* or *child.*
- A node may have zero or more *parents* and zero or more *children.*
- A node cannot be defined as both a *parent* and a *child* of a particular node.
- A node with no *parents* has a pre-defined default message. The node is said to be *pi conditioned* with this message.

The message a node receives from any of its parent nodes is defined as a *causal message.* A message from any child node is defined as a *diagnostic message.* External messages may also be received, a causal form defined as a *causal conditioning message* and a diagnostic form defined as a *diagnostic conditioning message.* The content of a message is entirely unconstrained, for instance a message may contain a vector of numbers representing a probability distribution. The principal message types are:

- Causal message.
- Causal (pi) condition message.

- Causal (pi) uncondition message.

- Diagnostic message.

- Diagnostic (lambda) condition message.

- Diagnostic (lambda) uncondition message.

Each node has a set of *attributes* which typically reflect the possible propositional states represented by that node. A *belief* vector is defined across the attributes, such that each attribute has a corresponding belief. Two other vectors, *lambda* and *pi* are also defined across the attributes. The lambda vector is associated with the influence of diagnostic messages on the attributes, and the pi vector is associated with the causal messages. For example, if a node's belief vector contains a set of numbers representing a probability distribution across its attributes, then the lambda and pi vectors might also be probability distributions representing respectively the combined diagnostic and combined causal evidence messages sent to the node. The belief vector would then be a combination of the lambda and pi vectors. The principal data associated with a node are:

name — The unique name of the node.

parents — A list of the names of the parents of this node and the most recent causal message that each parent has sent.

children — A list of the names of the children of this node and the most recent diagnostic message that each child has sent.

attributes — The states represented by the node.

belief — A vector across the attributes at this node.

pi — A vector across the attributes, associated with the causal messages.

lambda — A vector across the attributes, associated with the diagnostic messages.

Every node has an area of private storage for miscellaneous data. This list can be used to store any data that is necessary for the operation for a particular node type. Data can be placed in this slot either at creation time or dynamically at run-time, or both.

All nodes respond to a particular type of message by executing the same sequence of actions. For example, a causal message, containing the name of the originator of the message and the item of causal evidence, is processed using the following sequence of actions, shown in POP-11 pseudo-code:

```
defmethod causal_message(origin, evidence)
     update parents record to include new evidence
     unless pi conditioned
     do
          update pi
          update belief
          update display
          propagate diagnostic messages
          unless lambda conditioned
          do
               propagate causal messages
          endunless
     endunless
enddefmethod
```

The action sequences for the other methods are contained in Appendix B. The action sequences embody certain assumptions about propagation:

- A node will only propagate messages when it receives a message.
- A node that is lambda conditioned sends no message to its children.
- A node that is pi conditioned sends no message to its parents.
- A node cannot be simultaneously pi and lambda conditioned.
- Messages can only be propagated to the parents or children of a node.

Although all nodes respond to the same set of messages with the same sequence of actions, the effect of the actions is determined by the individual nodes. If a node performs the `update belief` action, for instance, the effects of that action and the interpretation of the resulting belief vector may be unique to that particular node. The common effects of the actions are:

update belief This function is used to update the belief vector of this node. It takes a single variable, `self` which is a reference to the node. The result it returns is stored in the belief vector.

update pi This function is used to update the pi vector of this node. It takes a single variable, `self`. The result it returns is stored in the pi vector. If the node is pi conditioned the result will be the conditioning vector.

update lambda This function is used to update the lambda vector of this node. It takes a single variable, `self`. The result it returns is stored in the lambda vector. If the node is lambda conditioned the result will be the conditioning vector.

propagate causal messages This function is used to calculate and send the appropriate causal messages. It takes two variables, `self` and either `false` to indicate that this propagation was not initiated in response to a diagnostic message, or the name of the child that sent the initiating diagnostic message. It returns nothing.

propagate diagnostic messages This function is used to calculate and send the appropriate diagnostic messages. It takes two variables, `self` and either `false` to indicate that this propagation was not initiated in response to a causal message, or the name of the parent that sent the initiating causal message. It returns nothing.

parents function This function is used to update the parents to reflect new causal messages. It takes three variables, `self`, the name of the originator of the causal message, and the value of the causal message. The result it returns is stored as the parents. The parents are updated even if the node is pi conditioned.

children function This function is used to update the children to reflect new diagnostic messages. It takes three variables, `self`, the name of the originator of the diagnostic message, and the value of the diagnostic message. The result it returns is stored as the children. The children are updated even if the node is lambda conditioned.

By allowing all the functional components of the message handling methods to be individually specified it is possible to create node types that can perform a variety of different behaviours. A small library of experimental node types has been be developed and in many cases a new node type will reuse some of the existing function definitions, as illustrated in the following node function definitions.

6.1.1 Pearl Type Node

This node type embodies the belief or Bayesian network propagation algorithm described by Pearl [Pearl 88b, chapter 4]. Initially this was the only node type supported and many of the design decisions were influenced by the requirements of the algorithm. It should be noted that there is no restriction on circular inference paths and there are no mechanisms for handling them. It is assumed that the underlying probability model is defined to ensure convergence (see Chapter 7).

A Pearl Node represents a propositional variable, with attributes that define exhaustive, mutually exclusive variable states. The parents of a Pearl Node are those nodes judged to be direct causal influences and the children are those nodes judged to be directly causally influenced by the node. Messages received from both parents and children are assumed to have a probabilistic interpretation consistent with that given by Pearl. Specifically the message is assumed to contain a vector of probabilities or likelihoods, depending on whether it is a causal or diagnostic message. The belief vector in this case represents an actual probability for each state represented by the node variable. The lambda message vector contains a relative likelihood value for each state at the node variable. The pi message vector contains probabilities in the form of the belief vector of the sending node, less the lambda contribution of the receiving node. These values are mapped to the states of the receiving node via conditional probability matrices stored at the receiving node.

The majority of the implementation of the Pearl Node is straightforward as there is a close match between the Node model and the function of a Pearl Node.

update belief The new belief vector is calculated as the normalised vector product of the lambda vector and the pi vector. This is actually defined as a standard function rather than specifically a Pearl Node function.

update pi This function simply returns a vector that has been stored in miscellaneous data by the node.

update lambda The standard function which calculates the vector product of the messages sent by the children of the node.

propagate causal messages A standard function is also used here. This function calculates and sends a message to all children except the originator of the initiating message if that node is a child. The message sent is the belief vector of the node, less the contribution made by the child node to the which the message is directed.

propagate diagnostic messages This function calculates and sends a message to all parents except the originator of the initiating message if that node is a parent. The message sent is the lambda vector of the node multiplied by a marginal probability matrix that is stored in the miscellaneous data at the node.

parents function This function first updates the parents record at the node to include the new evidence vector. It then recalculates the marginal tables for all the parents other than the originator, given a conditional probability matrix stored in the miscellaneous data, and then stores them in the miscellaneous data. A new pi vector is calculated and also stored in miscellaneous data. This rather complex sequence of actions is necessary as the links in the network model are unquantified, all notions of quantification must be implemented by the nodes themselves.

children function The standard function is used, it simply updates the children record at the node to include the new evidence vector.

6.1.2 Simple Classifier Type Node

In the course of the application developments a requirement to convert from a single number, *e.g.* a count of some quantity, to a probability over classes of quantity, *e.g.* {*inadequate, adequate*} as illustrated in figure 6.1, was identified. This function represents the conditional probability "matrix" for the two classes given the number. Two

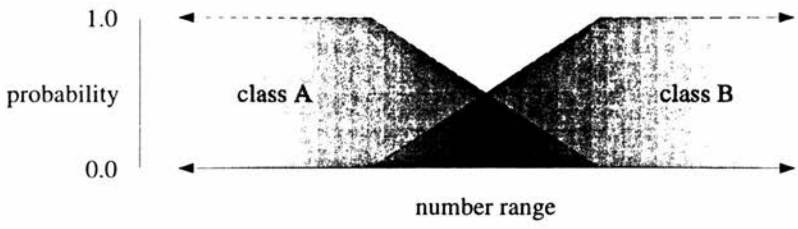


Figure 6.1: Conversion from a number to class membership

node types have been defined to perform this function, one taking the number from a single child node and one taking it from a single parent, only the first of these is described here. The node is given a classification model that specifies the incline slope and decline slope for each class membership range. The attributes are the classes over which membership is defined. In the application considered it was not valid to include information from the parent at the node.

update belief The new belief vector is the same as the lambda vector as the causal contribution of the parent is ignored.

update pi This function simply returns a vector that includes dummy characters to indicate that the data is ignored.

update lambda The classification model stored in miscellaneous data is used to calculate the class membership probabilities which are used as the lambda vector.

propagate causal messages Although it is assumed that there is no valid message to transmit to the child, a dummy control message is sent (this is explained below).

propagate diagnostic messages Unless the parent initiated the propagation, the current belief is propagated to the parent via a conditional probability matrix held in the node's miscellaneous data.

parents function The standard function is used, it updates the parents' record at the node to include the new evidence vector. This is not strictly necessary.

children function The standard function is used, it simply updates the children record at the node to include the new evidence vector.

The relationship between the classifier node and its child node is interesting as it shows the potential for including control messages. The child is a Number Display Type node which essentially just provides a way of displaying the numerical data that has been entered. The entered data is propagated directly to its parent. The single attribute is the quantity being measured, the lambda and belief vectors contain the quantity entered, the pi vector contains a dummy character. The data is entered at the node by lambda conditioning, if no data has been entered then the node should automatically use the default value provided in its miscellaneous data. In order to achieve this, the parents function checks to see if the node is lambda conditioned (*i.e.* data has been entered) if not then the node lambda conditions *itself* to its default value and propagates it to its parent. The causal message from its parent that results in this behaviour is otherwise ignored. This is obviously only a very simple example, but as there is no restriction on the contents of a message they could easily be used to convey genuine control information, or even routing information if an application required it.

6.2 Interface

FLAPNet provides a simple, application independent X-windows interface. The interface was intended to be a developers' interface only and it is therefore rather basic.

The primary display is the base display window which shows all the nodes in the network, identified by their unique name. Each node is assigned to a particular generation by the user, or is automatically assigned a generation number given its relationship to a reference node of a given generation, *i.e.* according to network topology. The generation numbers assigned by the user need not bear any relation to the structure of the network and so can be used to group network nodes into sets in any way the user chooses. Each generation has a user definable title and is displayed as a separate list in the base display window, as illustrated in figure 6.2. In the example shown, taken from the application

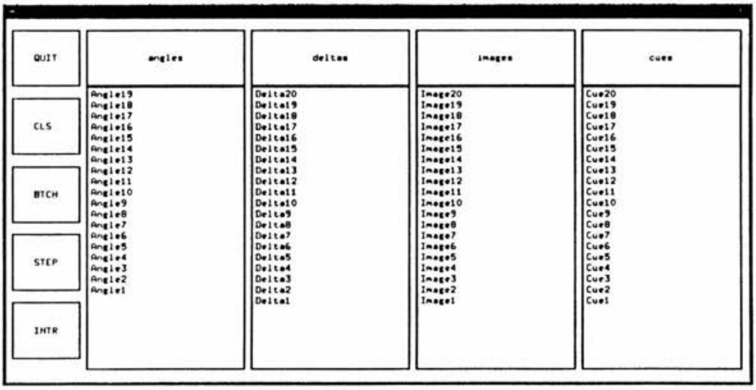


Figure 6.2: Base display window

discussed in Chapter 7, the generation groupings, angles, deltas, images and cues, have been imposed by the user to reflect the function of the individual nodes. Thus the column headed 'cues' contains a list of all the nodes of type cue because they share a common generation number. The individual names of the nodes are determined by the user, the fact that the names of the nodes in the cue list are all of the form CueXX is a feature of the application rather than the interface. The name of a nodes have no influence on the generation grouping into which it is placed.

The buttons on the side of the base display window have the following functions:

QUIT — Quits the display.

CLS — Clears all display windows other than the base window from the screen.

BTCH — A batch function for reading data from a file.

STEP — An incremental version of BTCH.

INTR — An interactive function for entering commands.

The way in which a BTCH, STEP or INTR function interprets its data is application dependent and specified by the user.

The lists of node names in the base display window are selectable. When a specific name is selected a node display window appears, figure 6.3. Each node has its own node display window, allowing several to be displayed simultaneously. The main area of a node

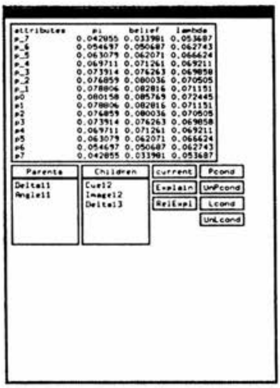


Figure 6.3: Node display window

display shows the attribute, pi, belief and lambda values for the node. In the example shown, the attributes have obscure names (p_7 and so on), but these are just a feature of the application. The Parents and Children lists contain the names of the parents and children of the node. These lists are selectable in the same way as the lists in the base display window and node display windows are created for the selected nodes.

The buttons on the node display window have the following functions:

current — As the network may be propagating values whilst a node display is active, the values at that node may change. Rather than constantly refreshing any node display window whose values have changed, the node display is marked when it is out of date. The 'current' label on this button is changed to *-OLD-* when the information display is out of date. The display can be brought up to date by clicking on this button, causing the node display to be refreshed.

Explain — See below.

RelExpl — See below.

Pcond — This button creates a dialogue that allows the user to enter a vector of values that is then used to pi condition the node.

UnPcond — This removes any pi conditioning currently in force at the node.

Lcond — This button creates a dialogue that allows the user to enter a vector of values that is then used to lambda condition the node.

UnLcond — This removes any lambda conditioning currently in force at the node.

The lambda and pi conditioning facilities are the principal means of entering data into the network. The BTCH, STEP and INTR functions use these mechanisms, as do the majority of the application dependent data entry methods.

Although this interface is very limited it displays the majority of the information a user requires. The notable exception to this is the display of node type specific data, such as the conditional probability matrices used by Pearl Type nodes. There is clearly a requirement for partial customisation of the node display according to node type. This extension would be fairly straightforward.

In addition to the experimentation with different node types, some simple explanation facilities were developed. Like the main display interface, these are text based. They give a flavour of the range of information that it is possible to offer the user for a relatively low computational overhead. The first of these facilities is shown in figure 6.4. Most of the

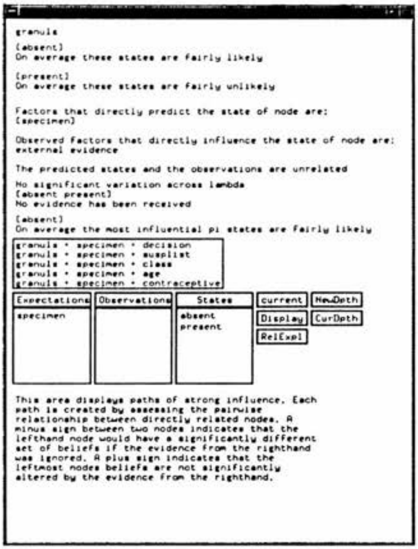


Figure 6.4: Explanation window

text in the window is selectable, generating further detail of the explanation. The area below the lists and buttons is used to display these dynamic messages. The functionality of this window is as follows:

granuls — The name of the node to which the explanation window relates. Selecting this item brings up a brief explanation of the purpose of this node.

[absent ... unlikely — Groups the states of the target nodes according to their approximate likelihood. These likelihoods are expressed in fuzzy verbal terms even though they have a precise definition within the system. Selecting these items brings up an explanation similar to this description.

Factors ... unrelated — Indicates which nodes are direct neighbours of the target node and provides an assessment of the degree of agreement between the evidence available from these neighbours. Selecting these items again brings up an explanation similar to this description.

No significant ... received — Identifies the most likely states based solely on the information from child nodes. In this case there are no children and no external evidence. Selection brings up an explanation similar to this description.

[absent ... likely — As above but for parent nodes. The likely states are categorised verbally.

granuls + ... contraceptive — (Shown selected) This displays paths of strong influence. The length of the path is selected by the user. Selecting this item brings up the display shown in the figure.

Expectations, Observations and States — Selecting these items brings up an explanation of the contents of the lists they head.

specimen — This is a link to the specimen node display. Selecting it causes the display to appear.

absent, present — These are the states of the granuls node. Selecting them causes a short explanation of the meaning of that state to be displayed.

The second explanation type is the relative explanation display, shown in figure 6.5. The relative explanation displays the relationship between the target node, granuls, and a second target node, loresc, selected by the user. The specimen node is included as it forms the converging join of the path between these two nodes. The display functions as follows:

granuls, specimen, loresc — (Specimen shown selected) Selecting these items brings up a brief explanation of the purpose of the selected node (non available in this example).

absent 0.716486 ... unknown 0.0 — These are the likelihoods of the states at each of the nodes. Selecting one of them brings up an explanation as to the meaning of that state.

Path — Selecting this brings up a brief description of the path, e.g. specimen is a common parent to both granuls and loresc.

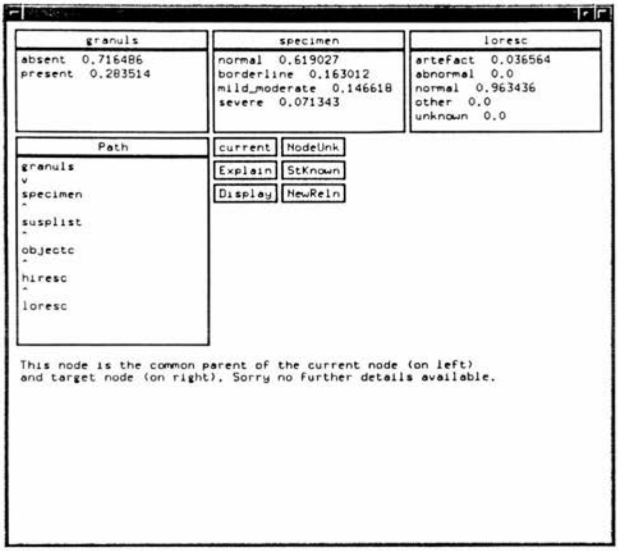


Figure 6.5: Relative explanation window

granuls v ... loresc — This shows the causal direction of the path between the selected target nodes. Selecting a node from this list brings up its display.

NodeUnk Allows the user to set one of the three displayed nodes to an unknown state. The effects of this on the likelihoods at the other two nodes is displayed.

StKnown — Allows the user to see the effects of setting one of the displayed node states with certainty. Neither this nor the NodeUnk function actually change the network permanently, rather a series of conditioning and unconditioning operations are used to record, simulate, and then restore the network.

These explanation facilities have been developed as part of the network environment and have not been used directly in the applications presented in this thesis. The explanation facilities work seamlessly within the Pearl type nodes, but they are not fully defined for the other node types, for which explanation facilities are an open problem.

Chapter 7

Application — Fetal Ultrasound

This application is a model-based matching system for two-dimensional boundaries in images. It takes as its starting point an estimate of the position of the boundary in the image. This initial cue is improved by the system on the basis of model constraints derived from training data. These constraints are currently of two types, a grey-level or image model in which a grey-level profile is matched to a learned model, and a geometric shape model for the local and global boundary. This is an example of the generic task of combining local and global constraints in an image interpretation problem. In principle the whole matching process is based on models derived from the known shape, size and variation of the fetus and in the measured characteristics. By basing the matching process on a probabilistic model it is then possible to establish true error estimates for the measurements required, *e.g.* the biparietal diameter, and to monitor fetal development.

7.1 Overview

The system attempts to modify an initial cue, in the form of an estimate of the boundary position, such that the geometric and image models are optimally matched. As the application is concerned with refinement of a cue which has been produced by some other process or in our case by the user, rather than the direct instantiation of a model, the search space can be limited to a region of the image local to the initial cue. The cue is defined by a number of equally spaced xy coordinates. At each of these coordinates an orthogonal to the boundary is placed, such that its midpoint lies on the boundary. This orthogonal defines the search space for matching the image profile and is dynamic, as the line is adjusted iteratively until an acceptable solution is found. The shape constraints operate outwith the reduced search space defined above, and *potentially* include the entire image. The two constraints each provide a best estimate of the new position to which each point on the boundary should be moved. These estimates are combined to give an overall best position. The line defined by the new set of points then serves as the current best estimate and the process is repeated. This iterative refinement is illustrated in figure 7.1.

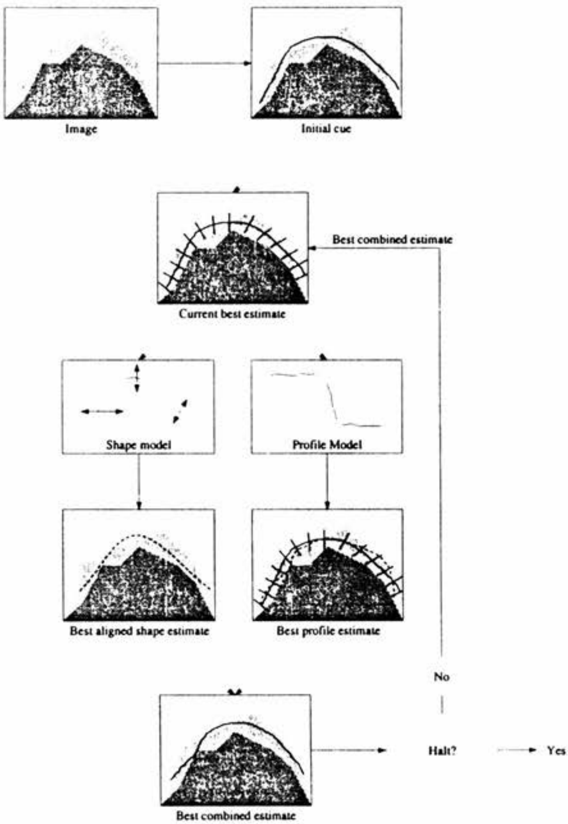


Figure 7.1: Processing cycle for iterative refinement of an hypothesised boundary

7.1.1 Probability Model

The boundary is defined by a set S of i boundary points $XY_{s,s=1...i}$. At each boundary point s an orthogonal is placed, such that the mid-point of the orthogonal lies at the intersection with the boundary. The orthogonal is defined by a set D of j matches or sample points. Thus each boundary point s is associated with a set of sample points XY_s^D. This is the current set of potential matches for the boundary point s and includes the point s itself. The task then is to generate a probability distribution over this set of potential matches on the basis of the information available from the image, other boundary points, trained models, and so on.

Local information can be defined as that which provides a probability distribution across XY_s^D for a particular s, irrespective of any information concerning other boundary points in S, *i.e.* treating s as an isolated point. In this application only a single local information source is considered, a measure of the match between the grey level profile at points in D and the grey level profile model derived from the training set. This provides a probability distribution over D. In a more developed application there may be several sources that need to be combined to give a single local probability distribution over D at boundary point s, $PL(XY_s^D)$. This combination can be modelled in a belief network by representing each boundary point s as a node, with states D, and representing the local information sources as parents or children of the node, depending on their causal relationship. The most probable d, $PL_{max}(XY_s^D)$ can be interpreted as the best match for the point s under consideration. In the absence of any non-local information the total probability distribution over D, $P(XY_s^D)$ is equal to $PL(XY_s^D)$. A matching process based solely on PL would locate a set of individual matches irrespective of non-local information such as the connectivity of the boundary or shape constraints.

The connectivity constraints of the boundary are modelled by linking nodes that represent adjacent points on the boundary and the relationship between adjacent nodes is represented by the conditional probability matrix. This provides a mechanism for combining the influence of the local information relating to $XY_{S\neg s}^D$ with $PL(XY_s^D)$. The resulting probability distribution at s reflects both the point specific local information and the local information at all the other points in S, mediated by their probabilistic relationship. The nature of the links between adjacent nodes is such as to impose a constraint on the curvature of the line. This is a *physical* constraint which acts only over a limited distance, some bending is permitted and sufficiently strong local matching information will always override this effect. Therefore the influence *decays* with respect to distance (path length) a feature we term *spatial* or *temporal* conditioning (temporal in the sense of causality). This is a type of probabilistic conditioning which allows a cyclic network to converge correctly and arises from the underlying physical model.

The construction of a network of this type raises questions about the nature of the causality embodied in the links. When considering a set of points there is no intuitive sense in which the position of any one point *causes* the position of any other point, rather the positions mutually constrain each other. Movement of a point, however, can be said to cause movement at adjacent points, though this implies *bidirectional* causality. In our model we arbitrarily impose a child/parent relationship between the nodes, such that each node has one adjacent point as a parent and the other as a child, resulting in a vertical chain between the point nodes. In addition to this we rely on the notion of spatial conditioning to control the propagation of inference in cyclic networks which are used for closed boundaries. In a cyclic network a path can be traced from a node back to itself in either direction and continuous propagation is possible even though the probabilities converge. Spatial conditioning implies that the further the propagation travels from the initial evidence source (in terms of the number of nodes or links traversed) the less effect that evidence has at a node. Therefore the effect will decay exponentially and when it falls below some resolution threshold, propagation can be terminated. An alternative solution is to let the network run continuously and then *sample* the probability values which will become stable with time. The choice between terminating on convergence or allowing a continuously running network is simply one of computational economy.

The result of propagating a piece of evidence is an assignment of probabilities across XY_S^D. As we are interested in matching the model to the image, we select the set of the most probable XY_S^D, $PS_{max}(XY_S^D)$ and move the current points to these new points. This displacement between the current points and the new, most probable points is a measure of the *stability* of the match. If there is no displacement then the match is stable. In the examples presented in this chapter, zero displacement stability will not be achieved because of noise arising from the image data and the image discretisation, so a stability threshold is applied.

The match domain described above is one-dimensional for a particular point, and *constrained* two-dimensional for the model as a whole. In order to introduce greater flexibility into the domain of potential matches we allow the shape constraints to consider matches over the entire image. In practice, of course, the refinement of a boundary's shape is constrained by the location of the boundary in the image and its current shape. The shape modelling system provides a single match for each point $G(XY_s)$, without an associated probability. We choose to interpret this as the most probable match given the shape constraints and the current boundary. This then provides a global constraint on the matching, with the most probable XY_S^D providing both the local and semi-local constraints on the matching.

We need to combine the boundary estimates from the shape and the image network, whilst maintaining probabilistic consistency. Conceptually we have a network as shown in figure 7.2, where the matrix is defined over the k pixels that make up the image. As

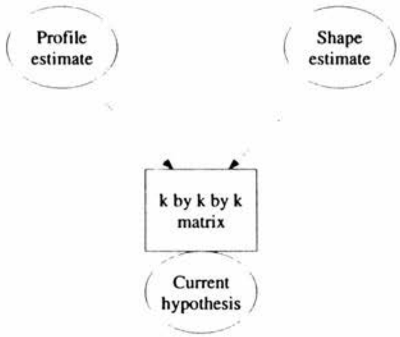


Figure 7.2: Conceptual network for boundary estimate combination

the profile estimate for a particular boundary point XY_s is defined only across the orthogonal set XY_s^D and the shape constraints provide only a single match point $G(XY_s)$, the majority of the probabilities, for the k points that theoretically make up the domain of each of the estimates, are undefined. We choose to solve this in terms of the maximum probability $PS_{max}(XY_s^D)$ and the single shape match point $G(XY_s)$. We can fit a probability distribution to each of the estimates such that it is maximal at that point and decreases in all directions with respect to the distance from the estimated point, reaching zero at infinity. If we do this for both estimates, we can construct a third distribution by taking the product of the distributions. The highest value in this distribution is the best combined hypothesis. The two estimates can be differentially weighted, for instance by altering the variance of a fitted normal distribution. In our work we have implemented this as a simple weighted average of the two estimates.

Although we have chosen to implement the shape constraints and the combination function outwith the belief network, and to reuse the network each iteration, there is in principle nothing to prevent the entire process being implemented as a single network. A simplified network is shown in figure 7.3, where each iteration is represented by a distinct time slice, dividing the network vertically across all boundary points, and each boundary point is represented horizontally in the network, dividing it across the iterations.

7.1.2 Grey-level Profile Model

In this application an average grey-level profile model for a given boundary is created, that is, a set of sample profiles along orthogonals to the boundary are combined to form a single average grey level profile model of the boundary. The set of sample profiles are derived from training examples in which a boundary is drawn by a user as an exemplar. This training set can be built up from a number of boundaries in a number of images in order to form a representative profile model. The number of orthogonals and the spacing and number of sample points along each orthogonal, which are determined by the user, are kept constant within the training set and between the training set and the cue. In cases where the boundary profile is asymmetrical the direction in which the boundary is drawn relative to the boundary profile must also be kept constant. In principle it would be possible to train a model with a different profile for each point if an average profile is unrepresentative.

To generate a set of likelihoods of the boundary being located at a particular sample point on an orthogonal, the Mahalanobis distance [Mahalanobis 36, Cootes & Taylor 92] is computed between the stored profile model and the actual profile across that orthogonal. This generates a set of likelihoods across the sample points at the orthogonal that can be input into the belief network.

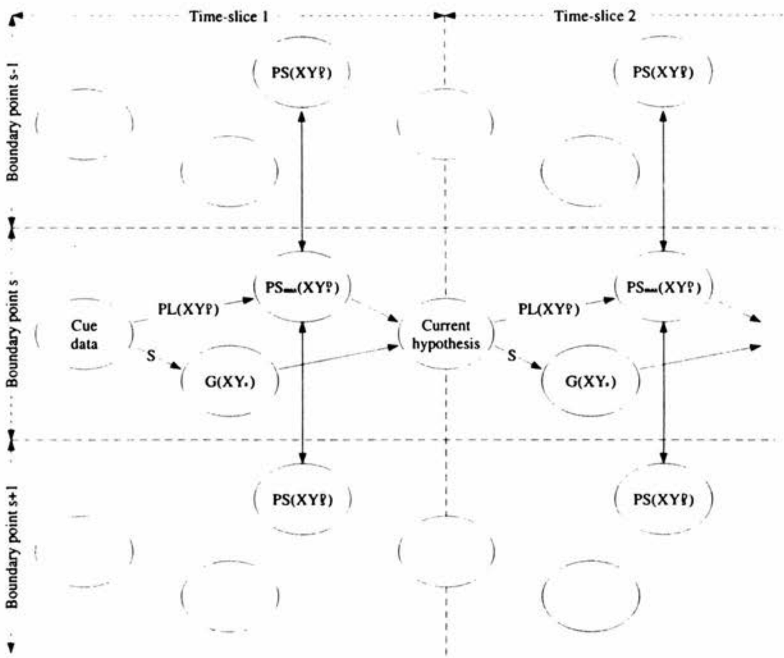


Figure 7.3: Time-slice network for iterative refinement

The belief network contains a node for every orthogonal in the model. The sample points are each represented by a state at the node. Probabilities in the network are therefore distributions across these points and it is these distributions that are propagated between nodes. The belief network outputs a likelihood for the sample points at each orthogonal and the set of the most likely points is the best estimate for the model. The profile constraints are a one dimensional model, the belief network links the results of a series of one dimensional matches to form an estimate based on the set of orthogonals. The probabilities at one orthogonal influence the probabilities of all the other orthogonals through network propagation.

7.1.3 Geometric Shape Model

The shape model is also derived from a set of training examples. The method used is the Point Distribution Model [Cootes *et al* 92]. This defines a shape by an ordered set of xy coordinates where each coordinate is associated with a particular feature on the boundary. In our work the points are evenly spaced along the boundary and the start point and boundary direction are kept constant. This ensures that there is a fixed relationship between the points in the shape model and the nodes in the belief network, which is crucial to the combination method selected. The point distribution models of each shape in the training set are scaled, translated and rotated until they correspond as closely as possible. A mean shape is calculated and principal component analysis is used to determine the major modes of variation. The variation is described by a set of parameters, each with a mean value and standard deviation, a set of weights that define the stability of each point in the line, and a mean line. It is these parameters we use to test and adjust the shape of a line with respect to a given shape model.

In addition to these shape parameters, a *scale range*, and *edge-definition* measure are calculated from the training set. The scale range imposes constraints on the acceptable size of the image line relative to the model line. The edge-definition measure is composed of the mean and standard deviations of the variance across every orthogonal in the *shape* training set. This measure is only really of use when the profile of the boundary is well defined. In these cases the edge-definition measure ensures that the hypothesised boundary is similarly well defined. It prevents the system accepting a boundary that satisfies the shape model and is maximal with respect to the limited search space defined by the orthogonals, but is not sufficiently well defined. It is expressed in terms of a number of standard deviations from the mean, defining an acceptable range within which the mean variance across the orthogonals on the boundary must lie.

To test the shape constraints for a given line, the translation, rotation and scaling that result in the best match to the mean model line are calculated. The set of parameter values that describe that line are extracted. If all the values lie within some pre-determined number of standard deviations from the mean parameter value, then the line satisfies the shape constraint. If a parameter value fails to satisfy its constraint, it is replaced with the mean parameter value. If the scaling factor lies within the acceptable range, then the line satisfies the size constraint. If the size constraint is not satisfied, then the closest acceptable scaling factor that satisfies the constraint is substituted for the scaling factor. If either constraint fails, the modified parameters are used to generate a new line and rotate, scale and translate it to best match the current line. The coordinates of this new line are taken as the best estimate based on shape and size constraints.

7.1.4 Combining the Best Estimates

If a line satisfies both the shape and size constraints then no shape or size constraints are applied to the best estimate which is determined solely by the grey-level profile constraints. If the shape and size constraints are not satisfied then the best estimate based on the grey-level profile constraints and the best estimate based on the shape and size constraints are combined to form a new best estimate. The combination function is a weighted average of the two estimates for each point on the boundary. The weight by which the grey-level profile constraint based estimate is multiplied is referred to as the *edge weight* and the shape and size constraint based estimate weight as the *model weight*. Once the coordinates of the new best estimate have been calculated a smooth curve is fitted between the points. The orthogonals are replaced at their new position so that their midpoint lies on the line. From this it can be seen that the relationship between the belief network and the image is dynamic, with the network following the refinement of the line.

7.1.5 Halting Conditions

The iterative refinement process continues until the matched line has stabilised. In order to be deemed stable the new set of matches must be within a specified distance of the current points. This distance, specified in the *stability* parameter is expressed as a number of pixels. The sum of the differences between the xy coordinates of each point on the current line and the xy coordinates of the same point in the new best estimate is calculated. This value, the *distance value*, is then compared with the stability threshold, if the distance value is less than or equal to the threshold then the constraint is met. The threshold for this is currently determined by the user, as we do not have a method for establishing it from the training set. In addition to having a low distance value, for a match to be considered stable the following must also be true:

- The shape constraints are satisfied.
- The size constraints are satisfied.
- The mean edge-definition is acceptably high.

The edge-definition constraint (described on page 136) sets the number of standard deviations from the training set mean edge-definition within which the mean edge-definition of the boundary must lie for the constraint to be satisfied. The purpose of this constraint is to ensure that the boundary is lying on a sufficiently well defined edge.

7.1.6 Interfaces

There are four main methods for interacting with the system, the model definition file `model.p`, the text interface, the image interface and the network interface. The relationship between these methods, the user and the network is illustrated in figure 7.4.

The file `model.p` contains a set of variables that determine the form of the network when it is created and the values of various parameters that govern its behaviour. This file is read each time a network is created.

The window from which the system is called becomes the text interface. It is used solely for messages from the system to the user.

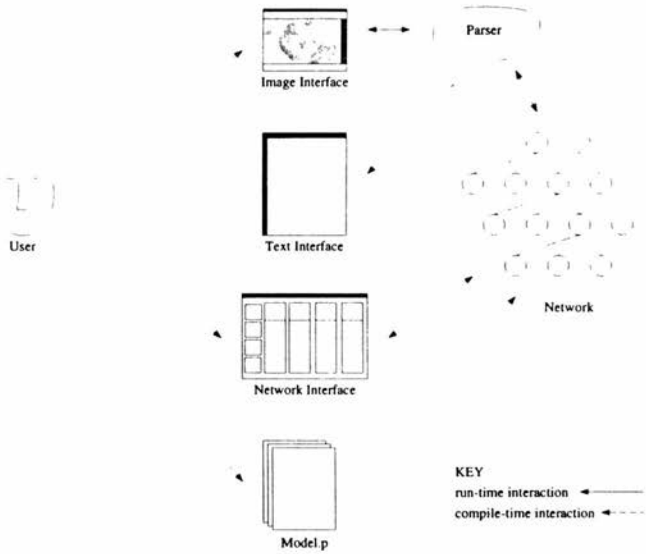


Figure 7.4: Interface overview

Calling the system creates the image interface. This is used to display the image and boundary lines. It provides facilities for the user to send data about the lines to the network, to receive results from the network, and to change the values of some of the parameters that govern the network's behaviour. Replies to image interface events are displayed in the text interface.

When a network is instantiated (created and sent data) the network interface becomes activated. This allows the user to examine the values of nodes within the belief network and to change those values.

7.2 Types of Network Model

Each orthogonal on the boundary is represented by a cluster of nodes in the network, including data entry nodes and nodes which calculate the the probability distribution over the points along the orthogonal. Each cluster is linked to the clusters representing the two (or one in the case of an end point) adjacent orthogonals on the boundary. One of the clusters is designated the *parent* and the other the *child* such that traversing the clusters in one direction along the boundary is always parent to child and in the other direction is always child to parent.

Currently there are four main variables in the model definition, the type of nodes, the type of line, the propagation direction and the propagation control or convergence criterion.

There are two types of node that can be used to calculate the probability distribution in the network, *delta nodes* and *composite delta nodes* (or composite nodes). These node types use two different algorithms to calculate the distributions which are propagated through the network to neighbouring clusters.

The type of line is either *open* or *closed* indicating that the boundary is a line with two distinct end points or a closed loop respectively. In a closed boundary every cluster has two adjacent clusters, in an open boundary the two end points have only one adjacent cluster.

Propagation direction (or type) controls the direction in which values are propagated through the network with respect to the parent child relationship.

Propagation limitation is used to control the distance over which probability distributions are propagated through the network. This may be desirable for reasons of efficiency, or necessary to halt propagation on convergence in closed boundaries.

By selecting different combinations of these variables and the other model parameters a variety of networks with different behaviours can be created.

7.2.1 Network Model

The network model is created from the duplication of four or five node clusters. Each orthogonal on the line has an associated cluster, as illustrated in figure 7.5. The nodes that make up a cluster are:

P - Image node — this node is a local evidence node. It represents the likelihoods across a single orthogonal due to the profile data. It has states corresponding to the sample points. Its lambda input is the likelihood vector across the sample points calculated using the Mahalanobis distance.

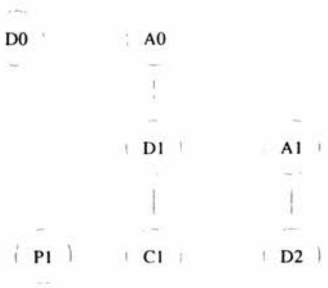


Figure 7.5: Single orthogonal node cluster with connections to adjacent orthogonals

C - Cue Node — this node is also a local evidence node and has states across the sample points. It was originally used to provide additional local evidence, for instance shape data or initial cue data. It is not used in the examples presented in this thesis and is set with a constant probability across all offsets.

A - Angle node — this node embodies the conditional probability matrix between delta nodes and between composite delta nodes. It is represented explicitly partially due to historical reasons and partially to explore the use of different node types. The function of the angle node is illustrated in figure 7.6. The vertical lines represent two neighbouring orthogonals, each with seven sample points (-3 to 3), the boundary line runs between the 0 sample points. The 'fan' between them indicates the strength of the probabilistic relationship, with the darker colour indicating a stronger relationship. Areas that are not coloured have zero strength. The angle nodes store the definition of the strength function. When propagating from one orthogonal to the next, the strength function is applied to each sample point in turn. The underlying model is that the line should be smooth with respect to the original cue and this is implemented as a tendency to maintain a small angle change at any node.

D - Delta node — this node calculates the distribution across the sample space of the point to which it relates. It is connected to delta nodes representing the two adjacent points on the boundary. It combines the local evidence for the sample space with evidence from the other delta nodes in the network. This node has states across the sample points. The inference algorithm used is that of Pearl [Pearl 88b,

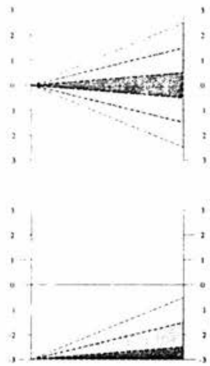


Figure 7.6: Angle node function, the upper image shows the function centered on sample point 0, the lower shows it centered on sample point -3

chapter 4] which calculates standard *a priori* probabilities given the evidence.

CD - Composite delta node — this node performs the same function as a delta node, except that it uses a different inference algorithm [Pearl 88b, chapter 5]. This algorithm is based on the notion of a *composite explanation* where a measure of *belief commitment* is used to identify an optimal set of jointly accepted propositions. The practical difference is the propagation of *maximum* probabilities rather than *average* probabilities.

The two types of delta node calculate different but related quantities. The normal delta nodes estimate the current individual degree of belief for each orthogonal independently, *i.e.* the maximum belief records the most probable position for that orthogonal given all the input data. The composite delta node on the other hand, records the best overall explanation of the data in terms of a set of orthogonals. Pearl [Pearl 88b, chapter 5] shows that for a set of nodes, the most likely explanation, *i.e. a set* of values (one for each node), is not in general the same as the set of values that maximises the belief at each node. This is essentially because of the conditional dependencies. Pearl also shows that the best explanation of all the data can be established using a modified form of the propagation mechanism which we have implemented as the composite delta node.

Either or both of these delta node types can be included in the network and can be used to determine the best fit for the model. It should be noted that the calculation of belief by the delta nodes is independent of the calculation made by the composite delta nodes. The relationship between these two networks is illustrated in figure 7.7, note that $P1$, for example, is the *same* node in both networks, there are not two nodes of the same name. Composite delta nodes send messages only to neighbouring composite delta

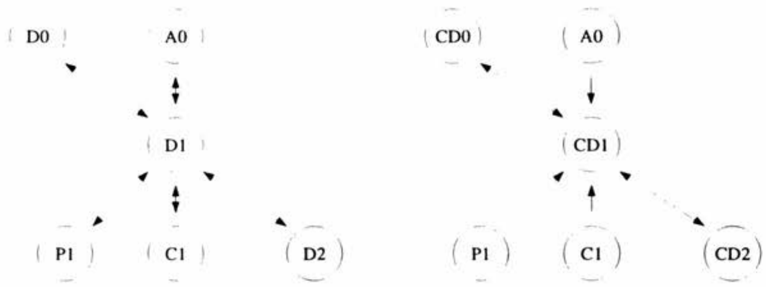


Figure 7.7: Delta and composite delta network relationship

nodes, the links from cue, image and angle nodes are one way only. This is purely for the sake of computational convenience as, for this application, we are interested only in the results at these nodes. There is no interaction between delta and composite delta nodes. Messages are only sent to a composite delta node if they carry new evidence from a cue or image node or a change in the relationship at an angle node.

This limited interaction is illustrated in figure 7.8, where solid nodes indicate new evidence, solid lines represent message passing in the indicated direction and dashed lines indicate no message passed. In the first pair of networks evidence entered at the cue node, *C1*, is propagated to both the delta node, *D1*, and the composite delta node, *CD1*. *D1* propagates the evidence to the image *P1* and angle *A0* nodes as well as through the delta network *D0* and *D2*. *CD1* propagates the evidence solely through the composite delta network *CD0* and *CD2*. In the second pair evidence from a delta node, *D0* is propagated through the delta network *D1* and *D2*, and to the image *P1*, angle *A0* and cue *C1* nodes. There is no propagation through the composite delta network. In the third pair of networks evidence from a composite delta node *CD0* is propagated through the composite delta network *CD1* and *CD2* only.

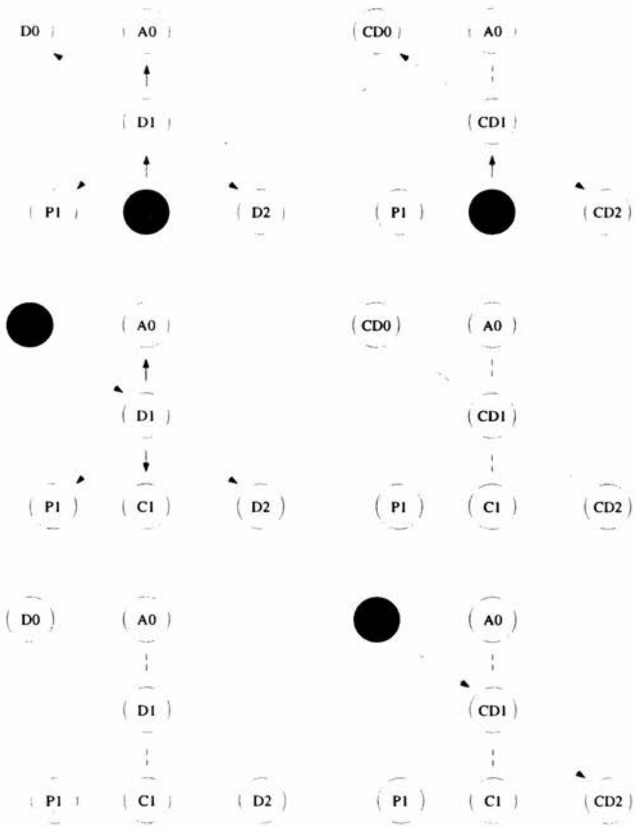


Figure 7.8: Propagation in delta and composite delta networks following new evidence

7.2.2 Line Types

Two types of model line can be selected, closed-line and open-line. An open-line model is used for a boundary that does not form a closed loop, its component nodes are shown in figure 7.9. The initial end-point, $D0$, has no angle constraint imposed upon it, and the final end-point, Dn imposes no angle constraint on any other point. In the closed-line model, shown in figure 7.10, the path between the deltas is continuous, there are no end-points. All points impose an angle constraint upon their direct successor and have an angle constraint imposed upon them by their direct predecessor. The closed-line model is identical to the open-line model, with the final end-point being connected to the initial end-point as its predecessor.

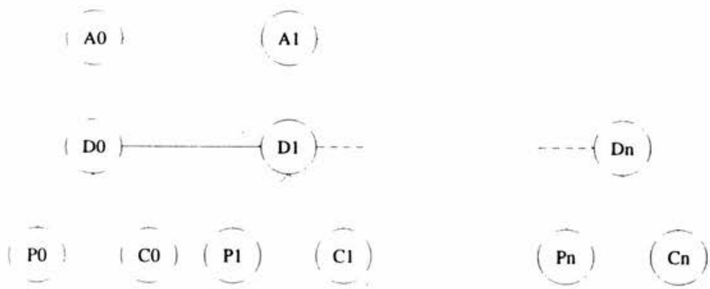


Figure 7.9: Open-line network model for non-closed boundaries

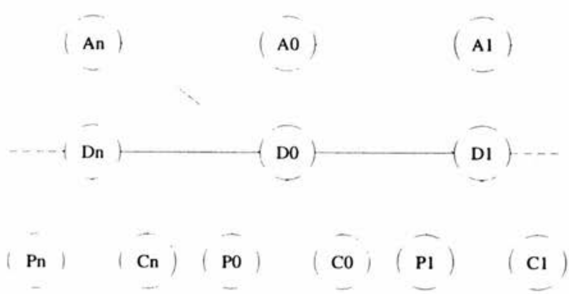


Figure 7.10: Closed-line network model for closed boundaries

Connections between composite delta nodes are the same as those of the delta nodes they shadow. Thus if the delta nodes form a closed-line, so will the composite delta nodes.

7.2.3 Directing Propagation

It is possible to select the direction in which messages will be propagated from one delta to the next. Messages can be passed from parent delta to child delta only, from child delta to parent delta only, in both directions excluding leaf and root nodes, or in all directions (this is the only completely consistent scheme for a belief network). As the probability distribution calculated at the delta node is the same under all these propagation schemes, bidirectional exclusive propagation is typically selected as it reduces message propagation overheads and the number of calculations that must be performed. These propagation schemes are illustrated in figure 7.11.

7.2.4 Limiting Propagation

Propagation limitation methods are employed for efficiency and to terminate infinite circular propagation on convergence when using closed-line models. Three types of limit on propagation from a node are available; decay, distance and circuits.

Decay continues to propagate messages from delta to delta, so long as there is sufficient change in belief. The old belief vector is compared to the new belief vector using the product moment correlation coefficient. If the value is greater than a user defined threshold then no propagation occurs.

Distance limits the propagation to a set number of delta nodes in every direction that propagation is allowed.

Circuits limits propagation to a set number of circuits (appropriate for closed-line models) in every direction that propagation is allowed. Again this is counted at the delta nodes.

It is possible to select any combination of limits, including none, but they are always checked in the same order, namely; decay, distance, circuits. If any limit has been reached then propagation is halted. There is an independent propagation limitation applied to composite deltas that functions in an identical fashion.

Limiting propagation for efficiency using arbitrary measures, such as distance and circuits above, could affect the probabilistic interpretation of the propagation results and

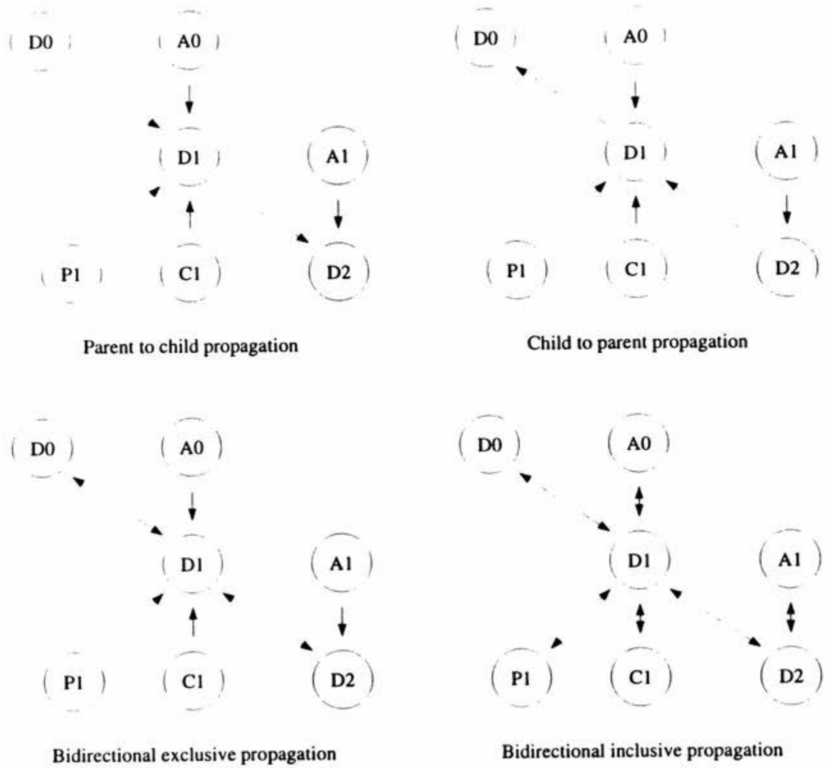


Figure 7.11: Propagation schemes

the results themselves. The decay parameter, based on a measure of change in information is dynamic, varies according to the information being propagated and maintains probabilistic integrity.

7.3 System Development

To develop and test the system, a number of ultrasound images with similar characteristics were selected. These images were then normalised to give an even distribution of pixels across the available grey-levels, and had their grey-levels inverted, resulting in dark features on a light background as shown in figure 7.12. In order to assess the

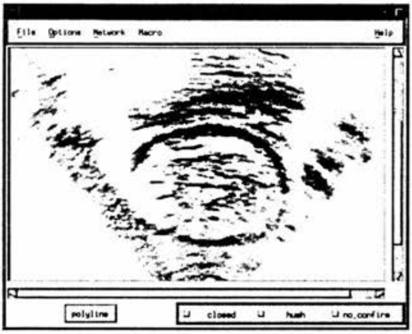


Figure 7.12: Ultrasound image of fetal skull

performance and the failure modes of the system we initially applied it to less complex images[1], microscopic images of sections of mouse embryo. The mouse embryo images are part of a large scale three dimensional reconstruction project which will require model based matching of boundaries to enable efficient manually guided segmentation. The techniques presented here are also likely to be of use for that project. The external edges in these images are well defined, allowing the system to be tested within a simpler environment. A typical mouse image with an initial cue is shown in figure 7.13.

A *grey-level profile model*, *profile map* and *distance map* for the line in figure 7.13 are shown in figure 7.14. The grey-level profile model is the average grey-level profile for the boundary edge calculated from a set of training examples. The profile map shows the

[1]The structure visible in the microscopic images is more complex than in the ultrasound images, however the signal to noise ratio is higher and therefore the image processing task is easier.

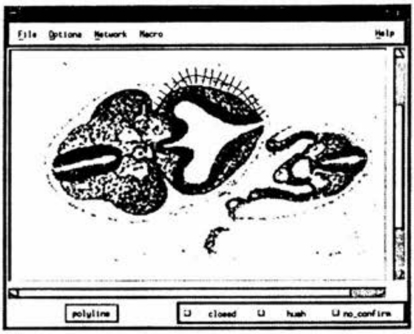


Figure 7.13: Mouse embryo image

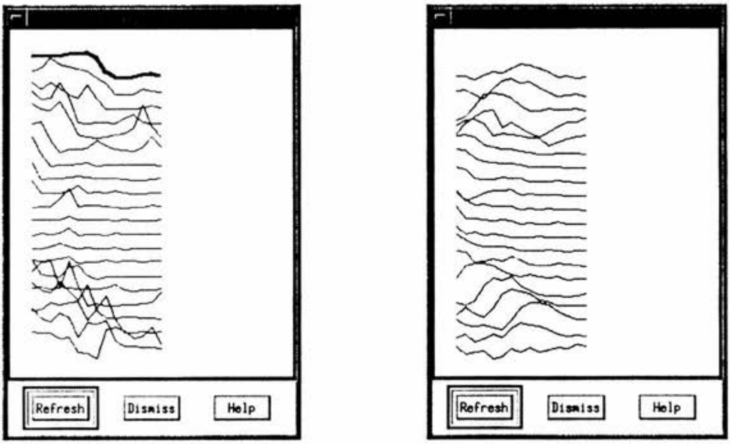


Figure 7.14: Grey-level profile model (bold line), profile map and distance map

actual grey-level profiles from the image along each orthogonal. The distance map shows the calculated Mahalanobis distances that result from applying the grey-level profile model to each of the profiles in the profile map. The distance is calculated for every sample point along the orthogonal.

The training set for the grey-level profile and shape models each comprised ten example lines similar to that shown in figure 7.15. The shape model was created after the grey-level model, using a different set of ten lines. The grey-level model must be constructed first as it is used in the calculation of the edge definition measure.

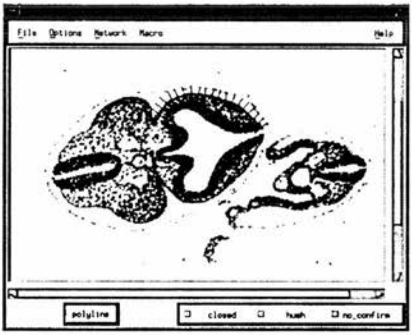


Figure 7.15: Mouse training line example

Except where stated otherwise, all the refinement results have been produced using the same grey level profile model and shape model. The system also has a number of parameters that can be defined, for the mouse images the values used for the most important parameters were as shown in table 7.1.

During testing we found that the best estimate line would occasionally become fixed at a point that did not satisfy the halting conditions, or would get caught in a cycle. In order to resolve these situations a random shift was introduced. Every point on the line is moved a bounded random distance in the x axis and an independent bounded random distance in the y-axis. The line is shifted whenever the current position of the line is judged to be the same as a previous position of the line and the halting conditions are not satisfied. Two lines are judged to be the same if the xy coordinates of the ends of the line (or the initial point if a closed line) are the same and the distance between orthogonals is also the same. These conditions would not distinguish between two lines that were a

parameter	*value*
node type	delta node only
line type	open
propagation type	bidirectional exclusive
propagation limitation	decay 0.95
	distance 25
edge weight	0.5
model weight	0.5
stability	$\leq$ 2 pixels
edge definition	3
orthogonals	20
sample points	15

Table 7.1: Mouse embryo network parameter values

reflection of each other about a straight line drawn between their co-occurring endpoints. Each time a shift is performed the record of previous line positions is discarded. Due to this random element, the refinement of a line that includes a shift will generally produce different results each time, even thought the starting conditions are the same. This can drastically change the number of iterations taken to reach a solution, and possibly whether or not a solution is reached at all. When describing a result, a pair of numbers will be used, *e.g.* (23-2) or (63$*$12), which denote the number of iterations performed and the number of times a shift was performed. An asterisk $*$ indicates that the process was stopped by the user before the halting conditions were met. As an illustration of the above, figures 7.16, 7.17 and 7.18 show three refinement runs on figure 7.13 under default conditions. In figure 7.22 a solution in (40-2) is shown, the shift occurred at (35-0), another refinement under the same conditions reached a solution in (102-11) indicating the effect that these random shifts can have on the refinement process.

We performed a number of trials using the mouse images to determine whether the system could adequately deal with the range of expected errors. In particular we deliberately generated cues that violated the grey-level and shape constraints by not placing the cue on the boundary, and by defining a shape different to those in the training set (figures 7.19 and 7.20). Cues that failed to satisfy the size constraint were also used (figure 7.21). The refinement process is able to cope with a variety of errors in the initial cue, including loops, figure 7.22.

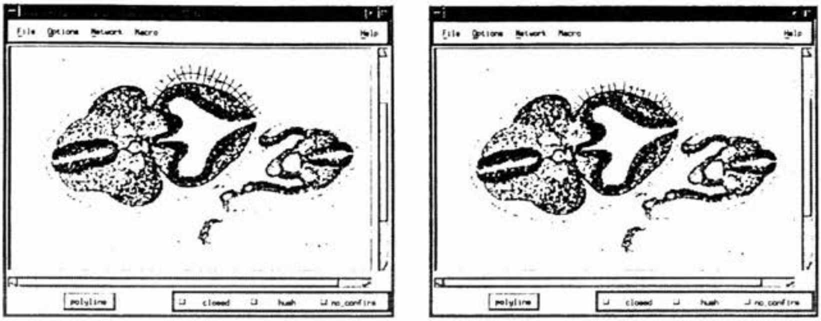


Figure 7.16: Initial cue and result (19-2)

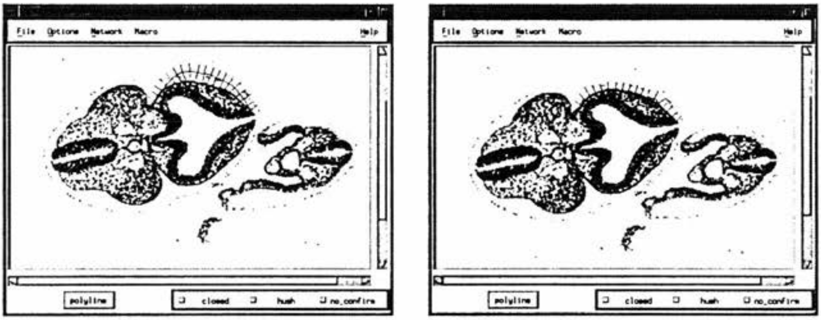


Figure 7.17: Initial cue and result (22-2)

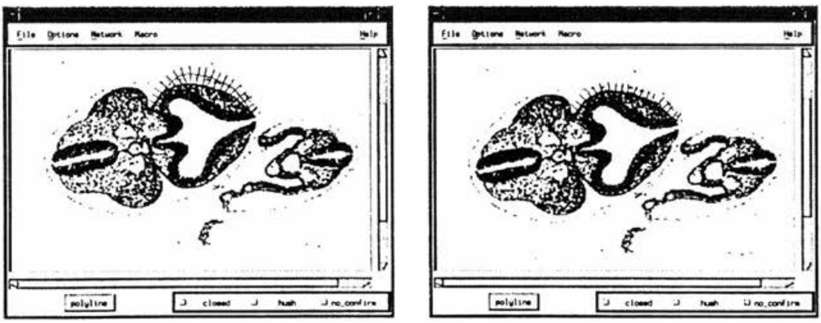


Figure 7.18: Initial cue and result (65*12)

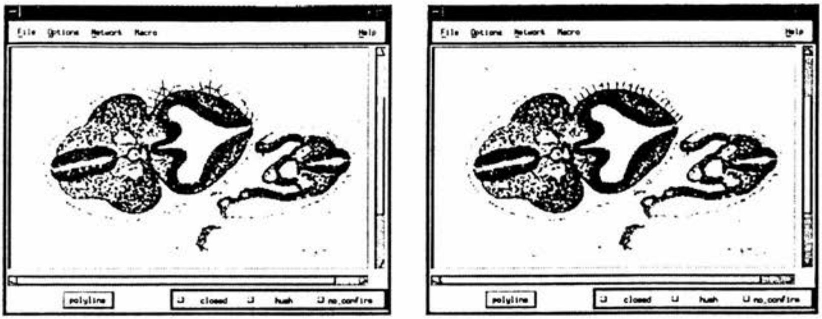


Figure 7.19: Initial cue and result (48-3)

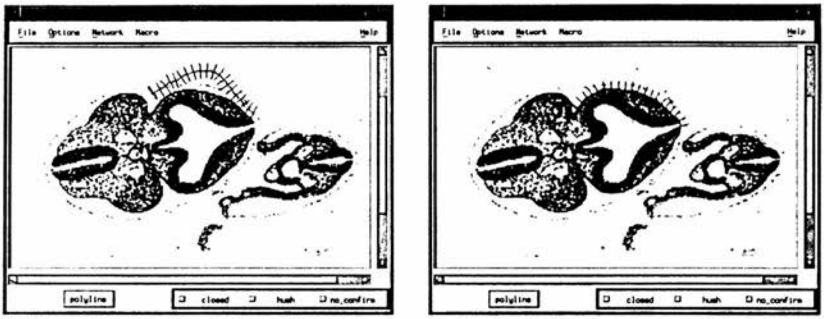


Figure 7.20: Initial cue and result (14-0)

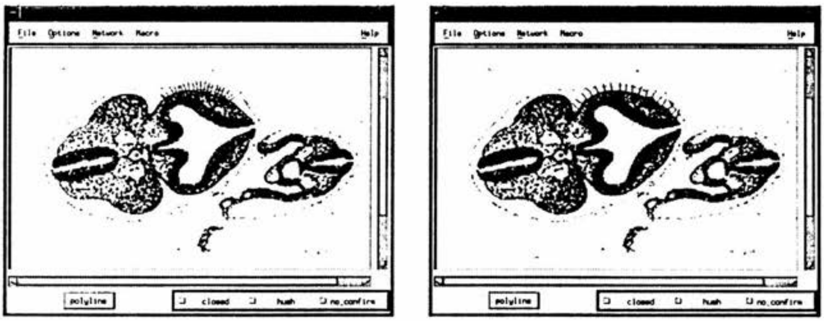


Figure 7.21: Initial cue and result (13-0)

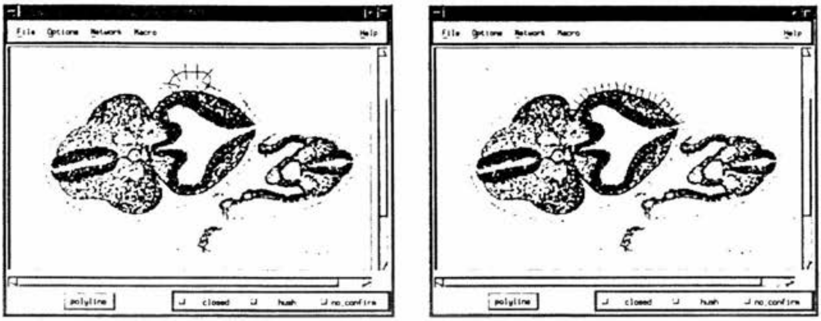


Figure 7.22: Initial cue and result (40-2)

The sensitivity of the refinement process to some of its default values was investigated using figure 7.16 as the test case. The variables considered were:

- Relaxing the definition of stability by increasing the distance value (pixel match)
- Varying the model weight used in the combination of the best estimates

The results of these tests are shown in figures 7.23 and 7.24. From the small samples used it is hard to identify any definite pattern in the results other than the correlation between the number of iterations and the number of shifts. It is also important to realise that these results are liable to be highly dependent on the type of images under consideration, specifically the utility of the grey-scale model relative to the shape model, the number of orthogonals and sample points, and the degree of fit required of the final line.

7.3.1 Initial Ultrasound Results

The initial application of the refinement system to ultrasound images used parameter settings identical to those used for the mouse images, except for those shown in table 7.2. The increase in the number of orthogonals and sample points was made on the basis of approximating the same density of points as used in the mouse images. The increased movement allowed reflected the increase in the total number of points.

A closed line model of the skull outline was created for both the shape and profile models, each being based on ten training lines similar to that shown in figure 7.25.

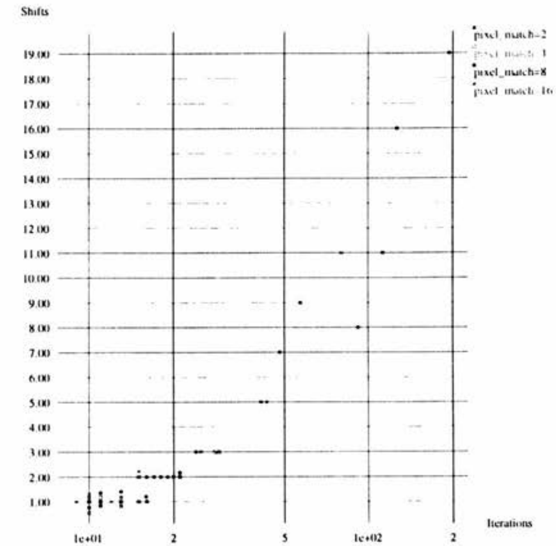


Figure 7.23: Pixel match sensitivity graph

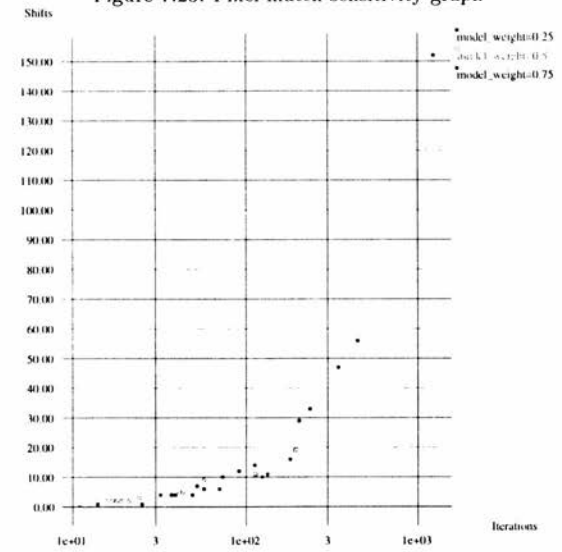


Figure 7.24: Model weight sensitivity graph

parameter	value
line type	closed
stability	$\leq$ 6 pixels
orthogonals	30
sample points	31

Table 7.2: Initial ultrasound network parameter values

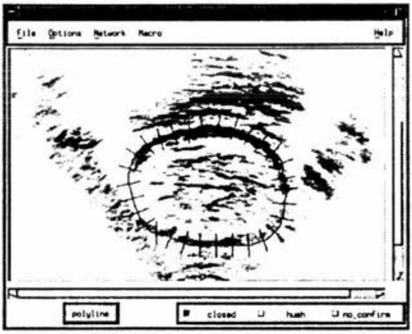


Figure 7.25: Fetal ultrasound training line

Initial results on these images were disappointingly poor, with the line typically wandering across the image, unable to fix on any position even when started close to the skull boundary, see figures 7.26 and 7.27 for example.

Possible causes of this poor performance included a breakdown of spatial conditioning for the closed-line model or grey level profiles providing insufficient information to anchor the boundary correctly. Tests conducted with open lines had similar or worse results to those using a closed-line, indicating that this was not the problem. In these ultrasound tests the propagation limits selected in fact prohibited feedback as the distance limitation is less than the number of nodes in the network. There are still advantages in representing a closed-line using a closed network rather than an open network as it places a constraint on the distance between the two end points that does not otherwise exist.

The attributes of the image and models were then considered. It was apparent that at least part of the problem was that the boundary of the fetal skull had significantly different grey-level profiles around its circumference, resulting in an average profile model that was a poor match for the boundary. The variation in the profiles is illustrated in

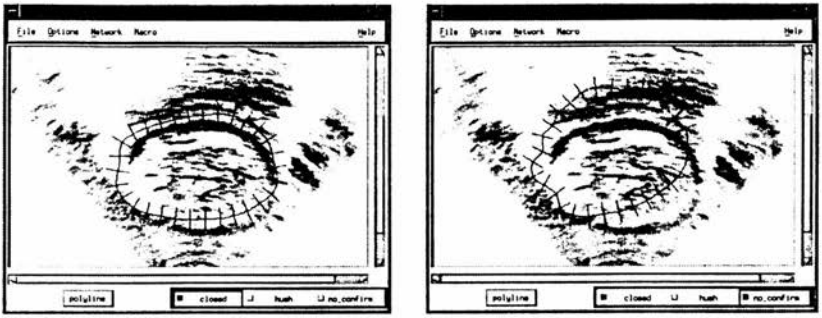


Figure 7.26: Initial cue and result (330*1)

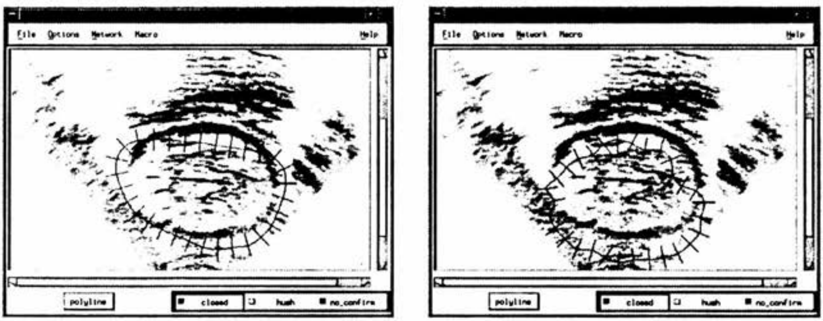


Figure 7.27: Initial cue and result (20*0)

figures 7.28 and 7.29, taken from the dark part of the skull (the top relative to the image) and the light part (the bottom) of the skull respectively. In both cases the profile model (shown in bold) is that generated from including profiles around the complete circumference. It can be seen that in neither case is the grey-level profile model a good match to the actual grey level profiles in the image. When the distances are compared with those from the mouse images (page 149), the relative lack of information in the ultrasound profile model becomes apparent.

Currently there is no provision in the system for the use of multiple grey-level profile models, so in order to make the grey-level profile information more effective, a new grey level profile model was created using only profiles taken from the dark part of the boundary. We hypothesised that the ability to identify even part of the boundary on the basis of its grey level profile should be sufficient to anchor the line in the image. Once the line is anchored, the shape constraints should ensure that the line is constrained to the region of the image that contains the less well defined boundary, enabling its poor profile match to provide the fine adjustment.

In addition to this change, several of the parameter values were varied in order to find a combination that would consistently produce acceptable results. The changes tried can be broadly classed as hard or soft. Hard changes were designed to force the line into the correct place, typically by increasing the model weight parameter. Soft changes generally involved relaxing the definition of a stable match by increasing the *pixel match* or the *edge definition* parameters. Various combinations of hard and soft changes were found to produce reasonably acceptable results, as illustrated in figures 7.30 to 7.33.

However, we were unable to find a single set of parameters that provided an acceptable refinement in all cases, as illustrated by figure 7.34. Similarly no set of parameters could be found that worked well when the test cues were applied to other images in the test set.

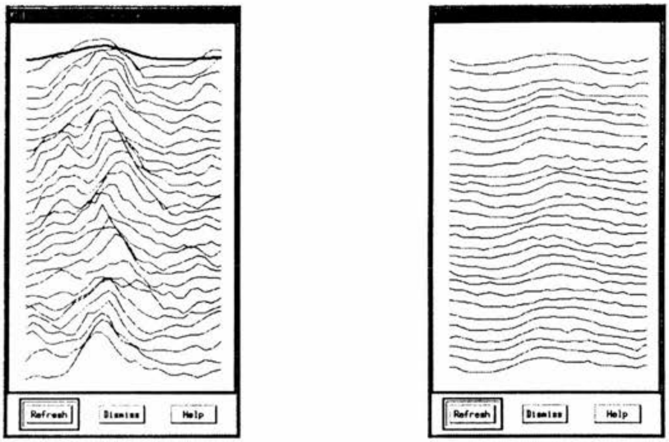


Figure 7.28: Dark grey-level profile map and Mahalanobis distances

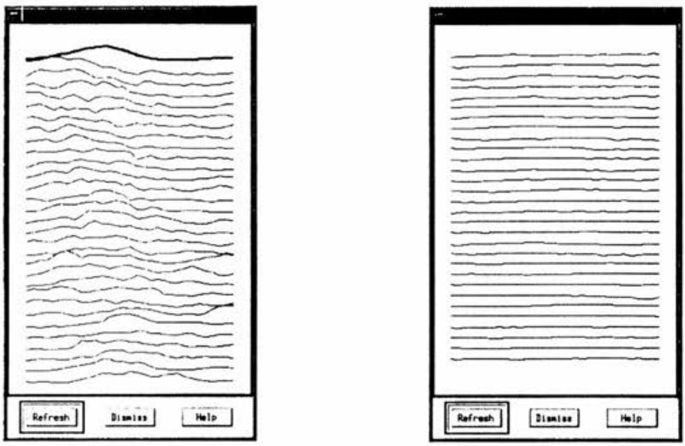


Figure 7.29: Light grey-level profile map and Mahalanobis distances

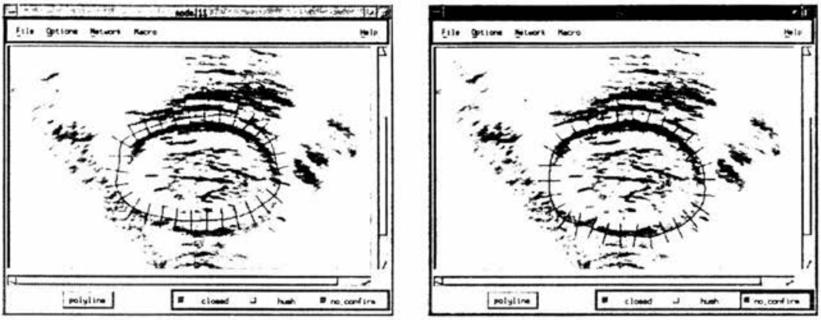


Figure 7.30: Initial cue and result (95-0) (model weight 0.75, pixel match 16)

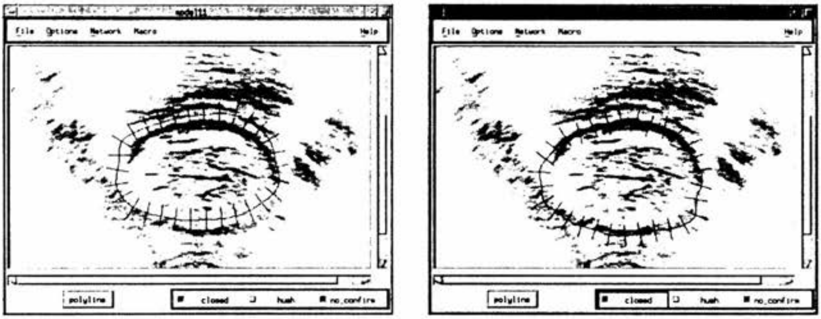


Figure 7.31: Initial cue and result (12-0) (model weight 0.7, pixel match 16)

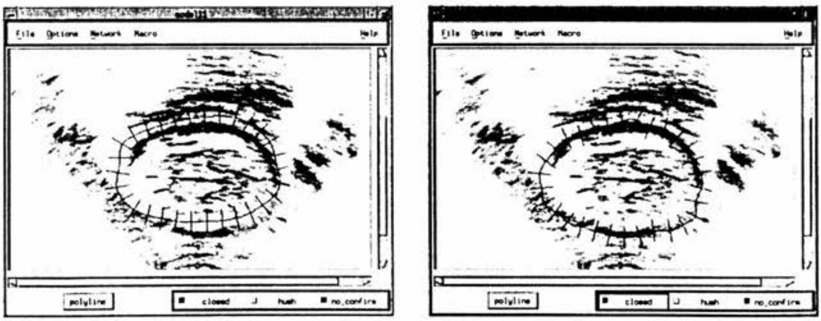


Figure 7.32: Initial cue and result (13-0) (model weight 0.7, pixel match 12)

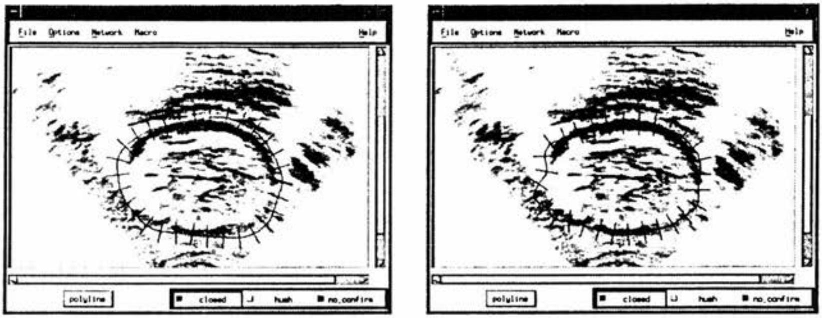


Figure 7.33: Initial cue and result (23-0) (model weight 0.5, pixel match 16)

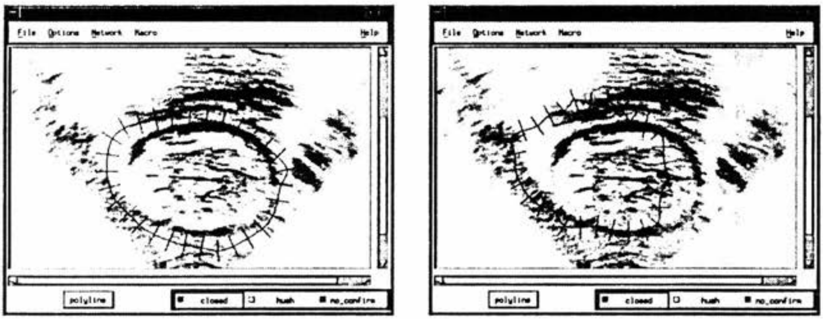


Figure 7.34: Initial cue and result (135*3) (model weight 0.5, pixel match 16)

7.4 Further Ultrasound Results

The failure to determine a set of parameter values that gave consistently acceptable results across a variety of initial lines and across different images required further consideration of the particular problems posed by ultrasound images. We first considered aspects of a single image and identified a number of features that could potentially cause the refinement process to fail, these are illustrated in figure 7.35.

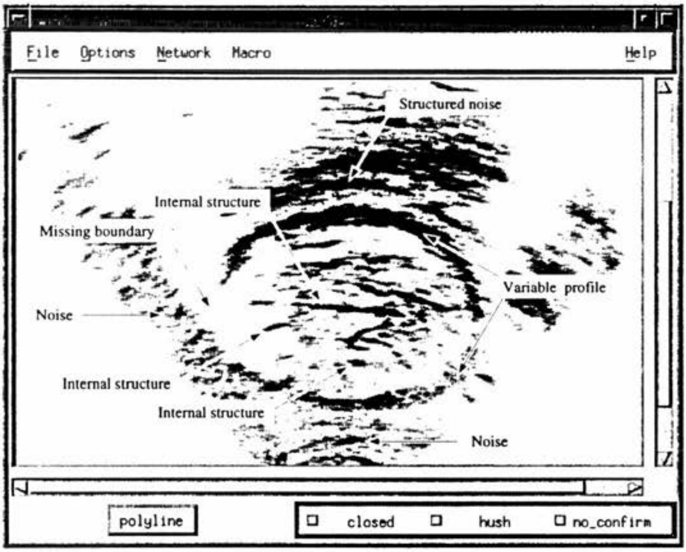


Figure 7.35: Fetal ultrasound image features

Noise The generally high noise level tended to obscure the true edge and generate spurious profile matches attracting the line in incorrect directions. Increasing the model weight could alleviate this to some degree, but in turn introduced other problems.

Structured Noise The presence of structured noise that mirrored not only the grey-level profile of the true edge, but also its shape was always going to be difficult to overcome. This problem was exacerbated by the use of only a partial profile model which made it difficult for the line to fix on the parts of the edge not well represented in the profile model.

Internal Structure This created similar problems to structured noise, presenting connected structure with a profile that reasonably approximates the profile model.

Missing boundary The total absence of a boundary at certain parts of the image often resulted in the line attaching itself to noise elements as the closest match for the profile model. Once this attachment occurred it was difficult or impossible for the rest of the line to exert enough influence to pull these points back again, particularly if the model weight was low.

Variable Profile This is probably the root of all the problems for this application, as we have discussed earlier. The necessary use of a partial boundary model greatly reduced the ability of the system to identify the correct boundary, particularly in view of the other features identified above.

In addition to these problems caused by the features of a single image, the consideration of different images introduced further problems due to the variability between them. Consider the test images shown in figure 7.36. Clearly these images vary considerably with respect to the problem areas identified above.

7.4.1 Revised Network Model

Following the analysis of the ultrasound images some changes were made to the grey-level profile modelling and to the function for combining the best estimates.

The change to the profile modelling allowed the creation of profile models that could be applied to profiles that were not of the same length as the model[2]. This allowed the use of long orthogonals to sample a larger area of the image. The problem with using correspondingly long profile models was that the grey-level profiles for the areas either side of the skull boundary were very different, both between orthogonals and between images. This change made it possible to construct profile models that spanned only the width of the skull boundary but still apply them to profiles that spanned a much greater part of the image. It was hoped that this would allow points lying some distance from their correct position to at least include their correct position within their local match domain.

[2]This was a technical constraint imposed for convenience in earlier versions of the program.

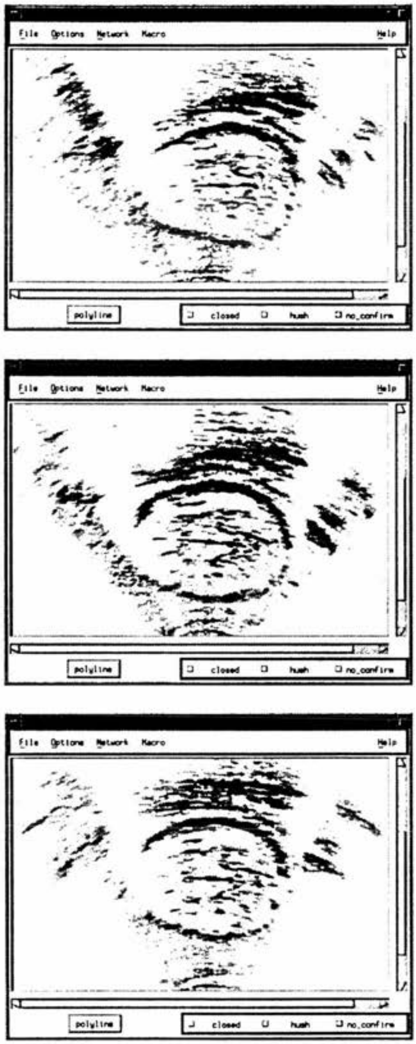


Figure 7.36: Similar ultrasound images

The changes to the combination function were primarily motivated by the recognition that in the ultrasound images it would be necessary to assume that the initial cue was reasonable. For the mouse embryo images the profile model provided such strong evidence that the system was able to cope with initial estimates that were very distant from the required result and in the mouse embryo refinement and the initial ultrasound tests this initial cue was discarded after the first iteration. If we assume that this initial estimate is in fact reasonable then discarding this information is clearly erroneous.

The second change in the combination function concerned refinement when the line satisfied the shape and size constraints. Previously the best estimate was simply that provided by the profile estimate. This was changed so that if the shape and size constraints were satisfied then the current position of the line was used as the best shape estimate. This was particularly relevant in the ultrasound domain where the profile data and the shape data were often at odds with each other and oscillations between a position estimated on the basis of profile only, and a position estimate including shape data was possible.

The new combination function was therefore a weighted average of the initial user line, the best profile line and the best shape line (either the current line or one derived from the imposition of the shape constraints). Additionally, separate shape model weights were provided for the case when the current line is used and for when the shape constraints are applied.

One further change was introduced in order to differentiate between the case where the most likely sample point according to the profile data was not a good match for the profile model (it is just the best match available), and the case where it was. To achieve this the most likely point location given the profile data was *weighted* by the Mahalanobis distance for that point (the distances lie between 0 and 1) as well as the edge weight. This allowed the shape model to move points that were poorly supported by the image data more easily than those that *appeared* to be on the true edge.

Using this new network it was possible to achieve reasonable results for both the delta and composite delta networks across the four test lines in all the three test images using the sets of parameter values shown in table 7.3. Some of the results of the experiments are presented in figures 7.37 to 7.42. A complete sets of results, including those shown

parameter	*delta value*	*composite value*
node type	delta node only	composite node only
line type	closed	closed
propagation type	bidirectional exclusive	parent to child
propagation limitation	decay 0.95	decay 0.95
	distance 25	distance 25
	circuits 2	circuits 2
initial cue weight	0.1	0.1
shape model weight	0.5	0.5
current shape weight	0.5	0.5
edge weight	1.0	1.0
stability	$\leq$ 16 pixels	$\leq$ 16 pixels
edge definition	3	3
orthogonals	30	30
sample points	59	59

Table 7.3: Revised ultrasound network parameter values

below, can be found in Appendix C.

The stability of these results arises from the full use of the probability information from the image matching and model prior knowledge.

We also ran some of the experiments again using the delta network, but with the only propagation constraint being decay (0.95). The results under the two regimes are shown in figures 7.43 to 7.45. Experiments comparing the limitation regimes using the composite delta network on the same examples are shown in figures 7.46 to 7.48. It should be noted that none of the refinements managed to satisfy the size constraint halting condition, regardless of the network model used. This is discussed below.

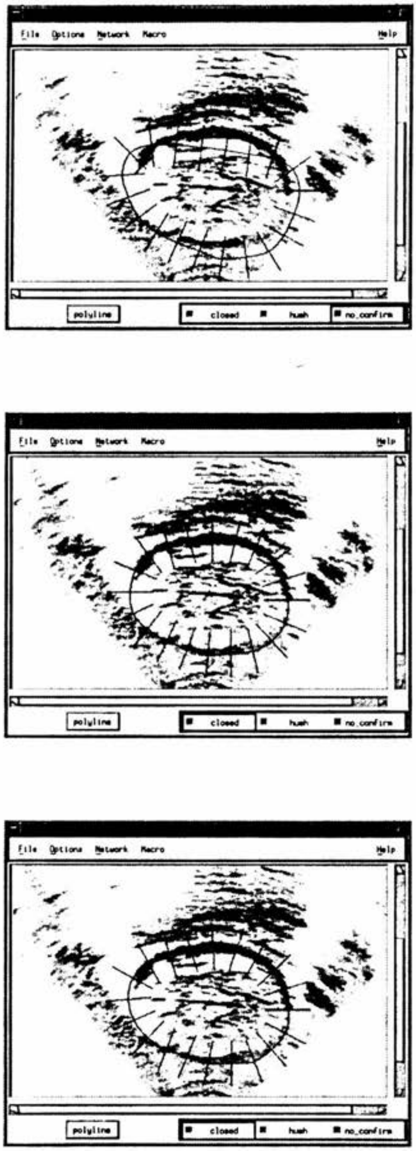


Figure 7.37: Initial image, result (149*13) and composite result (149*14)

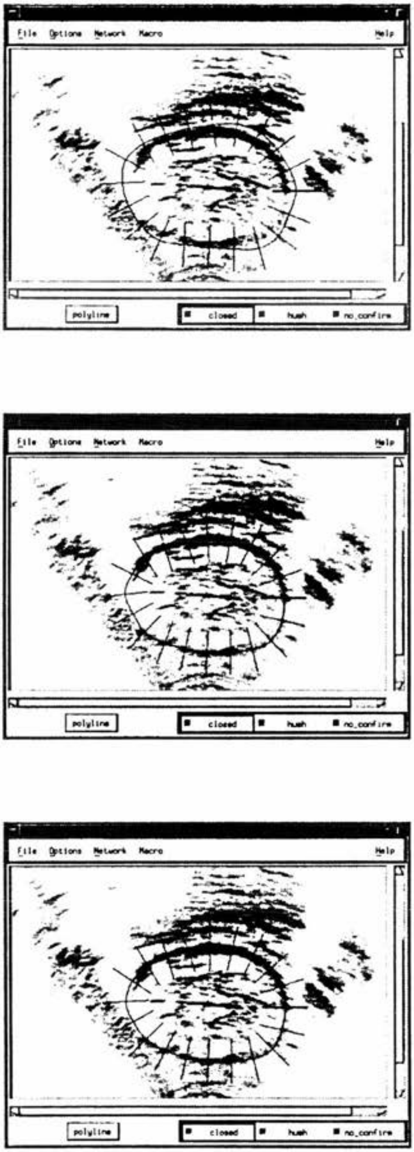


Figure 7.38: Initial image, result (29∗2) and composite result (29∗2)

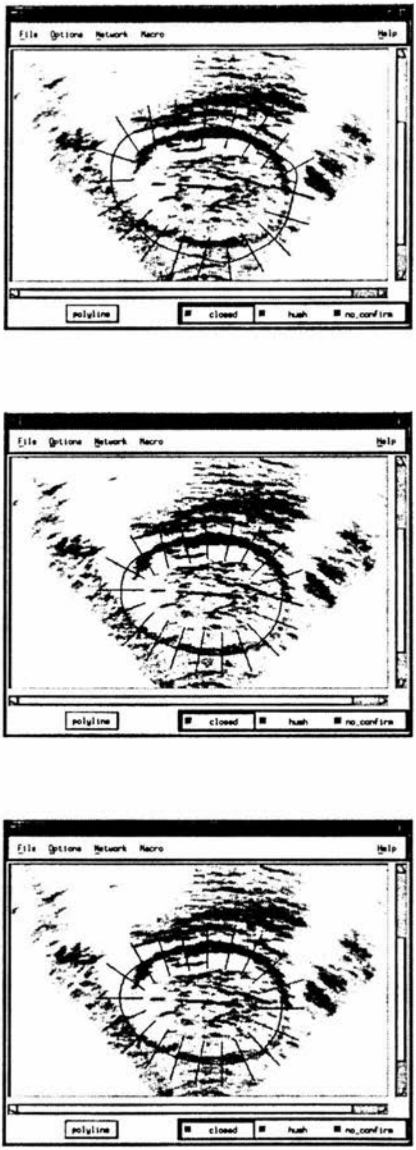


Figure 7.39: Initial image, result (659*59) and composite result (659*54)

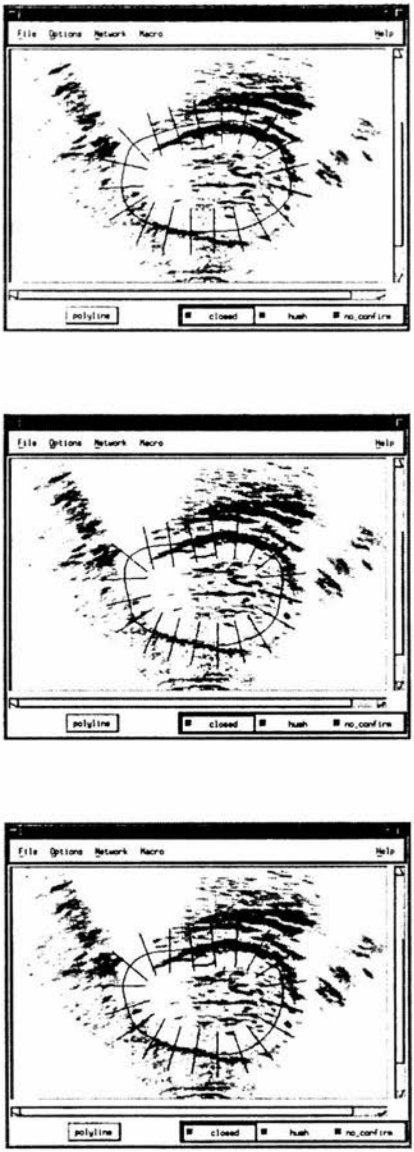
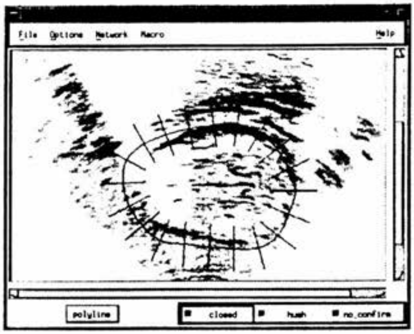
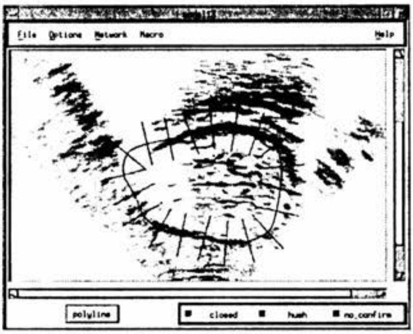


Figure 7.40: Initial image, result (134*4) and composite result (134*2)





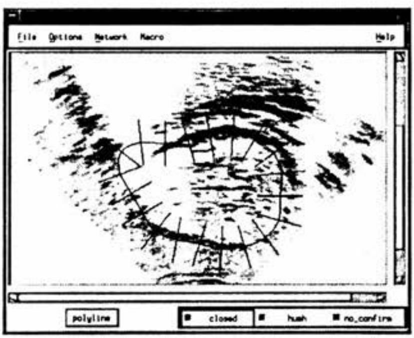


Figure 7.41: Initial image, result (209*7) and composite result (209*4)

Figure 7.42: Initial image, result (269*20) and composite result (269*10)

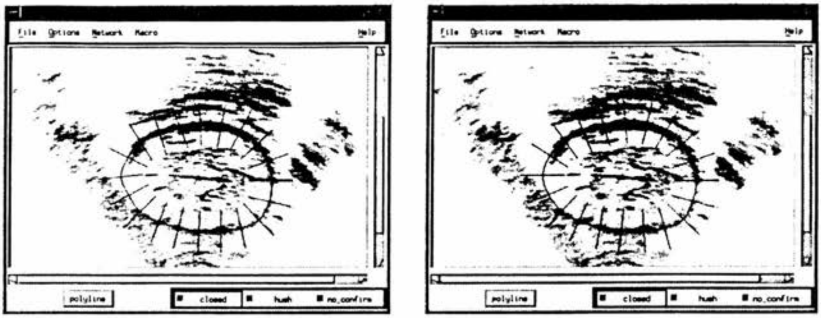


Figure 7.43: Delta network with arbitrary limitation (29*2) and decay only (29*4)

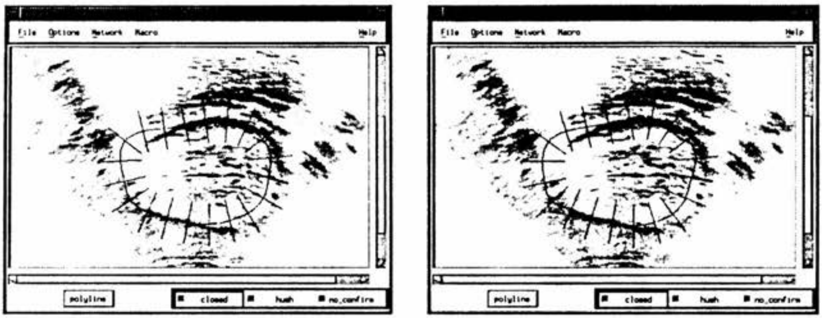


Figure 7.44: Delta network with arbitrary limitation (134*4) and decay only (134*4)

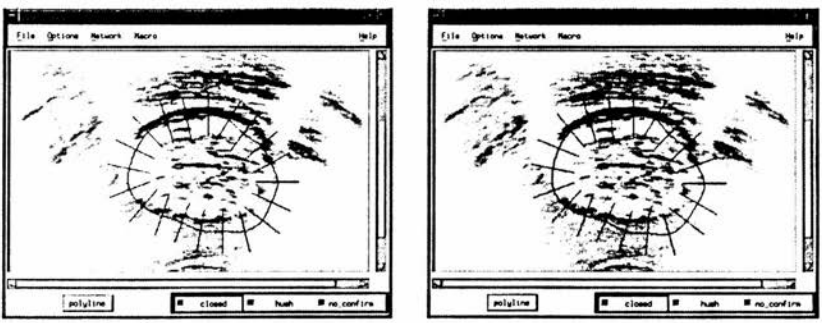


Figure 7.45: Delta network with arbitrary limitation (269*20) and decay only (269*16)

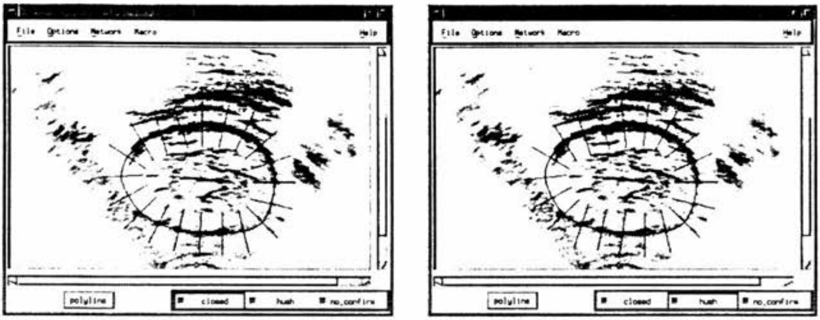


Figure 7.46: Composite network with arbitrary limitation (29*2) and decay only (29*1)

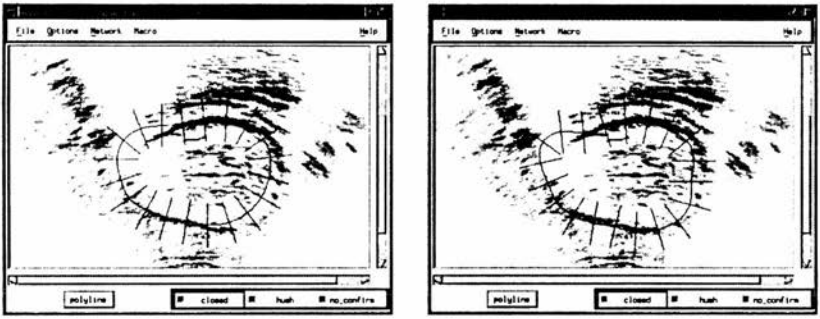


Figure 7.47: Composite network with arbitrary limitation (134*2) and decay only (134*3)

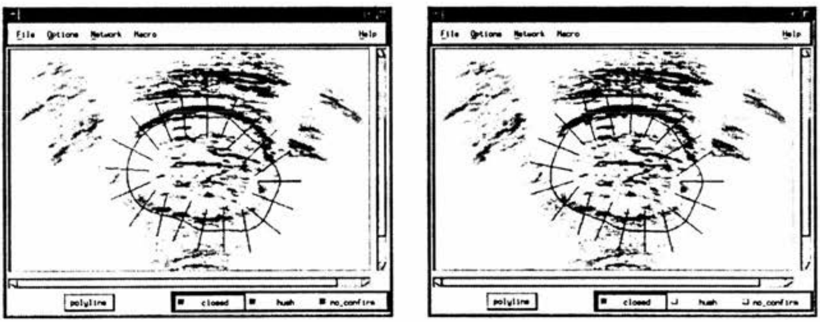


Figure 7.48: Composite network with arbitrary limitation (269*10) and decay only (269*10)

7.5 Discussion

The move from the mouse embryo images to the fetal ultrasound images demonstrated a number of weaknesses in the refinement system. Whilst the final results from the ultrasound images were better than the initial results suggested, the hoped for level of robustness was not achieved. However, the system does demonstrate the prime motivation for this work, that belief networks are capable of controlling complex image analysis tasks and can be modified easily and conveniently by the domain expert. Some particular points of concern with the *current* approach to the general problem include:

Network Propagation The current implementation has an undesirable feature which is best illustrated by an example, see figure 7.49. In the top image we see a line

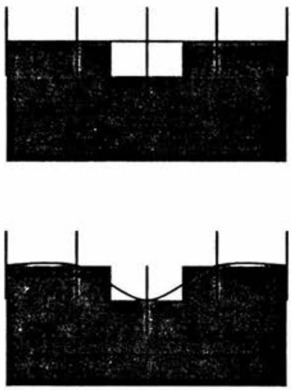


Figure 7.49: Nicked edge example

fitted to a boundary which has a nick in it. The orthogonal that lies across the nick is receiving conflicting information, on the one hand the profile information suggests it should move to the location shown in the lower image, on the other hand the influence of the neighbouring orthogonals suggests it should remain where it is. The influence of the neighbours is modelled such that a high probability at a particular sample point on one orthogonal is propagated as a distribution centred about the same sample point to the neighbouring orthogonals. In the top image this is the desired behaviour, as we are assuming that the edge is smooth and continuous we would expect the nick to be largely ignored.

In the lower image with the orthogonal in the nick, we might expect the neighbouring orthogonals to be suggesting that the orthogonal be moved to the position shown in the top image. In fact this is not the case, the reason for this is that *the network has no knowledge of the spatial relationship between orthogonals other than their ordering along the line.* As a result of this the neighbouring orthogonals in the lower image actually suggest that the orthogonal in the nick remain where it is, as the high probabilities of the centre sample points support each other.

On a related point, it is assumed that the orthogonals are approximately parallel to each other. This assumption is more or less false depending on the number of orthogonals and the degree of curvature along the line. It is not clear what effect this has in practice.

The profile model The current profile model is an average grey-level edge profile, formed from grey-level edge profiles taken along the extent of the edge. Three assumptions underlie this model:

1. The orthogonals are sufficiently long to incorporate the significant profile.
2. The profile is similar along the extent of the edge.
3. The averaging of the line does not destroy important fine-resolution detail.

Assumption 1 has occasionally proved to be problematic in the examples considered, particularly if the initial cue was poor.

Assumption 2 was violated in the ultrasound images, leading to the problems discussed earlier. In the ultrasound images it was possible to work around the problem to a limited degree by creating a profile model for part of the boundary and relying on the fact that the shape model is restrictive. In domains where assumption 2 was not met and the shape model was highly variable it is doubtful that the refinement could be made to succeed. In such domains it may be necessary to have multiple profile models, one for each distinct profile. These could then either be associated with only a subset of the orthogonals, or could provide competing likelihoods to the belief network.

Assumption 3 was not directly an issue in the examples considered though it is related to assumption 2 and a similar solution might prove appropriate. It may be possible to use profile models at different resolutions, the low resolution models being formed by averaging over a number of profiles and the high-resolution model being formed by maybe only a single profile. Initially the low resolution model would be used to position the line approximately, then the high resolution model would be used to perform the fine adjustment.

As the profile models will typically be constructed from a set of approximately aligned boundaries, it will never be an exact match to the edge and the final result of the alignment will always be dependent on the accuracy of the models with respect to the image data.

Parameter selection The system performance has proved to be highly variable depending on the parameter values used. This allows the system to be adapted to suit the particular images under consideration, as was shown in the ultrasound images. This information is part of the expertise within the image processing domain and needs to be elicited from the expert. In this case it may be necessary to introduce a learning technique, for example a neural network, to undertake this role.

Constraints The current constraints are embodied in the halting conditions as well as the belief network itself. The constraints are currently very simple, due to the limited sources of information available to the system, *i.e.* the grey level profile model and the shape model. The specific limitations of these two models are discussed above. More generally the fact that the refinement does not necessarily progress in the direction of the true boundary suggests that either the problem is underconstrained or that the available constraints are not being applied effectively. This is supported by the failure of the system to transfer well from the training images to new images.

The halting conditions do not necessarily define satisfactory refinements in the ultrasound images, though in the mouse embryo images they typically did. The edge definition measure has very wide bounds in the ultrasound images and is therefore easily satisfied by fairly arbitrary lines.

A richer variety of constraints may be necessary in order to overcome the problems associated with refinement in the ultrasound domain. Providing these constraints produce results defined in terms of probabilities, then they can be included using the belief network approach.

7.5.1 Results — Postscript

Subsequent investigations aimed at improving the results of the matching process have further emphasised the sensitivity of the system to its parameter values and to particular types of error in the initial cue. They have also indicated the source of the problem preventing convergence to a stable solution.

parameter	*value*
node type	composite node only
line type	closed
propagation type	parent to child
propagation limitation	decay 0.95
edge weight	1.0
model weight	0.5
initial weight	0.0
stability	$\leq$ 2 pixels
edge definition	3
shape standard deviation[3]	1
orthogonals	20
sample points	59

Table 7.4: Improved ultrasound network parameter values

The revised set of parameter values shown in table 7.4 was used to generate the results shown in figures 7.50 to 7.53. It should be noted that these results are better, they converge more quickly and all the constraints are satisfied. It appears that the *initial weight* parameter combined with a poor initial cue was preventing the refinement moving beyond a certain point and from satisfying the constraints. There is clearly further investigation required into the precise effects of different combinations of parameter values to see if a genuinely optimum set can be determined.

This revised parameter set was still unable to provide acceptable results in certain situations, as illustrated in figure 7.54. This behaviour is due to a combination of frag-

[3]The Shape Standard Deviation parameter defines the acceptable number of standard deviations each parameter in the shape model can be from the mean. In all previous experiments the value used was 3.

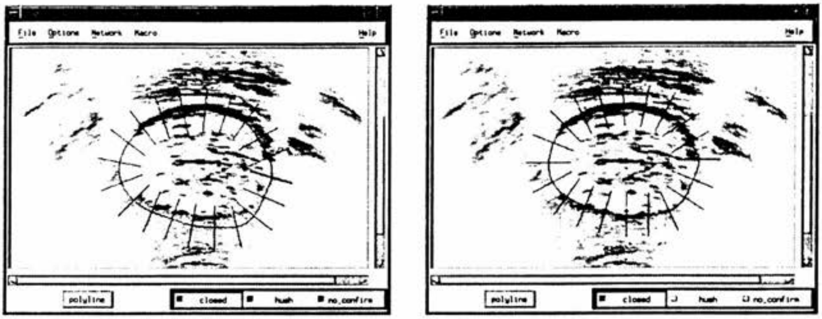


Figure 7.50: Composite delta network result with original parameters (59∗2) and with revised parameters (30-0)

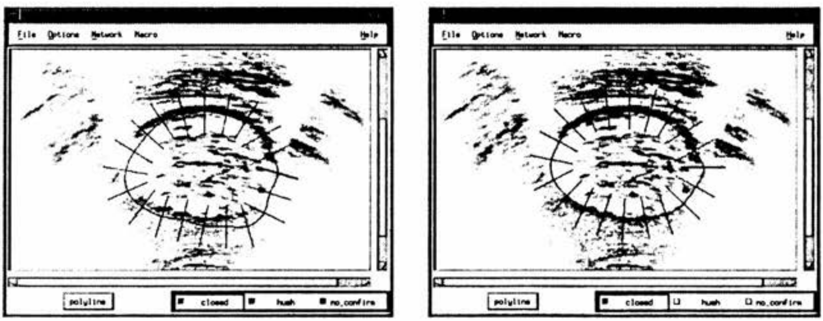


Figure 7.51: Composite delta network result with original parameters (44∗1) and with revised parameters (24-0)

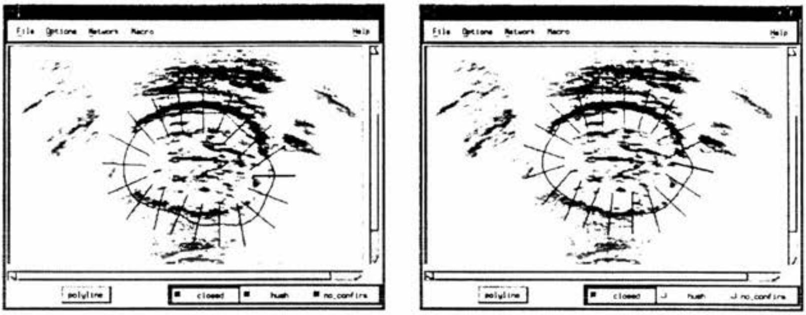


Figure 7.52: Composite delta network result with original parameters (269*10) and with revised parameters (20-0)

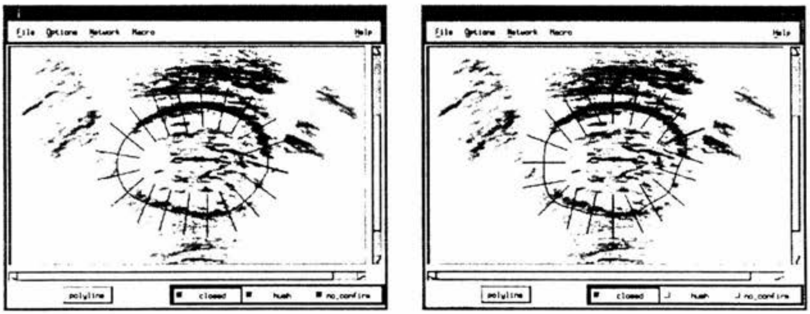


Figure 7.53: Composite delta network result with original parameters (104*6) and with revised parameters (49-0)

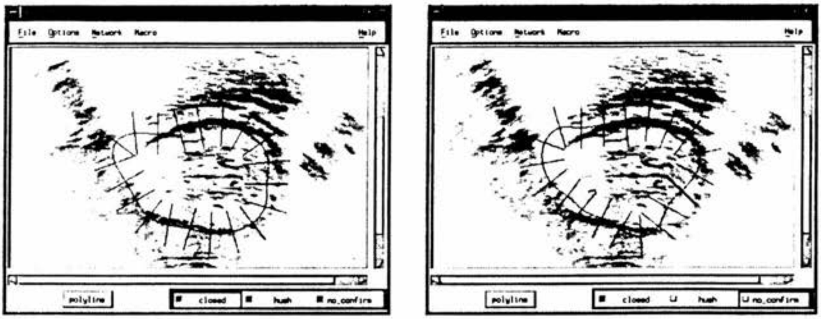


Figure 7.54: Composite delta network result with original parameters (210*4) and with revised parameters (210*0)

mentary or non-existent edge profiles at the easternmost and westernmost areas of the skull, noise providing spurious profile data, and very tight conditions for an acceptable solution. In cases where only one of these features is present, the shape model is typically able to impose its estimate as the edge data either supports (in the case of non-fragmentary edge profiles) or does not detract (in the case of no spurious profile data) from the estimate. It also depends on the positioning of the initial cue, as the process is better able to refine in an up/down direction than a left/right, due to the lack of profile data where the ultrasound beam is approximately tangential to the skull boundary.

7.5.2 Future Work

Despite some weaknesses in the specific implementation we adopted, the belief network approach to the ultrasound model matching tasks has proved both interesting and useful. On the basis of the work that has been completed, an outline for a new approach has been developed. This new approach addresses some of the invalid assumptions and poor design features of the current system.

The most significant change is to move from a belief network that shifts dynamically with the boundary, to one that is static with respect to the image. The boundary then moves with respect to both the image and the network. This creates an opportunity to incorporate more knowledge directly within the belief network.

In addition to providing an initial boundary cue, the user is required to provide an axis line running from front to back down the approximate middle of the skull. This should not present too many problems as it will follow the midline which is often well defined in the images. This axis line should be extended from the centre of the skull out to a point that is judged to contain both points of intersection between the skull and the axis line. This axis line is then used to place a set of orthogonals. The midpoint of the axis line is taken as the approximate centre of the skull, orthogonals[4] are placed so that they radiate from the centre towards the skull boundary. The length of each orthogonal and the number of sample points along each orthogonal can be calculated relative to the length of the axis line using knowledge about the distribution of skull shapes. The number of sample points per orthogonal will vary depending on the length of the orthogonal. This is illustrated schematically in figure 7.55.

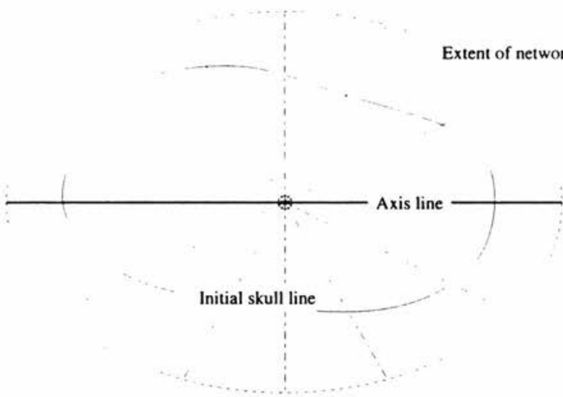


Figure 7.55: Schematic network model

The conditional probability matrices that determine the influence an orthogonal has on its neighbours can then be based on the degree of curvature, as the sample points on neighbouring orthogonals now have a fixed spatial relationship to each other. For example, in figure 7.56 if the curve being modelled is roughly symmetrical through orthogonals A, B and C, then a boundary passing through points 1 and 2 will have a greater probability of passing through point 4 than points 3 or 5. As the orientation and centre point

[4]These radial lines are no longer orthogonal to the boundary but have a similar function.

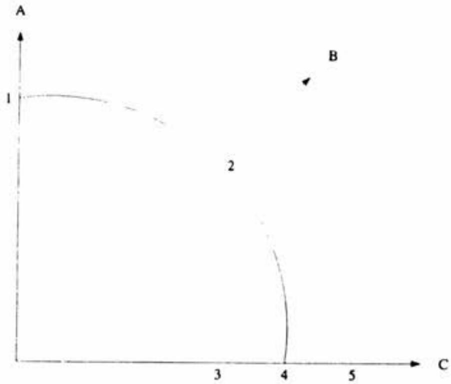


Figure 7.56: Conditional probabilities based on curvature

of the skull are approximately known it should be possible to derive the probabilities of different curvatures at different orthogonals from example images. Similarly it will be possible to produce prior probability distributions across each orthogonal based on a training set suitably scaled by the length of the axis line. These distributions would, for example, make sample points that are very close to the centre point highly unlikely, whereas those towards the outer edge will be more likely. The initial skull cue could be treated in a similar manner.

The shape model will need to be modified as it should only consider points lying within the extent of the network model. The issue of including multiple shape models requires further consideration.

7.5.3 Conclusion

The aim of this application was to demonstrate the suitability of belief networks as a tool for integrating local and global constraints for image interpretation. One of the distinguishing features of ultrasound images is the high level of noise. In addition to this, the noise may also be highly structured, mimicking genuine image features. The imaging technique also tends to produce ill-defined boundaries, both in terms of the discontinuity of the boundary under certain conditions and the blurring of the true position of a boundary. In addition to the difficulties arising from the imaging modality, the subject, *i.e.* the fetal head, introduces additional problems due to natural variations in size and shape and variations due to abnormality. These two factors combine to form an image interpretation task with a high degree of inherent uncertainty. The belief network approach provided on the one hand a principled mechanism for handling this uncertainly, and on the other hand an explicit model of constraint integration. Whilst the application concentrated only on the lowest level of the task, the identification of the fetal skull, this was always viewed as part of a larger system that would produce a high level analysis and diagnosis of the image. By adopting a probabilistic underpinning we hoped to maintain the validity of uncertain inferences from this low level image interpretation task through to the diagnosis itself, allowing for the possibility of this diagnosis then being used as a constraint at the image interpretation level as part of an iterative refinement process.

As a mechanism for handling uncertainty within this image interpretation task, the belief network has performed well. The probabilistic model employed was relatively straightforward, thereby avoiding some of the problems that can be encountered when determining an appropriate model. The most unusual feature of the model was the use of a circular network for propagating constraints between boundary points. While this model was a natural representation of the domain, it was at odds with the probabilistic theory underlying the belief network. Although the solution discussed earlier, namely propagation limitation, does not strictly adhere to the probabilistic theory, we believe that, in the context of the application, it provides an appropriate approximation. Whilst this may have implications for the results produced by our particular system, it does not undermine the probabilistic approach to image processing.

The application considered only a small number of constraints, but these were sufficient to demonstrate the principle behind the integration of constraints. Certain assumptions, such as an equal number of points in both estimates, were made but these were for programmatical convenience rather than any inherent limitation in the approach. A belief network appears to offer a flexible and powerful mechanism for combining disparate constraints under certain conditions.

Where possible the constraints used were based on training sets in order to capture the variation in the fetal skull and the way it appears in an ultrasound image. The image models derived from these training sets may themselves be probabilistic in nature, again suggesting that a belief network approach is appropriate.

As all the constraints within the application were concerned with the low level image processing task, the question of how the probabilities from this level would influence probabilities at higher levels of interpretation and diagnosis was never raised. This would depend to a certain extent on the types of evidence and the forms of domain model that were available. However, regardless of these factors, the belief network approach should be powerful enough to provide a coherent unifying framework.

One issue that was not addressed by this research was how estimates of the error associated with the boundary could be derived from the network. Ideally these estimates would have a statistical interpretation as they would be based on probabilities from the model.

Ultrasound image interpretation is a difficult task, typically performed by a trained expert. Although the specific task addressed by the application has concentrated only on fairly straightforward images, they have proved challenging given the constraints we have used. More important than the actual image interpretation was the demonstration of the basic approach to the task of image processing using belief networks. The general approach of combining multiple sources of evidence to aid image interpretation is widespread, belief networks offer an elegant mechanism for achieving this combination. The application demonstrates that belief networks can be used to provide probabilistic mechanisms for achieving low level image-based tasks. If uncertainty can be modelled within the system in terms of probabilities then belief networks can be used to provide a coherent and consistent mechanism for handling that uncertainty in a way that facilitates rational decision making at all levels of interpretation.

It may be that for image measuring tasks, a more traditional image processing approach, such as that taken by the Wayne State University group [Salari *et al* 90, Zador *et al* 91] mentioned earlier, may suffice. Where more flexible model matching and multi-level constraints are to be used, the belief network approach may be more appropriate.

There is evident similarity between the belief network approach and the active contour models discussed in Chapter 5, particularly the work of Cootes *et al* [Cootes *et al* 95], from where the Point Distribution Model was adopted. Both employ a form of active contour influenced by both local and global constraints. In fact, the belief network approach could be described as a form of active contour model using probability maximization in place of energy minimization. Both approaches share a common philosophy in terms of constraint integration and dynamic models, and some common limitations. The important difference between the belief network and active contours lies in the interpretations that can be applied to the numbers within the two approaches. The belief network numbers have a strong interpretation as *probabilities*, but it is not clear what interpretation can be applied to numbers within an active contour. This will have implications for any decision making that may be made on the basis of these numbers. It is also likely that belief networks will be more able to provide a common framework for all levels of a diagnostic task, rather than being limited to low level image processing tasks.

Chapter 8

Application — Cervical Screening

The potential for using belief networks in the task of cervical specimen classification has been recognised by at least two authors [Poole 91, Bartels & Weber 92, Poole 93]. Following on from the work by Poole, we have developed a belief network for classifying cervical specimens, implemented in FLAPNet. The intention is to illustrate and explore how a belief network can be used to integrate the results of low-level image processing functions with high level patient information in a diagnostic task. The goal of this chapter is to establish the behaviour and techniques for applying the belief network to an interactive diagnostic task.

8.1 Overview

The purpose of an automated prescreening system is to reduce the number of slides that must be inspected manually. Ideally a prescreening system divides slides into two groups, those that can be accepted as normal without manual inspection by a cytologist, and those which should be referred for manual inspection. Clearly this procedure is critically dependent on acceptable error rates. Whilst the effect of a high false-positive rate, *i.e.* manually inspecting normal slides, is essentially a question of efficiency, the false-negatives are potentially of great clinical significance. The determination of an acceptable false-negative rate is made more difficult by the absence of any reliable figures concerning the false-negative rates of cytologists.

The typical approach to automated pre-screening systems can be divided into three main stages; the location of objects on the slide, the measurement of distinguishing parameters of individual objects, and the combination of the individual measurements into a classification for the slide or specimen as a whole. On the basis of this classification, the slide is accepted as normal or referred for manual inspection. The tasks of locating objects and classifying them on the basis of measured features are relatively well understood [Eason]. The task of slide classification is still very much a research issue.

As far as the classification of the slide is concerned, we consider only four classes, *normal* (no abnormality), *borderline changes, mild/moderate dyskaryosis* and *severe dyskaryosis*[1]. These classes represent a continuous scale of change between a sample that has been taken from a healthy patient and a sample taken from one showing various degrees of severity of cervical cancer.

Although there are a number of risk factors that are believed to be involved in cervical cancer, the lack of understanding about the precise influence of these factors and the way in which they combine means that these factors are usually not incorporated in automated slide classification schemes. The influence these factors and other items of clinical and patient data have on manual classification is also not understood.

8.1.1 Probability Model

There are several sources of data that are potentially important in the classification of a cervical specimen. These include:

- Incidences of cervical cancer in the population.
- Items of patient-specific clinical data.
- Patient-specific risk factors.
- Sample quality of the slide.
- Particular clinical features of the slide as a whole.
- The classification of objects on the slide.

The population incidence rates can be taken as the base prior probability of cervical cancer for an individual selected at random from the population under consideration.

[1]Including severe dyskaryosis, carcinoma *in situ*, microinvasive carcinoma, invasive carcinoma and glandular.

This prior probability can then be refined to provide a patient specific prior probability by including relevant items of the patient's clinical and personal data.

The classification of a slide depends partly on the condition of the cells observed and partly on the number of cells observed to be in a particular condition, though the precise nature of the process is unclear. The belief network relies on the Cytoline-110 system [Poole *et al* 92, Poole 94, Eason] to locate and classify objects on a slide. The results generated by Cytoline are then used as inputs into the network. As the network has been constructed from an example set of data generated by Cytoline, it is necessarily tuned to the particular operating characteristics of Cytoline. Given a large enough set of example data from some other process providing object classifications, or some other diagnostic measure, this new process can be easily incorporated within the existing belief network model. Currently the network provides a slide classification which is all that is needed for a decision on manual inspection. The actual decision of *normal* (accept) or *suspicious* (refer for manual review) requires an analysis of the receiver operating characteristic curve and an assessment of the clinical and economic factors to set the decision point. This is not of direct interest here.

Cytoline produces a variety of information which may potentially be of diagnostic value. A first pass over a slide locates objects and classifies them as being being one of leukocyte, normal cell, suspicious cell, a clump of overlapping normal cells, or junk. It is generally believed that leukocytes and junk are not significant for diagnostic purposes. This initial classification is based on a number of low level image features. A count of the number of objects in each class and the average probability of all the objects of a particular class relative to the other classes is also produced. A certain number of the objects identified as being suspicious are then re-examined at low resolution to produce an individual set of probabilities across the classes for each specific object. Finally some of these objects are selected and scanned at a higher resolution to provide more detailed information. We refer to these three data sets as the *average data (AD), low resolution data (LR)* and *high resolution data (HR)* respectively. It is on the basis of these three sets of data that we must provide a classification of the specimen. One further source of information not directly considered here is the possible classification of the objects in the LR or HR sets by a trained human operative, though this is part of the Cytoline processing regime.

The diagnostic value of the data produced by Cytoline is still under investigation and there is no absolute benchmark against which a network model could be compared. This does not affect the purpose of this investigation which is to explore the use of belief networks and examine their operating characteristics.

Data sets produced by Cytoline are available, therefore the network model was derived from and tested on real data rather than a theoretical model. In addition to these data sets, a cytologist's clinical grading for each slide was also known, allowing a relationship to be established between Cytoline data and an estimate of the true classification. The clinical grading was defined over a slightly different set of classes to those we intended to use, but as the gradings were simply refinements (subsets) of the intended classes this presented no problem.

A total of 334 graded slides were available, these were divided arbitrarily into a training set, upon which the model was based, and a test set. The composition of these slides is shown in table 8.1. In view of the small number of slides in the severe class, it

slide grade	*total set*	*training set*	*test set*
normal	177	101	76
borderline	36	21	15
mild/mod.	104	47	57
severe	17	10	7
total	334	179	155

Table 8.1: Composition of training and test slide sets

was decided to combine the mild/moderate and severe classes.

An important benefit of using a belief network classifier is the ability to allow incremental addition of data so that the process can be stopped at any point to assess the current classification. For this reason initial investigations concentrated on trying to establish a coarse classification on the information contained in the AD (average data) alone. It was found that this information is insufficient to provide a satisfactory discrimination. Cytoline has since been modified to produce a finer grained representation of the probabilities for the initial pass in place of the average. Whether it will be possible to perform some form of discrimination on the basis of this improved data remains to be seen.

There were a number of ways in which the LR and HR data could have been used. Cytoline combines the low level image information (area, integrated optical density, shape, etc.) of each object and determines likelihoods across the classes for each object. We term the likelihood of the object being suspicious, an *index of suspicion*. This allows us to use the values assigned by Cytoline (or indeed by any process) with only minimal assumptions about the way in which the value has been calculated. We require only that the value assigned by the process varies monotonically across its range and that the ordering of specific cells by value is similar (we cannot expect any process to always be correct or accurate) to that produced by a cytologist ordering the cells relative to their degree of suspicion. A plot of the index of suspicion against likelihood (derived from the frequencies in the training data) on a per-class basis for the LR and HR are shown in figures 8.1 and 8.2 respectively.

A particular feature of the Cytoline data is that objects are ranked according to a calculated degree of suspicion, based on the degree of apparent abnormality due to measured features of the object. This allows us to test the incremental updating of the belief network which can then be used to determine if further analysis is required or if sufficient data has been collected. This may not be critical for this application but in the general medical case collecting further data may involve expensive or possibly hazardous medical procedures. The network can act both as a tool for reasoning and for planning.

The distributions of index of suspicion values for objects at different points, #1 $\ldots$ #n, in this order are very different, as shown in figure 8.3. The top graphs show the plot of index of suspicion values for the first ranked object (#1) on a per class basis. There is little difference between the class distributions as there are typically a small number of suspicious looking objects even on a normal slide. At the #20 object there are distinct differences between the distributions. Whilst the individual graphs are not particularly discriminating, by basing a classification on a series of them a more accurate discrimination should be possible. In this work we use the first 20 distinct likelihood distributions, representing #1 $\ldots$ #20.

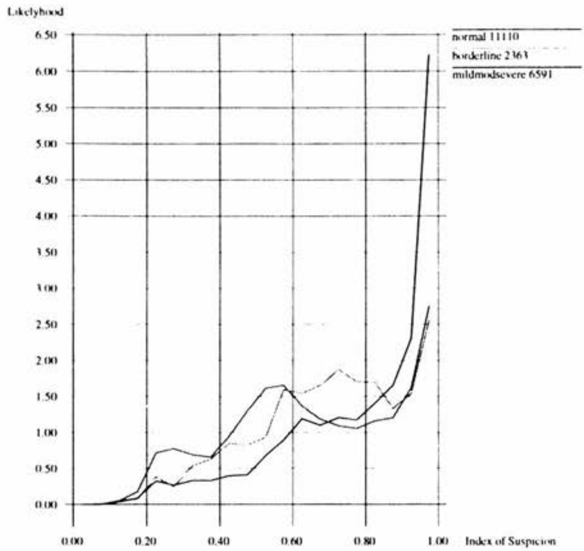


Figure 8.1: Index of suspicion likelihoods for LR

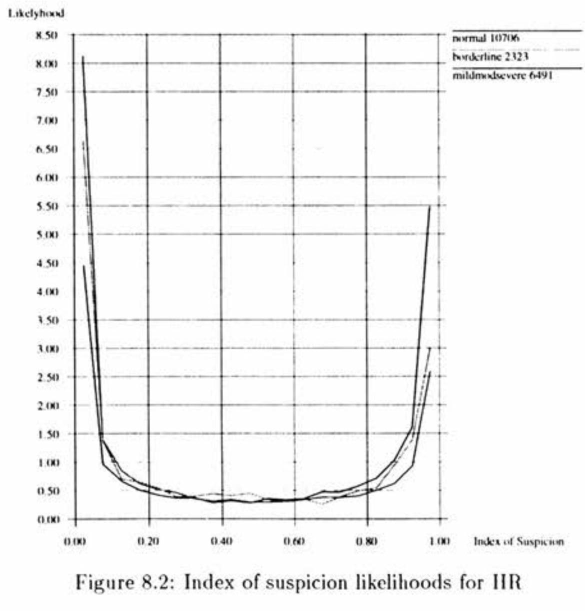


Figure 8.2: Index of suspicion likelihoods for HR

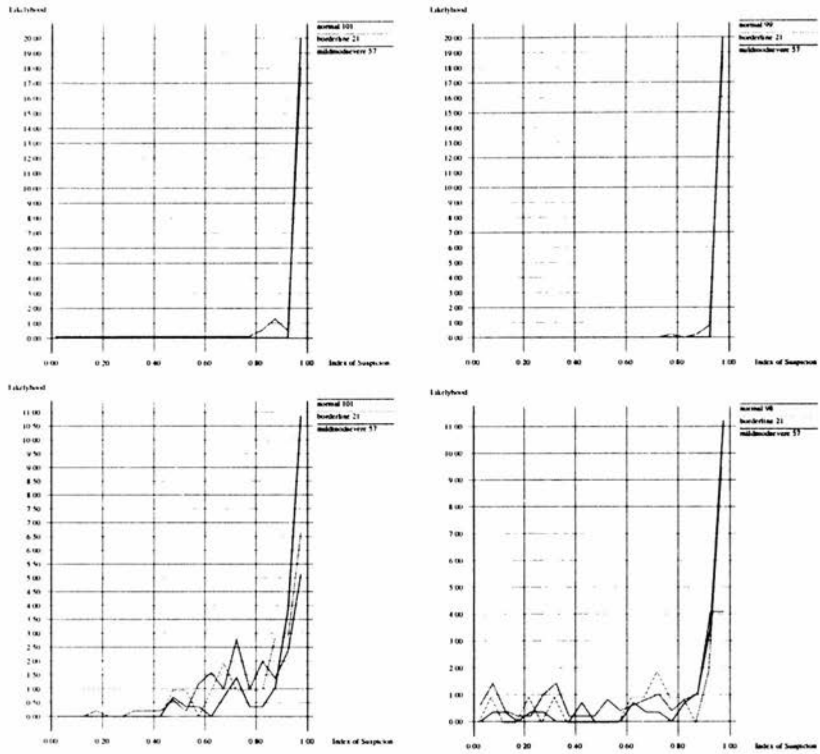


Figure 8.3: Likelihoods for #1 and #20, LR (*left*) and HR

Each of these graphs provides the likelihood of an object, o_i with a given index of suspicion $IS(o_i)$ belonging to each of the three classes, C, given its ordering $\#(o_i)$, $L(o_i = C \mid IS(o_i), \#(o_i))$. This likelihood depends also on whether the index of suspicion is based on the low resolution scan or the high resolution scan. The probabilities for the slide as a whole $P(S_j = C)$ can be calculated as the product of the likelihoods of all the object data multiplied by the prior probabilities across the classes.

8.2 Network Model

Although we have concentrated on the part of the model concerned with the classification of a slide on the basis of object data, extending the model to include, for instance, patient and population data is straightforward, as illustrated in figure 8.4. The precise nature

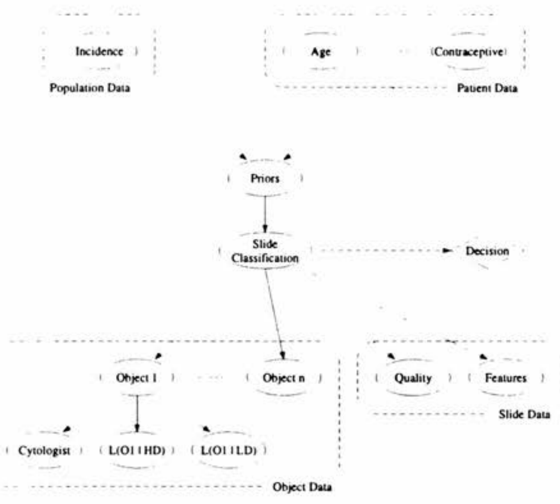


Figure 8.4: Schematic network model

of the relationship between the items of patient data, their combination with population data and the question of default values are not addressed in the work we have undertaken. In the network model used for the classification results presented in this chapter, we ignore slide data and assume that the priors for the slide classification are available from the training data. This simplified network is shown in figure 8.5. We have also created

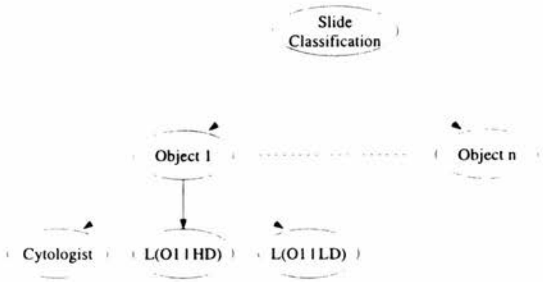


Figure 8.5: Reduced network model

some experimental networks that incorporate slide, patient and population data.

In this study we have tested two models for the conditional probability matrices. Both models assume a categorical relationship between the `Object` nodes and the `Slide Classification` node. In model 1 we assume that the matrices between the `LR` and `HR Likelihood` nodes and the `Object` node are also categorical, *i.e.* that there is no error in the classification. This is very unlikely to be true, but by including sufficient data the error introduced by this assumption may average out. In the second model we hypothesise that the conditional probability relationships between the `LR Likelihood` node and the `Object` node and the `HR Likelihood` node and the `Object` node are non-categorical. The relationship used is shown in table 8.2. Whilst the hypothesised relationship is

class	normal	borderline	mild/mod/sev
normal	0.9	0.1	0.0
borderline	0.1	0.8	0.1
mild/mod/sev	0.0	0.1	0.9

Table 8.2: Hypothesised non-categorical conditional probability matrix

certainly over simplified and inaccurate, it provides a comparison with the categorical hypothesis.

In the current situation we know the total number of objects in the LR and HR for a particular slide as they are taken from a pregenerated data set, however this would not be the case if Cytoline were supplying data at run-time directly to the network. To allow for the possibility of a direct run-time link to Cytoline, the network is created incrementally in response to the data it receives. Initially the network consists of only a single node, the `Slide Classification`, conditioned with the appropriate prior probabilities. All objects located by Cytoline are numbered, therefore the index of suspicion is linked to a specific object. If an index of suspicion datum is received for an object that is not represented in the current network, an `Object` node is added to the `Slide Classification` node. If it is a LR index then a `LR Likelihood` node is added with the distributions appropriate to the order in which the data has arrived. If an `Object` node but no `Likelihood` node exists an appropriate `Likelihood` node is added to the existing `Object` node. In this way the network contains no redundant nodes.

Each `Likelihood` node is in fact a pair of nodes, one that is used to input and display the raw index of suspicion and the other to convert the index into likelihoods. The conversion node holds an approximate, discrete representation of the three likelihood functions which it uses to make the conversion. This conversion is performed on the single value sent to it by its solitary child node. The resulting likelihoods are taken as the lambda value at the conversion node. The rest of the node functions as a Pearl Type node, except that no messages are sent to the child node. In the present model, each likelihood distribution is represented by twenty discrete classes across the range 0 to 1, corresponding to the range of possible index of suspicion values.

8.3 Results

Summaries of the results obtained from applying the network to the test slides are shown in the tables on page 198. Table 8.3 shows the classifications of the slides under the two models. We consider the most probable class at the `Slide Classification` node to be the network classification[2]. The *true classification* column indicates the grade allocated by the cytologist. These are offset to align with the appropriate *model classification* row. The figures are the number of slides of that clinical grade that were classified as the grade denoted in the *model classification* column. For each of the models both the intermediate result using only LR and the final result using both LR and HR are shown. The LR and HR results are based on the HR set *in addition* to the LR data, *i.e.* it is based on both sets of data, with the LR set being processed first.

[2]There were no cases in which there was not a single, most probable class

The actual slide classification results are not particularly good and certainly are not good enough to be included in any real classifier, though this was not the aim of the investigation. Consider, for example, the two extremes of the classification range, the normal and the severe samples. For an automated pre-screening system we would need to be able to identify normal samples efficiently, as these will form the bulk of the slides. The classifier correctly identifies these slides in only 48 to 68 % of the examples (depending on the model used) and classifies between 28 and 35 % of them as severe. In a screening system which handles a large number of samples, a high false-positive rate undermines the purpose of automation by requiring the cytologist to examine a significant number of normal slides. Similarly, and more importantly from a clinical point of view, 14 % of severe samples were misclassified as normal. False-negative classifications can have potentially grave clinical implications. To be usable, a classifier would need to have a false-negative rate of some fraction of 1 %.

Table 8.4 shows the change between the LR and the LR and HR classifications. Cross referencing the *LR* (intermediate) column and the *LR & HR* (revised) row gives the number of slides (of a particular clinical grade) that were given that intermediate grade, having that revised grade after the HR was added. For example, using the *model 1 LR* borderline column and the *true classification* normal section, *HR & LR* borderline row, gives the value 6. This means that 6 of the clinically normal slides that were classified as borderline using Model 1 and LR were still classified as borderline after the HR had been included.

Figure 8.6 shows the classification results for three slides, one each of normal, borderline and severe, for both the models. The graphs illustrate the way in which the probabilities of the classes change as data in the form of the index of suspicion values is made available. The *Data* axis of the graph shows the number of index of suspicion values that have been examined, 0 – 20 twice, once for the LR and once for the HR. In the normal and severe graphs the classification is not changed by the addition of the HR data, and settles after only about half of the LR data. Whilst it may be tempting to suggest that a classification can be based simply on the first ten LR index of suspicion values, this will not always be the case. Ideally a classifier would allow decisions to be made as to when the process had settled and further data was unnecessary.

true classification	*model classification*	*model 1* LR only	*model 1* LR & HR	*model 2* LR only	*model 2* LR & HR
normal	normal	41	52	39	37
	borderline	8	2	11	18
	severe	27	22	26	21
borderline	normal	5	6	5	6
	borderline	0	0	1	0
	severe	9	8	8	8
mild	normal	14	13	13	12
	borderline	0	1	1	3
	severe	20	20	20	19
moderate	normal	7	7	6	6
	borderline	1	1	3	2
	severe	14	14	13	14
severe	normal	1	1	1	1
	borderline	0	0	0	0
	severe	6	6	6	6

Table 8.3: Intermediate and revised classification results

true classification	*LR & HR classification*	*model 1 LR classification* normal	borderline	severe	*model 2 LR classification* normal	borderline	severe
normal	normal	41	1	2	34	1	2
	borderline	0	6	1	5	10	4
	severe	0	1	24	0	0	20
borderline	normal	5	0	2	5	0	2
	borderline	0	0	0	0	0	0
	severe	0	0	7	0	1	6
mild	normal	12	0	0	12	0	0
	borderline	0	0	1	1	1	1
	severe	2	0	19	0	0	19
moderate	normal	7	0	0	6	0	0
	borderline	0	1	0	0	2	0
	severe	0	0	14	0	1	13
severe	normal	1	0	0	1	0	0
	borderline	0	0	0	0	0	0
	severe	0	0	6	0	0	6

Table 8.4: Change between LR and LR & HR classifications

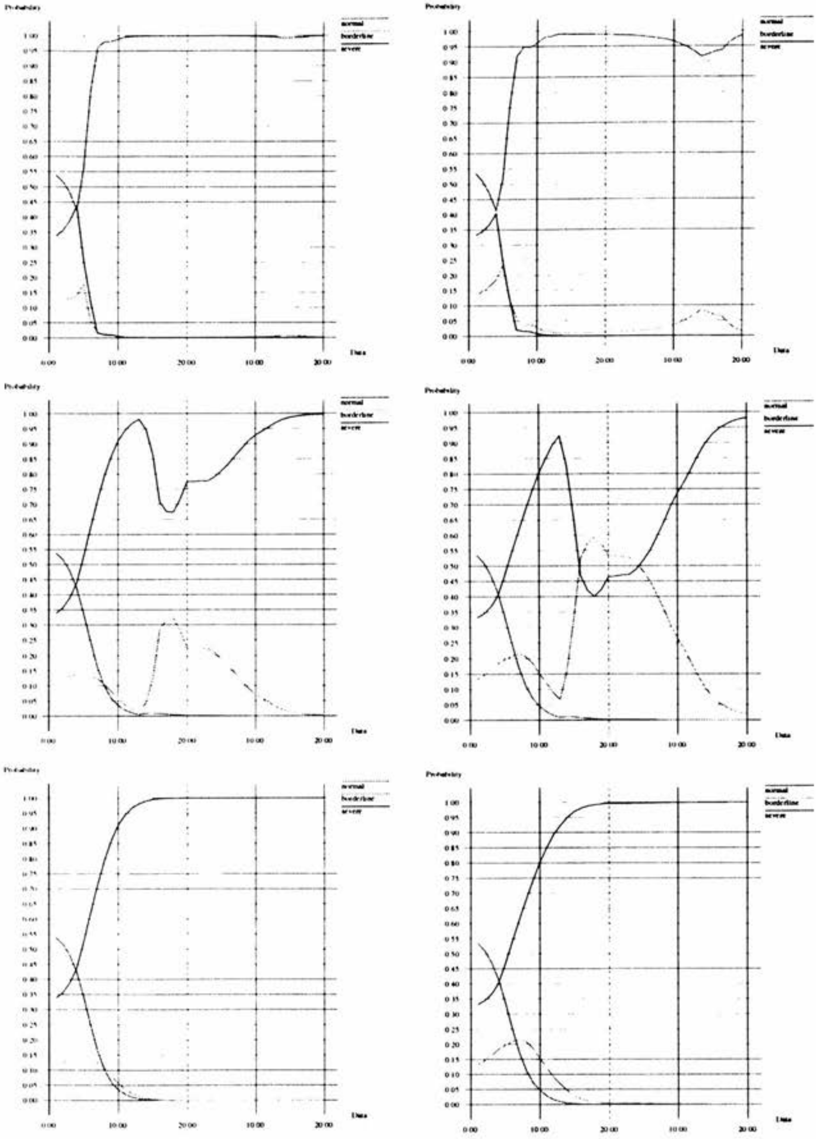


Figure 8.6: Example slide classifications, normal, borderline and severe (*top to bottom*), using Model 1 (*left*) and Model 2

8.4 Discussion

The network model presented in this chapter has focussed on the use of the Cytoline data sets for classifying cervical specimen slides. We have used the data in a way that makes a minimal number of assumptions about the precise meaning of the numbers Cytoline assigns. By adopting this kind of approach it is possible to include data with a less rigorously defined mathematical underpinning than Cytoline's, for instance cytologist gradings. Often when using human expert opinion in a classifier it is difficult to determine whether the numbers they assign should be treated as posterior probabilities, likelihoods, or in some other way. Belief networks provide a means of incorporating poorly defined values into a probabilistic framework.

One of the potential strengths of an explicit network model is its use as a testbed for theories about how the risk factors, for instance, are related both to each other and to the incidence of cancer. It has been argued [Boden 77, page 401] that the implementation of a theory as a computer model provides a good test of the validity and completeness of that theory. By experimenting using different models of the interaction between the risk factors it may be possible to improve on a classification based on object data alone. In our investigations we have experimented with simple models of the interaction between, for example, age, contraceptive usage and social class. The models are easy to construct and are entirely explicit once created allowing the expert user to follow the inference process. We have also experimented with quality control data, for instance by measuring the proportion of debris and cell objects located by Cytoline. This information can either be fed directly into the `Decision` or can be included at the `Slide Classification` node by expanding the classifications to include a 'quality reject' class. The fact that the underlying probability model is explicitly represented by the network structure and the conditional probability matrices, means they are ideally suited to this type of incremental, experimental model development. Different sub-parts of the model can be refined when an appropriate biomedical or mathematical model is available or represented only as an input if the underlying models are not understood or are not important.

Before any investigation into the way in which the risk factors affect classification, it will be necessary to improve the object based classifications. Once this improvement is achieved, the network can then be used to combine the results of the image processing, object based data, with expert opinion on how the risk factors and clinical data interact.

There are several factors which could be contributing to the current poor classification results, among these are:

- The index of suspicion may simply not be a powerful enough discriminator on its own. The results suggest that it may have some limited discriminatory power, but it is possible that this is an artifact of the available data. The whole question of identifying useful discriminators is still largely unanswered. There is a suggestion that the data on which a classification is to be based should be expanded to include more object classes, such as *inflammatory* and particularly *overlapping abnormals*. It is also possible that the special slide preparation method used for Cytoline is failing to reveal vital diagnostic information.

- The training data set was small and may therefore not be truly representative of the underlying distributions. The desire to maintain a totally independent test set also restricted the size of the training set.

- The number of slides in each of the classes varied greatly, even after the severe and mild/moderate classes had been combined. It is difficult to assess what effect this may have had, though it may be possible to compensate to some degree in the conditional probability matrices.

- In calculating the likelihood distributions, no attempt was made to compensate for missing data. Gaps in the data for a certain discrete range in the index of suspicion range were treated as genuine rather than a feature of the training data, which is more likely to be the case. Had some approximation of the missing data been attempted, it is possible the results may have been better.

- Rather than use the absolute ordering # of suspicious objects, some form of normalisation, based on the number of cells examined should be performed. Instead of examining the #20 cell we should examine, for instance, the #0.002 % cell.

- The use of only the first twenty index of suspicion values represents approximately one sixth of the index of suspicion data available for each slide. It is not clear if the discrimination would be improved or degraded by making use of the remainder of the index of suspicion and object data.

- The use of only twenty divisions to represent the index of suspicion discretely may have been too coarse, but this choice was influenced by the wish to avoid gaps in the data. A uniform division across the range should be replaced by a division with finer resolution at the lower end of the scale, on the basis of trying to identify slides that are "clearly" normal.

- The hypothesised conditional probability matrices and the use of the same matrix for the LR and HR data are certainly incorrect and discrimination may be improved by basing these matrices on real data.

8.4.1 Conclusion

Most of the factors that have been identified as possible causes of the poor classification results achieved concern the way in which the data was used and the quality of the data itself, rather than the belief network approach to the task. With a larger data set and a more thorough investigation of possible correlations between the items of data and the classification of the slides, the majority of these reservations could be addressed.

Even with significantly better classification results, the case for using a belief network instead of (or in addition to) the black-box statistical classifiers typically used is still to be made. The force of argument for the use of belief networks in this application rest less on their statistical validity (though this is still vital) and more on the explicit nature of belief network models and their ability to use weak, incremental evidence.

The importance of an explicit model is expressed well by Bartels and Weber when discussing prescreening procedures and systems [Bartels & Weber 92, page 1],

> Understanding the underlying assumptions requires knowledge and insights from different disciplines, most of them not in the area of the professional experience of cytopathologists and cytotechnologists. Yet, these are the professionals who will ultimately operate the devices, will have to trust the results, and whose profession will be affected by the process of automation.

A belief network can provide an explicit model in addition to a statistically sound inference mechanism for the combination of the measure or measures useful for classification with items of patient data.

The other attraction with the belief network approach is that it allows the cumulative aggregation of diagnostic information, where weak but consistent evidence can be utilized. It also provides a means for integrating diverse indicators. These features can potentially reduce the number of objects that must be examined or re-examined. If the uncertainty associated with the classification based on the current evidence is within limits a decision can be made without reference to further objects. Clearly this is useful in a process where through-put is a major concern. Bartels and Weber [Bartels & Weber 92, page 13] again summarise this well,

> As classifiers, inference networks offer capabilities very different from almost any other classification method: an inference network will allow utilization of discriminating information, even from feature distributions with substantial overlap of tolerance regions. Such features frequently can offer reliable diagnostic evidence, but the high classification error associated with each single such feature precludes their inclusion in the feature set of any statistical classifier, including neural net classification schemes.

These attributes of belief networks suggest that there is indeed a role for them within the cervical screening task. It may well be that the best overall approach is a hybrid of statistical classifiers at the lower, object orientated level and belief networks at the higher, patient orientated and decision making levels. A belief network provides a statistically sound framework for the integration of evidence from these different levels.

The two applications described in this thesis present different but complimentary perspectives on the potential uses of belief networks. Both applications are working within a common context, namely the formation of an expert diagnosis based, in part, on the interpretation of image data. Within this context, if all the processes involved can be placed on a probabilistic foundation we have the possibility of a completely probabilistic diagnosis, combining low level data-based information with high level expert knowledge within a consistent statistical model. The two applications should therefore not be considered as addressing separate tasks, but rather as addressing different levels of the same task.

The ultrasound application uses a belief network to implement a low level model matching task, operating directly on image data. The application is a constraint satisfaction system, with a static set of constraints embodied within the various models. In a complete system the constraints would vary dynamically, for instance the shape model may change if clinical evidence suggests that a malformation is likely. Similarly the detection of a malformed skull would influence the clinical diagnosis.

Within the image processing domain it is possible to imagine a suite of interpretation procedures, such as edge detection routines, designed to produce probabilistic results. Given suitable control information, a belief network based expert system could then select which procedures to execute in order to minimize resource expenditure relative to the expected result of the process and the quality of result required.

Clearly this is adding an entire additional level of functionality and whether this can be handled elegantly within the belief network model, perhaps through the use of decision theory, or whether a control element will act as an adjunct to a belief network is an open question.

The screening application focussed more on the higher level task of diagnosis, but again had an emphasis on the combination of information. Within this application the idea of belief networks as an explicit model, more accessible to the non-expert user than standard statistical methods, was also mooted. The ability of a belief network to produce incremental results as evidence becomes available was also illustrated. Control knowledge working at this level, based on information of this type could be used to control the lower level processes demonstrated within the ultrasound application.

In both applications doubts were raised as to the appropriateness of the network models and the quality of the data. In a real world development it would be important to ensure that both the model and the data were sufficiently powerful to enable a satisfactory diagnosis. It is possible that belief networks themselves could play a role in this initial validation phase. Within the screening application we discussed the possibility of creating network models in order to test theories regarding the application domain. We also noted that failure to achieve satisfactory results given an appropriate model, may point to a lack of diagnostic power in the evidence. In practice it may be difficult to separate weaknesses in the data from flaws in the model, and some other form of analysis may be more appropriate.

One of the greatest strengths of the belief network approach is that it is able to combine such diverse sources of information in a potentially powerful way. It is not necessary for any individual source to be a powerful discriminant, as noted by Bartels and Weber earlier. As a result of this it may be possible to apply expert systems to new areas, previously considered to lack the necessary quality of evidence. This possibility combined with our experience in developing these applications, suggests that there might be a particular class of problem suitable for solution using a belief network. Although it is difficult to be precise about the nature of this class of problem, it is likely to involve multiple, diverse, weakly suggestive information sources with a high degree of independence in the underlying model.

The two applications taken together suggest the outline form that a complete probabilistic expert system might take. We have identified various features of belief networks that make them suitable for such expert systems, but as yet there are few, if any, "industrial strength" belief network applications, despite an enthusiasm for them in research circles. Whether belief networks can make the transition, and what classes of application they are suited to, remains to be seen.

Chapter 9

Conclusion

At the beginning of this thesis we listed several broad areas that may contribute to the slow introduction of expert systems technology into biomedical domains, perspective, communication, assimilation, evaluation, and ethics and liability. When discussing expert systems and belief networks the topics included the representation of uncertain knowledge, the control of inference and the design of user interfaces. How then are these seemingly different outlooks to be reconciled?

The design perspective of an expert system will not be altered by the formalism employed for the expert system. Part of the problem is, perhaps, that the biomedical domain is *so* challenging it is ideal for research purposes but far from ideal from a commercial point of view. Often the research community is (maybe correctly) little interested in the practical application of its own research. Perhaps as some of the impediments are overcome the emphasis will shift from *what we can give the physician* to *what the physician wants.*

The underlying theme throughout this thesis is inference under uncertainty. As expert systems have moved from the laboratory into the real-world it has become increasingly apparent that uncertain reasoning is the norm rather than the exception. As we have seen, this has resulted in a number of diverse methodologies being developed, many of which claim to be measuring the same quantity. From the expert systems point of view, no methodology can truly claim to be more philosophically or cognitively valid than any other, and ultimately they must be judged on the support they provide for rational decision making. Cox and others have shown that in this respect, and under "common sense" assumptions, probability is both necessary and sufficient.

As we have stated, belief networks propagate inferences measured in terms of belief based on the axioms of probability theory. The fact that belief networks can perform correct inference that is justifiable as an accepted statistical value is important. Much has been made of the fact that people typically violate the axioms of probability theory, but this reflects more on the cognitive abilities of people than the validity of probability theory. Debates about the sufficiency of probability theory continue, but the importance of the debate must be clearly defined. It is one thing to claim that probability theory is the normative method for uncertain reasoning and another to claim that probability theory is an adequate inference method for certain types of uncertain inference. If we restrict the claims of sufficiency, we can ask more pertinent and more answerable questions, such as is probability theory sufficient for the general task of medical diagnosis? There are types of reasoning, such as reasoning by analogy, which will certainly fall outwith the belief network paradigm. It does, however appear that belief networks and probability theory provide a powerful tool for a particular type of reasoning, diagnostic inference and decision making. This class of reasoning is prevalent within the biomedical domain.

Probability theory does have a certain intuitive appeal in as much as the notion of probability is familiar to most people. Unfortunately this familiarity is also seductive in that the precision and validity of numbers is often accepted without question as they appear to carry some weight of scientific fact. People's general naivety about statistics combined with a respectful familiarity, can potentially make probability theory, and more generally any form of automated reasoning, a dangerous tool. People are well used to considering the reliability of an information source when deciding whether to act on the information, few people are ever familiar enough with the internal functioning of an expert system to form a reasonable assessment of its reliability.

Possibly the most important aspect of belief networks from the point of view of acceptance is the network representation itself. In many expert system representations the clarity of the representation degrades dramatically with the size of the model, for instance trying to debug a complex rule base is very difficult as the interactions between rules can be hard to trace. Belief networks rely on an analogue, pictorial representation of the causal relationships in the domain. Ideally the independencies in the domain result in a network that is relatively sparsely connected. The network explicitly shows all the relationships and conditional dependencies in the domain (assuming the model is correct and complete). As we mentioned earlier, Pearl suggests that it is the *structure* of the network that is important. This representation therefore allows the user to assess the reliability of an expert system on the basis of the causal model it is using, without necessarily being concerned with the precise probability values used.

An explicit model also enables the model to be expanded or modified more easily, without the problem of introducing inconsistencies. This structural representation is also an ideal basis for the generation of explanations.

This combination of an explicit analogue domain representation and a sound probabilistic inference mechanism makes belief networks appropriate both for use and design. It will remain to be seen if it is possible to develop the necessary extensions to this basic representation and inference model that will enable it to provide the full range of decision making, modelling and communication facilities that will be required. We have indicated that research is already underway into many of these important areas. Whilst none of these general points address specifically the areas highlighted as limiting the successful application of expert systems technology in the biomedical field, they all go some small way to making expert systems more transparent to the biomedical user and may therefore go some way to improving their acceptability.

We have also presented a gazetteer of biomedical applications ranging from small networks with only a handful of nodes, to those containing over a thousand. It is encouraging to find so many applications of a relatively new methodology to such a challenging domain. Many of the developments in belief networks are a direct result of domain constraints and requirements. As much as anything else the gazetteer shows the wide variety of problem types that are being addressed using belief networks, though diagnosis appears the most popular. The diversity of implementations of the basic approach indicates that belief networks are still a rather immature technology though we appear to be reaching a transition point where belief network shells are becoming available more widely, opening the way for true commercial applications.

We have introduced a network propagation tool, FLAPNet, and demonstrated applications in two very different domains. The cervical screening application is an example of a diagnostic task linking low-level processing with operator interaction. The ultrasound example is very different, using a belief network for part of an image processing task. These two examples serve to demonstrate that belief networks can span the entire range of inference required of a complex biomedical expert system. This principled propagation of inference is a general requirement that is common to a variety of applications, suggesting that belief networks will be employed in a variety of ways within expert systems and other computerised systems.

In the future belief networks and expert systems will inevitably become more widespread in the biomedical domain. The rate of acceptance will be critically dependent upon the willingness of expert systems designers to meet the needs of the biomedical user. Belief networks have a number of properties that suggest they will play an important role as a core inference mechanism in expert systems. Already belief networks are being applied to a variety of biomedical applications mainly for research purposes. Belief networks require further development in order to make them fully acceptable as expert systems rather than simply as inference mechanisms and research is currently underway to ad dress some of these developments.

Appendix A

Gazetteer

This appendix contains a gazetteer of belief network applications in biomedical domains. It is not exhaustive, as the intention was to give a view of the breadth of application of belief networks, and the increasing influence belief networks are exerting in the field of medical informatics.

For some projects it was difficult to determine whether or not they constituted belief networks or not (particularly some decision theoretical systems). In most cases those we have chosen to omit are mentioned elsewhere in the text, or are included in the bibliography.

Each entry comprises four parts, brief information, key points, references, and precis. In some cases the authors have not given their project a name, in which case we have given a name for reference purposes. These given names are identified by the dagger† superscript. The entries are in alphabetical order of project name and are indexed on page 211.

The dates give an approximate range for the *publications* associated with the project. The most recent entries are from approximately the first quarter of 1994. The list of publications concerning a particular project should not be considered complete.

Authors are listed in alphabetical order rather than any order of precedence. Authors names only are included, qualifications have been omited. The organisations are referenced with respect to the authors' affiliations where known.

The precis for each entry is deliberately brief and selective, the intent is to provide a overview of the project and any interesting features it may possess.

A.1 Gazetteer Index

ABDO

Project:
ABDO

Application:
Diagnosis

Domain:
Acute abdominal pain

Date:
1993

Authors:
J.R. Clarke[1]
G.M. Provan[2]

Organisations:
[1]Department of Surgery, Medical College of Pennsylvania, Philadelphia, USA.
[2]Computer and Information Science Department, University of Pennsylvania, Philadelphia, USA.

Software:
DYNASTY

Hardware:
Unknown

Key Points

- Use of decision theoretic models
- Dynamic construction and updating of models
- Interval-based temporal modelling

References:
[Provan & Clarke 93]

Precis

ABDO is a domain specific demonstration of the DYNASTY system, concerned with acute abdominal pain. The DYNASTY system dynamically constructs and updates influence diagrams with respect to temporal data.

A knowledge-base of domain-specific causal relationships is used to generate a belief network that models only the observations currently available and causal consequences or antecedents, at a particular time interval. The network is used to generate probabilities of diseases given the observations. It is assumed that the values of observations are constant over an interval and that an interval has a definite transition point. The model consists of over 50 findings nodes, 20 intermediate physiological states and 4 disease states. The data for the model was collected from several thousand real cases.

A subsequent time interval is then selected and a model created on the basis of the initial model. Sensitivity analysis is performed, comparing the decision-equivalence of the current interval model relative to the previous and subsequent interval models, to confirm that the model is valid. If necessary the network model topology is updated. This model is then augmented with decision theoretic nodes representing decisions and utilities of outcomes. Sensitivity analysis is also performed to validate the selected time interval relative to other time intervals, based on the decision options suggested by competing intervals.

Model updating includes the addition or deletion of nodes and the refinement or coarsening of nodes. The emphasis is on the development of the model over time where ever possible, with the creation of an entirely new model taking place only when it is necessary.

ALARM

Project:
ALARM

Application:
Patient monitoring and diagnosis

Domain:
Anesthesiology

Date:
1989

Authors:
I.A. Beinlich
R.M. Chavez
D.M. Gaba
H.J. Suermondt

Organisations:
Section on Medical Informatics and Department of Anesthesiology, Stanford University School of Medicine, Stanford, California, USA.

Software:
KNET

Hardware:
Macintosh II

Key Points
- Compared propagation algorithms for multiply connected networks

References:
[Beinlich & Gaba 89]
[Beinlich *et al* 89]

Precis

The ALARM prototype is designed to provide specific text warnings in order to advise the user of possible problems. A simple network of 8 diagnostic nodes, 16 evidence nodes and 13 intermediate nodes, containing loops is used. Continuous variables were represented categorically using discrete intervals, typically 3 to 5 depending on the situation. In experiments ALARM's top diagnosis was correct in 71% of test cases.

In time critical applications, propagation time is an important issue. Two propagation algorithms were tested, Pearl's with cutset conditioning and the clique-tree approach of Lauritzen and Spiegelhalter. The large cutset required and the fact that although several pieces of evidence arrived simultaneously they have to be processed sequentially, resulted in Pearl's algorithm being much slower. The accumulation of evidence reduces the size of the clique-tree, making it an ideal algorithm for this application.

Aspiration-Net

Project:
Aspiration-Net[†]

Application:
Diagnosis

Domain:
Fine needle aspiration cytology of the breast

Date:
1994

Authors:
N. Anderson[1]
P.H. Bartels[2]
P.W. Hamilton[1]
D. Thompson[2]

Organisations:
[1]Department of Pathology, The Queen's University of Belfast, Northern Ireland, UK.
[2]Optical Sciences Centre, University of Arizona, Tuscon, Arizona, USA.

Software:
C, algorithm based on [Morawski 89a, Morawski 89b]

Hardware:
Unknown

Key Points
- Standard belief network mechanism
- Use of measures of evidence conflict

References:
[Hamilton *et al* 94]

Precis

This is a simple network with a single diagnostic root node (Benign/Malignant) and ten evidence nodes. There are no intermediate nodes. The links are quantified by conditional-probability matrices, and standard belief propagation mechanisms are used. Evidence, in the form of likelihood ratios, is entered by the cytologist. One particularly interesting point is the use of equal priors in the diagnostic node modelling the cytologists reported assumptions, rather than the statistical priors that favour a benign diagnosis. This is reported to work well, though it is not clear how much experimentation was conducted.

The probabilities calculated at the diagnostic node are mapped onto four final diagnostic categories, if the probability of benign or malignant is greater than 0.9 then the final diagnosis is either benign or malignant. If the probability of benign is less than 0.9 but greater than 0.5 then the final diagnosis is equivocal (benign), and similarly with equivocal (malignant). In addition to this diagnosis, three other measures are calculated. These are derived from a cumulative probability graph which is plotted after each piece of evidence is entered. The evidence is entered according to an order determined by a cytologist to have decreasing impact on the final diagnosis. The measurements are P-score (number of peaks), T-score (number of troughs) and C-score (the number of intersections with the 0.5/0.5 threshold). These measure the extent of conflict in the information given by the evidence. It is planned to incorporate these measurements into the network when sufficient statistics can be gathered. Currently the graph is used to provide the cytologist with an overview which is useful, for example, for identifying evidence that appears to be aberrant and should be confirmed.

The system has been tested on forty cases, with a high degree of accuracy, the apparent misdiagnoses were all due to lesions known to be difficult to diagnose.

In the future the use of this system as a teaching tool and as part of an automatic diagnostic system will be considered.

BAYES

Project:
BAYES

Application:
Diagnosis

Domain:
Macroscopic melanocytic lesions

Date:
1994

Authors:
W. Abmayr
P.H. Bartels
D. Rehberg
W. Stolz
D. Thompson

Organisations:
Fachbereich Informatik/Mathematik, Fachhochschule München, Germany.
Dermatologische Klinik und Poliklinik, Universität Regensburg, Regensburg, Germany.
Optical Sciences Center, University of Arizona, Tuscon, Arizona, USA.

Software:
BAYES

Hardware:
Unknown

Key Points
- Standard belief network mechanism
- Simple two-level network model

References:
[Abmayr *et al* 94]

Precis
This is a simple, five node network used for classifying lesions. The network consists of a diagnosis node and four evidence nodes. There are no intermediate nodes. Good results are reported.

Cardio-Net

Project:
Cardio-Net†

Application:
Diagnosis

Domain:
Cardiovascular Hemodynamics

Date:
1989

Authors:
W. Long

Organisations:
Laboratory for Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

Software:
Unknown

Hardware:
Symbolics 3650

Key Points

- Domain model includes cycles
- Approximation through *noisy-OR* relationship
- Approximation through other archetypal relationships
- The modelling of qualitative terms
- The use of heuristic reasoning
- The use of abstracted notion of *probabilistic causality*
- The use of simple two-state nodes

References:
[Long 89]

Precis

This system provides physicians with a reasoned differential diagnosis on the basis of patient history, evidence, and a causal knowledge base. The domain is heart failure, a complex domain with a multitude of possible causes and strong compensatory mechanisms to consider.

The disease model must enable the system to distinguish between situations that are similar yet require different therapies. The hypotheses must also be complete and explicit. These requirements demand a detailed disease model and have lead to a system that considers hypotheses that are constructed from causal chains that link *primary causes* (those not requiring further causal justification) to the *findings* (evidence). The generation of possible hypotheses is far from trivial as a useful differential diagnosis must contain only those hypotheses that are comprehensive in their explanatory power, but that are also significantly different to other hypotheses in the set. This problem has been solved by designating certain nodes within the domain network model as *diagnostic nodes.* Hypotheses are considered to be different if the sets of diagnostic nodes they contain are different. This heuristic, based on high-level domain knowledge, has proved an effective solution to this particular problem.

The domain is modelled as a simplified network, based on the notion of *probabilistic causality*, each causal relation is summarised as the probability of the cause producing the effect. Causes are defined as preceding the effect and all nodes are either true or false. One particularly interesting feature of the selected domain is the occurrence of circularities or feedback loops. The problems typically encountered with such loops are avoided through the definition of the cause effect relationship, any time a feedback loop could be completed, the effect node will already be true, blocking the circularity.

The probabilities on the links are summaries of different types of causal relationship. In some cases such a representation is not appropriate and an equation is used instead. Qualitative descriptors are also important in the domain, typically these are divided into a limited number of partitions which reflect different causal roles or effects, whilst at the same time reduce the computational costs. For example, the level at which the effects of low cardiac output will start occurring varies from patient to patient. The program models this by specifying the probability that each qualitative partition produces a measured range in an associated parameter:

low: (range cardiac-index 0.7 2.3 0.2 2.5 0.1 2.7 0.0)

normal: (range cardiac-index 0.0 2.3 0.05 2.5 0.05 2.7 0.9)

thus 70% of low cardiac outputs have a cardiac index (cardiac output normalised for body size) below 2.3, 20% between 2.3 and 2.5 and so on. Similarly no normal cardiac outputs are below 2.3, but 5% are between 2.3 and 2.5.

Link probabilities may be fixed, dependent on patient parameters, or dependent on the diagnostic hypothesis. In the majority of cases there is insufficient knowledge to determine the situations in which the probability of a causal link should be changed, so the probability is generally represented as a fixed number. In cases where the link relationship is understood it can be modelled. For instance, pneumonia is less likely to produce a fever in the elderly than the young, the probability on the link can be defined as:

pneumonia: p(fever) = (0.9 (range age 0.95 70 0.9 80 0.8 90 0.7))

meaning if nothing is known about age, 0.9 should be used, if the patient's age is less than 70 then the probability of fever is 0.95, if the age is between 70 and 80, the probability is 0.9, and so on. These dependencies relating to patient parameters are incorporated by adjusting the relevant links before the computation of the differential diagnosis, *i.e.* the general model is tailored to produce a patient-specific model.

The *noisy-OR* relationship is used as a general combination mechanism, but two special cases have specific mechanisms. The first interaction models factors that increase the probability of a cause producing an effect, but cannot produce the effect in their own right, the probabilistic contribution of such factors are combined as if they were causes when another cause is present or zero otherwise. The second interaction is the reverse of this, factors that reduce the probability of a cause producing an effect. These factors are combined with existing causes, but multiplicatively decrease the causal probability. For instance, if the causes imply a 0.5 probability of producing an effect and there is a reducing factor that prevents the effect in 80% of cases, the resultant probability is 0.1. This collection of mechanisms were sufficient to represent the relationships in the domain as a probabilistic network that could be used as a knowledge base for diagnostic reasoning.

The network model cannot be used directly as a belief network as it contains both cycles and loops. Removing the cycles still resulted in a network with an infeasible number of loops, the cutset for the network contained over forty nodes. The network was instead used as a basis for heuristic reasoning. In addition to these heuristics, a number of computational tactics were employed in order to improve efficiency.

The system proved effective for generating the differential diagnosis in some 40 test cases.

CPNEDIT-ICU

Project:
CPNEDIT-ICU[1]

Application:
Patient monitoring

Domain:
Intensive care patients

Date:
1994

Authors:
P. Laursen

Organisations:
S & W Medico Teknik A/S, Albertslund, Denmark.

Software:
CPNEDIT

Hardware:
IBM-compatible PC

Key Points
- Mapping of continuous variables to discrete states

References:
[Laursen 94]

Precis

This is an investigative project concerned with the design of patient monitoring systems that raise alarms only in response to situations that threaten patient well being. Other causes of alarms in traditional systems include monitor malfunction and noise. In attempting to distinguish between genuine and spurious alarms, the inherent redundancy in the evidence is exploited.

An initial model has been developed that includes 10 evidence nodes, plus nodes which model monitor errors, whether intervention is in progress and whether a car diovascular event is occurring. The evidence comes from the monitored physiological parameter values. These continuous values are transformed into probabilities over the states [high, normal, low] and [rising, stable, falling]. Each parameter is transformed according to a specific rule, and it is suggested that these rules should be context dependent.

Due to lack of a suitable comparison there has been no formal testing, but subjective analysis suggests that the approach is promising.

Database-Net

Project:
Database-Net†

Application:
Database analysis/diagnosis

Domain:
Cytology

Date:
1990

Authors:
P.H. Bartels[1]
M. Bibbo[2]
H. Dytch[2]
D. Thompson[1]
G.L. Wied[2]

Organisations:
[1]Optical Sciences Centre and Department of Pathology, University of Arizona, Tucson, Arizona, USA.
[2]Section of Cytopathology, Departments of Pathology and of Obstetrics and Gynecology, University of Chicago, Chicago, Illinois, USA.

Software:
C, based on [Morawski 89a, Morawski 89b]

Hardware:
Unknown

Key Points
- Standard belief network

References:
[Weid *et al* 90]

Precis
Pearl's propagation algorithm is used to perform diagnosis based on cytologic data. It forms the uncertainty management component of a larger system designed to explore an analyse large cytologic databases. Other components include an associative network expert system, a neural network module and an unsupervised learning module.

Drug-Net

Project:
Drug-Net†

Application:
Treatment prediction

Domain:
Adverse drug reaction rates

Date:
1991 — 1993

Authors:
R.G. Cowell[1]
A.P. Dawid[1]
T.A. Hutchinson[2]
S. Roden[3]
D.J. Spiegelhalter[4]

Organisations:
[1]Department of Statistical Science, University College London, London, UK.
[2]Division of Clinical Epidemiology, Royal Victoria Hospital and McGill University, Montreal, Canada.
[3]Glaxo Group Research, Greenford, UK.
[4]MRC Biostatistics Unit, Cambridge, UK.

Software:
Unknown

Hardware:
Unknown

Key Points
- Use of clique-tree propagation algorithm
- Modelling of simple temporal data

References:
[Spiegelhalter *et al* 91]
[Cowell *et al* 93a]

Precis

This project is concerned with the modelling of suspected adverse drug reactions, particularly in cases where there is more than one drug to consider. The experimental nature of the models is emphasised.

The models typically contain loops, and the clique-tree propagation algorithm of Lauritzen and Spiegelhalter is used. Some very simple temporal relations are included in the models, though there is no attempt to model changes over time. The possibility of considering expert estimations of the model as starting points and refining these estimates over time as data accumulates is mentioned.

Although there has been no formal evaluation of the project, it is reported that experts consider the models to be reasonable.

In the later work the models are extended in order to predict future adverse reaction rates. This includes the modelling of the reporting procedure and clinical practices.

GAMEES

Project:
GAMEES

Application:
Therapy monitoring

Domain:
Uremic anemia. Cytotoxic chemotherapy in breast cancer

Date:
1991 - 1993

Authors:
R. Bellazzi[1]
C. Berzuini[1]
M. Leaning[2]
D. Spiegelhalter[3]
S. Quaglini[1]

Organisations:
[1]Dipartimento di Informatica e Sistemistica, Universitá di Pavia, Pavia, Italy.
[2]CORU, University College London, London, UK.
[3]MRC Biostatistics Unit, Cambridge, UK.

Software:
GAMEES [Bellazzi *et al* 91a]

Hardware:
Unknown

Key Points

- Use of Gibbs sampling for propagation
- Use of parameterised models
- The refinement of a patient-specific model over time
- Explicit modelling of a population model to estimate priors

References:
[Berzuini *et al* 91]
[Bellazzi *et al* 91b]
[Bellazzi 93]

Precis

This project is concerned with the general task of monitoring and predicting patient response to therapy over time. A domain independent model of a class of medical task is proposed.

A belief network model, using Gibbs sampling for inference, is designed to predict the future response of a patient undergoing treatment.

Patient responses are defined in terms of the parameterisation of a physiological model of the interaction between the treatment and a measurable symptom judged to be indicative of the effectiveness of that treatment. Population data in the form of a database of previous cases is used to provide mean parameter values. These parameter values can be further specialised to the current patient if it is possible to identify clusters within the population that share significant features with the current patient. These parameter estimates are combined with a patient model which contains previous details of the current patient in order to predict future response. The patient model is updated over time as more evidence becomes available, so in the longer term the patient specific parameter values will come to dominate those of the population model.

The same basic model has been tested in two domains, with good results. The first domain was the monitoring of white blood cell counts of breast cancer patients receiving chemotherapy. In the second domain, the treatment of uremic patients' anemia using r-HuEPO, a mathematical, compartmental model of the relationship between the treatment and the therapeutic goal (maintaining hemoglobin concentrations at certain levels) was developed.

Guardian

Project:
Guardian

Application:
Patient monitoring and therapy planning

Domain:
Ventilator-assisted patient monitoring

Date:
1989

Authors:
B. Hayes-Roth[1]
M. Hewett[1]
R. Hewett[1]
A. Seiver[2]
R. Washington[1]

Organisations:
[1]Knowledge System Laboratory, Stanford University, Palo Alto, California, USA.
[2]Palo Alto Veterans Administration, Medical Center, Palo Alto, California, USA.

Software:
Common LISP

Hardware:
TI Explorer

Key Points
- Use of belief network within black-board system

References:
[Hayes-Roth *et al* 89]

Precis
A belief network is used as a knowledge source in this complex black board based prototype system. The network is used to perform diagnoses. The system as a whole is able to combine opportunistic responses to unforeseen events and planned strategies. Guardian explicitly considers control of resource allocations as part of its processing cycle. Control knowledge and domain knowledge are represented independently.

Heart-Net

Project:
Heart-Net[†]

Application:
Therapy advisor

Domain:
Heart disease

Date:
1992

Authors:
F.J. Diez Vegas
J. Mira Mira

Organisations:
Departamento de Informática, Ciencias, UNED, Senda del Rey, Madrid, Spain.

Software:
Unknown

Hardware:
Unknown

Key Points
- Representation of domain independent metaknowledge
- Approximation through the *noisy-OR* relationship
- The use of simple two-state nodes

References:
[Vegas & Mira 92]

Precis

This system appears to be very much in the developmental stage. The final aim is the diagnosis of heart disease and therapy planning, with a particular emphasis on the interpretation of echocardiograms.

The choice of a belief network approach was motivated directly by the form in which domain experts described the heart disease domain.

The system defines three distinct levels of knowledge:

1. Domain knowledge for a specific field, containing anomalies, symptoms and signs and causal links. Constitutes the static (declarative) knowledge base for system.
2. Causal reasoning (high-level metaknowledge), concerning causes and effects. This is intended to be domain independent and therefore valid for all diagnostic problems.

3. Medical reasoning (low-level metaknowledge). In principal this is applicable to all medical domains. It provides a mapping between the domain knowledge and the causal reasoning metaknowledge as well as other general knowledge regarding medicine.

The two sources of metaknowledge combine to form a knowledge-based inference engine capable of acting on the domain knowledge, calculating probabilities, collecting data, explaining conclusions, and so on. The claimed advantages of this approach are its domain independence and its ability to produce explanations based on objectives and strategies.

The belief network itself appears straightforward, many of the variables have only two states, present or absent, and the standard mechanism for interaction is the *noisy-OR* relationship, except where this approximation is unacceptable. It includes a *noisy effectiveness* (a leak) associated with anomalies which models causes that are not explicitly represented in the network.

Among the points listed for future consideration are a method to resolve loops, extending the anomaly variables to include degrees of anomaly (absent, slight, moderate...), learning, and temporal reasoning.

MIDAS

Project:

MIDAS

Application:

Construction of individualised influence diagrams

Domain:

HIV with associated pulmonary disease

Date:

1994

Authors:

C.G. Hagerty[1]
C.A. Kulikowski[1]
F.A. Sonnenberg[2]

Organisations:

[1]Department of Computer Science, Rutgers University, Piscataway, New Jersey, USA.
[2]Division of General Internal Medicine, Department of Medicine, UMDNJ Robert Wood Johnson Medical School, New Brunswick, New Jersey, USA.

Software:

IDEAL

Hardware:

Apple Macintosh IIfx

Key Points

- Use of belief network to represent domain model

References:

[Sonnenberg *et al* 94]

Precis

The MIDAS system is intended to function as a domain independent tool for the automatic creation of patient-specific influence diagrams. Domain specific medical knowledge is represented separately from knowledge about decision model creation. A domain specific belief network is used to represent probabilistic dependencies between patient data and possible diagnoses. This information is then used in the construction of an influence diagram using the decision model creation knowledge.

MIDAS is able to create simple influence diagrams and formal evaluation is planned.

MUNIN

Project:
MUNIN

Application:
Diagnosis

Domain:
Electromyography

Date:
1986 — 1992

Authors:
S.K. Andersen[1,2,3]
S. Andreassen[1,2]
B. Falck[1,4]
Frank Jensen[1,5]
Finn V. Jensen[3,5]
U. Kjaerulff[3]
K.G. Olesen[1,5]
M. Woldbye[1]

Organisations:
[1]Nordjysk Udviklingscenter, Aalborg, Denmark.
[2]Institute of Electronic Systems, Aalborg University, Aalborg, Denmark.
[3]Judex Datasystemer A/S, Aalborg, Denmark.
[4]Department of Clinical Neurophysiology, Turku University Hospital, Turku, Finland.
[5]Department of Mathematics and Computer Science, Aalborg University, Aalborg, Denmark.

Software:
HUGIN [Andersen *et al* 87, Andersen *et al* 90, Jensen *et al* 91]

Hardware:
Unknown

Key Points
- Use of clique-tree based propagation algorithm
- Provision of simple user interface tools

References:
[Andersen *et al* 86b]
[Andreassen *et al* 87]
[Jensen *et al* 87a]
[Jensen *et al* 87b]
[Olesen *et al* 89]
[Andreassen *et al* 92]

Precis

This project investigates the use of belief network models in electromyography, the diagnosis of muscle and nerve diseases through the analysis of bioelectrical signal from muscle and nerve tissue.

The network model has undergone several transformations during the course of the project. Initially a small network consisting of three levels of nodes with multiple states was used. The levels represented diagnosis, pathophysiology and findings (measurements), with one, eight and fifteen nodes respectively. The diagnosis node represented eleven grades across three diseases, 'no disease' and 'other'. The 'other' state was generally diagnosed if the evidence was contradictory or erroneous. The network also contained continuous variables modelled by the normal distribution. The model was derived from textbooks, statistical data and expert opinion.

Although this initial network model contain loops, these were removed by clustering to provide a singly connected network and Pearl's propagation algorithm was used.

In addition to the diagnostic capabilities, simple tools that allowed the user to view graphically the influences between neighbouring nodes (see section 4.10 and offered guidance on evidence collection based on the consideration of entropy, were provided.

Later a more complex model of the median nerve, containing four levels of node (diseases, pathophysiological contributions, pathophysiology, and findings) was developed. This contained a larger number of loops and a propagation method based on the clique-tree propagation of Lauritzen and Spiegelhalter was used.

The most complex network developed, which modelled six muscles and eight nerves, contained over 1000 nodes, though this was divided into six subnetworks, each of which only considered a subset of the possible disorders. An informal evaluation of this model suggests that performance is at 'expert level' within the domain.

NESTOR

Project:
NESTOR

Application:
Diagnosis

Domain:
Hypercalcemia

Date:
1989

Authors:
G.F. Cooper

Organisations:
Medical Computer Science Group, Stanford University, Stanford, California, USA.

Software:
Unknown

Hardware:
Unknown

Key Points

- Importance of system/user interface acknowledged

References:
[Cooper 89]

Precis

NESTOR has a belief network knowledge-base containing seven diseases and around 100 symptoms and pathological states. The structural information was acquired from text books, this was then refined and quantified by a domain expert. A key design decision was to give the physician control over the computer interaction. NESTOR has been designed to criticise its hypotheses and explicitly record any assumptions made.

Pathfinder-2

Project:

Pathfinder-2

Application:

Diagnosis

Domain:

Lymph-node diseases

Date:

1990 — 1992

Authors:

G.F. Cooper[1]
D.E. Heckerman[2]
B.N. Nathwani[3]
H.J. Suermondt[1]

Organisations:

[1]Medical Computer Science Group, Stanford University, Stanford, California, USA.
[2]Department of Computer Science, University of California, Los Angeles, USA.
[3]Department of Pathology, University of Southern California, Los Angeles, USA.

Software:

SimNet

Hardware:

25 Megahertz 486 processor and maths co-processor

Key Points

- Use of *similarity networks* in belief network construction
- Propagation method based on clique-tree algorithm

References:

[Heckerman 90b]
[Suermondt *et al* 91]
[Heckerman & Nathwani 92]

Precis

This system is designed to perform diagnosis among more than sixty diseases of the lymph nodes. The belief network model for the domain is multiply connected and the clique-tree propagation algorithm of Lauritzen and Spiegelhalter provides the basis of the inference mechanism. This was later expanded to include cutset conditioning to further improve efficiency. The construction or updating of a diagnosis is achieved in less than one second.

The complexity and size of the domain meant that it was infeasible to construct the network directly due to the large number of conditional independence assertions that would need specifying. In order to construct the belief network, an intermediate model called a *similarity network* was developed. This representation allows the construction of the network to be decomposed into a number of smaller tasks. A similarity network consists of a *similarity graph* and a collection of local belief networks. The similarity graph is an undirected graph with nodes representing mutually exclusive diseases and edges connecting diseases that are judged to be similar or difficult to discriminate. Each edge is associated with a belief network that contains only those features judged relevant to the discrimination of the two diseases it connects. These local belief networks are typically small and easy to assess as the diseases are by definition, similar. From the similarity network a complete belief network can be constructed using graph union. Under relatively weak conditions this network is sound.

The Pathfinder-2 network was compared on the utility of its diagnoses with the Pathfinder-1 system which assumed all features were conditionally independent given each disease. Tests on 53 cases found an average increase in expected utility of $6,000 per case.

Prostate-Net

Project:
Prostate-Net†

Application:
Diagnosis

Domain:
Grading prostate lesions

Date:
1992 — 1993

Authors:

P.H. Bartels[1]
M. Bibbo[2]
R. Christen[2]
B. Fitzpatrick[2]
H. Galera-Davidson[3]
C. Minimo[2]
T. Pfeifer[4]
D. Thompson[1]
J.E. Weber[1]
J. Xiao[2]

Organisations:

[1]Optical Sciences Centre, University of Arizona, Tucson, Arizona, USA.
[2]Department of Pathology and Cell Biology, Jefferson Medical College, Thomas Jefferson University, Philadelphia, Pennsylvania, USA.
[3]Department of Pathology, University of Seville, Seville, Spain.
[4]Section of Cytopathology, Department of Pathology, University of Chicago, USA.

Software:
C, based on [Morawski 89a, Morawski 89b]

Hardware:
PC based

Key Points
- Use of Pearl's propagation algorithm
- Use of fuzzy membership functions to determine likelihood ratios

References:
[Bartels *et al* 92]
[Bibbo *et al* 93]
[Bibbo *et al* 94]

Precis
These papers address different aspects of the grading problem. The first paper discusses the application of shallow, standard belief networks to the control of image processing tasks, specifically thresholding and segmentation.

The second paper considers the problem of diagnosis. A simple network with a single diagnostic node, 13 evidence nodes and no intermediate nodes is used, along with standard propagation algorithms. Evidence from the physician given by a value on a scale are mapped to relative likelihood ratios, through the use of fuzzy membership functions. These membership functions are subjective to a certain degree, but the model appears to be robust in the event of small changes in these functions. The conditional probability matrices were estimated from data.

Test results using four sample areas from 64 consensus graded specimens, were correct in 94% of cases. The 6 areas that were incorrectly classified were due to the poor expression of classification features within that area. Had the results from all four areas been combined, each specimen would have been correctly graded.

The third paper discusses the implementation of the system on a PC hardware platform.

QMR-DT

Project:
QMR-DT (Quick Medical Research, Decision Theoretic)

Application:
Diagnosis and treatment planning

Domain:
Internal medicine

Date:
1991

Authors:
G.F. Cooper[1]
D.E. Heckerman[2,3]
M. Henrion[2]
E.J. Horvitz[2,4]
H.P. Lehmann[2]
B. Middleton[2]
M.A. Shwe[2]

Organisations:
[1]Section of Medical Informatics, University of Pittsburgh, Pittsburg, USA.
[2]Section on Medical Informatics, Stanford University, Stanford, California, USA.
[3] Departments of Computer Science and Pathology, University of Southern California, Los Angeles, USA.
[4]Palo Alto Laboratory, Rockwell International Science Center, Palo Alto, California, USA.

Software:
Unknown

Hardware:
Unknown

Key Points
- Inference using form of stochastic simulation
- Includes decision theoretic utility evaluation
- Use of *noisy-OR* gate

References:
[Heckerman & Horvitz 91]
[Middleton *et al* 91]
[Shwe *et al* 91]

Precis
QMR-DT is a reformulation of the QMR decision support tool which was developed from the INTERNIST-1 project. Both INTERNIST and QMR have been shown to work well in the domain of internal medicine. The QMR-DT project

aims to construct a system based on the INTERNIST-1 knowledge-base, but using probabilistic inference. The ultimate aim is to develop a system that can produce cost-effective test sequences and therapy plans.

The first stage was the expression of the INTERNIST-1 knowledge-base in the form of a belief network. The network has only two levels of nodes, disease nodes and symptom nodes. There are no intermediate nodes representing pathophysiological states. Each node has only two states, present and absent.

The network model is only approximate for a variety of reasons:

- The diseases are assumed to be marginally independent. This is not always the case. Dependence is reflected to a certain extent in one of the heuristics used to improve convergence.
- The symptoms are assumed to be conditionally independent, again this is not always the case.
- The use of two-state nodes restricts the accuracy of the representation.
- A *noisy-OR* gate is used to model the conditional probabilities. Leak probabilities, representing the probability that the symptoms could occur spontaneously or due to causes outwith the model area included.
- Historical data is modelled in a causally incorrect manner.

These assumptions greatly simplify the model and inference performed on it. The probabilities for the network were derived from both the INTERNIST-1 knowledge-base and available disease statistics. Propagation is achieved using a form of stochastic simulation called *likelihood weighting* in combination with two heuristics, *importance sampling* and *self-importance sampling*, which decrease convergence time.

Investigations into the performance and sensitivity of the QMR-DT model concluded, among other things, that:

- The model is insensitive to uniform prior probabilities, though this may be due to statistical errors.
- The model is sensitive to the value of the leak probabilities.
- Based on rank ordering, the QMR-DT diagnoses were not significantly different to those of QMR.

The model has been extended to include the consideration of treatment actions. An overall utility node, subvalue utility nodes and treatment nodes are added to the model. The treatment nodes indicate the presence or absence of a particular treatment. The subvalue nodes indicate intermediate utilities based on interactions between diseases and treatments. The overall utility node measures the total utility given the diseases and treatments. This complete set of treatments and utility nodes form a comprehensive decision model for the QMR-DT domain. For a particular case the comprehensive decision model is pruned on the basis of the underlying diagnosis, which in turn depends on the patients symptoms. This patient-specific model provides a tractable approximation to the comprehensive decision model.

QMR-DT is still under development, planned improvements include correcting the model by including intermediate nodes, representing conditional probabilities more accurately, and improving the decision model.

QUALQUANT

Project:
QUALQUANT

Application:
Image processing

Domain:
Endoscopy

Date:
1991 - 1994

Authors:
D.A. Gillies[1]
D.F. Gillies[2]
L.E. Sucar[2]

Organisations:
[1]Centre for Logic and Probability in Information Technology, King's College, London, UK.
[2]Department of Computing, Imperial College of Science, Technology and Medicine, Department of Computing, London, UK.

Software:
Unknown

Hardware:
Unknown

Key Points

- Approximate model based on *multitrees*
- Emphasises objective probabilities and testing
- Consideration of temporal modelling

References:
[Sucar *et al* 91]
[Sucar *et al* 93]
[Sucar & Gillies 94]

Precis

This application addresses some of the issues associated with the task of high-level vision, the representation of visual objects and the use of that representation for recognition. The approach taken views objects in the world as *causes* of their associated features in an image. In terms of a network model, the root nodes are hypotheses about the occurrence of objects in the image and the leaf nodes are evidence from low-level vision processes. Intermediate nodes represent entities (such as sub-parts, regions and so on) and relationships (such as above, near, etc.). This network is essentially hierarchical, composed of a number of layers each of which corresponds to a level of description or abstraction in the visual domain.

Within this network it is generally the case that intermediate objects and features are represented by separate nodes corresponding to different object hypothesis nodes. At the lowest level, evidence and feature nodes, containing information derived directly from the image, will be common to several competing object hypotheses. This leads to the definition of a *multitree*, a possibly multiply connected network in which only leaf nodes may have more than one parent.

Relational variables are assumed to be well defined, and the link between components and the relationship that exists among them is considered to be a categorical link, assuming the relationship can be uniquely determined from the related components. The relational variable still possesses a probabilistic link to its non-component parent, and is treated as an instantiated node for the purposes of propagation to that parent.

Temporal knowledge can be represented in two ways within the multitree, as *semi-static recognition* or *dynamic recognition*. In semi-static recognition an object can be identified from a single image, but observations from the previous image may prove useful evidence. The information from the previous image can be used as priors for the current interpretation task. In dynamic recognition it may be necessary to examine a number of images before object recognition can be performed, in this case different nodes may become instantiated after each image is interpreted.

Within the multitree there is a *recognition tree* for each object hypothesis. This tree is rooted at the object nodes and contains all the intermediate and leaf nodes that represent that object. It is argued that whilst this representation may appear restrictive, it is adequate for the task and facilitates efficient probability propagation.

The propagation mechanism developed for these multitrees relies on two assumptions about the properties of the image interpretation task:

1. Probability propagation is usually bottom-up, from image evidence to object hypotheses.
2. Leaf nodes typically correspond to instantiated variables and will therefore have fixed values obtained from low-level vision processes.

These assumptions make it possible to render the multitree singly connected by partitioning the network at the leaf nodes, giving a copy of the node to each of its parents. By definition every non-leaf node has only a single parent, therefore the network is rendered singly connected. The propagation mechanism then merely propagates up the tree to the root node, where the evidence is combined with the prior probabilities.

The actual construction is guided by three principles:

1. As far as possible only qualitative suggestions should be sought from the domain expert, and it should be left to the computer scientist to give this a more precise quantitative form.
2. Objective probabilities should be used whenever possible.
3. All assumptions should be tested and modified if they fail the test.

Initially a qualitative model of the network structure is provided by the domain expert. The parameters necessary to quantify the network can then be calculated from a set of examples. It is recognised that an expert assessment of the parameters could be used as a starting point, but it is pointed out that if these estimates are inaccurate, the parameters will take longer to converge on their true values. An algorithm for estimating the parameters of both observable nodes and unobservable nodes (*i.e.* those which are not evidence nodes and for which the expert cannot estimate a value) is presented.

The network structure can then be examined in order to check the independence assumptions. Each recognition tree is divided into subtrees consisting of a single root node and its immediate children. Correlation values between children are calculated to determine their independence. Whilst low correlation is not a guarantee of independence it suggests that the assumption is reasonable. A high correlation values indicates that the children are not independent, in which case each of the following is tried:

1. *Node elimination* — eliminate one of the dependent nodes and its associated subtree. This is a cheap action to perform.
2. *Node combination* — combine the two dependent nodes into a single node which incorporates the information provided by the two children. Link the two subtrees to new node and new node to parent. It will be necessary to recompute the probabilities on the link, based on the joint probability distribution of the two former children.
3. *Node creation* — create a new node between the parent and dependent children, link it to the parent, both children become its children. Hopefully children are now conditionally independent given their new parent. This will usually require an external agent to define the new variable and will also require the calculation of new parameters using the parameter learning algorithms.

Most of the structural testing can be conducted automatically and most areas of a well formed knowledge-base should not require modification. The testing and modification take place before the model is used by the end user.

A multitree model has been developed for use in a navigational aid for colonoscopy, with acceptable results.

Simulation-Net

Project:
Simulation-Net[†]

Application:
Simulation

Domain:
Abdominal pain

Date:
1994

Authors:
P. Macpherson[1]
R. Stamper[2]
B.S. Todd[2]

Organisations:
[1]The Nuffield Department of Obstetrics and Gynaecology, John Radcliffe Hospital, Oxford, UK.
[2]The Programming Research Group, Oxford University Computing Laboratory, Oxford, UK.

Software:
Unknown

Hardware:
Unknown

Key Points

- Conditional probabilities represented as sets of weighted inference rules
- Monte Carlo simulation used to generate cases

References:
[Todd *et al* 94]

Precis

This system is a tool for a larger research project investigating factors which limit accuracy in medical diagnosis. This system is designed to simulate cases of abdominal or lower back pain of suspected gynaecologicial cause. An accurate simulation would enable the generation of arbitrarily large, complete, representative data sets which could have a variety of uses in the project as a whole.

The basic model is a belief network, but the conditional probabilities within the network are represented as sets of weighted inference rules. These rules are based on a logistic model, using certainty factors defined over the interval [-1, 1], where -1 represents logical preclusion and 1 represents logical implication. This type of model is able to represent a wide range of supportive and inhibitory interactions, including the *noisy-OR*. Nodes in the network represent atomic propositions of the form 'v in U', which is true precisely when the value of the variable v lies in the set of values U, for example atomic propositions about the site of tenderness include:

1. site_of_tenderness in {generalized}
2. site_of_tenderness in {right_lower_quadrant, right, right_loin}
3. site_of_tenderness in {right_upper_quadrant, right, upper}

The belief network specifies a joint probability distribution over truth assignments to these atomic propositions.

The nodes are ordered causally, the relationship being expressed in the form of rules. As an example of the form these rules take, some of the rules used to determine whether the left adnexa appears abnormally enlarged under ultrasound examination are shown [Todd *et al* 94, page 86]:

Rule 1 previous_l_salpingectomy in {true} and previous_l_oophorectomy in {true} $\Rightarrow^{-1}$ ultrasound_l_adnexa in {enlarged, mass, cyst}

This first rule is categorical. If both the left fallopian tube and the left ovary have been previously removed then no left adnexal enlargement (of any kind) is possible.

Rule 2 true $\Rightarrow^{-0.908}$ ultrasound_l_adnexa in {enlarged, mass, cyst}

This reflects the fact that usually no abnormality of the left adnexa is seen on ultrasound examination.

Rule 3 left_ectopic_pregnancy in {unruptured, ruptured_into_mesosalpinx, ruptured_into_peritoneal_cavity} $\Rightarrow^{0.610}$ ultrasound_l_adnexa in {enlarged, mass, cyst}

The presence of a left ectopic pregnancy (whether ruptured or not) makes it much more likely that some form of enlargement of the left adnexa will be detected

A total of 2143 rules are used in the network, 571 of which are categorical. These describe the relationships between 178 propositions. The certainty factors of the non categorical rules were derived from a patient data set using standard optimi sation methods. Monte Carlo simulation is used to generate cases from the model.

In testing, the distributions of some of the variables proved to be incorrect, though not wholly unrealistic, so refinement of the rules may prove necessary. In spite of this, an expert was unable to distinguish between cases generated from the model and real-world cases, suggesting that the model is substantially correct and should prove useful.

SWAN

Project:
SWAN

Application:
Therapy monitoring and planning

Domain:
Glucose metabolism in diabetes, insulin adjustment

Date:
1990 — 1994

Authors:
S. Andreassen[1,2,3]
J.J. Benn[4]
E.R. Carson[3,4]
R. Hovorka[3,4]
U. Kjaerulff[5]
K.G. Olesen[1,2]

Organisations:
[1]Nordjysk Udviklingscenter, Aalborg, Denmark.
[2]Department of Medical Informatics and Image Analysis, Institute for Electronic Systems, Aalborg University, Denmark.
[3]Centre for Measurement and Information in Medicine, Systems Science, City University, London, UK.
[4]Department of Endocrinology, Chemical Pathology and Medicine, UMDS, St. Thomas' Hospital, London, UK.
[5]Judex Datasystemer A/S, Aalborg, Denmark.

Software:
HUGIN [Andersen *et al* 87, Andersen *et al* 90, Jensen *et al* 91]

Hardware:
SUN workstation

Key Points
- Discrete-time model of temporal changes
- Refinement of patient specific models
- Risk assessment and optimal therapy selection
- Clique-tree propagation algorithm

References:
[Andreassen *et al* 90]
[Hovorka *et al* 90]
[Andreassen *et al* 91a]
[Andreassen 92]
[Hovorka *et al* 92]
[Andreassen 94]
[Andreassen *et al* 94]

Precis

This project considers the prediction of patient well being under treatment and the recommendation of therapy plans. The domain examined is type 1 diabetes under insulin therapy.

A belief network model of the physiological processes of carbohydrate metabolism was created. This model takes as its inputs, insulin absorption, carbohydrates from meals, and current plasma glucose. The model predicts the plasma glucose levels that will be observed an hour later, given a measure of the patients sensitivity to insulin. The initial model structure was simplified by the addition of new nodes to facilitate computation. The links are quantified by a variety of relationships, including the linear addition or subtraction of glucose by processes. Continuous, qualitative values are mapped to discrete ranges that reflect significant differences. These discrete time models can be connected to provide a model of a particular time period.

If prior data is available on the patient, the model can be used to determine the value of the patients insulin sensitivity, thereby improving its predictive accuracy.

The model can also be used to determine optimal therapy regimes. In order to do this a risk model is defined for various plasma glucose level outcomes. A weighted mean of penalties can be computed on the basis of the probabilities of those out comes and summed over the 24 hour model. The therapeutic regime with the lowest associated risk is found by gradient decent methods.

The system has been tested on twelve patients, showing that the predictions are rea sonable and suggesting that an average risk reduction of 32% was possible against actual treatments. There is still work to be done on this project, particularly in refining the glucose metabolism model.

VentPlan

Project:

VentPlan

Application:

Therapy planning

Domain:

Ventilator management

Date:

1989

Authors:

I. Beinlich[1]
L. Fagan[1]
B. Farr[1]
G. Rutledge[1]
L. Sheiner[2]
G. Thomsen[1]

Organisations:

[1]Section on Medical Informatics, Department of Medicine, Stanford University, Stanford, California, USA.
[2]Division of Laboratory Medicine, University of California at San Fransisco, California, USA.

Software:

DxNet

Hardware:

Unknown

Key Points

- Propagation algorithm based on Lauritzen and Spiegelhalter
- Inclusion of input validation nodes

References:

[Rutledge *et al* 89]

Precis

The VentPlan system is designed to make recommendations about the setting of ventilator controls. Qualitative and quantitative data are used to develop a patient-specific model which is used for prediction and plan selection. VentPlan is composed of four modules, a belief network, a mathematical model, a plan evaluator and a control algorithm. The belief network is used to calculate probability distributions over physiological parameters used in the mathematical model. The network detects discrepancies between reported input measurements and their likely values given other data as well as modelling physiological relationships.

The belief network inference mechanism is based on the algorithms of Lauritzen and Spiegelhalter.

Appendix B

FLAPNet Methods

There is only a single class of node, and every node is able to receive the same set of messages. However, the content of the messages and the way that a node responds to a message is determined by procedures stored in instance slots (*i.e.* each instance may have a unique set of procedures that define its behaviour). The methods available to a node object can be divided into two main groups, those directly concerned with the implementation of the inference network, and those concerned with other areas, such as interface management. Only the inference methods will be described here. The principal inference methods respond to the following message types:

- Causal message.
- Diagnostic message.
- Lambda condition message (diagnostic).
- Pi condition message (causal).
- Lambda uncondition message.
- Pi uncondition message.

Each method runs through a fixed set of operations in a predetermined order. These operations are illustrated below in pseudo-code.

B.1 Causal Message Method

The causal message method handles all causal messages sent to a node by its parents. The message contains the name of the originator of the message and the item of causal evidence.

```
defmethod causal_message(origin, evidence)
    update parents record to include new evidence
    unless pi conditioned
    do
        update pi
        update belief
        update display
        propagate diagnostic messages
        unless lambda conditioned
        do
            propagate causal messages
        endunless
    endunless
enddefmethod
```

B.2 Diagnostic Message Method

The diagnostic message method handles all diagnostic messages sent to a node by its children. The message contains the name of the originator of the message and the item of diagnostic evidence.

```
defmethod diagnostic_message(origin, evidence)
    update children record to include new evidence
    unless lambda conditioned
    do
        update lambda
        update belief
        update display
        unless pi conditioned
        do
            propagate diagnostic messages
        endunless
        propagate causal messages
    endunless
enddefmethod
```

B.3 Lambda Condition Message Method

This method is called whenever a lambda condition message is received. The message contains the new lambda value.

```
defmethod lambda_condition(new_lambda)
    if pi conditioned
    then
        ERROR
    endif
    set lambda to new_lambda
    update belief
    update display
    propagate diagnostic messages
enddefmethod
```

B.4 Pi Condition Message Method

This method is called whenever a pi condition message is received. The message contains the new pi value.

```
defmethod pi_condition(new_pi)
    if lambda conditioned
    then
        ERROR
    endif
    set pi to new_pi
    update belief
    update display
    propagate causal messages
enddefmethod
```

B.5 Lambda Uncondition Message Method

This method is used whenever a lambda uncondition message is received. The message contains no data.

```
defmethod lambda_uncondition()
    if lambda not conditioned
    then
        ERROR
    endif
    calculate new lambda
    update belief
    update display
    propagate diagnostic messages
    propagate causal messages
enddefmethod
```

B.6 Pi Uncondition Message Method

This method is used whenever a pi uncondition message is received. The message contains no data.

```
defmethod pi_uncondition()
    if pi not conditioned or root node
    then
        ERROR
    endif
    calculate new pi
    update belief
    update display
    propagate diagnostic messages
    propagate causal messages
enddefmethod
```

Appendix C

Fetal Ultrasound Results

The complete set of results from applying the system to the four test cues and three test images are presented below. It should be noted that the test cues were not adjusted in any way when applied to different test images, therefore a good cue on one test image may be a poor cue on another. The three test images are shown in figure C.1.

The parameters used for the delta experiments and composite delta experiments are shown in the following table.

parameter	*delta value*	*composite value*
node type	delta node only	composite node only
line type	closed	closed
propagation type	bidirectional exclusive	parent to child
propagation limitation	decay 0.95	decay 0.95
	distance 25	distance 25
	circuits 2	circuits 2
initial cue weight	0.1	0.1
shape model weight	0.5	0.5
current shape weight	0.5	0.5
edge weight	1.0	1.0
stability	$\leq$ 16 pixels	$\leq$ 16 pixels
edge definition	3	3
orthogonals	30	30
sample points	59	59

The trials are presented in order, starting with test image one and cue one, followed by test image one with cue two, and so on. For each trial the initial image and cue are shown first, then the delta result, then the composite delta result.

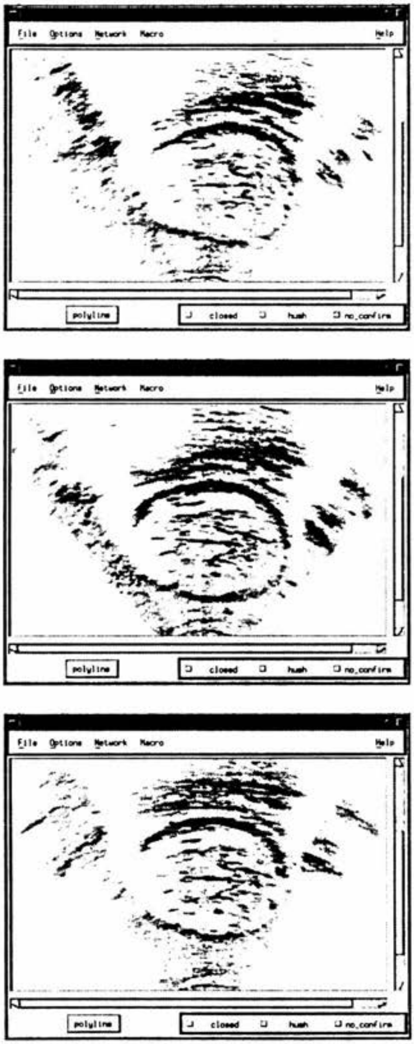


Figure C.1: Ultrasound test images

Figure C.2: Initial image, result (419*6) and composite result (419*4)

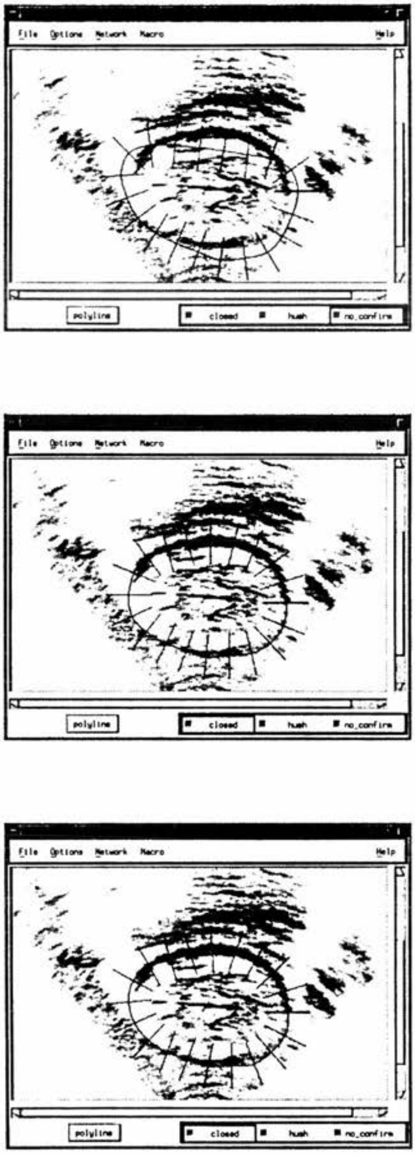


Figure C.3: Initial image, result (149*13) and composite result (149*14)

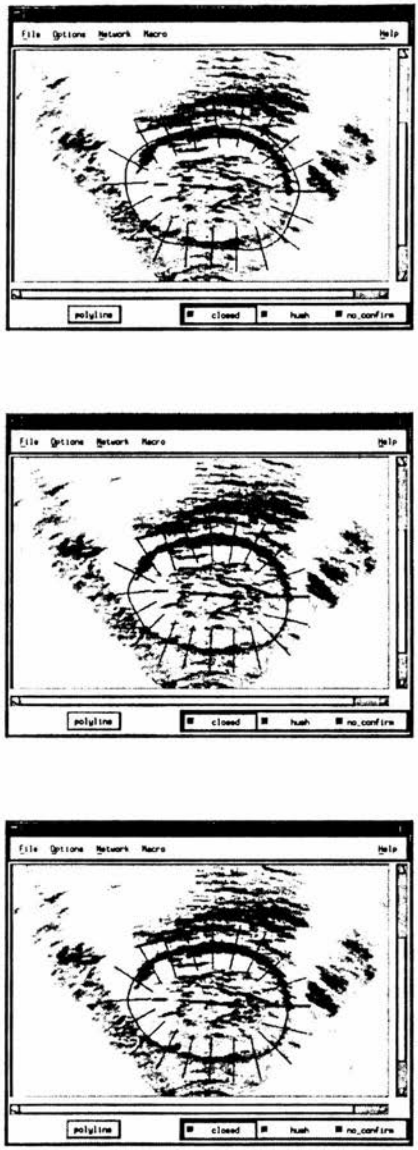


Figure C.4: Initial image, result (29*2) and composite result (29*2)

Figure C.5: Initial image, result (659*59) and composite result (659*54)

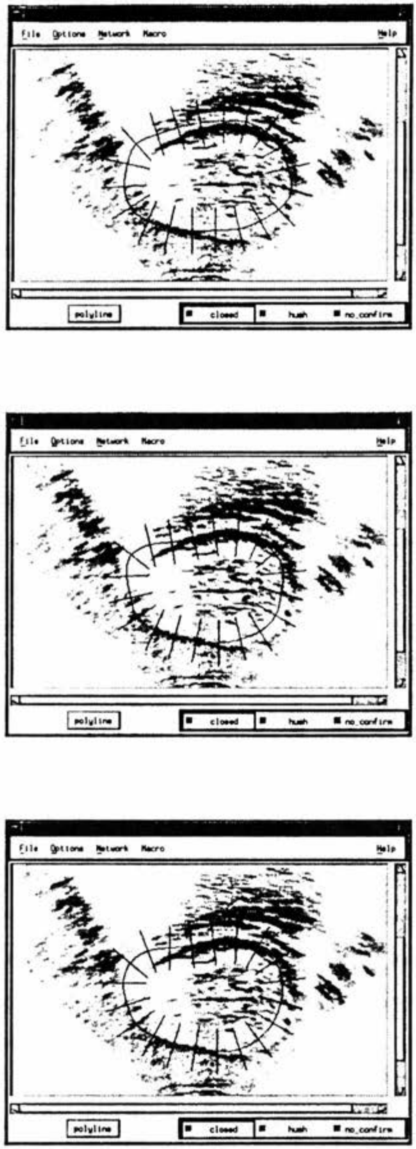


Figure C.6: Initial image, result (134*4) and composite result (134*2)

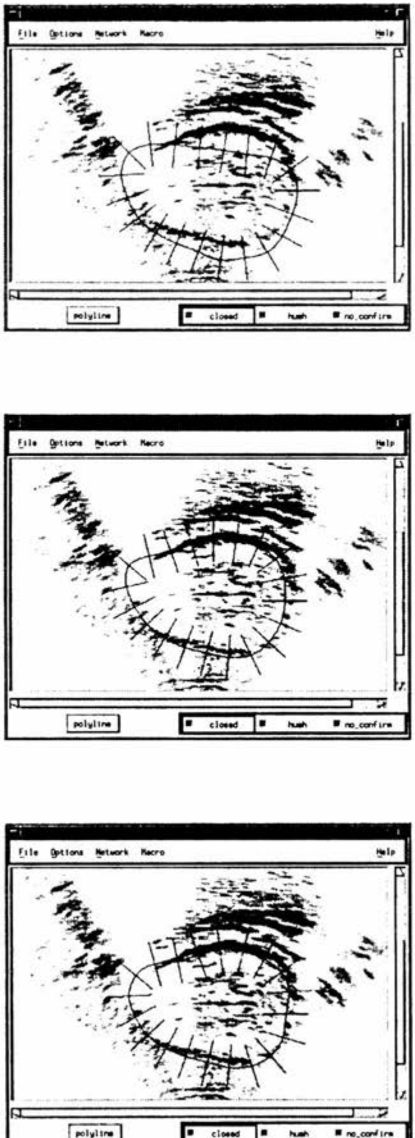


Figure C.7: Initial image, result (134*1) and composite result (134*1)

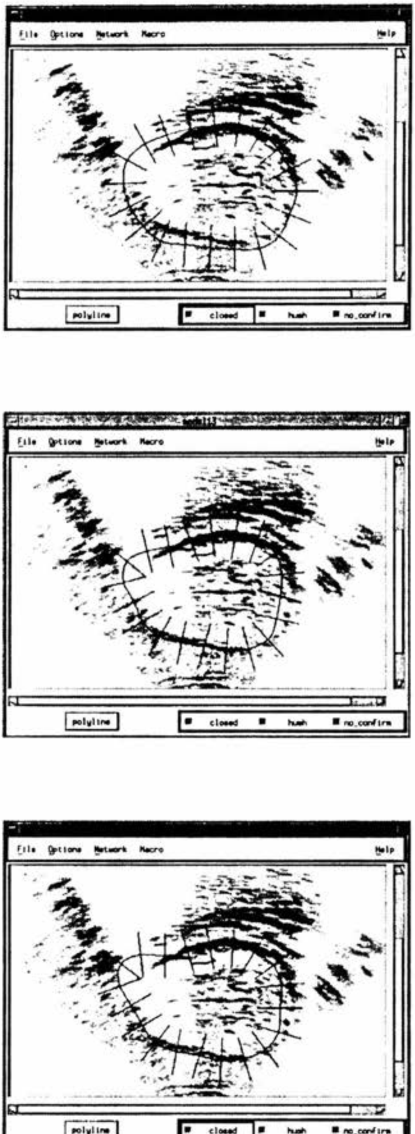


Figure C.8: Initial image, result (209*7) and composite result (209*4)

Figure C.9: Initial image, result (224∗11) and composite result (224∗3)

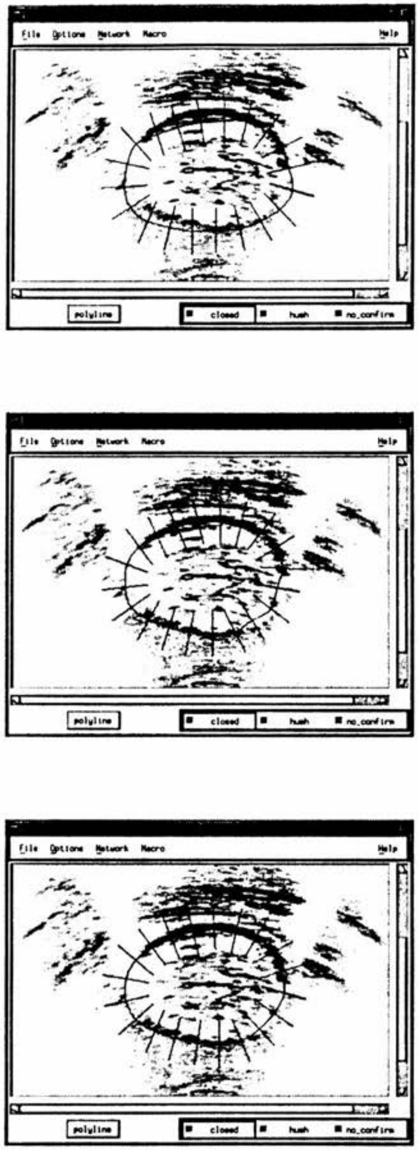


Figure C.10: Initial image, result (104∗2) and composite result (105∗6)

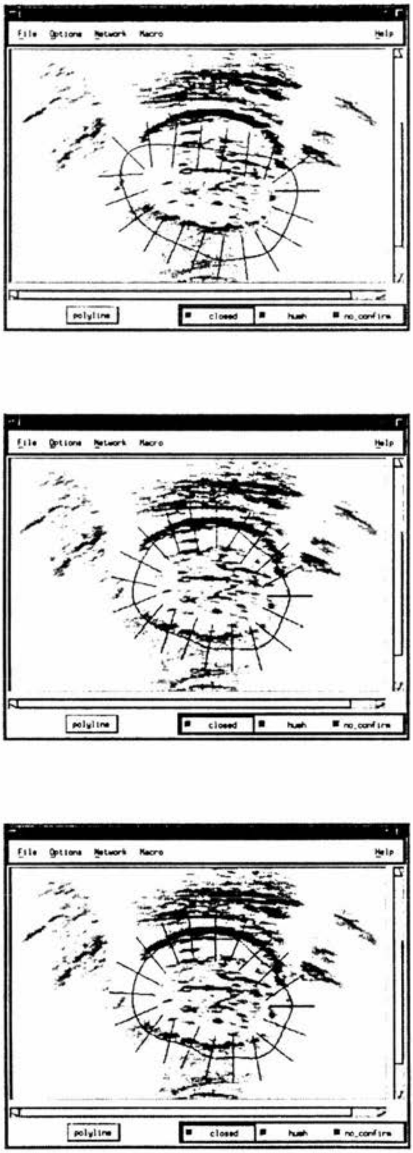


Figure C.11: Initial image, result (269*20) and composite result (269*10)

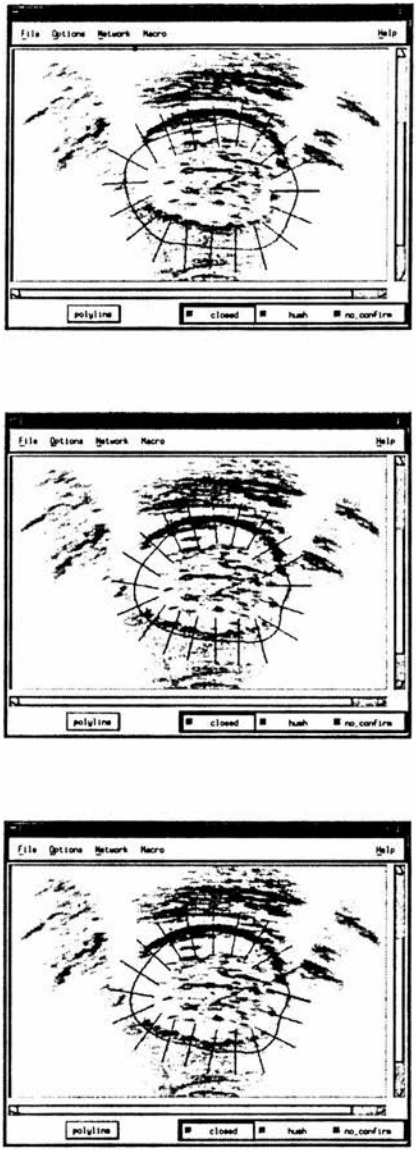


Figure C.12: Initial image, result (44*3) and composite result (44*1)

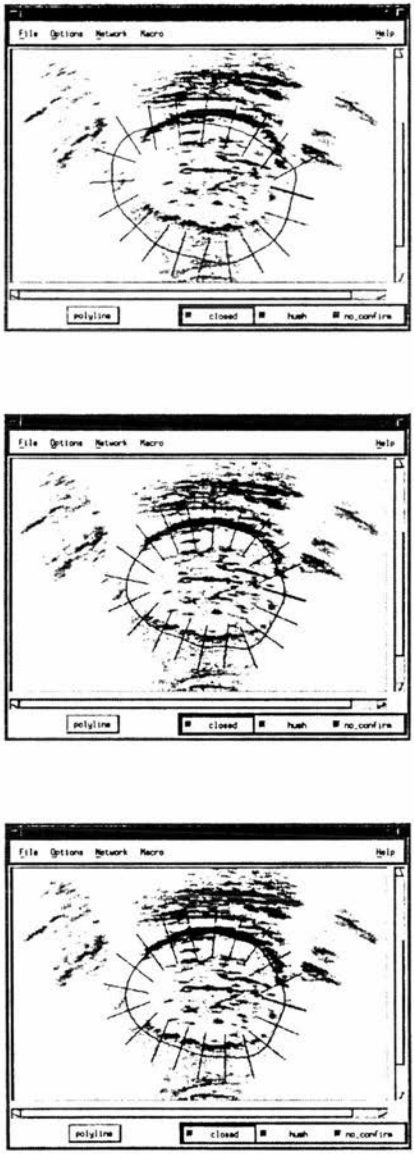


Figure C.13: Initial image, result (59∗7) and composite result (59∗2)

Bibliography

[Abel 88] S. Abel. The sum-and-lattice-points method based on an evidential-reasoning system applied to the real-time vehicle guidance problem. In J. F. Lemmer and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Abmayr *et al* 94] Wolfgang Abmayr, Wilhelm Stolz, Dirk Rehberg, Deborah Thompson, and Peter H. Bartels. Bayes Belief Network for Windows: A tool for computer aided diagnosis of macroscopic melanocytic lesions investigated by dermatoscopy. In George L. Wied, Peter H. Bartels, Dorothy L. Rosenthal, and Ulrich Schenck, editors, *Compendium on the Computerized Cytology and Histology Laboratory*, pages 96–106. Tutorials of Cytology, 1994.

[Abramson 91] Bruce Abramson. On knowledge representation in belief networks. *Lecture Notes in Computer Science*, 521:86–96, 1991.

[Agosta 90] John Mark Agosta. The structure of Bayes networks for visual recognition. In R. D. Shachter, T. S. Levitt, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 4*. North Holland, 1990.

[Aikins *et al* 84] Janice S. Aikins, John C. Kunz, Edward H. Shortliffe, and Robert J. Fallat. PUFF: An expert system for interpretation of pulmonary function data. In William J. Clancey and Edward H. Shortliffe, editors, *Readings in Medical Artificial Intelligence: The First Decade*. Addison Wesley, 1984.

[Al-Hajjaj & Bamgboye 92] Mohammad Saleh Al-Hajjaj and Elijah A. Bamgboye. Attitudes and opinions of medical staff towards computers. *Computers in Biology and Medicine*, 22(4):221–226, 1992.

[Andersen *et al* 86a] S. K. Andersen, S. Andreassen, F. V. Jensen, and M. Woldbye. Muscle-LOP — details for implementation. Technical Report C-NUC-860323, The University of Aalborg, Denmark, 1986.

[Andersen *et al* 86b] S. K. Andersen, S. Andreassen, and M. Woldbye. Knowledge representations for diagnosis and test planning in the domain of electromyography. In *Proceedings of the 7th European Conference on Artificial Intelligence*, pages 357–368, 1986. Also published in Advances in Artificial Intelligence 2, edited by B. Du Bulay, D. Hogg and L. Steels, North Holland, 1987.

[Andersen *et al* 86c] S. K. Andersen, F. Jensen, and F. V. Jensen. Theory behind belief and importance in LOP. Technical Report C-NUC-860522, The University of Aalborg, Denmark, 1986.

[Andersen *et al* 87] Stig K. Andersen, Finn V. Jensen, and Kristian G. Olesen. The HUGIN core — preliminary considerations on systems for fast manipulation of probabilities. In *Proceedings of the Workshop on Inductive Reasoning: Managing Empirical Information in AI-Systems*, 1987.

[Andersen *et al* 90] Stig K. Andersen, Kristian G. Olesen, Finn V. Jensen, and Frank Jensen. HUGIN — a shell for building Bayesian belief universes for expert systems. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Proceedings of the 10th International Joint Conference on Artificial Intelligence*, 1989.

[Anderson & Kettel 82] Robert M. Anderson and Louis J. Kettel. The impact of computers in medicine. *Arizona Medicine*, 39(3):193, 1982.

[Andreassen 90] Steen Andreassen. Decisions based on qualitative and quantitative reasoning. In Talmon and Fox, editors, *Knowledge-Based Systems in Medicine: Methods, Applications and Evaluations*, pages 68–79. Springer Verlag, 1990. Lecture Notes in Medical Informatics, volume 47.

[Andreassen 92] Steen Andreassen. Planning of therapy and tests in causal probabilistic networks. *Artificial Intelligence in Medicine*, 4(3):227–241, 1992.

[Andreassen 94] Steen Andreassen. Model-based biosignal interpretation. *Methods of Information in Medicine*, 33:103–110, 1994.

[Andreassen *et al* 87] Steen Andreassen, Marianne Woldbye, Bjorn Falck, and Stig K. Andersen. MUNIN — a causal probabilistic network for the interpretation of electromyographic findings. In *Proceedings of the 10th International Joint Conference on Artificial Intelligence*, pages 366–372, 1987.

[Andreassen *et al* 90] Steen Andreassen, Jonathan Benn, Ewart Carson, Roman Hovorka, Uffe Kjaerulff, and Kristian G. Olesen. A causal probabilistic network model of carbohydrate metabolism for insulin therapy adjustment. In *Proceedings of the 12th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, page 1011, 1990.

[Andreassen *et al* 91a] Steen Andreassen, Roman Hovorka, Jonathan Benn, Kristian G. Olesen, and Ewart R. Carson. A model-based approach to insulin adjustment. In M. Stefanelli, A. Hasman, M. Fieschi, and J. Talmon, editors, *Proceedings of AIME-91, 3rd Artificial Intelligence in Medicine Conference*, pages 239–248, 1991.

[Andreassen *et al* 91b] Steen Andreassen, Finn V. Jensen, and Kristian G. Olesen. Medical expert systems based on causal probabilistic networks. *International Journal of Biomedical Computing*, 28:1–30, 1991.

[Andreassen *et al* 92] Steen Andreassen, Bjorn Falck, and Kristian G. Olesen. Diagnostic function of the microhuman prototype of the expert system — MUNIN. *Electroencephalography and Clinical Neurophysiology*, 85:143–157, 1992.

[Andreassen *et al* 94] Steen Andreassen, Jonathan J. Benn, Roman Hovorka, Kristian G. Olesen, and Ewart R. Carson. A probabilistic approach to glucose prediction and insulin dose adjustment: Description of metabolic model and pilot evaluation study. *Computer Methods and Programs in Biomedicine*, 41:153–165, 1994.

[Baker & Boult 91] Michelle Baker and Terrance E. Boult. Pruning Bayesian networks for efficient computation. In P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 6*. Elsevier Science Publishers, 1991.

[Baldock & Towers 87] Richard Baldock and Simon Towers. Ultrasound image interpretation tasks for a medical image expert system. Technical Report Alvey/MMI-134/MRC/004, MRC Human Genetics Unit, Edinburgh, 1987.

[Baldock & Towers 88] Richard Baldock and Simon Towers. First steps towards a blackboard controlled system for matching image and model in the presence of noise and distortion. *Lecture Notes in Computer Science*, 301:429–438, 1988.

[Baldock 90] Richard Baldock. Probabilistic reasoning/reasoning under uncertainty. Technical Report MOBPRIM/MRC/DOC7/900110, MRC Human Genetics Unit, Edinburgh, 1990.

[Baldock *et al* 87] Richard Baldock, John Ireland, and Simon Towers. A pilot study of knowledge-based control for image processing. SBS Report 1, MRC Human Genetics Unit, Edinburgh, 1987.

[Banda-Gamboa *et al* 92] Hugo Banda-Gamboa, Ian Ricketts, Alistair Cairns, Kudair Hussein, James H. Tucker, and Nasseem Husain. Automation in cervical cytology: An overview. *Analytical Cellular Pathology*, 4:25–48, 1992.

[Barnett 84] Jeffrey A. Barnett. How much is control knowledge worth? A primitive example. *Artificial Intelligence*, 22:77–89, 1984.

[Barnett 90] Octo Barnett. Computers in medicine. *Journal of the American Medical Association*, 263(19):2631–2633, 1990.

[Bartels & Weber 92] Peter H. Bartels and Jean E. Weber. Decision strategies and methodologies applicable to cytological prescreening. In *Proceedings of the 2nd Annual International Symposium on Automated Cervical Cancer Screening*, 1992.

[Bartels *et al* 92] Peter H. Bartels, Deborah Thompson, Marluce Bibbo, and Jean E. Weber. Bayesian belief networks in quantitative histopathology. *Analytical and Quantitative Cytology and Histology*, 14(6):459–473, 1992.

[Barth & Norton 88] S. W. Barth and S. W. Norton. Knowledge engineering within a generalized Bayesian framework. In J. F. Lemmer and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Barthet & Hanachi 91] Marie-France Barthet and Chihab Hanachi. What kind of interface for expert systems? *Expert Systems with Applications*, 2:195–200, 1991.

[Bartoo *et al* 92] Grace T. Bartoo, James S. J. Lee, Peter H. Bartels, Nancy B. Kiviat, and Alan C. Nelson. Methods in laboratory investigation automated prescreening of conventionally prepared cervical smears: A feasibility study. *Laboratory Investigation*, 66(1):116–122, 1992.

[Batson 84] Eric Batson. Image processing: Where computers have had greatest medical impact. *Postgraduate Medicine*, 76(2):73–76, 1984.

[Bauman 81] William A. Bauman. The human side of computers in medicine. *Medical Instrumentation*, 15(2):115, 1981.

[Baumberg & Hogg 93] A. M. Baumberg and D. C. Hogg. Learning flexible models from image sequences. Research Reports Series 93.36, Division of Artificial Intelligence, School of Computer Studies, University of Leeds, 1993.

[Baumberg & Hogg 94] A. M. Baumberg and D. C. Hogg. An efficient method for contour tracking using active shape models. Research Reports Series 94.11, Division of Artificial Intelligence, School of Computer Studies, University of Leeds, 1994.

[Beinlich & Gaba 89] I. A. Beinlich and D. M. Gaba. The ALARM monitoring system — intelligent decision making under uncertainty. *Anesthesiology*, 71(3A):A337, 1989. Abstract only.

[Beinlich *et al* 89] Ingo A. Beinlich, H. J. Suermondt, R. Martin Chavez, and Gregory F. Cooper. The ALARM monitoring system: A case study with two probabilistic inference techniques for belief networks. In *AIME 89*, pages 247–256, 1989.

[Beischer *et al* 84] N. A. Beischer, D. A. Bell, and J. H. Drew. Intra-uterine growth retardation. In John Studd, editor, *Progress in Obstetrics and Gynaecology Volume 4*. Churchill Livingstone, 1984.

[Belknap *et al* 86] Robert Belknap, Edward Riseman, and Allen Hanson. The information fusion problem and rule-based hypotheses applied to complex aggregations of image events. In *IEEE Computer Vision and Pattern Recognition*, pages 227–234, 1986.

[Bellazzi 93] R. Bellazzi. Drug delivery optimization through Bayesian networks: An application to erythropoietin therapy in uremic anemia. *Computers and Biomedical Research*, 26:274–293, 1993.

[Bellazzi *et al* 91a] R. Bellazzi, S. Quaglini, C. Berzuini, and M. Stefanelli. GAMEES: A probabilistic environment for expert systems. *Computer Methods and Programs in Biomedicine*, 35:177–191, 1991.

[Bellazzi *et al* 91b] Riccardo Bellazzi, Carlo Berzuini, Silvana Quaglini, David Spiegelhalter, and Mark Leaning. Cytotoxic chemotherapy monitoring using stochastic simulation on graphical models. In *Proceedings of the AIME-91 Conference*, pages 227–238, 1991.

[Ben-Bassat & Teeni 84] Moshe Ben-Bassat and Dov Teeni. Human-orientated information acquisition in sequential pattern classification: Part 1 — single membership classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-14(1):131–138, 1984.

[Ben-Bassat 78] Moshe Ben-Bassat. Myopic policies in sequential classification. *IEEE Transactions on Computers*, C-27(2):170–174, 1978.

[Berry & Hart 91] Dianne Berry and Anna Hart. User interface standards for expert systems: Are they appropriate? *Expert Systems with Applications*, 2:245–250, 1991.

[Berzuini 90] Carlo Berzuini. Representing time in causal probabilistic networks. In M. Henrion, R. D. Shachter, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 5*. North Holland, 1990.

[Berzuini *et al* 91] Carlo Berzuini, Riccardo Ballazzi, and David Spiegelhalter. Bayesian networks applied to therapy monitoring. In Bruce D. D'Ambrosio, Phillipe Smets, and Piero P. Bonissone, editors, *Uncertainty in Artificial Intelligence, proceedings of the 7th conference*, pages 35–42. Morgan Kaufman, 1991.

[Bhatnagar & Kanal 86] Raj K. Bhatnagar and Laveen N. Kanal. Handling uncertain information: A review of numeric and non-numeric methods. In L. N. Kanal and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence*. North Holland, 1986.

[Bibbo *et al* 93] Marluce Bibbo, Peter H. Bartels, Tom Pfeifer, Deborah Thompson, Corrado Minimo, and Hugo Galera-Davidson. Belief network for grading prostate lesions. *Analytical and Quantitative Cytology and Histology*, 15(2):124–135, 1993.

[Bibbo *et al* 94] Marluce Bibbo, Corrado Minimo, John Xiao, Randoph Christen, Brendan Fitzpatrick, Hugo Galera-Davidson, and Peter H. Bartels. A workstation for objective grading of tumors. In George L. Wied, Peter H. Bartels, Dorothy L. Rosenthal, and Ulrich Schenck, editors, *Compendium on the Computerized Cytology and Histology Laboratory*, pages 89–95. Tutorials of Cytology, 1994.

[Binford *et al* 89] Thomas O. Binford, Tod S. Levitt, and Wallace B. Mann. Bayesian inference in model-based machine vision. In L. N. Kanal, T. S. Levitt, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 3*. North Holland, 1989.

[Biswas & Anand 89] Gautam Biswas and Tejwansh S. Anand. Using the Dempster–Shafer scheme in a mixed-initiative expert system shell. In L. N. Kanal and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 3*. North Holland, 1989.

[Black & Laskey 90] Paul K. Black and Kathryn B. Laskey. Hierarchical evidence and belief functions. In R. D. Shachter, T. S. Levitt, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 4*. North Holland, 1990.

[Black 37] Max Black. Vagueness: An exercise in logical analysis. *Philosophy of Science*, 4:427–455, 1937.

[Boden 77] Margaret A. Boden. *Artificial Intelligence and Natural Man*. The Harvester Press, 1977.

[Bonissone & Tong 85] Piero P. Bonissone and Richard M. Tong. Editorial: Reasoning with uncertainty in expert systems. *International Journal of Man-Machine Studies*, 22:241–250, 1985.

[Bonissone 83] Piero P. Bonissone. Coping with uncertainty in expert systems: A comparative study. In *Proceedings of the American Control Conference (ACC)*, pages 1230–1232, 1983.

[Bonissone 87] Piero P. Bonissone. Reasoning, plausible. In S. C. Shapiro, editor, *Encyclopedia of Artificial Intelligence, Volume 2*. John Wiley and Sons, 1987.

[Bonissone 90] Piero P. Bonissone. Summarizing and propagating uncertain information with triangular norms. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *International Journal of Approximate Reasoning*, 1:71–101, 1987.

[Booker & Hota 88] L. B. Booker and N. Hota. Probabilistic reasoning about ship images. In J. F. Lemmer and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Booker *et al* 90] Lashon B. Booker, Naveen Hota, and Connie Logia Ramsey. BaRT: A Bayesian reasoning tool for knowledge based systems. In M. Henrion, R. D. Shachter, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 5*. North Holland, 1990.

[Bowie & Andreotti 83] James D. Bowie and Rochelle Filker Andreotti. Estimating gestational age in utero. In P. W. Callen, editor, *Ultrasonography in Obstetrics and Gynecology*. W. B. Saunders Company, 1983.

[Brachman & Levesque 85] Ronald J. Brachman and Hector J. Levesque, editors. *Readings in Knowledge Representation*. Morgan Kaufmann, 1985.

[Breese & Fehling 90] John S. Breese and Michael R. Fehling. Control of problem solving: Principles and architecture. In R. D. Shachter, T. S. Levitt, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 4*. North Holland, 1990.

[Breese & Horvitz 91] John S. Breese and Eric J. Horvitz. Ideal reformulation of belief networks. In P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 6*. Elsevier Science Publishers, 1991.

[Brinkley 87] James F. Brinkley. The potential for intelligent three-dimensional ultrasound. Knowledge Systems Laboratory Report KSL 88-24, Knowledge Systems Laboratory, Computer Science Department, Stanford University, California, 1987.

[Brooks 90] Rodney A. Brooks. A robust layered control system for a mobile robot. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *IEEE Journal of Robotics and Automation*, RA-2:1, March, 1986.

[Buekens *et al* 93] F. Buekens, W. Ceusters, and G. De Moor. The explanatory role of events in causal and temporal reasoning in medicine. *Methods of Information in Medicine*, 32:274–278, 1993.

[Burns & Pearl 81] Michael Burns and Judea Pearl. Causal and diagnostic inferences: A comparison of validity. *Organizational Behaviour and Human Performance*, 28:379–394, 1981.

[Buxton 89] Richard Buxton. Modelling uncertainty in expert systems. *International Journal of Man-Machine Studies*, 31:415–476, 1989.

[Byrne *et al* 94] N. J. Byrne, A. M. Baumberg, and D. C. Hogg. Using shape and intensity to track non-rigid objects. Research Reports Series 94.14, Division of Artificial Intelligence, School of Computer Studies, University of Leeds, 1994.

[Campbell & Pearce 85] S. Campbell and J. M. F. Pearce. Ultrasound in obstetrics and gynaecology. In R. R. Macdonald, editor, *Scientific Basis of Obstetrics and Gynaecology*. Churchill Livingstone, 1985.

[Cayrol *et al* 80] Michel Cayrol, Henri Farreny, and Henri Prade. Possibility and necessity in a pattern-matching process. In *Proceedings of the 9th International Congress on Cybernetics*, pages 53–65, 1980.

[Chamberlain 89] J. Chamberlain. Reasons that some screening programmes fail to control cervical cancer. In M. Hakama, A. B. Miller, and N. E. Day, editors, *Screening for Cancer of the Uterine Cervix, IARC Scientific Publications number 76*, pages 161–168. International Agency for Research on Cancer, 1989.

[Chandrasekaran & Tanner 86] B. Chandrasekaran and Michael C. Tanner. Uncertainty handling in expert systems: Uniform vs task-specific formalisms. In L. N. Kanal and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence*. North Holland, 1986.

[Chang & Fung 89] Kou-Chu Chang and Robert Fung. Node aggregation for distributed inference in Bayesian networks. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pages 265–270, 1989.

[Chang & Fung 91] Kou-Chu Chang and Robert Fung. Refinement and coarsening of Bayesian networks. In P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 6*. Elsevier Science Publishers, 1991.

[Charniak & McDermott 85] Eugene Charniak and Drew McDermott. *Introduction to Artificial Intelligence*. Addison-Wesley, 1985.

[Charniak 83] Eugene Charniak. The Bayesian basis of common sense medical diagnosis. In *Proceedings of the AAAI*, pages 70–73, 1983.

[Charniak 91] Eugene Charniak. Bayesian networks without tears. *AI Magazine*, pages 50–63, Winter 1991.

[Chavez & Cooper 88] R. Martin Chavez and Gregory F. Cooper. KNET: Integrating hypermedia and Bayesian modelling. In *Proceedings of the 4th AAAI Workshop on Principles on Uncertainty in Artificial Intelligence*, pages 49–54, 1988.

[Chavez & Cooper 90] R. Martin Chavez and Gregory F. Cooper. An empirical evaluation of a randomized algorithm for probabilistic inference. In M. Henrion, R. D. Shachter, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 5*. North Holland, 1990.

[Cheeseman 84] Peter Cheeseman. Learning of expert systems from data. In *Proceedings of the Workshop on Principles of Knowledge-based Systems*, pages 115–122, 1984.

[Cheeseman 85] Peter Cheeseman. In defence of probability. In *Proceedings of the 9th International Joint Conference on Artificial Intelligence*, pages 1002–1009, 1985.

[Cheeseman 86] Peter Cheeseman. Probabilistic versus fuzzy reasoning. In L. N. Kanal and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence*. North Holland, 1986.

[Chen 88a] K. Chen. Learning to predict: An inductive approach. In J. F. Lemmer and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Chen 88b] S.-S. Chen. Some extensions of probabilistic logic. In J. F. Lemmer and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Chou & Brown 87] Paul B. Chou and Christopher M. Brown. Probabilistic information fusion for multi-modal image segmentation. In *Proceedings of the 10th International Joint Conference on Artificial Intelligence*, pages 779–782, 1987.

[Chudleigh & Pearce 86] Patricia Chudleigh and J. Malcolm Pearce. *Obstetrical Ultrasound: How, Why and When*. Churchill Livingstone, 1986.

[Clancey 83] William J. Clancey. The epistemology of a rule-based expert system — a framework for explanation. *Artificial Intelligence*, 20:215–251, 1983.

[Clark 90] Dominic A. Clark. Numerical and symbolic approaches to uncertainty management in AI. *Artificial Intelligence Review*, 4:109–146, 1990.

[Clarke 87] S. J. Clarke. Textural discrimination in fetal ultrasound - a project report. Technical report, Heriot-Watt University, Dept of Electrical and Electronic Engineering, 1987.

[Cobelli *et al* 84] C. Cobelli, E. R. Carson, L. Finkelstein, and M. S. Leaning. Validation of simple and complex models in physiology and medicine. *American Journal of Physiology*, 246(2):259–266, 1984.

[Cohen & Grinberg 83] Paul R. Cohen and Milton R. Grinberg. A theory of heuristic reasoning about uncertainty. *The AI Magazine*, pages 17–24, Summer 1983.

[Cohen & Howe 88] Paul R. Cohen and Adele E. Howe. How evaluation guides AI research. *The AI Magazine*, 9(4):35–43, 1988.

[Cohen 86] Paul R. Cohen. Numeric and symbolic reasoning in expert systems. In *Proceedings of the 7th European Conference on Artificial Intelligence*, pages 413–426, 1986.

[Cohen 89] Paul R. Cohen. Steps towards programs that manage uncertainty. In L. N. Kanal, T. S. Levitt, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 3*. North Holland, 1989.

[Cohen 90] Paul R. Cohen. The control of reasoning under uncertainty: A discussion of some programs. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *The Knowledge Engineering Review*, 1987.

[Cohen *et al* 85] M. S. Cohen, S. R. Watson, and E. Barrett. Alternative theories of inference in expert systems for image analysis. Technical report, Decision Science Consortium, 1985.

[Collins 90] Allan Collins. Fragments of a theory of human plausible reasoning. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Theoretical Issues in Natural Language Processing*, 1978.

[Collste 92] Goran Collste. Expert systems in medicine and moral responsibility. *Journal of Systems Software*, 17:15–22, 1992.

[Cooper & Herskovits 92] Gregory F. Cooper and Edward Herskovits. A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9:309–347, 1992.

[Cooper & Musen 90] Gregory F. Cooper and Mark A. Musen. Artificial Intelligence in medicine. *AI Magazine*, 11(3):27–28, 1990.

[Cooper 88] Gregory F. Cooper. A method for using belief networks as influence diagrams. In *Proceedings of the AAAI*, pages 53–63, 1988.

[Cooper 89] Gregory F. Cooper. Current research directions in the development of expert systems based on belief networks. *Applied Stochastic Models and Data Analysis*, 5:39–52, 1989.

[Cooper 90] Gregory F. Cooper. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, 42:393–405, 1990.

[Cootes & Taylor 92] T. F. Cootes and C. J. Taylor. Active shape models — 'smart snakes'. In David Hogg and Roger Boyle, editors, *Proceedings of the British Machine Vision Conference*, pages 266–275, 1992.

[Cootes *et al* 92] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Training models of shape from sets of examples. In David Hogg and Roger Boyle, editors, *Proceedings of the British Machine Vision Conference*, pages 9–18, 1992.

[Cootes *et al* 95] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models — their training and application. *Computer Vision and Image Understanding*, 61:38–59, 1995.

[Cover 74] Thomas M. Cover. The best two independent measurements are not the two best. *IEEE Transactions on Systems, Man, and Cybernetics*, pages 116–117, January 1974.

[Cowell *et al* 93a] R. G. Cowell, A. P. Dawid, T. A. Hutchinson, S. Roden, and D. J. Speigelhalter. Bayesian networks for the analysis of drug safety. *The Statistician*, 42:396–384, 1993.

[Cowell *et al* 93b] Robert G. Cowell, A. Philip Dawid, and David J. Speigelhalter. Sequential model criticism in probabilistic expert systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(3):209–219, 1993.

[Cox 61] Richard T. Cox. *The Algebra of Probable Inference*. The John Hopkins Press, 1961.

[Cox 90] Richard T. Cox. Probability, frequency and reasonable expectation. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *American Journal of Physics*, 14:1, 1946.

[Craddock & Browse 88] A. J. Craddock and R. A. Browse. Belief as summarization and meta-support. In J. F. Lemmer and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Cruickshank 84] P. J. Cruickshank. Computers in medicine: Patients' attitudes. *Journal of the Royal College of General Practitioners*, 34:77–80, February 1984.

[Dagum & Chavez 93] P. Dagum and R. M. Chavez. Approximating probabilistic inference in Bayesian belief networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(3):246–255, 1993.

[Dagum & Luby 93] Paul Dagum and Michael Luby. Approximating probabilistic inference in Bayesian belief networks is NP-hard. *Artificial Intelligence*, 60:141–153, 1993.

[Dalkey 88] N. C. Dalkey. Model vs inductive inference for dealing with probabilistic knowledge. In J. F. Lemmer and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[D'Ambrosio 90] Bruce D'Ambrosio. Process, structure, and modularity in reasoning with uncertainty. In R. D. Shachter, T. S. Levitt, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 4*. North Holland, 1990.

[Dasta *et al* 92] Joseph F. Dasta, Marianne L. Greer, and Stuart M. Speedie. Computers in healthcare: Overview and bibliography. *The Annals of Pharmacotherapy*, 26:109–117, January 1992.

[Davis 82] Randall Davis. Expert systems: where are we? and where do we go from here? Technical Report 665, MIT A.I. Laboratory, 1982.

[Dawes *et al* 89] Robyn M. Dawes, David Faust, and Paul E. Meehl. Clinical versus actuarial judgement. *Science*, 243:1668–1674, 1989.

[de Dombal 87] F. T. de Dombal. Ethical considerations concerning computers in the 1980s. *Journal of Medical Ethics*, 13(4):179–184, 1987.

[DeGroot 62] M. H. DeGroot. Uncertainty, information, and sequential experiments. *Annals of Mathematical Statistics*, 33:404–419, 1962.

[Dempster & Kong 90] A. P. Dempster and Augustine Kong. Uncertain evidence and artificial analysis. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Journal of Statistical Planning and Inference*, 20:355–368, 1988.

[Deter *et al* 83] Russell L. Deter, Frank M. Hadlock, and Ronald B. Harrist. Evaluation of normal fetal growth and the detection of intrauterine growth retardation. In P. W. Callen, editor, *Ultrasonography in Obstetrics and Gynecology*. W. B. Saunders Company, 1983.

[Detmar *et al* 78] Don E. Detmar, Dennis G. Fryback, and Kevin Gassner. Heuristics and biases in medical decision-making. *Journal of Medical Education*, 53:682–683, 1978.

[DeVore & Hobbins 79] Greggory R. DeVore and John C. Hobbins. Fetal growth and development: The diagnosis of intrauterine growth retardation. In John C. Hobbins, editor, *Clinics in Diagnostic Ultrasound, Volume 3, Diagnostic Ultrasound in Obstetrics*. Churchill Livingstone, 1979.

[Diez & Mira 94] F. J. Diez and J. Mira. Distributed inference in Bayesian networks. *Cybernetics and Systems*, 25:39–61, 1994.

[Dodson 90] David C. Dodson. Interaction with knowledge-based systems through connection diagrams: Where next? In Dianne Berry and Anna Hart, editors, *Expert Systems: Human Issues*. Chapman and Hall, 1990.

[Doyle 90] Jon Doyle. Methodological simplicity in expert system construction. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *The AI Magazine*, Summer, 39–43, 1983.

[Dubois & Prade 90a] Didier Dubois and Henri Prade. An introduction to possibilistic and fuzzy logics. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Non-Standard Logics for Automated Reasoning*, Academic Press, 1988.

[Dubois & Prade 90b] Didier Dubois and Henri Prade. Modelling uncertain and vague knowledge in possibility and evidence theories. In R. D. Shachter, T. S. Levitt, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 4*. North Holland, 1990.

[Duda *et al* 78] Richard O. Duda, Peter E. Hart, and Nils J. Nilsson. Semantic network representations in rule-based inference systems. In D. A. Waterman and F. Hayes-Roth, editors, *Pattern-Directed Inference Systems*. Academic Press, 1978.

[Duda *et al* 79] Richard O. Duda, John Gaschnig, and Peter Hart. Model design in the Prospector consultant system for mineral exploration. In Donald Michie, editor, *Expert Systems in the Micro-Electronic Age*. Edinburgh University Press, 1979.

[Duda *et al* 90] Richard O. Duda, Peter E. Hart, and Nils J. Nilsson. Subjective Bayesian methods for rule-based inference systems. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Proceedings of the National Computer Conference (AFIPS)*, 15, 1976.

[Durfee & Lesser 90] Edmund H. Durfee and Victor R. Lesser. Predictability versus responsiveness: Coordinating problem solvers in dynamic domains. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Proceedings of the 10th International Joint Conference on Artificial Intelligence*, 1987.

[Dutta 85] Amitava Dutta. Reasoning with imprecise knowledge in expert systems. *Information Sciences*, 37:3–24, 1985.

[Eason] Patricia Eason. A robotic slide preparation system for automated cervical cytology screening. Napier University PhD Thesis, to be submitted.

[Efron 90] B. Efron. Why isn't everyone a Bayesian? In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990.

[Einav & Fehling 91] David Einav and Michael R. Fehling. Computationally-optimal real-resource strategies for independent, uninterruptible methods. In P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 6*. Elsevier Science Publishers, 1991.

[Elstein *et al* 79] Arthur S. Elstein, Lee S. Shulman, and Sarah A. Sprafka. *Medical Problem Solving: An Analysis of Clinical Reasoning*. Harvard University Press, 1979.

[Engelbrecht *et al* 87] R. Engelbrecht, R. Schaaf, and W. Scholz. Pharmaceutical consultation system for physicians and pharmacists: Basis for an expert system. In *Proceedings of Medical Informatics 1987*, pages 1106–1115, 1987.

[Engle, Jr 92] Ralph L. Engle, Jr. Attempts to use computers as diagnostic aids in medical decision making: A thirty-year experience. *Perspectives in Biology and Medicine*, 35(2):207–219, 1992.

[Essin & Steen 85] Daniel J. Essin and Stephen N. Steen. Computers and the future of medical practice. *Archives of Internal Medicine*, 145(12):2171–2172, 1985.

[Evans 70] D. M. D. Evans, editor. *Cytology Automation*. E. and S. Livingstone Ltd, 1970.

[Evans *et al* 89] J. A. Evans, M. McNay, M. Gowland, and P. Farrant. BMUS ultrasonic fetal measurement survey. *BMUS Bulletin*, 52:14–17, 1989.

[Fairley 91] J. W. Fairley. Computers in medicine. *Journal of the Royal Society of Medicine*, 84:566–567, September 1991.

[Falkenhainer 88] B. Falkenhainer. Towards a general-purpose belief maintenance system. In J. F. Lemmer and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Feigenbaum 79] Edward A. Feigenbaum. Themes and case studies of knowledge engineering. In Donald Michie, editor, *Expert Systems in the Micro-Electronic Age*. Edinburgh University Press, 1979.

[Foreman 89] Julie Foreman. Computers in clinical medicine raise questions of liability. *Archives of Ophthalmology*, 107(1):25, 1989.

[Fox & Kempf 88] B. R. Fox and K. G. Kempf. Planning, scheduling, and uncertainty in the sequence of future events. In J. F. Lemmer and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Fox 86] John Fox. Three arguments for extending the framework of probability. In L. N. Kanal and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence*. North Holland, 1986.

[Franklin *et al* 88] Jude E. Franklin, Cora Lackey Carmody, Karl Keller, Tod S. Levitt, and Brandon L. Buteau. Expert systems technology for the military: Selected examples. *Proceedings of the IEEE*, 76(10):1327–1366, 1988.

[Gaines & Shaw 85] Brian R. Gaines and Mildred L. G. Shaw. From fuzzy logic to expert systems. *Information Sciences*, 36:5–16, 1985.

[Garrett 79] William J. Garrett. Ultrasound in determining normal fetal anatomy. In John C. Hobbins, editor, *Clinics in Diagnostic Ultrasound Volume 3: Diagnostic Ultrasound in Obstetrics*. Churchill Livingstone, 1979.

[Garvey *et al* 81] Thomas D. Garvey, John D. Lowrence, and Martin A. Fischler. An inference technique for integrating knowledge from disparate sources. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, pages 319–325, 1981.

[Geiger & Pearl 90] Dan Geiger and Judea Pearl. On the logic of causal models. In R. D. Shachter, T. S. Levitt, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 4*. North Holland, 1990.

[Geman & Geman 90] Stuart Geman and Donald Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6, 1984.

[Gemignani 91] Michael C. Gemignani. Some legal aspects of expert systems. *Expert Systems with Applications*, 2:269–283, 1991.

[Glymour 89] Clark Glymour. When less is more. In David A. Evans and Vimla L. Patel, editors, *Cognitive Science in Medicine: Biomedical Modeling*. MIT Press, 1989.

[Goddard 85] Helen Goddard. Artificial Intelligence in medicine. Technical Report W.142, Department of Computer Science, University of Exeter, 1985.

[Goldman & Charniak 91] Robert P. Goldman and Eugene Charniak. Dynamic construction of belief networks. In P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 6*. Elsevier Science Publishers, 1991.

[Goldman & Charniak 93] R. P. Goldman and E. Charniak. A language for construction of belief networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(3):196–208, 1993.

[Goldman & Rivest 88] S. A. Goldman and R. L. Rivest. A non-iterative maximum entropy algorithm. In J. F. Lemmer and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Gonzalez & Wintz 87] Rafael C. Gonzalez and Paul Wintz. *Digital Image Processing*. Addison Wesley, 1987.

[Goodman & Nguyen 85] Irwin R. Goodman and Hung T. Nguyen. *Uncertainty Models for Knowledge-Based Systems*. North Holland, 1985.

[Gordon & Shortliffe 85] Jean Gordon and Edward H. Shortliffe. A method for managing evidential reasoning in a hierarchical hypothesis space. *Artificial Intelligence*, 26:323–357, 1985.

[Gordon & Shortliffe 90] Jean Gordon and Edward H. Shortliffe. The Dempster-Shafer theory of evidence. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in G.G. Buchanan and E.H. Shortliffe, editors, *Rule-Based Expert Systems*. Addison Wesley, 1984.

[Graham & Jones 88] Ian Graham and Peter Llewelyn Jones. *Expert systems: Knowledge, uncertainty and decision*. Chapman and Hall Computing, 1988.

[Granum 79] Peter Granum. The placenta: A source of information. In John C. Hobbins, editor, *Clinics in Diagnostic Ultrasound, Volume 3, Diagnostic Ultrasound in Obstetrics*. Churchill Livingstone, 1979.

[Grosof 88] B. N. Grosof. Non-monotonicity in probabilistic reasoning. In J. F. Lemmer and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Gunn & Nixon 94] S. R. Gunn and M. S. Nixon. A model based dual active contour. In E. R. Hancock, editor, *Proceedings of the 5th British Machine Vision Conference, BMVC94*, pages 305–314, 1994.

[Hafner *et al* 89] Arthur W. Hafner, Audrey B. Filipowicz, and William P. Whitely. Computers in medicine: Liability issues for physicians. *International Journal of Clinical Monitoring and Computing*, 6:185–194, 1989.

[Hager & Durrant-Whyte 88] G. Hager and H. F. Durrant-Whyte. Information and multisensor coordination. In J. F. Lemmer and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Hall 82] P. Hall. A quarter of a century with computers in medicine. *Methods of Information in Medicine*, 21:107–108, 1982.

[Hamilton *et al* 94] P. W. Hamilton, N. Anderson, P. H. Bartels, and D. Thompson. Expert system support using Bayesian belief networks in the diagnosis of fine needle aspiration biopsy specimens of the breast. *Journal of Clinical Pathology*, 47:329–336, 1994.

[Hand 87] D. J. Hand. Artificial Intelligence and medicine: Discussion paper. *Journal of the Royal Society of Medicine*, 80:563–565, September 1987.

[Hasman 87] A. Hasman. Medical applications of computers: An overview. *International Journal of Bio-Medical Computing*, 20(4):239–251, 1987.

[Hawthorne 88] J. Hawthorne. A semantic approach to non-monotonic entailments. In J. F. Lemmer and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Hayes 87] Philip J. Hayes. Steps towards integrating natural language and graphical interaction for knowledge-based systems. In B. Du Bulay, D. Hogg, and L. Steels, editors, *Advances in Artificial Intelligence 2*. North Holland, 1987.

[Hayes-Roth 85] Barbara Hayes-Roth. A blackboard architecture for control. *Artificial Intelligence*, 26:251–321, 1985.

[Hayes-Roth *et al* 89] Barbara Hayes-Roth, Richard Washington, Rattikorn Hewett, Micheal Hewett, and Adam Seiver. Intelligent monitoring and control. In *Proceedings of the 11th Joint Conference on Artificial Intelligence*, pages 243–249, 1989.

[Heathfield & Wyatt 93] H. A. Heathfield and J. Wyatt. Philosophies for the design and development of clinical decision-support systems. *Methods of Information in Medicine*, 32:1–8, 1993.

[Heckerman & Horvitz 87] David Heckerman and Eric Horvitz. On the expressiveness of rule-based systems for reasoning with uncertainty. In *Proceedings of AAAI-87: 6th National Conference on Artificial Intelligence*, pages 121–126, 1987.

[Heckerman & Horvitz 88] David. E. Heckerman and E. J. Horvitz. The myth of modularity in rule-based systems for reasoning with uncertainty. In J. F. Lemmer and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Heckerman & Horvitz 91] David Heckerman and Eric Horvitz. Problem formulation as the reduction of a decision model. In P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 6*. Elsevier Science Publishers, 1991.

[Heckerman & Jimison 89] David E. Heckerman and Holly B. Jimison. A Bayesian perspective on confidence. In L. N. Kanal, T. S. Levitt, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 3*. North Holland, 1989.

[Heckerman & Nathwani 92] D. E. Heckerman and B. N. Nathwani. Toward normative expert systems: Part II Probability-based representations for efficient knowledge acquisition and inference. *Methods of Information in Medicine*, 31:106–116, 1992.

[Heckerman 88] David. E. Heckerman. An axiomatic framework for belief updates. In J. F. Lemmer and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Heckerman 90a] David E. Heckerman. Probabilistic interpretations for MYCIN's certainty factors. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in L.N. Kanal and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence*. North Holland, 1986.

[Heckerman 90b] David E. Heckerman. Probabilistic similarity networks. Technical Report STAN-CS-90-1316, Departments of Computer Science and Medicine, Stanford University, California, 1990. Submitted as thesis.

[Heckerman *et al* 91] David Heckerman, Eric Horvitz, and Blackford Middleton. An approximate nonmyopic computation for value of information. In Bruce D. D'Ambrosio, Phillipe Smets, and Piero P. Bonissone, editors, *Uncertainty in Artificial Intelligence, Proceedings of the 7th conference*, pages 135–141. Morgan Kaufman, 1991.

[Henkind & Harrison 88] Steven J. Henkind and Malcolm C. Harrison. An analysis of four uncertainty calculi. *IEEE Transactions on Systems, Man, and Cybernetics*, 18(5):700–714, 1988.

[Henkind 88] Steven. J. Henkind. Imprecise meanings as a cause of uncertainty in medical knowledge-based systems. In J. F. Lemmer and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Henrion & Druzdzel 91] Max Henrion and Marek J. Druzdzel. Qualitative propagation and scenario-based schemes for explaining probabilistic reasoning. In P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 6*. Elsevier Science Publishers, 1991.

[Henrion 87] Max Henrion. Uncertainty in Artificial Intelligence: Is probability epistemologically and heuristically adequate? In Jeryl L. Mumpower, Lawrence D. Phillips, Ortwin Renn, and V. R. R. Uppuluri, editors, *Expert Judgement and Expert Systems.* Springer Verlag, 1987. NATO ASI Series F: Computer and Systems Sciences, volume 35.

[Henrion 88] Max Henrion. Propagating uncertainty in Bayesian networks by probabilistic logic sampling. In J. F. Lemmer and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 2.* North Holland, 1988.

[Henrion 89] Max Henrion. Some practical issues in constructing belief networks. In L. N. Kanal, T. S. Levitt, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 3*, pages 161–173. North Holland, 1989.

[Henrion 90] Max Henrion. An introduction to algorithms for inference in belief nets. In M. Henrion, R. D. Shachter, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 5.* North Holland, 1990.

[Henrion *et al* 91] Max Henrion, John S. Breese, and Eric J. Horvitz. Decision analysis and expert systems. *AI Magazine*, 12:64–91, Winter 1991.

[Henry 90] John Bernard Henry. Computers in medical education: Information and knowledge management, understanding and learning. *Human Pathology*, 21(10):998–1002, 1990.

[Herskovits & Cooper 91a] E. H. Herskovits and G. F. Cooper. Algorithms for Bayesian belief-network precomputation. *Methods of Information in Medicine*, 30:81–89, 1991.

[Herskovits & Cooper 91b] Edward Herskovits and Gregory Cooper. Kutato: An entropy-driven system for construction of probabilistic expert systems from databases. In P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 6.* Elsevier Science Publishers, 1991.

[Hollnagel 90] Erik Hollnagel. Responsibility issues in intelligent decision support systems. In Dianne Berry and Anna Hart, editors, *Expert Systems: Human Issues.* Chapman and Hall, 1990.

[Holmes 94] Bob Holmes. Fantastic voyage into the virtual brain. *New Scientist*, 1932:26–29, 1994.

[Horvitz & Heckerman 86] Eric J. Horvitz and David Heckerman. The inconsistent use of measures of certainty in Artificial Intelligence. In L. N. Kanal, T. S. Levitt, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence.* North Holland, 1986.

[Horvitz & Rutledge 91] Eric Horvitz and Geoffrey Rutledge. Time-dependent utility and action under uncertainty. In Bruce D. D'Ambrosio, Phillipe Smets, and Piero P. Bonissone, editors, *Uncertainty in Artificial Intelligence, proceedings of the 7th conference*, pages 151–158. Morgan Kaufman, 1991.

[Horvitz 88] Eric J. Horvitz. Reasoning under varying and uncertain resource constraints. In *Proceedings of AAAI 1988*, pages 111–116, 1988.

[Horvitz 89] Eric J. Horvitz. Reasoning about beliefs and actions under computational resource constraints. In L. N. Kanal, T. S. Levitt, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 3*. North Holland, 1989.

[Horvitz 90] Eric J. Horvitz. Rational metareasoning and compilation for optimizing decisions under bounded resources. In F. Gardin and G. Mauri, editors, *Computational Intelligence II, Proceedings of the 1989 International Symposium on Computational Intelligence*, pages 205–216. Elsevier Science Publishers, 1990.

[Horvitz *et al* 86] Eric Horvitz, David Heckerman, Bharat Nathwani, and Lawrence Fagan. The use of a heuristic problem-solving hierarchy to facilitate the explanation of hypothesis directed reasoning. In *Proceedings of MEDINFO 1986*, pages 27–31, 1986.

[Horvitz *et al* 88] Eric J. Horvitz, John S. Breese, and Max Henrion. Decision theory in expert systems and Artificial Intelligence. *International Journal of Approximate Reasoning*, 2:247–302, 1988.

[Horvitz *et al* 89a] Eric J. Horvitz, Gregory F. Cooper, and David E. Heckerman. Reflection and action under scarce resources: Theoretical principles and empirical study. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pages 1121–1127, 1989.

[Horvitz *et al* 89b] Eric J. Horvitz, H. Jacques Suermondt, and Gregory F. Cooper. Bounded conditioning: Flexible inference for decisions under scarce resources. In *Uncertainty in Artificial Intelligence, Proceedings of the 5th Workshop, Windsor, Ontario*, pages 182–193, August 1989.

[Hovorka *et al* 90] R. Hovorka, S. Andreassen, J. J. Benn, E. R. Carson, U. Kjaerulff, L. D. Kristensen, and K. G. Olesen. Causal probabilistic network model to assist in insulin therapy adjustment. In *Proceedings of the 6th Annual Meeting on Expert Systems in Medicine*, pages 73–74, 1990.

[Hovorka *et al* 92] R. Hovorka, S. Andreassen, J. J. Benn, K. G. Olesen, and E. R. Carson. Causal probabilistic network modelling — an illustration of its role in the management of chronic diseases. *IBM Systems Journal*, 31(4):635–648, 1992.

[Hummel & Landy 88] R. Hummel and M. Landy. Evidence as opinions of experts. In J. F. Lemmer and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Hutchinson *et al* 89] S. A. Hutchinson, R. L. Cromwell, and A. C. Kak. Applying uncertainty reasoning to model based object recognition. In *Conference on Computer Vision and Pattern Recognition*, pages 541–548, 1989.

[Hwang *et al* 86] Vincent Shang-Shouq Hwang, Larry S. Davis, and Takashi Matsuyama. Hypothesis integration in image understanding systems. *Computer Vision, Graphics, and Image Processing*, 36:321–371, 1986.

[Ishizuka *et al* 81] Mitsuru Ishizuka, K. S. Fu, and James T. P. Yao. Inexact inference for rule-based damage assessment of existing structures. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, pages 837–842, 1981.

[Jason 84] Hilliard Jason. Will computers dehumanize medical care and education? *Journal of Clinical Laboratory Automation*, 4(4):226–227, 1984.

[Jeanty & Romero 83] Philippe Jeanty and Roberto Romero. *Obstetrical Ultrasound*. McGraw-Hill, 1983.

[Jensen 93] Finn V. Jensen. *Introduction to Bayesian Networks*. Hugin Expert A/S, 1993.

[Jensen *et al* 87a] Finn V. Jensen, Stig K. Andersen, Uffe Kjaerulff, and Steen Andreassen. MUNIN — on the case for probabilities in medical expert systems — a practical exercise. In *Proceedings of the 1st Conference of the European Society for Artificial Intelligence in Medicine*, pages 149–160, 1987.

[Jensen *et al* 87b] Finn V. Jensen, Stig K. Andersen, Uffe Kjaerulff, and Steen Andreassen. A causal network prototype in the domain of electromyography — an implementation of coherent probabilistic methods. In *Proceedings of the 10th International Joint Conference on Artificial Intelligence*, 1987.

[Jensen *et al* 90a] Finn V. Jensen, S. L. Lauritzen, and Kristian G. Olesen. Bayesian updating in causal probabilistic networks by local computations. *Computational Statistics Quarterly*, 4:269–282, 1990.

[Jensen *et al* 90b] Finn Verner Jensen, Jan Nielsen, and Henrik I. Christensen. Use of causal probabilistic networks as high level models in computer vision. Technical Report R 90-39, The University of Aalborg, Denmark, 1990.

[Jensen *et al* 90c] Finn Verner Jensen, Kristian G. Olesen, and Stig Kjaer Anderson. An algebra of Bayesian belief universes for knowledge-based systems. *Networks*, 20:637–659, 1990.

[Jensen *et al* 91] Finn Verner Jensen, Bo Chamberlain, Torsten Nordahl, and Frank Jensen. Analysis in HUGIN of data conflict. In P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 6*. Elsevier Science Publishers, 1991.

[Jimison 90] Holly B. Jimison. Generating explanations of decision models based on an augmented representation of uncertainty. In R. D. Shachter, T. S. Levitt, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 4*. North Holland, 1990.

[Jimison *et al* 92] Holly B. Jimison, L. M. Fagan, R. D. Shachter, and E. H. Shortliffe. Patient-specific explanation in models of chronic disease. *Artificial Intelligence in Medicine*, 4(3):191–205, 1992.

[Jones *et al* 91] R. B. Jones, L. M. Navin, J. Barrie, E. Hillan, and D. Kinane. Computer literacy among medical, nursing, dental and veterinary undergraduates. *Medical Education*, 25:191–195, 1991.

[Kak *et al* 90] A. C. Kak, K. M. Andress, C. Lopez-Abadia, M. S. Carroll, and J. R. Lewis. Hierarchical evidence accumulation in the PSEIKI system and experiments in model-driven mobile robot navigation. In M. Henrion, R. D. Shachter, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 5*. North Holland, 1990.

[Kalagnanam & Henrion 90] Jayant Kalagnanam and Max Henrion. A comparison of decision analysis and expert rules for sequential diagnosis. In R. D. Shachter, T. S. Levitt, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 4*. North Holland, 1990.

[Kanal & Lemmer 86] L. N. Kanal and J. F. Lemmer, editors. *Uncertainty in Artificial Intelligence*. North Holland, 1986.

[Kass *et al* 87] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. In *Proceedings of the 1st International Conference on Computer Vision*, pages 259–268, 1987.

[Kim & Pearl 83] Jin H. Kim and Judea Pearl. A computational model for causal and diagnostic reasoning in inference systems. In *Proceedings of the 8th International Joint Conference on Artificial Intelligence*, pages 190–193, 1983.

[Kingsland III 88] Lawrence C. Kingsland III. Evaluation of medical expert systems: Experience with the AI/RHEUM knowledge-based consultant system in rheumatology. In Perry L. Miller, editor, *Selected Topics in Medical Artificial Intelligence*. Springer-Verlag, 1988.

[Kjaerulff 90] Uffe Kjaerulff. Triangulation of graphs — algorithms giving small total state space. Technical Report R 90-09, Institute for Electronic Systems, University of Aalborg, Denmark, 1990.

[Kleinmuntz 92] Benjamin Kleinmuntz. Computers as clinicians: An update. *Computers in Biology and Medicine*, 22(4):227–237, 1992.

[Knapp *et al* 87] Rebecca G. Knapp, M. Clinton Miller III, and Jon Levine. Experience with and attitudes toward computers in medicine of students and clinical faculty members at one school. *Journal of Medical Education*, 62(4):344–346, 1987.

[Kremkau & Taylor 86] Frederick W. Kremkau and Kenneth J. W. Taylor. Artifacts in ultrasound imaging. *Journal of Ultrasound Medicine*, 5:227–237, April 1986.

[Kuipers & Kassirer 84] Benjamin Kuipers and Jerome P. Kassirer. Causal reasoning in medicine: Analysis of a protocol. *Cognitive Science*, 8:363–385, 1984.

[Kuipers 75] Benjamin J. Kuipers. A frame for frames: Representing knowledge for recognition. In D. G. Bobrow and A. Collins, editors, *Representation and Understanding*. Academic Press, 1975.

[Kuipers *et al* 90] Benjamin Kuipers, Alan J. Moskowitz, and Jerome P. Kassirer. Critical decisions under uncertainty: Representation and structure. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Cognitive Science*, 12, 1988.

[Kulikowski 84] Casimir A. Kulikowski. Artificial Intelligence methods and systems for medical consultation. In William J. Clancey and Edward H. Shortliffe, editors, *Readings in Medical Artificial Intelligence: The First Decade*. Addison Wesley, 1984.

[Kyburg, Jr 88] Henry E. Kyburg, Jr. Knowledge. In J. F. Lemmer and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Kyburg, Jr 89] Henry E. Kyburg, Jr. Higher order probabilities. In L. N. Kanal, T. S. Levitt, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 3*. North Holland, 1989.

[Kyburg, Jr 90] Henry E. Kyburg, Jr. Epistemological relevance and statistical knowledge. In R. D. Shachter, T. S. Levitt, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 4*. North Holland, 1990.

[Lane *et al* 86] C. D. Lane, Joan D. Walton, and E. H. Shortliffe. Graphical access to medical expert systems: II. Design of an interface for physicians. *Methods of Information in Medicine*, 25:143–150, 1986.

[Langlotz & Shortliffe 89] Curtis P. Langlotz and Edward H. Shortliffe. Logical and decision-theoretical methods for planning under uncertainty. *AI Magazine*, 10(1):39–47, 1989.

[Langlotz & Shortliffe 90] Curtis P. Langlotz and Edward H. Shortliffe. Pilot evaluation of a computer program that explains the results of a decision analysis. *Medical Decision Making*, 10:334, 1990. Abstract only.

[Langlotz 89] Curtis P. Langlotz. The feasibility of axiomatically-based expert systems. *Computer Methods and Programs in Biomedicine*, 30:85–95, 1989.

[Langlotz *et al* 86] Curtis P. Langlotz, Edward H. Shortliffe, and Lawrence M. Fagan. Computer-based explanation of decision analysis. *Medical Decision Making*, 6(4):277, 1986.

[Langlotz *et al* 87] Curtis P. Langlotz, Lawrence M. Fagan, Samson W. Tu, Branimir I. Sikic, and Edward H. Shortliffe. A therapy planning architecture that combines decision theory and Artificial Intelligence techniques. *Computers and Biomedical Research*, 20:279–303, 1987.

[Langlotz *et al* 88a] Curtis P. Langlotz, Edward H. Shortliffe, and Lawrence M. Fagan. Heuristic construction of a retorical defense for the results of a decision analysis. *Medical Decision Making*, 8(4):337, 1988. Abstract only.

[Langlotz *et al* 88b] Curtis P. Langlotz, Edward H. Shortliffe, and Lawrence M. Fagan. A methodology for generating computer-based explanations of decision-theoretic advice. *Medical Decision Making*, 8(4):290–303, 1988.

[Langlotz *et al* 90] Curtis P. Langlotz, Lawrence M. Fagan, Samson W. Tu, Branimir I. Sikic, and Edward H. Shortliffe. A therapy planning architecture that combines decision theory and Artificial Intelligence techniques. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Computers in Biomedical Research*, 20:279–303, 1987.

[Laskey & Lehner 90] Kathryn B. Laskey and Paul E. Lehner. Belief maintenance: An integrated approach to uncertainty management. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Proceedings of the National Conference of Artificial Intelligence*, 1988.

[Lauritzen & Spiegelhalter 90] S. L. Lauritzen and D. J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Proceedings of the Royal Statistical Society, London, Series B50*, 2:157–224, 1984.

[Laursen 94] P. Laursen. Event detection on patient monitoring data using causal probabilistic networks. *Methods of Information in Medicine*, 33:111–115, 1994.

[Laxminarayan & Kristol 92] Swamy Laxminarayan and David Kristol. Computers in medicine. *IEEE Engineering in Medicine and Biology Magazine*, 11(1):24, 1992.

[Leal & Pearl 77] Antonio Leal and Judea Pearl. An interactive program for conversational elicitation of decision structures. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-7(5):368–376, 1977.

[Lee & Evans 91] Chong-Yen Lee and Martha Evans. Recommending tests in a multimembership Bayesian diagnostic expert system. In *Proceedings of the 4th Annual IEEE Symposium on Computer-based Medical Systems*, pages 28–35, 1991.

[Lehmann & Shortliffe 91] Harold P. Lehmann and Edward H. Shortliffe. THOMAS: Building Bayesian statistical expert systems to aid in clinical decision making. *Computer Methods and Programs in Biomedicine*, 35:251–260, 1991.

[Lemmer & Kanal 88] John. F. Lemmer and Laveen. N. Kanal, editors. *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Levitt 86] Tod S. Levitt. Model-based probabilistic situation inference in hierarchical hypothesis spaces. In L. N. Kanal and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence*. North Holland, 1986.

[Levitt 88] Tod S. Levitt. Bayesian inference for radar imagery based surveillance. In J. F. Lemmer and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Levitt *et al* 90a] Tod S. Levitt, John M. Agosta, and Thomas O. Binford. Model-based influence diagrams for machine vision. In M. Henrion, R. D. Shachter, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 5*. North Holland, 1990.

[Levitt *et al* 90b] Tod S. Levitt, Thomas O. Binford, and Gil J. Ettinger. Utility-based control for computer vision. In R. D. Shachter, T. S. Levitt, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 4*. North Holland, 1990.

[Li & Uhr 88] Ze-Nian Li and Leonard Uhr. Evidential reasoning in a computer vision system. In John F. Lemmer and Laveen N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Lin & Goebel 91] Dekang Lin and Randy Goebel. Integrating probabilistic, taxonomic and causal knowledge in abductive diagnosis. In P. P.

Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 6*. Elsevier Science Publishers, 1991.

[Linder 92] James Linder. Automation in cytopathology. *American Journal of Clinical Pathology*, 98:S47–S51, 1992.

[Lindley 87] Dennis V. Lindley. The probability approach to the treatment of uncertainty in Artificial Intelligence and expert systems. *Statistical Science*, 2(1):17–24, 1987.

[Llacer *et al* 91] J. Llacer, E Veklerov, and J Nunez. Preliminary examination of the use of case specific medical information as prior in Bayesian reconstruction. In *Proceedings of the 12th International Conference on Information Processing in Medical Imaging*, pages 81–93, 1991.

[Long 89] William Long. Medical diagnosis using a probabilistic causal network. *Applied Artificial Intelligence*, 3:367–383, 1989.

[Loui 88] R. P. Loui. Computing reference classes. In John F. Lemmer and Laveen N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Lowrance & Garvey 82] John D. Lowrance and Thomas D. Garvey. Evidential reasoning: A developing concept. In *Proceedings of the International Conference on Cybernetics and Society*, pages 6–9, 1982.

[Lowrance *et al* 90] John D. Lowrance, Thomas D. Garvey, and Thomas M. Strat. A framework for evidential-reasoning systems. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Proceedings of the National Conference on Artificial Intelligence*, 1986.

[MacKay 83] D. R. MacKay. The place of computers in medicine. *The New Zealand Medical Journal*, 96(740):722–724, 1983.

[Mahalanobis 36] P. C. Mahalanobis. On the generalized distance in statistics. In *Proceedings of the National Institute of Science (India)*, volume 2, pages 49–55, 1936.

[Mann & Binford 92] Wally Bishop Mann and Thomas O. Binford. An example of 3-D interpretation of images using Bayesian networks. In *Proceedings of the DARPA Image Understanding Workshop*, pages 793–801, 1992.

[Manning *et al* 89] Frank A. Manning, Savas Menticogulou, and Chris Harman. Fetal assessment by biophysical methods: Ultrasound. In Sir Alec Turnbull and G. Chamberlain, editors, *Obstetrics*. Churchill Livingstone, 1989.

[Marin *et al* 93] Roque Marin, Maria Taboada, Jose Mira, Alvaro Barreiro, and Ramon P. Otero. Design and integration of a graphic interface for an expert system in oncology. *International Journal of Biomedical Computing*, 33:25–43, 1993.

[Marquardt, Jr 93] Victor C. Marquardt, Jr. Artificial intelligence and decision-support technology in the clinical laboratory. *Laboratory Medicine*, 24(12):777–782, 1993.

[Mars & Miller 88] Nicolaas J. I. Mars and Perry L. Miller. Knowledge acquisition and verification tools for medical expert systems. In Perry L. Miller, editor, *Selected Topics in Medical Artificial Intelligence*. Springer-Verlag, 1988.

[McLeish 88] M. McLeish. Probabilistic logic: Some comments and possible use for nonmonotonic reasoning. In John F. Lemmer and Laveen N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[McNay 87] M. B. McNay. Diagnostic ultrasonography. *Bailliere's Clinical Obstetrics and Gynaecology*, 1(1), March 1987.

[McNeil *et al* 75] Barbara J. McNeil, Emmett Keeler, and S. James Adelstein. Primer on certain elements of medical decision making. *The New England Journal of Medicine*, 293(5):211–215, 1975.

[Middleton *et al* 91] B. Middleton, M. A. Shwe, D. E. Heckerman, M. Henrion, E. J. Horvitz, H. P. Lehmann, and G. F. Cooper. Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base: II. Evaluation of diagnostic performance. *Methods of Information in Medicine*, 30:256–267, 1991.

[Miller & Fisher 88] Perry L. Miller and Paul R. Fisher. Causal models for medical Artificial Intelligence. In Perry L. Miller, editor, *Selected Topics in Medical Artificial Intelligence*. Springer-Verlag, 1988.

[Miller 86] Perry L. Miller. The evaluation of Artificial Intelligence systems in medicine. *Computer Methods and Programs in Biomedicine*, 22:5–11, 1986.

[Miller 88] Perry L. Miller. Artificial Intelligence in medicine: An emerging discipline. In Perry L. Miller, editor, *Selected Topics in Medical Artificial Intelligence*. Springer-Verlag, 1988.

[Miller 89] A. B. Miller. Evaluation of the impact of screening for cancer of the cervix. In M. Hakama, A. B. Miller, and N. E. Day, editors, *Screening for Cancer of the Uterine Cervix, IARC Scientific Publications number 76*, pages 149–160. International Agency for Research on Cancer, 1989.

[Minsky 80] Marvin Minsky. A framework for representing knowledge. In D. Metzing, editor, *Frame Conceptions and Text Understanding*. Walter de Gruyter, 1980.

[Montanari 74] Ugo Montanari. Networks of constraints: Fundamental properties and applications to picture processing. *Information Sciences*, 7:95–132, 1974.

[Morawski 89a] Paul Morawski. Programming Bayesian belief networks. *AI Expert*, pages 74–79, August 1989.

[Morawski 89b] Paul Morawski. Understanding Bayesian belief networks. *AI Expert*, pages 44–48, May 1989.

[Musman *et al* 90] S. A. Musman, L. W. Chang, and L. B. Booker. A real time control strategy for Bayesian belief networks with application to ship classification problem solving. In *Proceedings of the 2nd International IEEE Conference on Tools for Artificial Intelligence*, pages 738–744, 1990.

[Neapolitan & Kenevan 91] Richard E. Neapolitan and James R. Kenevan. Investigation of variances in belief networks. In Bruce D. D'Ambrosio, Phillipe Smets, and Piero P. Bonissone, editors, *Uncertainty in Artificial Intelligence, proceedings of the 7th conference*, pages 232–241. Morgan Kaufman, 1991.

[Neapolitan 90] R. E. Neapolitan. *Probabilistic Reasoning in Expert Systems.* Wiley, 1990.

[N.E.F.J. 72] N.E.F.J. The prevention of cancer of the cervix. In John Wakefield, editor, *Seek Wisely to Prevent: Studies of Attitudes and Action in a Cervical Cytology Programme*, pages 38–43. HMSO, 1972. Taken from The North-West England Faculty Journal.

[Nelson & Pretorius 92] T. R. Nelson and D. H. Pretorius. Three-dimensional ultrasound of fetal surface features. *Ultrasound in Obstetrics and Gynecology*, 2:166–174, 1992.

[Neufeld & Poole 89] Eric Neufeld and David Poole. Towards solving the multiple extension problem: Combining defaults and probability. In L. N. Kanal, T. S. Levitt, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 3*. North Holland, 1989.

[Neufeld 90] Eric Neufeld. Defaults and probabilities; extensions and coherence. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Proceedings of the 1st International Conference on Principles of Knowledge Representation and Reasoning*, 1989.

[Nii 86a] H. Penny Nii. Blackboard systems, part one. *The AI Magazine*, Summer:38–53, 1986.

[Nii 86b] H. Penny Nii. Blackboard systems, part two. *The AI Magazine*, August:82–106, 1986.

[Nilsson 90] Nils J. Nilsson. Probabilistic logic. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Artificial Intelligence*, 28:71–87, 1986.

[North 90] D. Warner North. A tutorial introduction to decision theory. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *IEEE Transactions on Systems Science and Cybernetics*, SCC-4:3, September, 1968.

[Norton 88] S. W. Norton. An explanation mechanism for Bayesian inferencing systems. In John F. Lemmer and Laveen N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Nykanen *et al* 91] P. Nykanen, S. Chowdhury, and O. Wigertz. Evaluation of decision support systems in medicine. *Computer Methods and Programs in Biomedicine*, 34:229–238, 1991.

[Olesen 93] Kristian G. Olesen. Causal probabilistic networks with both discrete and continuous variables. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(3):275–279, 1993.

[Olesen *et al* 89] Kristian G. Olesen, Uffe Kjaerulff, Frank Jensen, Finn V. Jensen, Bjorn Falck, Steen Andreassen, and Stig K. Andersen. A MUNIN network for the median nerve — a case study in loops. *Applied Artificial Intelligence*, 3:385–403, 1989.

[Oppel 91] Ulrich G. Oppel. Every complex system can be determined by a causal probabilistic network without cycles and every such network determines a Markov field. *Lecture Notes in Computer Science*, 548:262–266, 1991.

[Osborn 82] John J. Osborn. Computers in critical care medicine: Promises and pitfalls. *Critical Care Medicine*, 10(12):807–810, 1982.

[Patil & Senyk 88] Ramesh S. Patil and Oksana Senyk. Compiling causal knowledge for diagnostic reasoning. In Perry L. Miller, editor, *Selected Topics in Medical Artificial Intelligence*. Springer-Verlag, 1988.

[Patil 86] Ramesh S. Patil. Review of causal reasoning in medical diagnosis. In *Proceedings of the 10th Annual Symposium on Computer Application in Medical Care*, pages 11–16, 1986.

[Patil 87] Ramesh S. Patil. Causal reasoning in computer programs for medical diagnosis. *Computer Methods and Programs in Biomedicine*, 25:117–124, 1987.

[Patil *et al* 81] Ramesh S. Patil, Peter Szolovits, and William B. Schwartz. Causal understanding of patient illness in medical diagnosis. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, pages 893–899, 1981.

[Pearce & Campbell 84] J. Malcolm Pearce and Stuart Campbell. The use of ultrasound in the diagnosis of fetal anomalies. In John Studd, editor, *Progress in Obstetrics and Gynaecology, Volume 4*. Churchill Livingstone, 1984.

[Pearl & Dechter 89] Judea Pearl and Rina Dechter. Learning structure from data: A survey. In *Proceedings of COLT 1989*, pages 230–244, 1989.

[Pearl 85] Judea Pearl. How to do with probabilities what people say you can't. Technical Report CSD-850031 R-49, Cognitive Systems Laboratory, UCLA, 1985. Presented to the 2nd Conference on Artificial Intelligence Applications, 1985.

[Pearl 86a] Judea Pearl. A constraint-propagation approach to probabilistic reasoning. In L. N. Kanal and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence*. North Holland, 1986.

[Pearl 86b] Judea Pearl. Evidential reasoning using stochastic simulation of causal models. Technical Report R-68, Cognitive Systems Laboratory, UCLA Computer Science Department, 1986.

[Pearl 87a] Judea Pearl. Bayesian decision methods. In S. C. Shapiro, editor, *Encyclopedia of Artificial Intelligence, Volume 1*. John Wiley and Sons, 1987.

[Pearl 87b] Judea Pearl. Distributed revision of composite beliefs. *Artificial Intelligence*, 33:173–215, 1987.

[Pearl 88a] Judea Pearl. Distributed revision of belief commitment in composite explanations. In John F. Lemmer and Laveen N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Pearl 88b] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann Publishers Inc., 1988.

[Pearl 90a] Judea Pearl. Bayesian and belief-functions formalisms for evidential reasoning: A conceptual analysis. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Proceedings of the 5th Israeli Symposium on Artificial Intelligence*, 1989.

[Pearl 90b] Judea Pearl. Bayesian decision methods. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Encyclopedia of AI*. Wiley Interscience, 1987.

[Pearl 90c] Judea Pearl. Embracing causality in default reasoning. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Artificial Intelligence*, 35:259–271, 1988.

[Pearl 90d] Judea Pearl. Fusion, propagation, and structuring in belief networks. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Artificial Intelligence*, 29:241–288, 1986.

[Pearl 90e] Judea Pearl. On evidential reasoning in a hierarchy of hypotheses. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Artificial Intelligence*, 28:9–15, 1986.

[Pearl 90f] Judea Pearl. Probabilistic semantics for nonmonotonic reasoning: A survey. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Proceedings of the 1st International Conference on Principles of Knowledge Representation and Reasoning*, 1989.

[Pearl 90g] Judea Pearl. Which is more believable, the probably provable or the provably probable? In *Proceedings of CSCSI-90, the 8th Canadian Conference on Artificial Intelligence*, pages 1–7, 1990.

[Pearl 93] Judea Pearl. Belief networks revisited. *Artificial Intelligence*, 59:49–56, 1993.

[Pearl *et al* 90] Judea Pearl, Dan Geiger, and Thomas Verma. Conditional independence and its representations. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Kybernetika*, 25:2, 1989.

[Perkin 92] Reg L. Perkin. Computers in medicine. *Canadian Family Physician*, 38:220, January 1992.

[Pinto 86] Javier Andres Pinto. Relevance based propagation in Bayesian networks. Unpublished M.Sc. thesis, UCLA, 1986.

[Poole & Provan 91] David Poole and Gregory M. Provan. What is the most likely diagnosis? In P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 6*. Elsevier Science Publishers, 1991.

[Poole 91] Ian Poole. Outline of a Bayesian belief network approach to cervical specimen classification. Technical report, MRC Human Genetics Unit, Edinburgh, 1991.

[Poole 93] Ian Poole. A statistical model for classifying cervical monolayers. Research Note RN94_003, MRC Human Genetics Unit, Edinburgh, 1993.

[Poole 94] Ian Poole. Plan and outline protocol for the cytoline development trial. Internal Note IN94_005, MRC Human Genetics Unit, Edinburgh, 1994.

[Poole *et al* 92] Ian Poole, Jim Piper, Denis Rutovitz, and Margaret Stark. Cytoline — software design. Internal Note IN94_002, MRC Human Genetics Unit, Edinburgh, 1992.

[Prade 83] Henri Prade. A synthetic view of approximate reasoning techniques. In *Proceedings of the 6th International Joint Conference on Artificial Intelligence*, pages 130–136, 1983.

[Preece 90] Alun D. Preece. DISPLAN: Designing a usable medical expert system. In Dianne Berry and Anna Hart, editors, *Expert Systems: Human Issues*. Chapman and Hall, 1990.

[Preston, Jr 79] Kendall Preston, Jr. Computer processing of biomedical images. *Computer*, 9:54–68, 1979.

[Provan & Clarke 93] Gregory M. Provan and John R. Clarke. Dynamic network construction and updating techniques for the diagnosis of acute abdominal pain. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(3):299–307, 1993.

[Provan 90] Gregory M. Provan. The application of Dempster Shafer theory to a logic-based visual recognition system. In M. Henrion, R. D. Shachter, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 5*. North Holland, 1990.

[Pylyshyn 86] Zenon W. Pylyshyn. *Computation and Cognition: Toward a Foundation for Cognitive Science*. The MIT Press, 1986.

[Quaglini *et al* 89] Silvana Quaglini, Carlo Berzuini, Riccardo Bellazzi, Mario Stefanelli, and Giovanni Barosi. Therapy planning by combining AI and decision theoretic techniques. In *AIME 89*, pages 125–134, 1989.

[Quaglini *et al* 92] Silvana Quaglini, Riccardo Bellazzi, Carlo Berzuini, Mario Stefanelli, and Giovanni Barosi. Hybrid knowledge-based systems for therapy planning. *Artificial Intelligence in Medicine*, 4(3):207–226, 1992.

[Reggia 88] James A. Reggia. Evaluation of medical expert systems: Case study in performance assessment. In Perry L. Miller, editor, *Selected Topics in Medical Artificial Intelligence*. Springer-Verlag, 1988.

[Reiter 80] J. Reiter. *AL/X: An Expert System Using Plausible Inference*. Intelligent Terminals Ltd, Oxford, 1980.

[Reiter 90] Raymond Reiter. Nonmonotonic reasoning. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Annual Review of Computing Science*, 2:147–186, 1987.

[Rich 83] Elaine Rich. *Artificial Intelligence*. McGraw-Hill, 1983.

[Richter 87] Michael M. Richter. Some abstract problems in knowledge representation. In Jeryl L. Mumpower, Lawrence D. Phillips, Ortwin Renn, and V. R. R. Uppuluri, editors, *Expert Judgement and Expert Systems*. Springer Verlag, 1987. NATO ASI Series F: Computer and Systems Sciences, volume 35.

[Rimey & Brown 92] Raymond D. Rimey and Christopher M. Brown. Control of selective perception using Bayes nets and decision theory. Submitted to International Journal of Computer Vision, Special Issue on Active Vision, October 1992.

[Rollinger 83] Claus-Rainer Rollinger. How to represent evidence — aspects of uncertain reasoning. In *Proceedings of the 8th International Joint Conference on Artificial Intelligence*, pages 358–361, 1983.

[Rowe & Watkins 92] Alan J. Rowe and Paul R. Watkins. Beyond expert systems — reasoning, judgement, and wisdom. *Expert Systems with Applications*, 4:1–10, 1992.

[Rubenstein 94] Roy H. Rubenstein. Smooth navigation in virtual worlds. *New Electronics*, pages 12–14, June 1994.

[Rubin 75] Andee Rubin. The role of hypotheses in medical diagnosis. In *Proceedings of the 4th International Joint Conference on Artificial Intelligence*, pages 856–862, 1975.

[Rumelhart *et al* 86] David E. Rumelhart, James L. McClelland, and the PDP Research Group. *Parallel Distributed Processing: Exploration in the Microstructure of Cognition, Volume 1: Foundations*. The MIT Press, 1986.

[Rutledge *et al* 89] G. Rutledge, G. Thomsen, I. Beinlich, B. Farr, L. Sheiner, and L. M. Fagan. Combining qualitative and quantitative computation in a ventilator therapy planner. In *Proceedings of the 13th Annual Symposium on Computer Applications in Medical Care*, 1989.

[Sabbagha 79] Rudy E. Sabbagha. The use of ultrasound in defining gestational age. In John C. Hobbins, editor, *Clinics in Diagnostic Ultrasound Volume 3: Diagnostic Ultrasound in Obstetrics*. Churchill Livingstone, 1979.

[Saffiotti 87] Alessandro Saffiotti. An AI view of the treatment of uncertainty. *The Knowledge Engineering Review*, 2(2):75–97, 1987.

[Salari *et al* 90] V. Salari, I. Zador, L. Chik, and R. Sokol. Automated measurements of fetal head from ultrasound images. In *Proceedings on Medical Imaging, International Society of Optical Imaging*, volume 2, pages 213–216, 1990.

[Sarkar & Boyer 93] S. Sarkar and K. L. Boyer. Integration, inference, and management of spatial information using Bayesian networks: Perceptual organisation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(3):256–274, 1993.

[Saul 94] Helen Saul. Screening without meaning? *New Scientist*, 1917:14–15, 1994.

[Savage 90] Leonard J. Savage. The foundations of statistics reconsidered. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Proceedings of the 4th Berkely Symposium on Mathematics and Probability*, 1, 1961.

[Schocken 88] S. Schocken. On the rational scope of probabilistic rule-based inference systems. In J. F. Lemmer and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Schwartz 70] William B. Schwartz. Medicine and the computer: The promise and problems of change. *The New England Journal of Medicine*, 283(23):1257–1264, 1970.

[Schwartz *et al* 87] William B. Schwartz, Ramesh S. Patil, and Peter Szolovits. Artificial Intelligence in medicine: Where do we stand? *The New England Journal of Medicine*, 316(11):685–688, 1987.

[Schwartz *et al* 88] S. M. Schwartz, J. Baron, and J. R. Clarke. A causal Bayesian model for the diagnosis of appendicitis. In John F. Lemmer and Laveen N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Senyk *et al* 89] Oksana Senyk, Ramesh S. Patil, and Frank A. Sonnenberg. Systematic knowledge base design for medical diagnosis. *Applied Artificial Intelligence - special issue on causal modelling*, 3(2 and 3):249–274, 1989.

[Shachter & Heckerman 88] Ross D. Shachter and D. Heckerman. A backwards view for assessment. In John F. Lemmer and Laveen N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Shachter 88] Ross D. Shachter. DAVID: Influence diagram processing system for the Macintosh. In John F. Lemmer and Laveen N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Shachter 90] Ross D. Shachter. Evaluating influence diagrams. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Operations Research*, 33:6, November–December, 1986.

[Shachter *et al* 91] Ross D. Shachter, Stig K. Andersen, and Kim L. Poh. Directed reduction algorithms and decomposable graphs. In P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 6*. Elsevier Science Publishers, 1991.

[Shafer & Srivastava 90] Glenn Shafer and Rajendra Srivastava. The Bayesian and belief-function formalisms: A general perspective for auditing. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Auditing: A Journal of Practice and Theory*, 1990.

[Shafer & Tversky 90] Glenn Shafer and Amos Tversky. Languages and designs for probability judgement. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Cognitive Science*, 9, 1985.

[Shafer 87] Glenn Shafer. Probability judgement in Artificial Intelligence and expert systems. *Statistical Science*, 2(1):3–16, 1987.

[Shafer 90] Glenn Shafer. Savage revisited. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Statistical Science*, 1:4, 1986.

[Shenoy & Shafer 90] Prakash P. Shenoy and Glenn Shafer. Axioms for probability and belief-function propagation. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in R. Shachter, T. S. Levitt, J. F. Lemmer and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 4*. North Holland, 1990.

[Shenoy *et al* 88] Prakash P. Shenoy, Glen Shafer, and K. Mellouli. Propagation of belief functions: A distributed approach. In John F. Lemmer and Laveen N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Shingleton & Orr, Jr 87] Hugh M. Shingleton and James W. Orr, Jr. *Cancer of the Cervix: Diagnosis and Treatment*. Churchill Livingstone, 1987.

[Shortliffe & Buchanan 90] Edward H. Shortliffe and Bruce G. Buchanan. A model of inexact reasoning in medicine. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Edited version of paper originally published in *Mathematical Biosciences*, 23:351–379, 1975.

[Shortliffe & Clancey 84] Edward H. Shortliffe and William J. Clancey. Anticipating the second decade. In William J. Clancey and Edward H. Shortliffe, editors, *Readings in Medical Artificial Intelligence: The First Decade*. Addison Wesley, 1984.

[Shortliffe & Fagan 82] Edward H. Shortliffe and Lawrence M. Fagan. Expert systems research: modelling the medical decision making process. Technical Report HPP-82-3, Computer Science Department, Stanford University, 1982.

[Shortliffe 90] Edward H. Shortliffe. Computer programs to support clinical decision making. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Journal of the American Medical Association*, 258:1, 1987.

[Shortliffe *et al* 84] Edward H. Shortliffe, Bruce G. Buchanan, and Edward A. Feigenbaum. Knowledge engineering for medical decision making: A review of computer-based clinical decision aids. In William J. Clancey and Edward H. Shortliffe, editors, *Readings in Medical Artificial Intelligence: The First Decade*. Addison Wesley, 1984.

[Shwe *et al* 91] M. A. Shwe, B. Middleton, D. E. Heckerman, M. Henrion, E. J. Horvitz, H. P. Lehmann, and G. F. Cooper. Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base: I. the probabilistic model and inference algorithms. *Methods of Information in Medicine*, 30:241–255, 1991.

[Sieghart 84] Paul Sieghart. Medical confidence, the law, and computers: Discussion paper. *Journal of the Royal Society of Medicine*, 77:656–662, 1984.

[Simon & Kadane 75] Herbert A. Simon and Joseph B. Kadane. Optimal problem-solving search: All-or-none solutions. *Artificial Intelligence*, 6:235–247, 1975.

[Sloman 79] Aaron Sloman. Epistemology and Artificial Intelligence. In Donald Michie, editor, *Expert Systems in the Micro-Electronic Age*. Edinburgh University Press, 1979.

[Smith 82] Richard Smith. Computers in medicine: Searching for the rainbow and the crock of gold. *British Medical Journal*, 284:1859–1860, 1982.

[Smith *et al* 88] R. Smith, M. Self, and P. Cheeseman. Estimating uncertain spatial relationships in robotics. In John F. Lemmer and Laveen N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Smithson 89] Michael Smithson. *Ignorance and Uncertainty*. Springer-Verlag, 1989.

[Sonnenberg *et al* 94] Frank A. Sonnenberg, C. Greg Hagerty, and Casmir A. Kulikowski. An architecture for knowldege-based construction of decision models. *Medical Decision Making*, 14(1):27–39, 1994.

[Spiegelhalter & Cowell 92] David J. Spiegelhalter and Robert G. Cowell. Learning in probabilistic expert systems. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics 4*. Oxford University Press, 1992.

[Spiegelhalter & Lauritzen 90a] David J. Spiegelhalter and Steffen L. Lauritzen. Sequential updating of conditional probabilities on directed graphical structures. *Networks*, 20:579–605, 1990.

[Spiegelhalter & Lauritzen 90b] David J. Spiegelhalter and Steffen L. Lauritzen. Techniques for Bayesian analysis in expert systems. *Annals of Mathematics and Artificial Intelligence*, 2:353–366, 1990.

[Spiegelhalter 87] David J. Spiegelhalter. Probabilistic expert systems in medicine: Practical issues in handling uncertainty. *Statistical Science*, 2(1):25–30, 1987.

[Spiegelhalter 90] David J. Spiegelhalter. A statistical view of uncertainty in expert systems. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in W. Gale, editor, *Artificial Intelligence and Statistics*. Addison Wesley, 1986.

[Spiegelhalter *et al* 91] David J. Spiegelhalter, A. Philip Dawid, Tom A. Hutchinson, and Robert G. Cowell. Probabilistic expert systems and graphical modelling: A case study in drug safety. *Philosophical Transactions of the Royal Society, London, Series A*, 337:387–405, 1991.

[Stead, Jr & Stead 85] Eugene A. Stead, Jr and William W. Stead. Computers and medical practice: Old dreams and current realities. *M. D. Computing*, 2(6):26–31, 1985.

[Stefik *et al* 83a] Mark Stefik, Janice Aitkins, Robert Balzer, John Benoit, Lawrence Birnbaum, Frederick Hayes-Roth, and Earl Sacerdoti. Basic concepts for building expert systems. In Frederick Hayes-Roth, Donald A. Waterman, and Douglas B. Lenat, editors, *Building Expert Systems*. Addison-Wesley, 1983.

[Stefik *et al* 83b] Mark Stefik, Janice Aitkins, Robert Balzer, John Benoit, Lawrence Birnbaum, Frederick Hayes-Roth, and Earl Sacerdoti. The architecture of expert systems. In Frederick Hayes-Roth, Donald A. Waterman, and Douglas B. Lenat, editors, *Building Expert Systems*. Addison-Wesley, 1983.

[Stephanou & Sage 87] Harry E. Stephanou and Andrew P. Sage. Perspectives on imperfect information processing. *IEEE Transactions on Systems, Man and Cybernetics*, 17(5):780–798, 1987.

[Stillman 91] Jonathan Stillman. On heuristics for finding loop cutsets in multiply connected belief networks. In P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 6*. Elsevier Science Publishers, 1991.

[Strat 87] Thomas M. Strat. The generation of explanations within evidential reasoning systems. In *Proceedings of the 10th International Joint Conference on Artificial Intelligence*, pages 1097–1104, 1987.

[Sucar & Gillies 94] L. Enrique Sucar and Duncan F. Gillies. Probabilistic reasoning in high-level vision. *Image and Vision Computing*, 12(1):42–60, 1994.

[Sucar *et al* 91] L. Enrique Sucar, Duncan F. Gillies, and Donald A. Gillies. Handling uncertainty in knowledge-based computer vision. *Lecture Notes in Computer Science*, 548:328–332, 1991.

[Sucar *et al* 93] L. E. Sucar, D. F. Gillies, and D. A. Gillies. Objective probabilities in expert systems. *Artificial Intelligence*, 61:187–208, 1993.

[Suermondt & Cooper 89] H. Jacques Suermondt and Gregory F. Cooper. Initialization for the method of conditioning in Bayesian belief networks. Technical Report KSL-89-61, Knowledge Systems Laboratory, Stanford University, 1989.

[Suermondt & Cooper 90] H. Jacques Suermondt and Gregory F. Cooper. Probabilistic inference in multiply connected belief networks using loop cutsets. *International Journal of Approximate Reasoning*, 4:283–306, 1990.

[Suermondt & Cooper 93] Henry J. Suermondt and Gregory F. Cooper. An evaluation of explanations of probabilistic inference. *Computers and Biomedical Research*, 26:242–254, 1993.

[Suermondt *et al* 91] H. Jacques Suermondt, Gregory F. Cooper, and David E. Heckerman. A combination of cutset conditioning with clique-tree propagation in the Pathfinder system. In P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 6*. Elsevier Science Publishers, 1991.

[Sullivan & Cohen 90] Michael Sullivan and Paul R. Cohen. An endorsement-based plan recognition program. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Proceedings of the National Conference of Artificial Intelligence*, 1985.

[Swartout & Smoliar 88] William R. Swartout and Stephen W. Smoliar. Explaining the link between causal reasoning and expert behaviour. In Perry L. Miller, editor, *Selected Topics in Medical Artificial Intelligence*. Springer-Verlag, 1988.

[Swartout 81] William R. Swartout. Explaining and justifying expert consulting programs. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, pages 815–822, 1981.

[Swartout 87] William R. Swartout. Explanation. In S. C. Shapiro, editor, *Encyclopedia of Artificial Intelligence, Volume 1*. John Wiley and Sons, 1987.

[Swartout *et al* 91] William Swartout, Cecile Paris, and Johanna Moore. Design for explainable expert systems. *IEEE Expert*, 6(3):58–64, 1991.

[Szolovits & Pauker 78] Peter Szolovits and Stephen G. Pauker. Categorical and probabilistic reasoning in medical diagnosis. *Artificial Intelligence*, 11:115–144, 1978.

[Szolovits & Pauker 90] Peter Szolovits and Stephen G. Pauker. Categorical and probabilistic reasoning in medical diagnosis. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Artificial Intelligence*, 11:115–144, 1978.

[Szolovits 79] Peter Szolovits. Artificial Intelligence and clinical problem solving. Technical Report MIT/LCS/TM-140, MIT Laboratory for Computer Science, 1979.

[Szolovits *et al* 86] Peter Szolovits, Jerome P. Kassirer, William J. Long, Alan J. Moskowitz, Stephen G. Pauker, Ramesh S. Patil, and Michael P. Wellman. An Artificial Intelligence approach to clinical decision making. Technical Report MIT/LCS/TM-310, MIT Laboratory for Computer Science, 1986.

[Teach & Shortliffe 81] Randy L. Teach and Edward H. Shortliffe. An analysis of physician attitudes regarding computer-based clinical consultation systems. *Computers and Biomedical Research*, 14:542–558, 1981.

[Terzopoulos *et al* 87] Demetri Terzopoulos, Andrew Witkin, and Michael Kass. Symmetry-seeking models for 3D object reconstruction. In *Proceedings of the 1st International Conference on Computer Vision*, pages 269–276, 1987.

[Thompson 85] Terence R. Thompson. Parallel formulation of evidential-reasoning theories. In *Proceedings of the 8th International Joint Conference on Artificial Intelligence*, pages 321–327, 1985.

[Todd *et al* 94] Bryan S. Todd, Richard Stamper, and Paul Macpherson. The design and construction of a medical simulation model. *Computer Methods and Programs in Biomedicine*, 42:77–91, 1994.

[Tong & Appelbaum 88] R. M. Tong and L. A. Appelbaum. Experiments with interval-valued uncertainty. In J. F. Lemmer and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Toronto, ONT 90] Toronto, ONT. Computers + medicine = future. *Canadian Family Physician*, 36:2254–2255, December 1990.

[Touretzky 90] David S. Touretzky. Implicit ordering of defaults in inheritance systems. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Proceedings of the National Conference of Artificial Intelligence*, 1984.

[Towers & Baldock 88] Simon Towers and Richard Baldock. Application of a knowledge-based system to the interpretation of ultrasound images — preliminary studies. In *Proceedings of the 9th International Conference on Pattern Recognition*, pages 107–110, 1988.

[Tversky & Kahneman 90a] Amos Tversky and Daniel Kahneman. Judgement under uncertainty: Heuristics and biases. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Science*, September 1974.

[Tversky & Kahneman 90b] Amos Tversky and Daniel Kahneman. Rational choice and the framing of decisions. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Journal of Business*, 59, 1986.

[Uckun 92] Serdar Uckun. Model-based reasoning in biomedicine. *Critical Reviews in Biomedical Engineering*, 19(4):261–292, 1992.

[Ursic 88] S. Ursic. Generalizing fuzzy logic probabilistic inferences. In John F. Lemmer and Laveen N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Vegas & Mira 92] F. J. Diez Vegas and J. Mira Mira. Causal Bayesian reasoning in medicine. *Cybernetics and Systems: An International Journal*, 23:417–429, 1992.

[Verma & Pearl 90] Thomas Verma and Judea Pearl. Causal networks: Semantics and expressiveness. In R. D. Shachter, T. S. Levitt, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 4*. North Holland, 1990.

[Wang *et al* 91] Lin Wang, Alan L. Porter, and Scott Cunningham. Expert systems: Present and future. *Expert Systems with Applications*, 3:383–396, 1991.

[Weid *et al* 90] George L. Weid, Harvey Dytch, Marluce Bibbo, Peter H. Bartels, and Deborah Thompson. Artificial Intelligence-guided analysis of cytologic data. *Analytical and Quantitative Cytology and Histology*, 12:417–428, 1990.

[Weiss *et al* 84] Sholom M. Weiss, Casimir A. Kulikowski, Saul Amarel, and Aran Safir. A model-based method for computer-aided medical decision making. In William J. Clancey and Edward H. Shortliffe, editors, *Readings in Medical Artificial Intelligence: The First Decade*. Addison Wesley, 1984.

[Wellman 90] Micheal P. Wellman. Qualitative probabilistic networks for planning under uncertainty. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Revised version of a paper published in J. F. Lemmer and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North-Holland, 1988.

[Wen 91] Wilson X. Wen. Optimal decomposition of belief networks. In P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 6*. Elsevier Science Publishers, 1991.

[Wesley 86] Leonard P. Wesley. Evidential knowledge-based computer vision. *Optical Engineering*, 25(3):363–379, March 1986.

[W.H.O. 88] W.H.O. *Cytological Screening in the Control of Cervical Cancer: Technical Guidelines*. World Health Organisation, 1988.

[Wied *et al* 76] George L. Wied, Gunter F. Bahr, and Peter H. Bartels. *The Automation of Uterine Cancer Cytology*. Tutorials of Cytology, 1976.

[Winograd 90] Terry Winograd. Extended inference modes in reasoning by computer systems. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Artificial Intelligence*, 13:5–26, 1980.

[Wise 88] B. P. Wise. Experimentally comparing uncertain inference systems to probability. In John F. Lemmer and Laveen N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Woods 75] William A. Woods. What's in a link: Foundations for semantic networks. In D. G. Bobrow and A. Collins, editors, *Representation and Understanding*. Academic Press, 1975.

[Worden *et al* 87] R. P. Worden, M. H. Foote, J. A. Knight, and S. K. Andersen. Co-operative expert systems. In B. Du Bulay, D. Hogg, and L. Steels, editors, *Advances in Artificial Intelligence 2*. North Holland, 1987.

[Wright 94] Jim Wright. Computed reality can illuminate research. *Scientific Computing*, pages 28–30, July 1994.

[Yadrick *et al* 88] R. M. Yadrick, B. M. Perrin, D. S. Vaughan, P. D. Holden, and K. G. Kempf. Evaluation of uncertain inference models I: PROSPECTOR. In John F. Lemmer and Laveen N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Yager 88] R. R. Yager. On implementing usual values. In John F. Lemmer and Laveen N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*. North Holland, 1988.

[Yen 89] John Yen. Implementing evidential reasoning in expert systems. In L. N. Kanal, T. S. Levitt, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 3*. North Holland, 1989.

[Young 84] D. W. Young. What makes doctors use computers?: Discussion paper. *Journal of the Royal Society of Medicine*, 77:663–667, August 1984.

[YunLeong 91] Tze Yun Leong. Representation requirements for supporting decision model formulation. In Bruce D. D'Ambrosio, Phillipe Smets, and Piero P. Bonissone, editors, *Uncertainty in Artificial Intelligence, proceedings of the 7th conference*, pages 212–219. Morgan Kaufman, 1991.

[Zadeh 65] Lofti A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[Zadeh 75a] Lofti A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning — I. *Information Sciences*, 8:199–249, 1975.

[Zadeh 75b] Lofti A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning — II. *Information Sciences*, 8:301–357, 1975.

[Zadeh 75c] Lofti A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning — III. *Information Sciences*, 9:43–80, 1975.

[Zadeh 83] Lofti A. Zadeh. The role of fuzzy logic in the management of uncertainty in expert systems. Technical Report Memorandum UCB/ERL M83/41, Electronics Research Laboratory, University of California, Berkeley, 1983.

[Zadeh 86] Lofti A. Zadeh. Is probability theory sufficient for dealing with uncertainty in AI: A negative view. In Laveen N. Kanal and John F. Lemmer, editors, *Uncertainty in Artificial Intelligence*. North Holland, 1986.

[Zador & Sokol 92] I Zador and R. J. Sokol. Perinatal computing design criteria for high tech support for clinical image services. *Early Human Development*, 29:199–202, 1992.

[Zador *et al* 91] I. E. Zador, V. Salari, L. Chik, and R. J. Sokol. Ultrasound measurement of the fetal head: Computer versus operator. *Ultrasound in Obstetrics and Gynecology*, 1:208–211, 1991.

[Zarley *et al* 90] Debra Zarley, Yen-Teh Hsia, and Glenn Shafer. Evidential reasoning using DELIEF. In Glenn Shafer and Judea Pearl, editors, *Readings in Uncertain Reasoning*. Morgan Kaufmann, 1990. Originally published in *Proceedings of the National Conference of Artificial Intelligence*, 1988.

[Zimmermann 90] H. J. Zimmermann. Problems and tools to model uncertainty in expert and decision support systems. In *Proceedings of the 7th International Conference on Mathematical and Computer Modelling*, pages 8–20, 1990.