Extended Incidence Calculus and Its Comparison with Related Theories

Weiru Liu

Department of Artificial Intelligence

University of Edinburgh

1995

Abstract

This thesis presents a comprehensive study of incidence calculus, a probabilistic logic for reasoning under uncertainty which extends two-value propositional logic to a multiple-value logic. There are three main contributions in this thesis.

First of all, the original incidence calculus is extended considerably in three aspects: (a) the original incidence calculus is generalized; (b) an efficient algorithm for incidence assignment based on generalized incidence calculus is developed; (c) a combination rule is proposed for the combination of both independent and some dependent pieces of evidence. *Extended incidence calculus* has the advantages of representing information flexibly and combining multiple sources of evidence.

Secondly, a comprehensive comparison between extended incidence calculus and the Dempster-Shafer (DS) theory of evidence is provided. It is proved that extended incidence calculus is equivalent to DS theory in representing evidence and combining independent evidence but superior to DS theory in combining dependent evidence.

Thirdly, the relations between extended incidence calculus and the assumption-based truth maintenance systems are discussed. It is proved that extended incidence calculus is equivalent to the ATMS in calculating labels for nodes. Extended incidence calculus can also be used as a basis for constructing probabilistic ATMSs.

The study in this thesis reveals that extended incidence calculus can be regarded as a bridge between numerical and symbolic reasoning mechanisms.

I DECLARE THAT THIS THESIS HAS BEEN COMPOSED BY MYSELF AND THAT THE WORK DESCRIBED IN IT IS MY OWN.

(Weiru Liu)

Acknowledgements

I am in debt to many people who have helped me in different way while I was working on my thesis.

- Professor Alan Bundy, my principle supervisor, has influenced my research greatly in many aspects. Without his theory, there would not be my thesis. Without his constant criticism, my work would not have gone so far. From him I have learned how to do research, how to present my ideas in papers clearly, and many many more ... . I have been and will be greatly benefited from these in my academic carrer.

- Dave Robertson, my second supervisor, has helped me in numerous ways and acted as my consultant on many issues over the years. I am very grateful to both Alan and Dave for their excellent supervision, encouragement and patience.

- Professor Jim Howe and Dr Paul Brna, helped me to get the scholarship and the ORS award which gave me the opportunity to study in the Department of Artificial Intelligence at the University of Edinburgh. I am also grateful to Paul for his patience to answer my countless questions on LaTex.

- Members of mathematical reasoning group and my roommates in F11 helped me in their own ways. Particularly, Dr Jason Gallagher , who read through Chapter 2 of my thesis with pages of comments.

- My thanks also to many members of the department, especially, Dr. Peter Ross, Ms. Janet Lee, Ms. Carole Douglas, Ms. Jean Bunten et al for their help.

- Some people in the Faculty of Informatics at the University of Ulster, especially, Professor David Bell, Professor John Hughes, Professor Michael McTear, had interesting discussions with me on the topics of my research and have provided me with an excellent working environment at the final stage of my PhD work. I would also like to thank Professor Philippe Smets at Belgium who acted as my 'extra' supervisor on a few of issues regarding the Dempster-Shafer theory of evidence and gave me many valuable comments on my work.

- Finally, I am grateful to my parents and my husband Dr. Jun Hong for their constant support, encouragement and love in this special period of time. Thanks to Jun for his patience in reading through my thesis among and between lines.

Table of Contents

1. Introduction **1**

1.1 Knowledge Representing and Reasoning Patterns 2

1.2 Incidence Calculus: Two Level Structures 5

1.2.1 From propositional logic to probabilistic logic 5

1.2.2 What is incidence calculus 8

1.2.3 Incidence calculus theories 8

1.2.4 Symbolic level 9

1.2.5 Numerical level 9

1.2.6 Incidence calculus is a bridge 10

1.3 Related Theories 11

1.3.1 The Dempster-Shafer theory of evidence 13

1.3.2 The ATMS 14

1.3.3 Probabilistic logic 17

1.4 Contributions of the Thesis 19

1.4.1 Extending the original incidence calculus 19

1.4.2 The main contributions of the thesis 20

1.5 Related Work 21

1.5.1 Bacchus's work 21

1.5.2 Laskey and Lehner's work 23

1.6 Thesis Structure 25

2. Incidence calculus **27**

2.1 Incidence Calculus Theories 27

2.2 The Legal Assignment Finder 30

2.3 Examples of Inference 33

2.4 Constraint Sets 38

2.5 Termination Decision of the Inference Procedure 39

2.6 Assigning Incidences to Formulae 40

2.7 Summary of Incidence Calculus 43

3. Extended Incidence Calculus **45**

3.1 Generalized Incidence Calculus 46

3.1.1 Generalized incidence calculus theories 46

3.1.2 Basic incidence assignments 50

3.1.3 Implementations of basic incidence assignments 56

3.2 Incidence Assignments 57

3.2.1 Assigning incidences to formulae in generalized incidence calculus 57

3.2.2 An example of incidence assignments 61

3.2.3 Generating multiple consistent incidence assignments 65

3.2.4 Estimating lower bounds of probabilities 67

3.2.5 Implementation of incidence assignments 69

3.3 Combining Evidence 70

3.3.1 Effects of new information 71
3.3.2 The Combination Rule in incidence calculus 73
3.3.3 DS-independent information 76
3.3.4 Implementation of combination rule 81
3.4 Summary of Extended Incidence Calculus 83

4. The Dempster-Shafer theory of evidence **86**

4.1 Basic Concepts in Dempster-Shafer Theory 87
4.2 Probability Background of Mass Functions 90
4.2.1 Dempster's probability prototype of mass functions 90
4.2.2 Deriving mass functions from probability spaces 93
4.3 Problems with Dempster's Combination Rule 95
4.3.1 The condition for using Dempster's combination rule 98
4.3.2 Examples 100
4.4 Some Other Aspects of DS Theory 106
4.4.1 Computational complexity problems in DS theory 106
4.4.2 Heuristic knowledge representation in DS theory 108
4.4.3 Open world assumptions by Smets 110
4.5 Summary of DS theory 112
4.5.1 Our argument in this chapter 112
4.5.2 Summary 113

5. A comprehensive comparison between generalized incidence calculus and the Dempster-Shafer theory of evidence **114**

5.1 Comparison I: Representing Evidence 115

5.2 Comparison II: Combining DS-Independent evidence 121

5.3 Comparison III: Combining Dependent Evidence 126

5.4 Analysing Examples 130

5.5 Comparison IV: Some Other Aspects Between The Two Theories . 134

5.5.1 Similar but different mapping relations in the two theories . 134

5.5.2 Recovering mass functions 135

5.6 Summary 138

6. Assumption-based truth maintenance systems **141**

6.1 The Reasoning Mechanism in The ATMS 141

6.2 Non-Redundant Justification Sets and Environments 144

6.3 Probabilistic Assumption Sets 147

6.4 Summary 152

7. On the relations between extended incidence calculus and the ATMS **153**

7.1 Incidence Calculus Review 154

7.1.1 Essential semantic implication sets in incidence calculus . . . 154

7.1.2 Similarities of the reasoning models in extended incidence calculus and the ATMS 159

7.2 Constructing Labels and Calculating Beliefs in Nodes Using Extended Incidence Calculus 160

7.2.1 An example 160

7.2.2 The algorithm of equivalent transformation from an ATMS to extended incidence calculus 165

7.2.3 Formal proof 167

7.2.4 Comparison with Laskey and Lehner's work 173

7.3 Implementing Incidence Calculus Using an ATMS 177

7.3.1 Examples 178

7.3.2 Transforming a set of generalized incidence calculus theories into an ATMS 183

7.4 Extended Incidence calculus Can Provide Justifications For The ATMS 186

7.5 Summary 188

8. Conclusions **190**

8.1 Introduction 190

8.2 Contributions of The Thesis 190

8.2.1 Extended incidence calculus 191

8.2.2 Relations between extended incidence calculus and DS theory191

8.2.3 Relations between extended incidence calculus and the ATMS192

8.3 Issues of Implementation 193

8.4 Limitations of Extended Incidence Calculus 194

8.5 Future Work 195

Appendix: Publications related to the Thesis **197**

Bibliography **199**

Mathematical Notation

P: a set of propositions.

q or q_j: an atomic proposition.

$\mathcal{A}t$: the basic element set formed from P.

$\mathcal{L}(P)$: the language set formed from P.

$\mathcal{W}$: a set of possible worlds.

μ: a probability distribution on a set.

i: an incidence function.

ii: a basic incidence assignment.

p_*: lower-bound of a probability distribution.

p^*: upper-bound of a probability distribution.

p: a probability distribution.

X or S, a set or space.

χ: a σ-algebra of a set X.

Θ: a frame of discernment.

bel: a belief function on a frame.

pls: a plausibility function on a frame.

m: a mass function on a frame.

A_{DS}: a set containing all the focal elements of a belief function.

Γ: a multivalued mapping function.

G: an incidence assignment.

$\otimes$: set product. $X_1 \otimes X_2 = \{< x_{1i}, x_{2j} > | x_{1i} \in X_1, x_{2j} \in X_2\}$

$\psi \models \phi$: formula $\psi \rightarrow \phi$ is valid (a tautology)

$\psi = \phi$: when $\psi \models \phi$ and $\phi \models \psi$.

Chapter 1

Introduction

Incidence calculus is a probabilistic logic developed by Bundy [Bundy, 1985]. This thesis is concerned with the further development of incidence calculus with an emphasis on both of its symbolic and numerical reasoning features. This thesis has the following main contributions.

- Extending the original incidence calculus
- Comparing the extended incidence calculus with the Dempster-Shafer theory of evidence (DS theory)
- Comparing the extended incidence calculus with assumption-based truth maintenance systems (the ATMS)

Since DS theory is a numerical reasoning mechanism and the ATMS is a symbolic reasoning mechanism, and incidence calculus can be compared with both of these theories, this suggests that incidence calculus must possess features of both. Therefore in this chapter, I will first discuss the main features of these forms of symbolic and numerical reasoning and explore the connections between them via incidence calculus. This discussion gives the insight that incidence calculus can be used as a bridge between symbolic and numerical reasoning. This insight has motivated my work in this thesis. As a consequence, I have developed the extended incidence calculus which captures the advantages of DS theory and the ATMS. I will then give a brief review of classical propositional logic and three closely related theories, namely, the Dempster-Shafer theory of evidence (DS theory) [Shafer76], the ATMS [de Kleer, 1986] and probabilistic logic [Nilsson, 1986]. Related work will also be discussed. Finally I will give an overview of the thesis.

1.1 Knowledge Representing and Reasoning Patterns

Representing and reasoning with knowledge and evidence in intelligent systems has been one of the major research topics in artificial intelligence (AI). Many mechanisms for representing and reasoning with knowledge and evidence have been developed so far. These mechanisms can, in general, be divided into the following two categories.

Symbolic reasoning: those mechanisms which represent and reason with precise (or certain) information belong to this category, such as propositional logic, first order logic, default logic, the ATMS and so on. In this type of reasoning system, information (including knowledge and evidence) and solutions are all represented in the form of symbols.

The normal province of symbolic reasoning is the derivation from initial precise information to a precise conclusion. Although such a conclusion is understood to be tentative (it may have to be retracted after new information has been added). The updated conclusion is still represented in symbolic form. No numerical values are used to assess the truth value of any statement.

For example, in the classical propositional logic, it is perfectly correct to say that if statements q_1 and q_2 are true, and the conjunction of q_1 and q_2 logically implies q_3, then q_3 is true.

$$q_1 \wedge q_2 \rightarrow q_3$$

In an ATMS, if the collected information says that fact A supports hypothesis q_1 and fact B supports hypothesis q_2, and we know that $q_1 \wedge q_2 \rightarrow q_3$, then it is possible to infer that $A \wedge B$ supports hypothesis q_3 if there is no contradictions between them.

Although these approaches are powerful in many aspects, such as logical soundness, they still suffer from some problems. One problem is that it is difficult to represent vague information. For instance, it is not possible to represent the sentences such as 'q_1 is possibly true' or 'q_1 is true with probability 0.7'.

Numerical reasoning: those mechanisms which represent and reason with vague or uncertain information such as 'probabilities' or 'possibilities' belong to this category. Numerous different approaches for managing uncertain information have been proposed, such as, the certainty factor model in MYCIN [Shortliffe, 1976], Bayes' rule based reasoning model in PROSPECTOR [Duda etal, 1976], the Dempster-Shafer theory of evidence [Shafer76], Fuzzy logic [Zadeh, 1975], probabilistic logic [Nilsson, 1986], belief networks [Pearl, 1988] and so on [Kruse et al, 1992].

The common feature of these uncertainty mechanisms is to model or describe vague or incomplete information explicitly and use it to make further judgements. In this category of reasoning, information is mainly characterized by numerical uncertain values[1]. A reasoning procedure involves both propagating and calculating uncertain values on hypotheses.

---

[1]I include the symbolic approaches designed for representing uncertainties such as Cohen's endorsements [Cohen, 1985] in this category. In such systems, although numbers are not used to represent uncertain information, some linguistic uncertain words, such as 'possible', 'probably', 'certain', are still used to describe vague information. This kind of approach is called symbolic-oriented uncertainty approach which is fundamentally different from the symbolic approach defined in the first category.

For example, if instead of saying that the hypotheses q_1 and q_2 are supported by A and B, the collected information tells us that hypotheses q_1 and q_2 are true with probabilities 0.6 and 0.7 respectively and their chances of being true are independent, then the probability of q_1 and q_2 are both true is 0.42. Without having any other information, q_3 is then true with probability 0.42.

Both symbolic and numerical approaches in the two categories have advantages and limitations. An intensive survey and discussion is provided by Dubois and Prade in [Dubois and Prade, 1994] between classical logic and the Bayesian networks. It is concluded that 'the deficiencies of classical logic and of Bayesian networks with respect to the plausible reasoning endeavour are not the same. The overriding ambition for knowledge representation and reasoning in the domain of plausible inference, is to identify a logic which combines the advantages of Bayesian networks with those of classical logic.' If we take the Bayesian network and classical logic as the representatives of numerical reasoning and symbolic reasoning techniques, this analysis also reveals the limitations of these two categories of reasoning patterns.

It is also addressed that 'numerical and symbolic (means logic) approaches to uncertainty should not be considered as completing models. It is far more interesting and fruitful to display their underlying coherence.' This could help to explain our motivation of the thesis work.

Besides, in any numerical reasoning mechanism, after a few steps of propagation and fusion, the meaning of numerical degrees is very difficult to interpret [Strat, 1987]. While in a symbolic reasoning mechanism, such as the ATMS, the degrees of being true of some statements can always be interpreted using other statements, like assumptions in the ATMS.

Incidence calculus, in some sense, unifies these two reasoning mechanisms. The comprehensive exploration into incidence calculus may provide some useful and valuable ideas to the researchers on both sides and reveal the underlying connections between the two reasoning mechanisms.

The real world problems which require both strong logical expression and numerical measurements may need such an combination.

Since the symbolic and numerical approaches are, to some extent, complementary not exclusive, attempts have been made to link them together in order to solve complex problems. [Rich, 1983] proposed a likehood-based interpretation of default rules using certainty-factors calculus. [Ginsberg, 1984], [Baldwin, 1987], [McLeith, 1988] used DS theory to describe Default theory. [d'Ambrosio, 1988], [Laskey and Lehner, 1989] attempted to encode DS theory into an ATMS.

The aim underlying their work is to construct a model which can effectively integrate numerical and symbolic reasoning into one structure. After having examined incidence calculus [Bundy, 1985], [Bundy, 1992], I discovered that it seems already to be an integration of these two forms. Therefore, in the next section I will concentrate on the structural analysis of incidence calculus.

1.2 Incidence Calculus: Two Level Structures

In this section, I will first introduce propositional logic and the basics of incidence calculus, and then explore its relations with the two reasoning types discussed above.

1.2.1 From propositional logic to probabilistic logic

Classical propositional logic is introduced in almost every artificial intelligence text book. Here I only briefly review its very basic features.

Propositional logic symbols

The symbols of propositional logic are the propositional symbols:

$$q, q_1, q_2, \ldots$$

Truth symbols:

$$true, false$$

and connectives:

$$\wedge, \vee, \neg, \rightarrow, \leftrightarrow$$

Propositional symbols, denote ***propositions***, or ***statements*** that may be either true or false, such as "it is raining" or "the car is white".

Sentences (or formulae) in the propositional logic are formed from these atomic symbols based on the following rules.

- Every atomic proposition and truth symbol is a sentence.
- The negation ($\neg$) of a sentence is a sentence, such as $\neg q$ from q.
- The conjunction ($\wedge$) of two sentences is a sentence, such as $q_1 \wedge q_2$.
- The disjunction ($\vee$) of two sentences is a sentence, such as $q_1 \vee q_2$.
- The implication ($\rightarrow$) of a sentence for another is a sentence, such as $q_1 \rightarrow q_2$.
- The equivalence of two sentences is a sentence, such as $q_1 \leftrightarrow q_2$.

Given two formulae ϕ and ψ, notation $\psi \models \phi$ is used when $\psi \rightarrow \phi$ is valid (a tautology) and $\psi = \phi$ means that sentence $\psi \leftrightarrow \phi$ is valid, that is $\psi \models \phi$ and $\phi \models \psi$.

A sentence (or formula) may be either true or false given some state of the world. The truth value assignment to sentences is called an *interpretation*, an assertion about their truth in some *possible world*. Formally, an interpretation of a sentence is a mapping from the propositional symbols into the set $\{T, F\}$ which are different from the symbols *true* and *false*.

The truth assignments of compound propositions can be calculated solely from their parts. For example, the truth value of $q_1 \wedge q_2$ is T if both q_1 and q_2 are T

and the truth value of $q_1 \wedge q_2$ is F if either q_1 or q_2 or both of them are F. This feature is called truth functional.

The classical propositional logic, like any other symbolic reasoning mechanism, lacks the ability to deal with uncertain information. Some work has been done on the generalization of propositional logic to probabilistic logic, that is, to extend a two-value logic to a multi-value logic by assigning uncertain values to sentences. In general, there are two ways to generalize the classical propositional logic to a probabilistic logic: direct encoding and indirect encoding.

Direct encoding : assigning probabilities to sentences directly, such as $prob(q_1 \rightarrow q_2) = 0.7$. Nilsson's probabilistic logic [Nilsson, 1986] uses this approach.

The disadvantage of this approach is that it is difficult to propagate probabilities based on a initial probability assignment. For instance, if we know that $prob(q_1) = 0.7$ and $prob(q_1 \rightarrow q_2) = 0.5$, it is not possible to calculate the probability of q_2, we can only know that the probability of q_2 lies between $prob(q_1) + prob(q_1 \rightarrow q_2) - 1$ and $prob(q_1 \rightarrow q_2)$. For more complicated cases, it is even difficult to tell the lower or upper bounds of probabilities on sentences.

Indirect encoding : assigning probabilities to sentences via a set of possible worlds, such as $prob(q_1 \rightarrow q_2) = prob(W_1) = 0.7$.

The advantage of this approach is that it is easy to propagate probabilities (or their bounds). Bundy's incidence calculus uses this method.

For example, if we are told that a set of possible worlds W_1 support q_1, and another set of possible worlds W_2 support $q_1 \rightarrow q_2$, then it is possible to say that the lower bound of the support set for q_2 is $W_1 \cap W_2$. It is then easy to calculate the probability of $W_1 \cap W_2$. Calculating probabilities through a set of possible worlds is the main component of indirect encoding and is the key idea in incidence calculus.

In the following, I use *formula* to name either an atomic proposition or a compound sentence.

1.2.2 What is incidence calculus

Incidence calculus is a probabilistic logic developed from propositional logic by associating probabilities with formulae indirectly. In incidence calculus, for a formula, denoted as ϕ, instead of saying ϕ is true or false, we say the probability of ϕ being true is x where x is any number between $[0, 1]$. If for every sentence in incidence calculus, its probability is either 0 or 1, then the theory reduces to the traditional propositional logic case.

To understand what we mean by saying the probability of a formula, a set of possible worlds, on which probability distributions are known, are employed to provide the explanation. Assume we talk about the truth value of a sentence in a particular domain, denoted as $\mathcal{W}$, which contains events (called samples in probability theory) related to the sentence. Any event in $\mathcal{W}$ will either support the sentence (make it true) or be against the sentence (make it false). If we put all the events supporting the sentence together, called W_1, then we get a subset of $\mathcal{W}$. This set is called the *incidence set* of this sentence. The probability of a sentence is defined as the probability of this subset. These events are called *possible worlds* in incidence calculus.

1.2.3 Incidence calculus theories

A piece of information is represented using an incidence calculus theory in incidence calculus.

An incidence calculus theory is defined formally as the quintuple

$$< \mathcal{W}, \mu, P, \mathcal{A}, i >$$

where $\mathcal{W}$ is a set of possible worlds or events, μ is a discrete probability distribution on $\mathcal{W}$, P is a finite set of propositions, $\mathcal{A}$ is a set of axioms (formulae) from the language set of P and i is a mapping function between $\mathcal{W}$ and $\mathcal{A}$. For a formula $\phi \in \mathcal{A}$, $i(\phi)$ is a subset of $\mathcal{W}$ containing those possible worlds which support the statement (or hypothesis) ϕ.

Since incidence calculus is a probabilistic logic achieved by assigning probabilities to formulae indirectly, the best way to describe and understand incidence calculus is to break its structure into two parts (the symbolic and numerical) and characterise each part.

1.2.4 Symbolic level

If we choose three elements $\mathcal{W}$, $\mathcal{A}$, and i from an incidence calculus theory to form a structure $< \mathcal{W}, \mathcal{A}, i >$, then this structure is purely symbolic as shown in Figure 1.1.

For instance, if we know that the incidence sets of $q_1 \rightarrow q_2$ and $q_2 \rightarrow q_3$ are W_1 and W_2, meaning W_1 and W_2 support statements $q_1 \rightarrow q_2$ and $q_2 \rightarrow q_3$, respectively, then logically, $W_1 \cap W_2$ should support $(q_1 \rightarrow q_2) \wedge (q_2 \rightarrow q_3)$, and further $q_1 \rightarrow q_3$. In this way, the incidence set of $q_1 \rightarrow q_3$ is changed from unknown to at least $W_1 \cap W_2$. This is the crucial point in propagating supporting sets and obtaining incidences in incidence calculus.

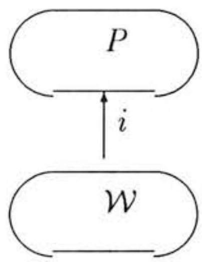

Figure 1.1. Symbolic support relation in IC

1.2.5 Numerical level

Don't forget that apart from the three elements we used in the symbolic level, there is another element in an incidence calculus theory, namely μ. Considering a structure $< \mathcal{W}, \mu, \mathcal{A} >$, what more can we know about statements in $\mathcal{A}$ from it? If it is known that W_1 supports statement $q_1 \rightarrow q_2$ and the probability weight of subset W_1 is $\mu(W_1) = \Sigma_{w \in W_1} \mu(w)$, then it is a natural consequence that the probability of the statment being true is $\mu(W_1)$. Similarly, the probability that $q_2 \rightarrow q_3$ is true is $\mu(W_2)$.

Moreover, as the relation $((q_1 \rightarrow q_2) \wedge (q_2 \rightarrow q_3)) \rightarrow (q_1 \rightarrow q_3)$ holds logically, it is believed that $W_1 \cap W_2$ should support $q_1 \rightarrow q_3$. Under the condition that we only have this logical relation to infer $q_1 \rightarrow q_3$, it is, once again, reasonable to say that the probability of $q_1 \rightarrow q_3$ being true is at least $\mu(W_1 \cap W_2)$.

Obviously, this time, we have used probabilities (or numerical uncertain values) to assess the truth value of a formula instead of using possible worlds to qualify it. This approach should be regarded as a numerical reasoning mechanism as shown in Figure 1.2.

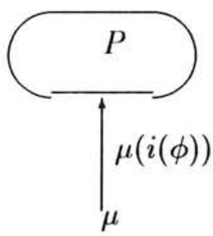

Figure 1.2. Numerical support relation in IC

1.2.6 Incidence calculus is a bridge

Where should we put incidence calculus, symbolic category or numerical slot? It seems that it fits both of them, but none of them covers it. Putting incidence calculus in a single category will make it lose the features of another. To make things even more clear, we'd better examine Figure 1.3.

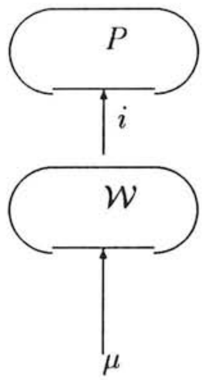

Figure 1.3. Numerical Reasoning via Symbolic Reasoning

A set of possible worlds (events) is located at the middle level in this structure. It acts as a bridge between formulae and probabilities. Through these events, numerical uncertainty values are assigned to hypotheses. Figure 1.2 shows that if we make the bridge invisible, then the theory looks like a numerical theory completely. This bridge, as we will see later, is very important in making the links between two different reasoning patterns.

It is not surprising, therefore, that we say that incidence calculus is different from the existing theories and it is expected that incidence calculus has the features of both numerical and symbolic reasoning systems. The following section serves this purpose.

1.3 Related Theories

Considering the fact that there are many different reasoning mechanisms both in numerical and symbolic reasoning patterns, it is not possible to examine the relations between incidence calculus and other mechanisms one by one. It is better to select a representative from each category respectively and to examine their relations with incidence calculus.

Bayesian networks are currently the most popular and most widely implemented model of reasoning under uncertainty. They are particularly successful from a computational point of view. In the theory of Bayesian nets, as presented by Pearl [Pearl, 1988], Bayesian networks certainly look like a very powerful tool for the efficient encoding of any complex multivariate probability distribution. However, as pointed out by Dubois and Prade in [Dubois and Prade, 1994] that there are a number of objections to the Bayesian approach: the results of the approach rely heavily on the independence assumptions encoded in the topology of the graph; the network building method never produces inconsistencies. The expert is asked exactly the amount of data required for ensuring the unicity of the underlying distribution.

One of the main advantages of extended incidence calculus is the ability to combine dependent information. This is a weakness of Bayesian nets. In addition, Bayesian approach does not have a correct representation of partial ignorance while extended incidence calculus does. So it is unlikely to show an equivalence between extended incidence calculus and Bayesian nets. Therefore, we didn't take Bayesian inference method as an example of numerical approaches to compare with extended incidence calculus. We intended to choose a theory which bears some resemblance to extended incidence calculus. The Dempster-Shafer theory of evidence satisfies this condition.

Constrained by this guideline, I have chosen the Dempster-Shafer theory of evidence (DS theory) and the ATMS as delegates to represent numerical and symbolic reasoning patterns respectively. The main reason for me to choose DS theory is that this theory is well known as a generalized probability theory and its relations with other numerical uncertainty methods have been intensively studied (for example, [Kruse et al, 1992], [Pearl, 1988]). The main reason for me to choose the ATMS in another category is that the ATMS is regarded as the *foundation for implementing various kinds of default reasoning* [de Kleer, 1986]. Default reasoning is a typical example among a few nonmonotonic reasoning mechanisms. Therefore, the analysis on the relations between incidence calculus and purely numerical or symbolic reasoning theories is focused on its comparisons with DS theory and the ATMS. This analysis reflects the general relations between incidence calculus and numerical and symbolic reasoning theories.

1.3.1 The Dempster-Shafer theory of evidence

DS theory was mainly developed by Shafer [Shafer76] based on Dempster's early work [Dempster, 1967]. DS theory is also well known as 'belief function theory' or 'evidential reasoning theory'. This theory has attracted a lot of attention since the early eighties [Lowrance, et al 1981], [Yager, Fadrizzi, Kacprzyk, 1994].

The advantages of DS theory, often addressed in the uncertainty community, are its abilities in representing ignorance and in combining different bodies of evidence using Dempster's combination rule.

In this theory, there are several basic concepts as follows.

Frame of discernment : a frame of discernment (or simply a frame) is a set which contains mutually exclusive and exhaustive explanations of a problem. That is, if this frame consists of all the possible answers to a question, then one and only one answer is correct at any one time.

Mass function : a mass function m gives a mapping between a frame of discernment and [0,1]. If S is a frame, then a mass function m on S satisfies the conditions $\Sigma_{A \subseteq S} m(A) = 1$ and $m(\emptyset) = 0$.

Belief function : a function bel defined by

$$bel(B) = \Sigma_{A \subseteq B} m(A)$$

is a belief function on S when m is a mass function on S.

Dempster's combination rule : Dempster's combination rule combines two mass functions on the same frame to produce the third one. Given that m_1 and m_2 are two mass functions, if they are allowed to be combined by the rule, then the result is $m = m_1 \oplus m_2$

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

where $\oplus$ means Dempster's combination rule is used and A, B, A', B' are subsets of S which is a frame.

A piece of evidence produces a belief function on a frame which, in many papers, is proved to be the same as an inner measure (or a lower bound) of a probability distribution [Fagin and Halpern, 1989b].

In incidence calculus, if we only look at the set P and its numerical measure derived from the possible worlds, then this numerical measure is exactly the same as a belief function if we let the two theories concern the same domain and the same information [Correa da Silva and Bundy, 1990b] and [Liu and Bundy, 1994].

Therefore, incidence calculus is able to simulate the reasoning procedure in DS theory and these two theories are equivalent in this aspect.

1.3.2 The ATMS

Among various symbolic approaches, I choose an ATMS as a basis for examining the symbolic feature of incidence calculus as far as this thesis is concerned. The detailed examination of the relations between incidence calculus and the ATMS is carried out in chapters 6 and 7. Discussions about the relations between incidence calculus and other mechanisms are open for the future.

Assumption based truth maintenance systems (ATMS) [de Kleer, 1986] were stimulated by Doyle's work on truth maintenance systems (TMS) [Doyle, 1979]. In such a system, dependent relations on statements are explicitly recorded and maintained. The main difference between an assumption based truth maintenance system and a truth maintenance system is that in the former only a specific set of statements (called assumptions) are qualified to support other statements while in the latter there is no such a distinction. Two tables below are used to show this intuitively.

Line	Statement	Dependencies	Justification
1.	$q_1 \rightarrow q_2$	$\{1\}$	Premise
2.	$q_2 \rightarrow q_3$	$\{2\}$	Premise
3.	q_1	$\{3\}$	Hypothesis
4.	q_2	$\{1, 3\}$	MP 1, 3
5.	q_3	$\{1, 2, 3\}$	MP 2, 4
6.	$q_1 \rightarrow q_3$	$\{1, 2\}$	Discharge 3, 5

Table 1.1. A TMS Example

Premises and hypotheses depend on themselves. The other lines depend on the set of premises and hypotheses derived from their justifications, which represent reasons for beliefs. Here MP means the application of Modus Ponens on premises and hypotheses.

This example can be rewritten in an ATMS as follows.

Node	Statement	Label	Justification
N_1	$q_1 \rightarrow q_2$	$\{\{A\}\}$	$\{(A)\}$
N_2	$q_2 \rightarrow q_3$	$\{\{B\}\}$	$\{(B)\}$
N_3	q_1	$\{\}$	$\{\}$
N_4	q_2	$\{\{A\}\}$	$\{(N_1, N_3)\}$
N_5	q_3	$\{\{A, B\}\}$	$\{(N_1, N_2, N_3)\}$
N_6	$q_1 \rightarrow q_3$	$\{\{A, B\}\}$	$\{(N_1, N_2)\}$

Table 1.2. An ATMS Example Converted from Table 1.1

In this table, nodes are used to replace the lines. All the capital letters are *assumptions* which are assumed to be true without requiring any extra information when there are no conflicts. q_1 is still a hypothesis (or called a premise) which is observed to be true. There is no longer a column labelled as *dependencies* in this table. Instead a new column called *label* is used to recode the dependencies between statements and assumptions. The purpose of having justifications is to provide routes to derive label sets for nodes. In a TMS, the dependency set of a statement records those statements on which this statement depends. Those statements can be any statements in this system (except itself usually). In an ATMS, the label set of a statement records those statements on which this statement depends. Those statements can only be assumptions. So the main difference between TMS dependencies and ATMS dependencies is that in a TMS any statement can appear in a dependency set while in an ATMS only assumptions can appear in a label set. For example, in Table 1.1, if we want to see whether q_3 holds, we will have to see whether lines 1, 2, and 3 hold, and if lines 1, 2, and 3 are dependent on other statements, we will have to continue this procedure until we reach those statement which are premises or hypotheses before we decide the truth of statement q_3. But in Table 1.2 if we want to know whether node N_5 holds, we only need to check whether A and B hold.

Therefore, dependent relations among statements in a TMS are restricted to the dependent relations between statements and assumptions. This is regarded as the advantage of the ATMS over the original TMS. The main target in such an ATMS is to manipulate label sets for statements through justifications.

For statement q_3 or $q_1 \rightarrow q_3$, the label set $\{\{A, B\}\}$ means that when both A and B hold and when there is no conflict to this fact, then q_3 or $q_1 \rightarrow q_3$ is derivable (or believed). So we have $A \wedge B$ supports q_3 or $q_1 \rightarrow q_3$.

In Table 1.2, the fourth line $< N_3, q_1, \{\}, \{\} >$ means that q_1 is observed to be true or N_3 holds universally.

Syntactically, if we put A and B in set theory, the conjunction between A and B should be replaced by intersection $\cap$, and A and B should be explained as the subsets of a certain set. In this sense, the meaning of A and B has been extended from individual assumptions to sets. I will consider both the ATMS and incidence calculus from this perspective and we shall find that these two theories share the same idea in their fundamental reasoning principles as shown in Table 1.3.

Formula	Incidences	Implication relation
$q_1 \rightarrow q_2$	W_1	
$q_2 \rightarrow q_3$	W_2	
q_1	$\mathcal{W}$	
q_2	W_1	$(q_1 \rightarrow q_2) \wedge q_1$
q_3	$W_1 \cap W_2$	$q_1 \wedge (q_1 \rightarrow q_2) \wedge (q_2 \rightarrow q_3)$
$q_1 \rightarrow q_3$	$W_1 \cap W_2$	$(q_1 \rightarrow q_2) \wedge (q_2 \rightarrow q_3)$

Table 1.3. Inference Procedure in Incidence Calculus.

Here I need to point out that the incidence set of q_2 should be $W_1 \cap \mathcal{W}$ which is equal to W_1. The same principle applies to the incidence set of q_3. Comparing Table 1.2 with Table 1.3, it is possible to draw the following mappings between them. The column 'Label' in Table 1.2 is equivalent to the column 'Incidences' in Table 1.3 if we imagine that A and B are the subsets W_1 and W_2. The 'Justification' column in Table 1.2 is equivalent to the 'Implication relation' column in Table 1.3 if we use node names to replace node numbers. It is necessary to point out that the incidence set of q_1 is the whole set of possible worlds because that q_1 holds universally and any event in $\mathcal{W}$ should support its occurrence.

Therefore, it is not difficult to see that the core parts in these two reasoning mechanisms are identical, *i.e.*, using logical implication relations to propagate the support environment. To be more explicit, justifications in an ATMS are functionally equivalent to the implication relations in incidence calculus.

1.3.3 Probabilistic logic

Probabilistic logic was introduced by Nilsson in 1986 [Nilsson, 1986]. The main contribution of the paper is to generalize the standard logic to the probabilistic case. That is, the truth value of a sentence is a probability value rather than just two values (truth or false). At the beginning of the paper, Nilsson provides an explanation of what it means to say the probability of a sentence. Given a sentence, say ϕ, if we start with a set of samples (as required in probability theory), this sentence can be either true or false on one sample. These samples are called possible worlds. So there are two sets of possible worlds, ϕ is true in one of them and is false in another. In general, if there are more sentences, we have more sets of possible worlds. If there are L sentences, then we might need as many as 2^L sets of possible worlds. Typically, far fewer sets of possible worlds are required.

Therefore, Nilsson concludes that 'the probability of a sentence is the sum of the probabilities of the sets of possible worlds in which that sentence is true'. This statement is identical to the idea of calculating probabilities of sentences through possible worlds in incidence calculus.

In probabilistic logic, although Nilsson uses sets of possible worlds to explain the source from which probabilities of sentences are obtained, probabilities are associated with sentences directly. A set of sentences, denoted as Π, each of which is associated with a probability, is called a base. From this base set, a new set of sentences is deducible with proper probabilities. The impact of the added information is considered as conditional probability updating in this theory. Therefore the main problem in probabilistic logic is to propagate probabilities to other sentences based on the base. In this propagation procedure, sets of possible worlds are dismissed.

As a consequence, it is not difficult to see the difference between probabilistic logic and incidence calculus. Both theories extend the traditional logic to probabilistic cases, but incidence calculus keeps sets of possible worlds (at an intermediate level) to achieve this purpose while probabilistic logic omits the use of possible worlds in the procedure of inferring more probabilities on sentences. The difference in their theoretical structures will have significant impact on their usages in practice.

In [Fagin and Halpern, 1989a], [Fagin and Halpern, 1989b], it is proved that probabilistic logic is covered by DS theory. This is explained as that given any base set with probabilities on sentences in the base (usually called Nilsson's probability structure), in DS theory there are a proper frame and a belief function on this frame which represent the same information.

As we have proved in [Liu and Bundy, 1994] generalized incidence calculus and DS theory have the same ability to represent evidence, *i.e.*, whatever DS theory can represent, generalized incidence calculus can represent it as well. It is natural to conclude that for any Nilsson's probability structure, there is an equivalent generalized incidence calculus theory to represent the same set of information.

1.4 Contributions of the Thesis

1.4.1 Extending the original incidence calculus

As we have seen, incidence calculus can be regarded as a bridge between symbolic reasoning and numerical reasoning mechanisms. This is determined by the theoretical structure of the theory. However, the original incidence calculus has three weaknesses which block its applications. These three shortcomings are: limited ability in representing evidence, limited ability in combining evidence and limited ability in assigning incidences given probabilities. These abilities, in many application domains, play very important roles in dealing with uncertain information.

In order to carry forward the structural advantage of incidence calculus and to make the theory be an even more powerful integration of the two reasoning patterns, we first have to extend the original incidence calculus. Therefore, the first part of my work in this thesis is concerned with extending the original incidence calculus to obtain an advanced mechanism which maintains the advantage of the theory but overcomes the three weaknesses.

The first limitation of the original incidence calculus is overcome by generalizing the conditions on incidence functions. I give fewer constraints on this crucial function. Generalized incidence calculus theories have the ability to model uncertain information flexibly, such as representing ignorance. The second limitation of the original incidence calculus is overcome by proposing a new combination rule to combine generalized incidence calculus theories. The new combination mechanism combines multiple pieces of evidence in their symbolic form first and then calculates numerical values. In this way, the new combination rule can combine both dependent and independent pieces of evidence. The third limitation is overcome by designing an efficient algorithm for incidence assignments from probability assignments.

Therefore, extended incidence calculus not only possesses the two advantages that DS theory has but also is superior to DS theory in combining dependent evidence.

Extended incidence calculus is a nonmonotonic reasoning mechanism. It shares this fundamental similarity with the ATMS. The advantage of extended incidence calculus over the original ATMS is that the former is able to calculate degrees of belief apart from performing inference at the symbolic level. This advantage can be used as a basis for constructing a probabilistic ATMS and for providing some theoretical explanations to the results in [Laskey and Lehner, 1989] (see next section and chapter 7), which is concerned with transforming DS belief functions into a probabilistic ATMS.

Therefore, we conclude that extended incidence calculus is a powerful reasoning mechanism unifying both symbolic and numerical reasonings. My work in this thesis has not only developed such a mechanism but also proved the important relations between extended incidence calculus and other theories.

1.4.2 The main contributions of the thesis

There are three main contributions in the thesis. First of all, I have extended the original incidence calculus in the following aspects: (i). generalized the definitions of incidence calculus theory; (ii). developed an efficient incidence assignment algorithm; (iii). proposed a combination rule to combine multiple pieces of evidence. So the original incidence calculus has been extended from three dimensions and the new, advanced mechanism is called *extended incidence calculus*. Chapter 3 contains the material for the extensions. Secondly, I investigated the relations between the extended incidence calculus and the Dempster-Shafer theory of evidence, their similarities and differences. Chapter 4 and 5 contribute to this investigation. Finally, I studied the relations between extended incidence calculus and the ATMS. I compared these two reasoning mechanisms at both the symbolic level and the numerical level. I proved that these two mechanisms are functionally equivalent. Chapter 6 and 7 have the details of this study.

1.5 Related Work

1.5.1 Bacchus's work

Bacchus in [Bacchus, 1988] and [Bacchus, 1990] extended the propositional logic and first order logic to probabilistic logic. His main aim, as he said himself, was to try to show that probabilities have an important role to play in the design of intelligent systems in general. Bacchus extended propositional logic to propositional probabilities, discussed statistical probabilities and combined these two types of probabilities into one form. He also addressed the application of statistical knowledge to default inference. I am, however, more interested in the first part of his work, *i.e.* propositional probabilities, which is closely related to my work.

The basic idea of extending propositional logic to propositional probabilities in [Bacchus, 1990] is the same as that in Nilsson's work, that is, changing the two-value assignments of propositions (or assertions) to be real values in [0,1]. This change can be done in two means, assigning probabilities to propositions directly or assigning probabilities to propositions via a set of possible worlds which have some formal links with the propositions.

Obviously, he preferred the second approach to the first one and criticized that the first approach suffers from the shortcoming of providing a unified language for assertions and probabilities over those assertions. By means of possible worlds, Bacchus defined the following structure for propositional probabilities (p41,[Bacchus, 1990]).

"**Propositional probability structures**: we define the following structure which we use to interpret the formulae of our language for propositional probabilities.

$$M =< \mathcal{O}, S, \vartheta, \mu >$$

where:

a) $\mathcal{O}$ is a set of individuals representing objects of the domain that one wishes to describe in the logic. $\mathcal{O}$ corresponds to the domain of discourse in the ordinary usage of first-order logic.

b) S is a set of possible worlds.

c) ϑ is a function that associates an interpretation of the language with each world.

d) μ is a discrete probability function on S. That is, μ is a function that maps the elements of S to the real interval [0,1] such that $\Sigma_{s \in S}\mu(s) = 1$."

The interpretation of the formula is explained as follows.

"In sum, the truth value assigned to a formula will depend on three items: the semantic structure or model M (which determines the probability distribution μ, the interpretation function ϑ, and the domain of objects $\mathcal{O}$); the current world s; and the variable assignment function v. We now give the inductive specification of the truth assignment, writing $(M, s, v) \models \alpha$ if the formula α is assigned a truth value **true** by the triple and writing $t^{(M,v)}$ for the individual denoted by the term t in the triple.

For every formula α, the term created by the probability operator $prob(\alpha)$ is given the interpretation

$$(prob(\alpha))^{(M,v)} = \mu\{s' \in S : (M, s', v) \models \alpha\}$$

"

So the probability of a formula is interpreted as the probability of the set of possible worlds which satisfy that formula.

The propositional probability structure given here is very similar to the incidence calculus theory structure except in an incidence calculus theory, a set of formulae (axioms) is particularly specified in $\mathcal{A}$. Both structures have used the probability of a set of possible worlds to interpret the probability of a formula.

However, although the two structures are similar in their appearance, there is a significant difference in their probability propagation procedures. In incidence calculus, possible worlds remain to be the main material in the propagation of probabilities, that is, the probability of a formula is calculated through its incidence set $i(\phi)$; while in Bacchus's structure, this seems not to be the case. In other words, in [Bacchus, 1990] possible worlds are used to represent information, their functions in further evidence propagation are not clear. For instance, in (p.45, [Bacchus, 1990]), an example is given as

$$prob(\exists x.has - cancer - type(John, x)) > 0.5$$

to represent the sentence 'John probably has some type of cancer'.

A set of possible worlds is associated with formulae in such a structure rather than separating them from formulae. This is the key difference between Bacchus structure [Bacchus, 1990] and incidence calculus.

The advantage of Bacchus structure is that a framework for representing first order logic has been defined. This part of work in incidence calculus remains to be done.

1.5.2 Laskey and Lehner's work

Embedding a proper numerical reasoning mechanism into an ATMS has been discussed by many researchers. Among them, Laskey and Lehner's work [Laskey and Lehner, 1989] is widely recognized. Their work is about translating a list of belief functions (which are given on a set of formulae) to a proper ATMS structure and then using this ATMS structure to carry out the inference. The main tasks in the translation procedure are to create assumptions and to calculate probabilities on formulae.

For example, if there are two mass functions m_1 and m_2 where

$$m_1(q_1 \rightarrow q_2) = 0.6 \qquad m_1(S) = 0.4$$

$$m_2(q_2 \rightarrow q_3) = 0.8 \qquad m_2(S) = 0.2$$

where S is the frame containing all formulae related to this question.

Then two assumptions A, B are created to separate mass values from formulae. The assumption A is used to support formula $q_1 \rightarrow q_2$ and the assumption B is used to support formula $q_2 \rightarrow q_3$. Furthermore, two probability sets $\{A, \neg A\}$ and $\{B, \neg B\}$ are created to associate probabilities with assumptions (and their negations). In this example, $prob(A) = 0.6, prob(\neg A) = 0.4$ and $prob(B) = 0.8, prob(\neg B) = 0.2$.

When it is assumed that the two mass functions are specified by two distinct pieces of evidence and the label set of $q_1 \rightarrow q_3$ is $\{\{A, B\}\}$, the probability of $q_1 \rightarrow q_3$ is calculated as $prob(A \wedge B) = 0.6 \times 0.8 = 0.48$.

In general, when the label set of a formula is $label(\alpha)$, it is not so easy to calculate the probability of α because of the overlapping of different parts in the label set. An algorithm has been proposed to deal with this case in [Laskey and Lehner, 1989].

The problem with their work is that some of the main results were given without proofs. The discussion about the relations between the extended incidence calculus and the ATMS will supply the necessary proofs [Liu and Bundy, 1993] and [Liu, Bundy and Robertson, 1993b].

1.6 Thesis Structure

This thesis is concerned with extending incidence calculus as a symbolic and numerical approach for uncertainty reasoning and its comparison with related theories. The thesis consists of eight chapters. The abstract of each chapter is given below.

Chapter 1 reviews the up-to-date techniques for numerical and symbolic uncertainty reasoning, explores the position of incidence calculus [Bundy, 1985] in these two major reasoning categories. This brief review provides a general background for exploring the potential applications of incidence calculus and the significance of its development. Because incidence calculus can make inference at both the symbolic and numerical levels, it can be regarded as a bridge between numerical and symbolic reasoning mechanisms.

Chapter 2 introduces Bundy's incidence calculus in great detail. This includes original definitions of incidence calculus, the Legal Assignment Finder for calculating lower and upper bounds of incidence and incidence assignment approaches.

Chapter 3 concentrates on how to generalize the original incidence calculus developed by Bundy [Bundy, 1992] to a more general case. This is done by dropping some of the conditions on incidence function i in the original incidence calculus. The generalized incidence calculus has the ability to represent ignorance. A fast algorithm for incidence assignment is designed and implemented based on generalized incidence calculus. A combination mechanism is proposed in generalized incidence calculus which can combine both dependent and independent pieces of evidence. Generalized incidence calculus with its alternative combination rule forms an advanced reasoning mechanism called *extended incidence calculus*.

Chapter 4 continues the discussion in [Halpern and Fagin, 1992] and [Voorbraak, 1991]. This chapter contains my contributions to the clarification of the problems with Dempster's combination rule in the Dempster-Shafer theory of evidence [Shafer76]. I considered not only Dempster's combination rule but also the original idea suggested by Dempster in his fundamental paper [Dempster, 1967].

Chapter 5 gives a comprehensive comparison between incidence calculus and DS theory. The result reveals that (i) both theories have the same ability to represent evidence; (ii) they have the same ability to combine DS-independent evidence and achieve the same result; (iii) incidence calculus is superior to DS theory in combining dependent evidence [Liu and Bundy, 1994].

Chapter 6 reviews the ATMS [de Kleer, 1986] and extends the original ATMS into a probabilistic oriented structure. This chapter is necessary for the discussion in the next chapter.

Chapter 7 focuses on the relations between incidence calculus and the ATMS. Because of its symbolic feature, incidence calculus is proved to be equivalent to the ATMS [Liu, Bundy and Robertson, 1993a], [Liu and Bundy, 1993], [Pearl, 1988]. In addition, incidence calculus provides a basis for constructing probabilistic based ATMSs and supplying justifications for the ATMS automatically.

Chapter 8 concludes the main issues in the thesis and my main contributions. I will also discuss the further work in the chapter.

This thesis can either be read in the order of chapters 1, 2, 3, 4, 5, 8, if one is interested in incidence calculus and DS theory; or in the order of chapters 1, 2, 3, 6, 7, 8, if one is interested in incidence calculus and the ATMS.

Chapter 2

Incidence calculus

We have had a rough idea about what incidence calculus looks like. In this chapter, we introduce the original incidence calculus developed by Bundy [Bundy, 1985] in detail. We will discuss its main features, its Legal Assignment Finder for deriving lower and upper bounds of incidences and the incidence assignments methods. Some limitations of the original incidence calculus will also be explored.

2.1 Incidence Calculus Theories

Incidence calculus was introduced in [Bundy, 1985]. It aims to overcome the problems which arose from applying purely numerical uncertainty reasoning techniques. In this new probabilistic reasoning model, probabilities are associated with sets of possible worlds and these sets are associated with formulae. These sets are called *incidences* of formulae. The reasoning procedure consists of calculating incidence sets (or their lower and upper bounds) of formulae and obtaining probabilities (or their lower and upper bounds) of the formulae. A simple introduction is given in [Bundy, 1992].

Definition 2.1: *Propositional Language*

- *P is a finite set of atomic propositions.*

- *$\mathcal{L}(P)$ is the propositional language formed from P.*

  $true, false \in \mathcal{L}(P)$,

  if $q \in P$, then $q \in \mathcal{L}(P)$, and

  if $\phi, \psi \in \mathcal{L}(P)$ then $\neg\phi \in \mathcal{L}(P)$, $\phi \wedge \psi \in \mathcal{L}(P)$, $\phi \vee \psi \in \mathcal{L}(P)$, and $\phi \rightarrow \psi \in \mathcal{L}(P)$.

  That is, $\mathcal{L}(P)$ is closed under the operations negation ($\neg$), disjunction ($\vee$), conjunction ($\wedge$) and implication ($\rightarrow$).

Definition 2.2: *Basic element set*

Assume that P is a finite set of propositions $P = \{q_1, q_2, ..., q_n\}$, an item δ, defined as $\delta = q'_1 \wedge ... \wedge q'_n$ where q'_j is either q_j or $\neg q_j$, is called a basic element. The collection of all the basic elements, denoted as $\mathcal{A}t$ is called the basic element set from P. Any formula ψ in the language set $\mathcal{L}(P)$ can be represented as

$$\psi = \delta_1 \vee ... \vee \delta_k$$

where $\delta_j \in \mathcal{A}t$.

Definition 2.3: *Incidence Calculus Theories*

An incidence calculus theory is a quintuple

$$< \mathcal{W}, \mu, P, \mathcal{A}, i >$$

where

- *$\mathcal{W}$ is a finite set of possible worlds.*
- *For all $w \in \mathcal{W}$, $\mu(w)$ is the probability of w and $\mu(\mathcal{W}) = 1$, where $\mu(I) = \Sigma_{w \in I} \mu(w)$.*
- *P is a finite set of propositions. $\mathcal{A}t$ is the basic element set of P. $\mathcal{L}(P)$ is the language set of P.*

- *$\mathcal{A}$ is a distinguished set of formulae in $\mathcal{L}(P)$ called the axioms of the theory.*

- *i is a function from the axioms in $\mathcal{A}$ to $2^{\mathcal{W}}$, the set of subsets of $\mathcal{W}$. $i(\phi)$ is to be thought of as the set of possible worlds in $\mathcal{W}$ in which ϕ is true, i.e. $i(\phi) = \{w \in \mathcal{W} |\ w \models \phi\}$. $i(\phi)$ is called the incidence of ϕ.*

i is extended to a function from $\mathcal{L}(\mathcal{A})$ to $2^{\mathcal{W}}$ by the following defining equations of incidence.

$$\begin{aligned}
&i(true) = \mathcal{W} \\
&i(false) = \{\} \\
&i(\neg\phi) = \mathcal{W} \setminus i(\phi) \\
&i(\phi \wedge \psi) = i(\phi) \cap i(\psi) \\
&i(\phi \vee \psi) = i(\phi) \cup i(\psi) \\
&i(\phi \rightarrow \psi) = \mathcal{W} \setminus i(\phi) \cup i(\psi)
\end{aligned}$$

Such an incidence calculus theory is *truth functional*, *i.e.*, the incidence set of a formula (if it exists) can be calculated purely from its parts.

It seems that there are no restrictions on i when it is initially defined on $\mathcal{A}$. However, as i can be extended to be a function on $\mathcal{L}(\mathcal{A})$ on which i should be truth functional, we can also assume that the incidence sets of formulae in $\mathcal{A}$ should satisfy these conditions as well. Given an incidence calculus theory, we should first check whether i satisfies these conditions on $\mathcal{A}$ before calculating the lower or upper bounds on a formula. For instance, if we know that

$$\phi = (\psi_1 \vee \psi_2) = (\psi_3 \vee \psi_4)$$

and $i(\psi_j), j = 1, .., 4$ are known, then $i(\psi_1) \cup i(\psi_2)$ must be the same as $i(\psi_3) \cup i(\psi_4)$, otherwise this incidence assignment is not consistent.

For a formula in $\mathcal{L}(P) \setminus \mathcal{L}(A)$, the lower and upper bounds of its incidence set are defined as:

$$i_*(\phi) = \bigcup_{\psi \in \mathcal{L}(A)} \{i(\psi) \mid i(\psi \to \phi) = \mathcal{W}\} \tag{2.1}$$

$$i^*(\phi) = \bigcap_{\psi \in \mathcal{L}(A)} \{i(\psi) \mid i(\phi \to \psi) = \mathcal{W}\} \tag{2.2}$$

The corresponding lower and upper bounds of probabilities of this formula are $p^*(\phi) = \mu(i^*(\phi))$ and $p_*(\phi) = \mu(i_*(\phi))$. For a formula, if $i_*(\phi) = i^*(\phi) = i(\phi)$, then $p(\phi)$ is defined as $p_*(\phi)$. We say $p(\phi)$ is the probability of formula ϕ.

It is easy to see that for a formula in $\mathcal{L}(A)$, $i_*(\phi) = i^*(\phi) = i(\phi)$ and $p(\phi) = p_*(\phi) = p^*(\phi)$.

For formulae ϕ and ψ in $\mathcal{L}(\mathcal{A})$, the conditional probability of ϕ given ψ is defined as

$$p(\phi \mid \psi) = \frac{p(\phi \wedge \psi)}{p(\psi)} \tag{2.3}$$

2.2 The Legal Assignment Finder

When an incidence calculus theory is specified, it guarantees that the known evidence gives a probability distribution on $\mathcal{L}(\mathcal{A})$ and gives lower and upper bounds of probabilities on other formulae. However, it is still the case that a piece of evidence may only specify the lower bounds and upper bounds of incidences on some formulae without giving the incidence function explicitly. Assume that the lower and upper bounds of incidences on a formula ϕ in a subset S of $\mathcal{L}(\mathcal{A})$ are $inf(\phi)$ and $sup(\phi)$ respectively. Then (inf, sup) can be extended to be two mappings from $\mathcal{L}(A)$ to $2^{\mathcal{W}}$ by assigning $inf(\psi) = \{\}$ and $sup(\psi) = \mathcal{W}$ for $\psi \in \mathcal{L}(A) \setminus S$. The lower and upper bounds of incidences on a formula given in this way are not tight enough. The approach for finding tighter bounds is called the *Legal Assignment Finder* in [Bundy, 1985], [Bundy, 1986]. Here we briefly introduce this approach. More details can be found in [Bundy, 1985], [Bundy, 1986], [McLean, 1992].

Definition 2.4: *Assignments*

An assignment G consists of a pair of mappings (inf_G, sup_G) from a subset S of $\mathcal{L}(P)$ to $2^{\mathcal{W}}$. inf_G is called the lower incidence bound and sup_G the upper incidence bound of formula ϕ in S.

inf_G and sup_G can be extended as mappings from $\mathcal{L}(P)$ to $2^{\mathcal{W}}$ by assigning $inf_G(\psi) = \{\}$ and $sup_G(\psi) = \mathcal{W}$ when $\psi \notin S$.

Definition 2.5: *Canonical Form*

A formula ϕ in $\mathcal{L}(P)$ is in canonical form if it has the form $\bigwedge_{j=1}^{m}(\neg \bigwedge_{l=1}^{n} r_j^l)$ where $r_j^l = q$ or $r_j^l = \neg q$ for $q \in P$.

For each formula ϕ, it is always possible to transform it into canonical form. We first transform ϕ into conjunctive normal form and then use de Morgan's law to turn each conjunct $\bigvee_{l=1}^{m} r_j^l$ into $\neg \bigwedge_{l=1}^{m} \neg r_j^l$ and then cancel all double negations. Note that a formula can either be rewritten as disjunctions of some basic elements in $\mathcal{A}t$ or be rewritten in canonical form. In the rest of this chapter, we use the canonical form of a formula more than we use the other form, especially in the inference rules below. However from the next chapter, we will mainly describe a formula using disjunctions of the basic elements.

Definition 2.6: *The Inference Rules in the Legal Assignment Finder*

A rule of inference is a mapping from assignments to assignments. If G_1 is the assignment before a rule fires and G_2 is the assignment after the rule works, then this rule changes the assignments from G_1 status to G_2 status.

There are following basic inference rules in this procedure.

$$
\begin{array}{lrcl}
\mathrm{Not}_1: & sup_{G_2}(\phi) & = & (\mathcal{W} \setminus inf_{G_1}(\neg\phi)) \cap sup_{G_1}(\phi) \\
\mathrm{Not}_2: & inf_{G_2}(\phi) & = & (\mathcal{W} \setminus sup_{G_1}(\neg\phi)) \cup inf_{G_1}(\phi) \\
\mathrm{Not}_3: & sup_{G_2}(\neg\phi) & = & (\mathcal{W} \setminus inf_{G_1}(\phi)) \cap sup_{G_1}(\neg\phi) \\
\mathrm{Not}_4: & inf_{G_2}(\neg\phi) & = & (\mathcal{W} \setminus sup_{G_1}(\phi)) \cup inf_{G_1}(\neg\phi) \\
\mathrm{And}_1: & sup_{G_2}(\phi) & = & (sup_{G_1}(\phi \wedge \psi) \cup (\mathcal{W} \setminus inf_{G_1}(\psi))) \cap sup_{G_1}(\phi) \\
\mathrm{And}_2: & inf_{G_2}(\phi) & = & inf_{G_1}(\phi \wedge \psi) \cup inf_{G_1}(\phi) \\
\mathrm{And}_3: & sup_{G_2}(\psi) & = & (sup_{G_1}(\phi \wedge \psi) \cup (\mathcal{W} \setminus inf_{G_1}(\phi))) \cap sup_{G_1}(\psi) \\
\mathrm{And}_4: & inf_{G_2}(\psi) & = & inf_{G_1}(\phi \wedge \psi) \cup inf_{G_1}(\psi) \\
\mathrm{And}_5: & sup_{G_2}(\phi \wedge \psi) & = & sup_{G_1}(\phi) \cap sup_{G_1}(\psi) \cap sup_{G_1}(\phi \wedge \psi) \\
\mathrm{And}_6: & inf_{G_2}(\phi \wedge \psi) & = & (inf_{G_1}(\phi) \cap inf_{G_1}(\psi)) \cup inf_{G_1}(\phi \wedge \psi)
\end{array}
$$

Some more inference rules for disjunction and implication are introduced in [McLean, 1992].

$$
\begin{array}{lrcl}
\mathrm{Or}_1: & inf_{G_2}(\phi) & = & (inf_{G_1}(\phi \vee \psi) \cap (\mathcal{W} \setminus sup_{G_1}(\psi))) \cup inf_{G_1}(\phi) \\
\mathrm{Or}_2: & sup_{G_2}(\phi) & = & sup_{G_1}(\phi \vee \psi) \cap sup_{G_1}(\phi) \\
\mathrm{Or}_3: & inf_{G_2}(\psi) & = & (inf_{G_1}(\phi \vee \psi) \cap (\mathcal{W} \setminus sup_{G_1}(\phi))) \cup inf_{G_1}(\psi) \\
\mathrm{Or}_4: & sup_{G_2}(\psi) & = & sup_{G_1}(\phi \vee \psi) \cap sup_{G_1}(\psi) \\
\mathrm{Or}_5: & inf_{G_2}(\phi \vee \psi) & = & inf_{G_1}(\phi) \cup inf_{G_1}(\psi) \cup inf_{G_1}(\phi \vee \psi) \\
\mathrm{Or}_6: & sup_{G_2}(\phi \vee \psi) & = & (sup_{G_1}(\phi) \cup sup_{G_1}(\psi)) \cap sup_{G_1}(\phi \vee \psi) \\
\mathrm{Imp}_1: & sup_{G_2}(\phi) & = & (\mathcal{W} \setminus inf_{G_1}(\phi \rightarrow \psi) \cup sup_{G_1}(\psi)) \cap sup_{G_1}(\phi) \\
\mathrm{Imp}_2: & inf_{G_2}(\phi) & = & (\mathcal{W} \setminus sup_{G_1}(\phi \rightarrow \psi)) \cup inf_{G_1}(\phi) \\
\mathrm{Imp}_3: & sup_{G_2}(\psi) & = & sup_{G_1}(\phi \rightarrow \psi) \cap sup_{G_1}(\psi) \\
\mathrm{Imp}_4: & inf_{G_2}(\psi) & = & (inf_{G_1}(\phi \rightarrow \psi) \cap inf_{G_1}(\phi)) \cup inf_{G_1}(\psi) \\
\mathrm{Imp}_5: & sup_{G_2}(\phi \rightarrow \psi) & = & ((\mathcal{W} \setminus inf_{G_1}(\phi)) \cup sup_{G_1}(\psi)) \cap sup_{G_1}(\phi \rightarrow \psi) \\
\mathrm{Imp}_6: & inf_{G_2}(\phi \rightarrow \psi) & = & (\mathcal{W} \setminus sup_{G_1}(\phi)) \cup inf_{G_1}(\psi) \cup inf_{G_1}(\phi \rightarrow \psi)
\end{array}
$$

It is proved in [Bundy, 1986] that the exhaustive application of these rules will terminate. It is also proved in [Correa da Silva and Bundy, 1990a] and in [Bundy, 1986] that the final assignment gives the same lower and upper incidence bounds on formulae as equations (2.1) and (2.2) do.

2.3 Examples of Inference

We will show in this section how to use the inference rules introduced above to derive lower and upper incidence bounds on formulae. In the first example, incidence function i on a subset of $\mathcal{L}(P)$ is specified, so the lower and upper bounds of the incidence of a formula can be obtained by either using the equations in Section 2.1 or using the Legal Assignment Finder. In the second example, we are only given the lower and upper bounds of incidence on formulae in a subset S of $\mathcal{L}(P)$, so only the Legal Assignment Finder method is adequate for inferring incidence bounds.

Example 2.1

Suppose that the set of axioms for a given incidence calculus theory is $\mathcal{A} = \{\phi, \phi \rightarrow \psi\}$, the set of possible worlds is $\mathcal{W} = \{a, b\}$ and $i(\phi) = \{a\}$, $i(\phi \rightarrow \psi) = \{a, b\}$. As ψ is not in $\mathcal{L}(A)$, we can only infer the lower and upper bounds of its incidence set.

There are two approaches to obtain the lower and upper bounds of incidence set of ψ. The first approach is to use equations (2.1) and (2.2) to get them while the second approach is to use the Legal Assignment Finder to do so.

Approach 1: Using the equations for lower and upper bounds

As $(\phi \wedge (\phi \rightarrow \psi)) \rightarrow \psi$ holds, applying equations (2.1) we have

$$i_*(\psi) = i(\phi \wedge (\phi \rightarrow \psi)) = i(\phi) \cap i(\phi \rightarrow \psi) = \{a\}$$

Similarly, as $i(\psi \rightarrow (\phi \rightarrow \psi)) = \mathcal{W}$, applying (2.2) on formula ψ, we have

$$i^*(\psi) = i(\phi \rightarrow \psi) = \{a, b\}$$

The basic idea of using the Legal Assignment Finder is to obtain tighter bounds from those currently known on some formulae. This requires the initial assignment of bounds on some relevant formulae from the given incidence assignment and then using the inference rules. For a formula ϕ, if $i(\phi)$ is known then the initial assignment G_1 is given as $inf_{G_1}(\phi) = sup_{G_1}(\phi) = i(\phi)$. Otherwise, we define $inf_{G_1}(\phi) = \{\}$ and $sup_{G_1}(\phi) = \mathcal{W}$.

Firing each inference rule normally leads from one assignment to another. This procedure terminates when the tightest bounds are found, that is, applying rules produce no new bounds, or when inconsistency is encountered, that is, a lower bound is larger than a upper bound.

For this example, from the incidence function i, we initialize the first assignment as (inf_1, sup_1) where $inf_1(\phi) = sup_1(\phi) = i(\phi)$, $inf_1(\neg(\phi \wedge \neg\psi)) = sup_1(\neg(\phi \wedge \neg\psi)) = i(\neg(\phi \wedge \neg\psi))$. Of course, this initialized bounds are not the tightest bounds yet. Only after applying the inference rules exhaustively, can the bounds be the tightest. Here $\neg(\phi \wedge \neg\psi) = \phi \rightarrow \psi$. Apart from the bounds on these two formulae, we also need to have the bounds on relevant formulae, such as, the bounds on ψ, $\neg\psi$, $\phi \wedge \psi$ and so on, in order to apply the rules. Following this basic inference principle, in order to infer the lower and upper bounds on ψ, we initialize the first two mapping functions inf_1, sup_1 on some relevant formulae as shown in Table 2.1.

Formula	inf_1	sup_1
ϕ	$\{a\}$	$\{a\}$
ψ	$\{\}$	$\{a, b\}$
$\neg\psi$	$\{\}$	$\{a, b\}$
$\phi \wedge \neg\psi$	$\{\}$	$\{a, b\}$
$\neg(\phi \wedge \neg\psi)$	$\{a, b\}$	$\{a, b\}$

Table 2.1. Initial Assignment of bounds on relevant formulae

Here we have ignored those formulae which have no effect on this inference procedure and all the formulae used in Table 2.1 are rewritten in the Canonical Form. After applying inference rules Not_1, And_3 and Not_2, Table 2.2 is obtained which contains the tightest bounds for formula ψ. That is, it is not necessary to apply any rules any more.

$$
\begin{aligned}
sup_2(\phi \wedge \neg\psi) &= (\mathcal{W} \setminus inf_1(\neg(\phi \wedge \neg\psi))) \cap sup_1(\phi \wedge \neg\psi) && \text{by rule Not}_1 \\
&= (\{a,b\} \setminus \{a,b\}) \cap \{a,b\} \\
&= \{\} \\
sup_3(\neg\psi) &= (sup_2(\phi \wedge \neg\psi) \cup (\mathcal{W} \setminus inf_2(\phi))) \cap sup_2(\neg\psi) && \text{by rule And}_3 \\
&= (\{\} \cup (\{a,b\} \setminus \{a\})) \cap \{a,b\} \\
&= \{b\} \\
inf_4(\psi) &= (\mathcal{W} \setminus sup_3(\neg\psi)) \cup inf_3(\psi) && \text{by rule Not}_2 \\
&= (\{a,b\} \setminus \{b\}) \cup \{\} \\
&= \{a\}
\end{aligned}
$$

The inference result can be summarized as shown in Table 2.2. The result obtained here is the same as that obtained with the first approach.

Formula	inf_4	sup_4
ϕ	$\{a\}$	$\{a\}$
ψ	$\{a\}$ (Not$_2$)	$\{a,b\}$
$\neg\psi$	$\{\}$	$\{b\}$ (And$_3$)
$\phi \wedge \neg\psi$	$\{\}$	$\{\}$ (Not$_1$)
$\neg(\phi \wedge \neg\psi)$	$\{a,b\}$	$\{a,b\}$

Table 2.2. Final assignment of upper and lower bounds after applying inference rules.

Example 2.2

Assume that the initial inf and sup are given for each formula in a subset $S = \{\delta_1 \vee \delta_2, \delta_1 \vee \delta_3, \delta_1, \delta_3\}$ of $\mathcal{L}(P)$ as

$$inf(\delta_1 \vee \delta_2) = \{w_1, w_4\}$$
$$inf(\delta_1 \vee \delta_3) = \{w_1, w_2\}$$
$$inf(\delta_1) = \{\}$$
$$inf(\delta_3) = \{\}$$
$$sup(\delta_1 \vee \delta_2) = \mathcal{W}$$
$$sup(\delta_1 \vee \delta_3) = \mathcal{W}$$
$$sup(\delta_1) = \{w_1, w_2, w_3\}$$
$$sup(\delta_3) = \{w_3, w_5\}$$

where $\mathcal{W} = \{w_1, w_2, ..., w_n\}$.

This example differs from Example 2.1 in that inf and sup are defined from S to $2^{\mathcal{W}}$ without giving any incidence function. In this case the lower and upper bound equations, (2.1) and (2.2), cannot be used directly.

Assume that we want to know the incidence bounds of formulae

$$S' = \{\delta_1, \delta_2, \delta_3, \delta_1 \vee \delta_2, \delta_1 \vee \delta_3, \delta_2 \vee \delta_3\}$$

we can extend the inf and sup to an initial assignment as shown in Table 2.3.

Formula	inf_1	sup_1
δ_1	$\{\}$	$\{w_1, w_2, w_3\}$
δ_2	$\{\}$	$\mathcal{W}$
δ_3	$\{\}$	$\{w_3, w_5\}$
$\delta_1 \vee \delta_2$	$\{w_1, w_4\}$	$\mathcal{W}$
$\delta_1 \vee \delta_3$	$\{w_1, w_2\}$	$\mathcal{W}$
$\delta_2 \vee \delta_3$	$\{\}$	$\mathcal{W}$

Table 2.3. Initial assignment of lower and upper incidence bounds.

In this initial lower and upper bound assignment, we have added $\{\}$ and $\mathcal{W}$ as lower and upper incidence bounds to some formulae in order to carry out the inference. A later lower and upper assignment must be a refinement of these.

Starting from the initial assignment, we applying the following Or inference rules after which no further inferences are possible. Table 2.4 is then obtained.

$$
\begin{array}{rcll}
inf_2(\delta_1) & = & (inf_1(\delta_1 \vee \delta_3) \cap (\mathcal{W} \setminus sup_1(\delta_3)) \cup inf_1(\delta_1) & \text{by rule } \mathrm{Or}_1 \\
 & = & (\{w_1, w_2\} \cap (\mathcal{W} \setminus \{w_3, w_5\})) \cup \{\} & \\
 & = & \{w_1, w_2\} & \\
inf_3(\delta_2) & = & (inf_2(\delta_1 \vee \delta_2) \cap \mathcal{W} \setminus sup_2(\delta_1)) \cup inf_2(\delta_2) & \text{by rule } \mathrm{Or}_3 \\
 & = & (\{w_1, w_4\} \cap (\mathcal{W} \setminus \{w_1, w_2, w_3\})) \cup \mathcal{W} & \\
 & = & \{w_4\} & \\
inf_4(\delta_1 \vee \delta_2) & = & inf_3(\delta_1) \cup inf_3(\delta_2) \cup inf_3(\delta_1 \vee \delta_2) & \text{by rule } \mathrm{Or}_5 \\
 & = & \{w_1, w_2\} \cup \{w_4\} \cup \{w_1, w_4\} & \\
 & = & \{w_1, w_2, w_4\} & \\
sup_5(\delta_1 \vee \delta_3) & = & (sup_4(\delta_1) \cup sup_4(\delta_3)) \cap sup_4(\delta_1 \vee \delta_3) & \text{by rule } \mathrm{Or}_6 \\
 & = & (\{w_1, w_2, w_3\} \cup \{w_3, w_5\}) \cap \mathcal{W} & \\
 & = & \{w_1, w_2, w_3, w_5\} & \\
inf_6(\delta_2 \vee \delta_3) & = & inf_5(\delta_2) \cup inf_5(\delta_3) \cup inf_5(\delta_2 \vee \delta_3) & \text{by rule } \mathrm{Or}_5 \\
 & = & \{w_4\} \cup \{\} \cup \{\} & \\
 & = & \{w_4\} &
\end{array}
$$

Formula	inf_6	sup_6
δ_1	$\{w_1, w_2\}$ (Or_1)	$\{w_1, w_2, w_3\}$
δ_2	$\{w_4\}$ (Or_3)	$\mathcal{W}$
δ_3	$\{\}$	$\{w_3, w_5\}$
$\delta_1 \vee \delta_2$	$\{w_1, w_2, w_4\}$ (Or_5)	$\mathcal{W}$
$\delta_1 \vee \delta_3$	$\{w_1, w_2\}$	$\{w_1, w_2, w_3, w_5\}$ (Or_6)
$\delta_2 \vee \delta_3$	$\{w_4\}$ (Or_5)	$\mathcal{W}$

Table 2.4. Final assignment of lower and upper incidence bounds after applying Or inference rules.

2.4 Constraint Sets

Assume that the set of axioms to which lower and upper bounds are assigned at the beginning is $\mathcal{A}$ and the set of theorems whose incidences (or bounds) we are interested in calculating is S. Then usually we need to extend the assignment of initial lower and upper bounds to a certain set in order to use the Legal Assignment Finder. This set is called the *constraint set* and is defined as follows.

Definition 2.7: *Constraint Set*

Let $\mathcal{A}$ be the set of axioms of a theory and $\mathcal{A}'$ be the set of formulae whose bounds of incidences we are interested in calculating.

Let $sf(S)$ be the set of subformulae of the formulae S, i.e.

if $\phi \in S$ then $\phi \in sf(S)$;
if $\neg\phi \in sf(S)$ then $\phi \in sf(S)$;
if $\phi \wedge \psi \in sf(S)$ then $\phi, \psi \in sf(S)$;
if $\phi \vee \psi \in sf(S)$ then $\phi, \psi \in sf(S)$;
if $\phi \rightarrow \psi \in sf(S)$ then $\phi, \psi \in sf(S)$.

The constraint set is $sf(\mathcal{L}'(\mathcal{A} \cup \mathcal{A}'))$ where $sf(\mathcal{L}'(\mathcal{A} \cup \mathcal{A}'))$ is the subset of $sf(\mathcal{L}(\mathcal{A} \cup \mathcal{A}'))$ that are in canonical form.

So given the initial assignment inf and sup, we only need to extend them as mappings from $sf(\mathcal{L}(\mathcal{A} \cup \mathcal{A}'))$ to $2^{\mathcal{W}}$ to get the bounds of any formula in $\mathcal{A}'$.

For Example 2.1, the constraint set is $\{\phi, \psi, \neg\psi, \phi \wedge \neg\psi, \neg(\phi \wedge \neg\psi)\}$ given that $\mathcal{A}' = \{\psi\}$.

2.5 Termination Decision of the Inference Procedure

It was proved in [Bundy, 1986] that the application of the above inference rules to an initial set of incidence assignments will eventually terminate. This procedure may or may not produce a consistent incidence assignment. Suppose that G is the assignment after some rules are fired and $inf_G(\phi)$ and $sup_G(\phi)$ are the lower and upper bounds assigned to ϕ.

1. If there exists a formula ψ where $inf_G(\psi) \not\subseteq sup_G(\psi)$ then the initial incidence assignment is not consistent. Terminate the inference procedure.

2. If for all formulae ψ in $\mathcal{L}(A)$, $inf_G(\psi) = sup_G(\psi)$, then the incidence assignment is consistent. Terminate the procedure.

3. For all other cases, a consistent incidence assignment may or may not exist.

For a case in situation 3, Bundy in [Bundy, 1985] and [Bundy, 1986] designed, proved and implemented a procedure called *case splitting* to continue the searching for a consistent assignment. It is said in [Bundy, 1985] that "Case splitting is necessary when the inference process runs out of rules to fire without specializing the assignment to a total or contradictory one. It is done by picking a point (means a possible world) and considering the two cases (1) that is is not and (2) that it is in the incidence of a sentence."

A more detailed discussion of case splitting, inconsistency checking and features of the inference procedure can be found in [Bundy, 1985] and [Bundy, 1986].

2.6 Assigning Incidences to Formulae

In incidence calculus, we use incidence calculus theories to represent pieces of evidence. In each of these theories, an incidence function is defined which is used to link probabilities on formulae. However, it may be the case that ***probabilities*** are directly associated with a set of axioms without specifying the incidence functions. In this case we need to recover the incidence assignment on axioms from the probability assignment. This procedure is called *Assigning incidences to formulae.* In [Bundy, 1992], it is assumed that a set of possible worlds is fixed, for instance 100 possible worlds, and the probability on each of them is equally distributed, *e.g.*, 1/100 for every w. Under this assumption, the Monte Carlo method [Corlett and Todd, 1985] is used to divide these possible worlds into groups to suit the given probability distribution on the set of axioms. This method has further been developed in [McLean, 1992]. Alternatively, a depth first approach on an incidence assignment tree has also been developed in [McLean, 1992] to solve the same problem under the same assumptions. Using these approaches, multiple consistent incidence assignments may be found given an initial probability distribution. Here we only briefly introduce these two approaches developed by [McLean, 1992]. A detailed discussion and comparison of the two approaches can be found in [McLean, 1992].

Depth First Incidence Assignment Algorithm. A depth first search is performed on the incidence assignment tree until either a consistent assignment is found or all leaves have been checked without a suitable assignment being discovered.

For example, if the numerical assignment on axioms x and y are $\mu(x) = 0.8$ and $\mu(y) = 0.2$, and if we assume that this unknown set of possible worlds W contains 10 elements, then there are 10!/(8!2!)=45 possible choices for $i(x)$ with $\mu(i(x)) = \mu(x)$. Similarly there are 45 possible choices for $i(y)$ with $\mu(i(y)) = \mu(y)$. Thus there are in total 2025 incidence functions compatible with the given uncertainty.

Monte Carlo Incidence Assignment Method. McLean further developed the approach introduced in [Bundy, 1992] and [Corlett and Todd, 1985] and proposed a Monte Carlo Incidence Assignment Method. This approach also performs on an incidence assignment tree as in the above method. This algorithm, as McLean discussed, suffers from the same problems shown below as the Depth First Method does.

Example 2.3

Suppose that the initial numerical assignment is

$$\mu(a \wedge b \wedge c) = 0.2$$
$$\mu(a \wedge c) = 0.4$$
$$\mu(b \wedge c) = 0.2$$
$$\mu(a) = 0.8$$
$$\mu(b) = 0.6$$
$$\mu(c) = 0.5$$

with a fixed set of possible worlds $W = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$.

Using the Depth First Method on axioms $\{a \wedge b \wedge c, a \wedge c, b \wedge c, a, b, c\}$ (in this order), it takes 10.700 (milliseconds) to find the first consistent assignment at case 6. However if the axioms are in the order $\{a, b, c, a \wedge c, b \wedge c, a \wedge b \wedge c\}$, then it tries 52 cases before a consistent assignment is found and the time is 70.600.

Using the Monte Carlo assignment method on this example has also been tested by McLean, but he didn't give the precise runtime using Prolog.

Example 2.4

A more complicated example used by McLean is as follows

$$\mu(a \wedge b \wedge c) = 0.18$$
$$\mu(a \wedge b) = 0.52$$
$$\mu(a \wedge c) = 0.35$$
$$\mu(b \wedge c) = 0.22$$
$$\mu(a) = 0.760$$
$$\mu(b) = 0.640$$
$$\mu(c) = 0.480$$

with the assumption that W has 100 elements.

Using the Depth First Method, it takes 374.520 to find the first consistent assignment and it takes 355.060 using the Monte Carlo Incidence Assignment Method. So the experimental results show that the algorithms are rather slow.

The execution of the examples suggested the following limitations of both of the methods.

(1). It is possible to have two leaves of the assignment tree each of which is compatible with the given uncertainties but for which the intervals

$$(\mu(lower\_bound1(\phi)), \mu(upper\_bound1(\phi)))$$

and

$$(\mu(lower\_bound2(\phi)), \mu(upper\_bound2(\phi)))$$

are disjoint.

(2). Using a depth first incidence assignment method to search for all possible assignments which are compatible with some given uncertainties may be very inefficient, since time may be wasted on discovering a large number of assignments which are permutations of a few basic ones.

(3). A search for a single consistent leaf may terminate with an unrepresentative sample which will then lead to poor estimates of uncertainties.

It was also pointed out by McLean that the efficiency of both algorithms are affected by the order in which the axioms are considered when assigning incidences.

2.7 Summary of Incidence Calculus

Incidence calculus is a logic for probabilistic reasoning intended to overcome the weaknesses in purely numerical uncertainty approaches. Its distinctive feature over other numerical methods is the indirect encoding of probabilities via incidence sets. This feature allows incidence calculus to be *truth functional*, that is the incidence of a compound formula can be calculated directly from the incidences of its parts. This is the basis for applying the inference rules in the Legal Assignment Finder.

The property $i(\neg\phi) = \mathcal{W} \setminus i(\phi)$ of i requires that the elements in $\mathcal{W}$ must be distributed into either $i(\phi)$ or $i(\neg\phi)$. If both $i(\phi)$ and $i(\neg\phi)$ are specified respectively, then $i(\phi) \cup i(\neg\phi)$ should be the whole set $\mathcal{W}$.

The property $i(\phi_1 \vee \phi_2) = i(\phi_1) \cup i(\phi_2)$ says that if $i(\phi_1)$, $i(\phi_2)$ and $i(\phi_1 \vee \phi_2)$ are all known, the possible worlds in $i(\phi_1 \vee \phi_2)$ can be split into two groups (not necessarily disjoint) $i(\phi_1)$ and $i(\phi_2)$. $i(\phi_1 \vee \phi_2)$ carries no more information than the union of $i(\phi_1)$ and $i(\phi_2)$.

If a real situation fails to meet either of the above two properties, incidence calculus theories cannot be used to describe it.

For any formula ϕ in $\mathcal{L}(\mathcal{A})$, we have $i(\phi) = i_*(\phi) = i^*(\phi)$ which results in $p(\phi) = p_*(\phi) = p^*(\phi)$. Therefore when the set of axioms is specified, the known evidence gives a probability distribution on $\mathcal{L}(\mathcal{A})$ and gives lower and upper bounds of probabilities on other formulae. When the effect of a piece of evidence fails to provide a probability distribution on this *core* part (e.g. $\mathcal{L}(\mathcal{A})$ is empty), incidence calculus cannot represent it as it is impossible to construct a proper incidence calculus theory.

Example 2.5

For instance, consider the function i' defined as:

$$i'(rainy) = \{mon, tues\}$$

$$i'(windy) = \{wed\}$$

$$i'(rainy \vee windy) = \{mon, tues\} \cup \{wed\} \cup \{thur\}$$

This function i' is *not* an incidence function as $i'(rainy \vee windy) \neq i'(\phi) \cup i'(\psi)$. We only know that on $thur$, it will be $rainy\, or\, windy$, but we don't *definitely* know it will be $rainy$ or be $windy$. So we cannot put $thur$ in either $i'(rainy)$ or $i'(windy)$. There is no corresponding incidence calculus theory to represent this piece of information. Our intention in the next chapter is to generalize the original incidence calculus described above to apply to a wider range of cases by weakening the conditions on incidence functions. Generalized incidence calculus keeps the feature of indirect encoding of uncertainties while losing the truth functional feature.

In the original incidence calculus, it is not clear how to cope with multiple sources of information. The basic inference mechanism in the theory is to ultimately infer the incidence bounds for a formula. It lacks the ability to combine the impact of several sources of information.

Chapter 3

Extended Incidence Calculus

At the end of the last chapter, we briefly discussed the problems with the original incidence calculus. It was pointed out that the original incidence calculus has limited abilities in representing and combining evidence. It was also pointed out that the methods for incidence assignments are not efficient. In order to overcome these weaknesses in the original incidence calculus, we extended it in the following three aspects.

- Generalization of incidence calculus theories: to weaken the definition of incidence calculus theories, in particular, the conditions on incidence functions, in order to model a wider range of cases, such as Example 2.5.

- New algorithm for incidence assignments: to provide a new incidence assignment algorithm based on generalized incidence calculus theories.

- New combination rule: to propose a new combination rule for combining multiple pieces of evidence which are in the form of generalized incidence calculus theories.

After we have made these three improvements on the original incidence calculus, we obtain an advanced reasoning mechanism which is called *Extended incidence calculus*. The crucial point in achieving the Extended Incidence Calculus is the new definition of incidence functions. In this chapter, I will discuss the three extensions in detail.

3.1 Generalized Incidence Calculus

3.1.1 Generalized incidence calculus theories

In order for incidence calculus to have the ability to represent a situation which an original incidence calculus theory is not suitable to represent, we generalize the original incidence calculus by dropping some of the conditions on it.

A mapping function $i' : \mathcal{A} \rightarrow 2^{\mathcal{W}}$ maps each formula ϕ in $\mathcal{A}$ to a subset of $\mathcal{W}$. $\mathcal{W}$ is interpreted as a set consisting of possible answers to a question Q. For $w \in \mathcal{W}$, w is a answer to Q. We still call $\mathcal{W}$ a set of possible worlds in this thesis. $w \in i'(\phi)$ means that if w is the answer to the question Q, then formula ϕ is true. We also require that $i'(false) = \{\}$ and $i'(true) = \mathcal{W}$. For a possible world $w \in \mathcal{W}$, if $w \notin i'(\phi)$, it doesn't necessarily mean that $w \in i'(\neg\phi)$. So if both $i'(\phi)$ and $i'(\neg\phi)$ are known, $i'(\phi) \cup i'(\neg\phi)$ may be just a subset of $\mathcal{W}$. This can be explained as that the current information says w supports neither ϕ nor $\neg\phi$. In other words, it is not known whether w supports ϕ or $\neg\phi$. This phenomenon is usually called *ignorance*. A mechanism which can model this phenomenon is said having the ability to represent ignorance.

Moreover, if $i'(\phi)$, $i'(\psi)$ and $i'(\phi \vee \psi)$ are all specified, it is possible that $i'(\phi) \cup i'(\psi) \subset i'(\phi \vee \psi)$ is valid. For instance, suppose that there are ten delegates elected to attend a meeting. The meeting will be held some day next week for which all the delegates are asked to give their preferences. The meeting will be held on the day which is preferred by most of the delegates. Suppose that delegates 1 to 4, denoted as a_1, ..., a_4, prefer mon, delegate 5, a_5, prefers mon or $tues$, the rest prefer $tues$. Then a mapping function i' could be defined as

$$i'(q_1) = \{a_1, a_2, a_3, a_4\}$$

$$i'(q_2) = \{a_6, a_7, a_8, a_9, a_{10}\}$$

and

$$i'(q_1 \vee q_2) = \{a_1, ..., a_5, ..., a_{10}\}$$

where q_1 stands for 'The meeting is held on Monday', q_2 for 'The meeting is on Tuesday'. Obviously, we have $i'(q_1) \cup i'(q_2) \subset i'(q_1 \vee q_2)$ because a_5 cannot be put into either $i'(q_1)$ or $i'(q_2)$.

A mapping function i' which has the ability to represent the above two phenomena is called a generalized incidence function.

For any two formulae ϕ, ψ in $\mathcal{A}$, if $i'(\phi), i'(\psi)$ and $i'(\phi \wedge \psi)$ are all known, then it can be proved that $i'(\phi \wedge \psi) = i'(\phi) \cap i'(\psi)$.

In fact, we have

$w \in i'(\phi) \cap i'(\psi) \Longleftrightarrow$
$w \in i'(\phi)$ and $w \in i'(\psi) \Longleftrightarrow$
ϕ is true when w is the answer and ψ is true when w is the answer $\Longleftrightarrow$
both ϕ and ψ are true when w is the answer $\Longleftrightarrow$
$w \in i'(\phi \wedge \psi)$

If we use $\wedge(\mathcal{A})$ to denote the language set which contains $\mathcal{A}$ and all the possible conjunctions of its elements. then a generalized incidence function can be extended to any formula in this set by defining $i'(\wedge\phi_j) = \cap_j i'(\phi_j)$, if $\wedge_j\phi_j$ is not given initially. Therefore, the domain of i', the set of axioms $\mathcal{A}$, can always be extended to a set which is closed under the operator $\wedge$.

Thus, whenever we have a set of axioms $\mathcal{A}$ on which a generalized incidence function i' is defined, this set of axioms can always be extended to another set which is closed under the operator $\wedge$. In the following, we always assume that the set of axioms $\mathcal{A}$ has already been extended and is closed under $\wedge$.

In particular, if $i'(\wedge_j\phi_j) = \{\}$, it doesn't matter whether this formula is in $\wedge(\mathcal{A})$ as this formula has no effect on further inferences. However if $\wedge_j\phi_j = \perp$, then $i'(\wedge_j\phi_j) = \cap_j i'(\phi_j)$ must be empty; otherwise the information for constructing the function i' is contradictory.

In the following, we use i to stand for a generalized incidence function, and from now on we will refer to it simply as an incidence function. Where any confusion could arise we will make clear the distinction between original and generalized incidence functions.

Definition 3.1: *Generalized Incidence Calculus Theories*

A quintuple $< \mathcal{W}, \mu, P, \mathcal{A}, i >$ *is called a generalized incidence calculus theory if the incidence function* i *satisfies the following three conditions*

$$i(false) = \{\} \qquad i(true) = \mathcal{W}$$

$$i(\phi_1 \wedge \phi_2) = i(\phi_1) \cap i(\phi_2) \ \ for \ \ \phi_1, \phi_2 \in \mathcal{A}$$

where $\mathcal{W}$*,* μ*,* P *and* $\mathcal{A}$ *are the same as defined in definition 2.3.*

In Example 2.5 given in Chapter 2, the function i' is a generalized incidence function. It is also easy to define a generalized incidence calculus theory as:

$$< W, \mu, P, \mathcal{A}, i >$$

where $\mathcal{W} = \{mon, tues, ..., sun\}$. $\mathcal{A} = \{rainy, windy, rainy \vee windy\}$, $i(rainy) = \{mon, tues\}$, $i(windy) = \{wed\}$, and $i(rainy \vee windy) = \{mon, tues, wed, thur\}$. As $i(rainy \wedge windy) = \{\}$, it is not necessary to put $rainy \wedge windy$ in $\mathcal{A}$ as an axiom.

Definition 3.2 *Representing Total Ignorance*

Given a generalized incidence calculus theory $< \mathcal{W}, \mu, P, \mathcal{A}, i >$*, if* $\mathcal{A} = \{true, false\}$*, then we say that this generalized incidence calculus theory represents total ignorance.*

Proposition 1 *If* ϕ *is a formula in a set of axioms* $\mathcal{A}$ *and both* $i(\phi)$ *and* $i(\neg\phi)$ *are known, then*

$$i(\phi) \cap i(\neg\phi) = \{\}$$

$$i(\phi) \cup i(\neg\phi) \subseteq \mathcal{W}$$

Proposition 2 *If ϕ and ψ are two formulae in a set of axioms $\mathcal{A}$ and $i(\phi)$, $i(\psi)$ and $i(\phi \vee \psi)$ are known, then*

$$i(\phi) \cup i(\psi) \subseteq i(\phi \vee \psi)$$

For a formula $\varphi \in \mathcal{L}(P) \backslash \mathcal{A}$, we can only define both the upper and lower bounds of its incidences using the functions i^* and i_* respectively. For all $\phi \in \mathcal{L}(P)$ these are defined as follows:

$$i_*(\phi) = \bigcup_{\psi \in \mathcal{A}, \psi \models \phi} i(\psi) \tag{3.1}$$

$$i^*(\phi) = \mathcal{W} \setminus i_*(\neg \phi) \tag{3.2}$$

For any $\phi \in \mathcal{A}$, we have $i_*(\phi) = i(\phi)$.

The lower bound represents the set of possible worlds which make ϕ true and the upper bound represents the set of possible worlds which fails to make $\neg\phi$ true. Function $p_*(\phi) = \mu(i_*(\phi))$ gives the degree of our belief in ϕ and function $p^*(\phi) = \mu(i^*(\phi))$ represents the degree we fail to believe in $\neg\phi$. For any formula ϕ in $\mathcal{A}$, if $p_*(\phi) = p^*(\phi)$, then $p(\phi)$ is defined as $p_*(\phi)$ and called the probability of this formula. In this case, for any ϕ and ψ in $\mathcal{A}$, let $p(\phi \mid \psi)$ be the conditional probability of ϕ given ψ, we define

$$p(\phi \mid \psi) = \frac{p(\phi \wedge \psi)}{p(\psi)} \tag{3.3}$$

When a generalized incidence calculus theory reduces to be an original incidence calculus theory, it is proved in [Correa da Silva and Bundy, 1990a] and [Bundy, 1986] that equations (3.1) and (2.1) produce the same result, so do (3.2) and (2.2).

It is necessary to notice that for any $\phi \in \mathcal{A}$, $i_*(\phi)$ and $i^*(\phi)$ are not the same in most cases. So function p on $\mathcal{A}$ cannot be defined. Therefore, only a function p_*, the lower bound of a probability distribution, is defined on $\mathcal{A}$. Given a formula

$\phi \in \mathcal{L}(P)$, it is more natural to describe the lower bound of its probability than its probability. In the following, when we mention a numerical assignment on a language set, we always mean a lower bound of an unknown probability measure.

3.1.2 Basic incidence assignments

For a formula ϕ in its disjunctive normal form $\delta_1 \vee ... \vee \delta_l$, we define a subset A of $\mathcal{A}t$ as $A = \{\delta_1, ..., \delta_l\}$ and denote formula ϕ as ϕ_A. ϕ_A means that the disjunction of the elements in A is the disjunctive normal form of formula ϕ. $i(\phi_A)$ contains both the possible worlds which make ϕ_A true and the possible worlds which make ϕ_B true for $B \subset A$. So some of these possible worlds may only make ϕ_A true without making any of ϕ_B (for $B \subset A$) true.

For instance, suppose that we have two propositions q_1 and q_2 in P, then there are four basic elements in $\mathcal{A}t$ as $\delta_1 = q_1 \wedge q_2$, $\delta_2 = q_1 \wedge \neg q_2$, $\delta_3 = \neg q_1 \wedge q_2$ and $\delta_4 = \neg q_1 \wedge \neg q_2$.

If we are given that $i(\phi_{\{\delta_1\}}) = \{w_1\}$, $i(\phi_{\{\delta_1,\delta_2\}}) = \{w_1, w_2\}$ and $i(\phi_{\{\delta_1,\delta_3\}}) = \{w_1, w_3\}$, then w_2 makes only $\phi_{\{\delta_1,\delta_2\}} = q_1$ true without making $q_1 \wedge q_2$ true. Similarly w_3 makes q_2 true without making $q_1 \wedge q_2$ true.

In general, the subset of $i(\phi_A)$ which contains the possible worlds only making ϕ_A true without making any of ϕ_B true ($B \subset A$) is denoted as $ii(\phi_A)$ and the notation ii is called the **basic incidence assignment**. When a set $ii(\phi)$ is empty, we don't think it carries any significant message for further inference, so usually we only consider those formulae of which $ii(*)$ is not empty. In order to show the relation between i and ii, we first look at an example. Suppose there are two propositions, $P = \{rainy, windy\}$, and seven possible worlds, $\mathcal{W} = \{sun, mon, tues, wed, thus, fri, sat\}$. Assume that each possible world is equally probable, *i.e.* occurs 1/7 of the time. Through a piece of evidence, we learn that four possible worlds *fri, sat, sun, mon* make *rainy* true, and three possible worlds *mon, wed, fri* make *windy* true. Therefore the incidence sets of these two propositions are:

$$i(rainy) = \{fri, sat, sun, mon\}$$
$$i(windy) = \{mon, wed, fri\}$$

As $i(rainy \wedge windy) = i(rainy) \cap i(windy)$, we also have $i(rainy \wedge windy) = \{fri, mon\}$. So the set of axioms $\mathcal{A}$ is $\mathcal{A} = \{rainy, windy, rainy \wedge windy\}$ which is closed under $\wedge$. The corresponding incidence calculus theory is

$$< \mathcal{W}, \mu, P, \mathcal{A}, i >$$

and the $\mathcal{A}t$ of P is $\mathcal{A}t = \{rainy \wedge windy, rainy \wedge \neg windy, \neg rainy \wedge windy, \neg rainy \wedge \neg windy\}$. A basic incidence assignment ii could be naturally defined as:

$$ii(rainy \wedge windy) = \{fri, mon\}$$
$$ii(rainy) = \{sat, sun\}$$
$$ii(windy) = \{wed\}$$

For any other formula ϕ except $true$, $ii(\phi)$ is empty. It is easy to see that from ii, the incidence function can be recovered as:

$$i(rainy \wedge windy) = ii(rainy \wedge windy)$$

$$i(rainy) = ii(rainy) \cup ii(rainy \wedge windy)$$

$$i(windy) = ii(windy) \cup ii(rainy \wedge windy)$$

Definition 3.3: *Basic Incidence Assignment*

Given a set of axioms $\mathcal{A}$, a mapping function ii: $\mathcal{A} \rightarrow 2^{\mathcal{W}}$ is called a basic incidence assignment if ii satisfies the following conditions:

$$\begin{array}{ll} ii(\phi) \neq \{\} & \phi \in \mathcal{A} \\ ii(\phi) \cap ii(\psi) = \{\} & \phi \neq \psi \\ ii(false) = \{\} & \\ ii(true) = \mathcal{W} \setminus \bigcup_j ii(\phi_j) & \phi_j \in \mathcal{A} \end{array}$$

where $\mathcal{W}$ is a set of possible worlds.

Proposition 3 *Given a set of axioms $\mathcal{A}$ with a basic incidence assignment ii, then function i defined by equation (3.4) below is an incidence function on $\mathcal{A}$.*

$$i(\phi) = \bigcup_{\phi_j \in \mathcal{A}, \phi_j \models \phi} ii(\phi_j) \tag{3.4}$$

PROOF

First of all, because $ii(true) = \mathcal{W} \setminus \cup_j ii(\phi_j)$, we have $i(true) = ii(true) \cup (\cup_j ii(\phi_j)) = \mathcal{W}$. As $ii(false) = \{\}$, it is straightforward to infer that $i(false) = \{\}$.

Next we are going to prove that $i(\phi \wedge \psi) = i(\phi) \cap i(\psi)$ for ϕ and $\psi \in \mathcal{A}$.

Suppose that $i(\phi) \cap i(\psi) = \mathcal{W}' \neq \{\}$,

$$\begin{array}{ll}
\forall w \in \mathcal{W}', & w \in i(\phi) \cap i(\psi) \Longleftrightarrow \\
\exists \phi_0, & w \in ii(\phi_0) \;\; (\, \phi_0 \models \phi,\, \phi_0 \models \psi) \Longleftrightarrow \\
\exists \phi_0, & w \in ii(\phi_0) \;\; (\, \phi_0 \models \phi \wedge \psi) \Longleftrightarrow \\
 & w \in i(\phi \wedge \psi)
\end{array}$$

So $i(\phi) \cap i(\psi) = i(\phi \wedge \psi)$.

When $i(\phi) \cap i(\psi) = \{\}$, it is still easy to prove that $i(\phi) \cap i(\psi) = i(\phi \wedge \psi)$. Therefore function i defined by (3.4) is an incidence function.

QED

Proposition 4 *Given a generalized incidence calculus theory $< \mathcal{W}, \mu, P, \mathcal{A}, i >$, there exists a basic incidence assignment ii on $\mathcal{A}$ from which the incidence function i in the theory can be derived using equation (3.4).*

PROOF

This proof procedure is actually to construct a basic incidence assignment ii from the given incidence function.

From the theory $< \mathcal{W}, \mu, P, \mathcal{A}, i >$, we have

$$i(\phi \wedge \psi) = i(\phi) \cap i(\psi)$$

where $\phi, \psi \in \mathcal{A}$.

The requirement on i leads us to the conclusion that if $\psi \rightarrow \phi$ is a tautology then $i(\psi) \subseteq i(\phi)$. As we assume that P is finite, then $\mathcal{A}t, \mathcal{L}(P)$ and $\mathcal{A}$ are all finite (we assume that all the formulae in $\mathcal{L}(P)$ are in the form of disjunctions of the basic elements).

A subset $\mathcal{A}_0$ of $\mathcal{A}$ can be defined as $\mathcal{A}_0 = \{\psi_1, ..., \psi_n\}$ where $\mathcal{A}_0$ satisfies the condition that

$$\forall \psi_j \in \mathcal{A}_0, \forall \phi \in \mathcal{A},\ if\ \ \phi \neq \psi_j\ \ then\ \ \phi \not\models \psi_i$$

Therefore, $\mathcal{A}_0$ contains the *smallest (or prime implicant in lattice theory)* formulae in $\mathcal{A}$ and $\mathcal{A}_0$ is not empty[1]. In fact, we can get $\mathcal{A}_0$ using the following procedure. For a formula $\psi \in \mathcal{A}$, if $\exists \phi \in \mathcal{A}$, $\phi \neq \psi$ and $\phi \rightarrow \psi$ is valid, then we use ϕ to replace ψ and repeat the same procedure until we obtain a formula ϕ_l and we cannot find any formula which makes ϕ_l true, then ϕ_l will be in $\mathcal{A}_0$. For instance, the set $\mathcal{A}_0$ in Example 2.5 is $\mathcal{A}_0 = \{rainy \wedge windy\}$.

For any two formulae $\psi_l, \psi_j \in \mathcal{A}_0$, when $\psi_l \neq \psi_j$ we have

$$i(\psi_l) \cap i(\psi_j) = \{\}$$

In fact if $i(\psi_l) \cap i(\psi_j) = \mathcal{W}' \neq \{\}$, then

[1]We require that $\mathcal{A}$ contains at least one more element except $false$ and $true$.

$$\mathcal{W}' = i(\psi_l \wedge \psi_j) \Longrightarrow$$

$$\exists \psi \in \mathcal{A} \qquad \psi = \psi_l \wedge \psi_j, \qquad \psi \not\models \perp \Longrightarrow$$

$$\psi \models \psi_l, \ \psi \models \psi_j \Longrightarrow$$

$$\psi \neq \psi_l \ \text{ or } \ \psi \neq \psi_j \ \text{ as } \psi_i \neq \psi_j \Longrightarrow$$

$$\psi_l \notin \mathcal{A}_0 \text{ or } \psi_j \notin \mathcal{A}_0$$

Contradictory! So we have $i(\psi_i) \cap i(\psi_j) = \{\}$.

For any formula ϕ_l in $\mathcal{A} \setminus \mathcal{A}_0$, there are ψ_{l1}, ..., $\psi_{lm} \in \mathcal{A}_0$ where $\psi_{lj} \models \phi_l$. So $i(\psi_{lj}) \subseteq i(\phi_l)$ and $(\bigcup_j i(\psi_{lj})) \subseteq i(\phi_l)$.

Algorithm A

From a function i, we can obtain another function ii using the following procedure:

Step 1: for every formula $\psi \in \mathcal{A}_0$, define $ii(\psi) = i(\psi)$.

Step 2: define $\mathcal{A}'$ as $\mathcal{A} \setminus \mathcal{A}_0$.

Step 3: using the same method as we define $\mathcal{A}_0$ on p51, we can choose a formula ϕ_l in $\mathcal{A}'$ which satisfies the requirement that for any $\phi_j \in \mathcal{A}'$, if $\phi_j \neq \phi_l$, then $\phi_j \not\models \phi_l$ and there must be a list of formulae ψ_{l1}, ..., $\psi_{lm} \in \mathcal{A}_0$ where $\psi_{lj} \to \phi_l$ is valid.

Define $ii(\phi_l) = i(\phi_l) \setminus \bigcup_j ii(\psi_{lj})$.

Delete ϕ_l from $\mathcal{A}'$ and update $\mathcal{A}_0$ as $\mathcal{A}_0 \cup \{\phi_l\}$ when $ii(\phi_l) \neq \{\}$.

Step 4: If $\mathcal{A}'$ is empty then redefine $\mathcal{A}$ as $\mathcal{A}_0$ and terminate the procedure otherwise go to step 3.

Further defining $ii(true) = \mathcal{W} \setminus \cup_j ii(\phi_j)$. If $ii(true) \neq \{\}$ then $ii(true)$ represents those possible worlds which make only formula $true$ true and $true$ is added into $\mathcal{A}$. We also define $ii(\perp) = \{\}$, so that function $ii : \mathcal{A} \to 2^{\mathcal{W}}$ is defined. Now we need to prove that ii is a basic incidence assignment. That is, for ϕ_l and $\phi_j \in \mathcal{A}$, we need to prove

$$ii(\phi_l) \cap ii(\phi_j) = \{\} \quad when \;\; \phi_l \neq \phi_j$$

Suppose that $ii(\phi_l) \cap ii(\phi_j) = W' \neq \{\}$, we have the following inference procedure.

$$\begin{aligned}
&w \in ii(\phi_l) \cap ii(\phi_j) \Longrightarrow \\
&w \in i(\phi_l) \;\; and \;\; w \in i(\phi_j) \Longrightarrow \\
&w \in i(\phi_l) \cap i(\phi_j) \Longrightarrow \\
&w \in i(\phi_l \wedge \phi_j) \Longrightarrow \\
&\exists \phi \neq \perp \;\; \wedge \;\; w \in i(\phi) \;\; and \;\; \phi = \phi_l \wedge \phi_j \Longrightarrow \\
&w \notin i(\phi_l) \setminus i(\phi) \;\; or \;\; w \notin i(\phi_j) \setminus i(\phi) \;\; as \;\; \phi_l \neq \phi_j \Longrightarrow \\
&w \notin ii(\phi_l) \cap ii(\phi_j)
\end{aligned}$$

Contradiction.

So the equation $ii(\phi_l) \cap ii(\phi_j) = \{\}$ holds for any two distinct elements ϕ_l and ϕ_j in $\mathcal{A}$. As we also have $ii(true) = \mathcal{W} \setminus \cup_j ii(\phi_j)$ and $i(false) = ii(false) = \{\}$, ii is a basic incidence assignment.

QED

Given a generalized incidence calculus theory with a set of axioms as $\mathcal{A}$, there is an unique basic incidence assignment ii on set $\mathcal{A}_0$ matching to the incidence function i on $\mathcal{A}$. However, different generalized incidence calculus theories may generate the same basic incidence assignment. That is, one basic incidence assignment matches to a family of generalized incidence calculus theories. All the generalized incidence calculus theories in the family produce the same bounds of incidences for any formula as the basic incidence assignment does.

3.1.3 Implementations of basic incidence assignments

A basic incidence function ii is separated from an incidence function i. So given a generalized incidence calculus theory, it should be possible to obtain the corresponding basic incidence assignment from i. This has been described in Algorithm A which was implemented using Sicstus Prolog.

For example, suppose that the incidences of axioms in a set

$$\mathcal{A} = \{a, b, a \wedge b, c, a \wedge c, c \wedge d\}$$

are

$$i(a) = \{1, 2, 3, 4\}$$

$$i(b) = \{1, 2, 3\}$$

$$i(a \wedge b) = \{1, 2, 3\}$$

$$i(c \wedge d) = \{5, 6, 7\}$$

$$i(c) = \{4, 5, 6, 7\}$$

$$i(a \wedge c) = \{4\}$$

then the corresponding basic incidence assignment is

$$ii(a \wedge c) = \{4\}$$

$$ii(a \wedge b) = \{1, 2, 3\}$$

$$ii(c \wedge d) = \{5, 6, 7\}$$

Using this basic incidence assignment, it is possible to obtain an incidence function i_1 on a set of axioms $\mathcal{A}' = \{a, a \wedge b, c, a \wedge c, c \wedge d\}$.

Although set $\mathcal{A}'$ is slightly different from the set of axioms $\mathcal{A}$ initially specified (axiom b is in $\mathcal{A}$ but not in $\mathcal{A}'$), generalized incidence calculus theory

$< \mathcal{W}, \mu, P, \mathcal{A}_1, i_1 >$ produces the same incidence bounds on all formulae as generalized incidence calculus theory $< \mathcal{W}, \mu, P, \mathcal{A}, i >$ does. Axiom b in $\mathcal{A}$ makes no more contributions than $a \wedge b$ does. So it is not necessary to include it into the set of axioms.

This algorithm has been tested on several examples. The program gives immediate response either for a large set of possible worlds, such as with 20 possible worlds, or for a large set of axioms, such as with 11 axioms. The summary of tested examples is listed in Table 3.1.

No. of Axioms in $\mathcal{A}$	No. of axioms with ii	runtime
7	5	1.309
11	7	1.509
11	9	1.999

Table 3.1 Test result of Algorithm A

3.2 Incidence Assignments

3.2.1 Assigning incidences to formulae in generalized incidence calculus

Given an incidence calculus theory, we can infer lower and upper bounds of probabilities on formulae. Incidence functions are crucial in the inference procedure. However, sometimes numerical assignments, particularly lower bounds of probabilities, are given on some formulae directly without defining any incidence calculus theories. We are interested in how to build incidence calculus theories in these cases. In this section, we show a way to recover incidence functions from lower bounds of probabilities in these circumstances.

Formally our problem can be described as: given an assignment of lower bound of probabilities on a set of axioms $\mathcal{A}$, our objective is to find an incidence function i, a set of possible worlds $\mathcal{W}$ and the discrete probability distribution μ on $\mathcal{W}$ which produces the lower bound of probabilities on $\mathcal{A}$.

In order to reach this goal, we will construct a function ii first and then form i.

In a similar way to that we described in the above section, a special set $\mathcal{A}_0$ can be constructed from $\mathcal{A}$ which satisfies the condition

$$\forall \phi \in \mathcal{A}_0, \forall \phi' \in \mathcal{A}, \phi' \not\models \phi, if \;\; \phi \neq \phi' \tag{3.5}$$

Suppose that an incidence function i and a basic incidence assignment ii associated with $\mathcal{A}$ are known, then $w_1 = ii(\phi_l)$ and $w_2 = ii(\phi_j)$ must be two disjoint subsets of an unknown $\mathcal{W}$ because of the property $ii(\phi_l) \cap ii(\phi_j) = \{\}$ when $\phi_l, \phi_j \in \mathcal{A}_0, \phi_l \neq \phi_j$.

The following algorithm gives the procedure for determining the incidence function i, its basic incidence assignment ii and the set of possible worlds with its probability distribution.

Algorithm B

Given $\mathcal{A}$ and an assignment of lower bound of probabilities p_* on $\mathcal{A}$, determine a basic incidence assignment and an incidence function.

Step 1: Assume that $\mathcal{A}_0$ is a subset of $\mathcal{A}$ as defined above in (3.5). If there are n elements in $\mathcal{A}_0$, then n elements in $\mathcal{W}$ can be defined from $\mathcal{A}_0$ and define $\mu(w_j) := p'_*(\phi_j)$, $p'_*(\phi_j) := p_*(\phi_j)$ for $j = 1, ..., n, \phi_i \in \mathcal{A}_0$. Further define $ii(\phi_j) = \{w_j\}$ and $\mathcal{A}' := \mathcal{A} \setminus \mathcal{A}_0$.

Step 2: Using the same method as we define $\mathcal{A}_0$ on page 51, we can choose a formula ψ from $\mathcal{A}'$ which satisfies the condition that $\forall \psi' \in \mathcal{A}'$, $\psi' \not\models \psi$ if $\psi' \neq \psi$[2].

---

[2]In this step, there may be more than one alternative formula qualifies this condition, so runing this algorithm each time may have different formula being chosen, but the final result remains the same.

Then define $p'_*(\psi) := p_*(\psi) - \Sigma_{\phi_j \in \mathcal{A}_0, \phi_j \models \psi} p'_*(\phi_j)$.

If $p'_*(\psi) > 0$ then add an element w_{n+1} to $\mathcal{W}$ and define

$$ii(\psi) = \{w_{n+1}\}$$
$$\mu(w_{n+1}) := p'_*(\psi)$$
$$\mathcal{A}_0 := \mathcal{A}_0 \cup \{\psi\}$$
$$\mathcal{A}' := \mathcal{A}' \setminus \{\psi\}$$
$$n := n + 1$$

If $p'_*(\psi) = 0$, define $ii(\psi) = \{\}$.

If $p'_*(\psi) < 0$, this assignment is not consistent, stop the procedure.

Repeat this step until $\mathcal{A}'$ is empty.

Step 3: Finally, if $\Sigma_j(p'_*(\phi_j)) < 1$, then add an element w_{n+1} to $\mathcal{W}$ and define

$$\mu(w_{n+1}) = 1 - \Sigma_j p'_*(\phi_j)$$
$$ii(true) = \{w_{n+1}\}$$

We also define $ii(false) = \{\}$.

Step 4: Eventually, the set of possible worlds is $\mathcal{W} = \{w_1, w_2, ..., w_{n+1}\}$ and redefine $\mathcal{A}$ as $\mathcal{A}_0$. The probability distribution is μ and $\Sigma_j \mu(w_j) = 1$ where $\phi_j \in \mathcal{A}$. Two functions ii and i are defined as $ii(\phi_j) = \{w_j\}$ and $i(\phi) = \cup_{\phi_j \models \phi} ii(\phi_j)$, $\phi_j \in \mathcal{A}$.

Proposition 5 *Given a set of axioms $\mathcal{A}$ which is closed under $\wedge$ and an assignment of lower bound of probabilities p_* on $\mathcal{A}$. Functions i and ii obtained after applying Algorithm B on $(\mathcal{A}, p_*)$ are an incidence function and a basic incidence assignment. The corresponding generalized incidence calculus theory $< \mathcal{W}, \mu, P, \mathcal{A}, i >$ will produce p_* on $\mathcal{A}$.*

PROOF

For any two formulae in $\mathcal{A}$, we have

$$ii(\phi) \cap ii(\psi) = \{\} \quad when \quad \phi \neq \psi$$

and

$$ii(true) = \{w_{n+1}\} = \mathcal{W} \setminus \cup_{\phi_j \in \mathcal{A}} ii(\phi_j)$$

$$ii(false) = \{\}$$

So ii is a basic incidence assignment. Therefore $i(\psi) = \cup_{\phi_j \models \psi} ii(\phi_j)$ is an incidence function based on Proposition 3.

The corresponding generalized incidence calculus theory is

$$< \mathcal{W}, \mu, P, \mathcal{A}, i >$$

For any $\psi \in \mathcal{A}$, we can calculate the lower bound of its probability, denoted as p_{i*} (in order to distinguish it from p_*), as follows.

$$\begin{aligned}
p_{i*}(\psi) &= \mu(i_*(\psi)) \\
&= \mu(\cup_{\phi_j \in \mathcal{A}, \phi_j \models \psi} i(\phi_j) \\
&= \mu(\cup_{\phi_j \in \mathcal{A}, \phi_j \models \psi} \cup_{\phi_{jl} \in \mathcal{A}, \phi_{jl} \models \phi_j} ii(\phi_{jl})) \\
&= \mu(\cup_{\phi_{jl} \in \mathcal{A}, \phi_{jl} \models \psi} ii(\phi_{jl})) \\
&= \mu(\cup_{\phi_{jl} \in \mathcal{A}, \phi_{jl} \models \psi, \phi_{jl} \neq \psi} ii(\phi_{jl})) + \mu(ii(\psi)) \\
&= \Sigma_{\phi_{jl} \in \mathcal{A}, \phi_{jl} \models \psi} \mu(ii(\phi_{jl})) + \mu(ii(\psi)) \\
&= \Sigma_{\phi_{jl} \in \mathcal{A}, \phi_{jl} \models \psi} p'_*(\phi_{jl}) + p'_*(\psi)) \\
&= p_*(\psi)
\end{aligned}$$

So this theory produces the same lower bounds of probabilities for those formulae in $\mathcal{A}$ as P_*.

QED

If there are N elements in $\mathcal{A}$ then there are at most $N + 1$ elements in $\mathcal{W}$.

This algorithm is entirely based on the result that $ii(\phi) \cap ii(\psi) = \{\}$ and is different from the methods used in incidence assignment introduced in the original incidence calculus in Section 2.6. In algorithm B, for a formula ϕ, we keep deleting those portions in $p_*(\phi)$ which can be carried by a formula ψ, where $\psi \models \phi$, until we obtain the last bit which must be carried by ϕ itself. This last portion will only be contributed by its basic incidence set. This algorithm is relatively fast on the sets tested so far.

3.2.2 An example of incidence assignments

In this section we use Example 2.4 to demonstrate algorithm B described above. The example is reconstructed from [Kyburg, 1991].

Example 3.1

Assume that we know the lower bound of a probability distribution on a set of axioms of formulae. We want to create a set of possible worlds and its probability distribution and to define an incidence function from the set of axioms to this set. The created set of possible worlds and the incidence function can, in turn, produce the lower bound of probability distribution on the set of axioms.

Suppose that we have P, $\mathcal{L}(P)$ and a set of axioms $\mathcal{A} = \{a, b, c, a \wedge b, a \wedge c, b \wedge c, a \wedge b \wedge c\}$ with the lower bound of a probability distribution as

$$p_*(a) = 0.760$$
$$p_*(b) = 0.640$$
$$p_*(c) = 0.480$$
$$p_*(a \wedge b) = 0.525$$
$$p_*(a \wedge c) = 0.350$$
$$p_*(b \wedge c) = 0.225$$
$$p_*(a \wedge b \wedge c) = 0.165$$

The set $\mathcal{A}$ is closed under the operator $\wedge$. Following Algorithm B, an incidence function is defined by the following steps.

Step 1. The set $\mathcal{A}_0$ is $\{a \wedge b \wedge c\}$ which contains the smallest formula in $\mathcal{A}$. So we know that at least one possible world, w_1, validating formula $a \wedge b \wedge c$ and $\mu(w_1) = 0.165$. We also have

$$\begin{aligned}
&p'_*(a \wedge b \wedge c) = p_*(a \wedge b \wedge c) = 0.165 \\
&ii(a \wedge b \wedge c) = \{w_1\} \\
&\mathcal{A}' := \mathcal{A} \setminus \mathcal{A}_0 \\
&n := 1
\end{aligned}$$

Step 2. Choose a formula $a \wedge b$ from $\mathcal{A}'$. Because only formula $a \wedge b \wedge c$ has the property that $a \wedge b \wedge c \to a \wedge b$, we have

$$p'_*(a \wedge b) := p_*(a \wedge b) - p'_*(a \wedge b \wedge c) = 0.525 - 0.165 = 0.36$$

Because $p'_*(a \wedge b) > 0$, we define

$$\begin{aligned}
&ii(a \wedge b) = \{w_2\} \\
&\mu(w_2) = p'_*(a \wedge b) \\
&\mathcal{A}_0 := \mathcal{A}_0 \cup \{a \wedge b\} \\
&\mathcal{A}' := \mathcal{A}' \setminus \{a \wedge b\} \\
&n := n + 1
\end{aligned}$$

Repeat this step for all the remaining elements in $\mathcal{A}'$, we get

$$\begin{aligned}
&ii(a \wedge c) = \{w_3\} \qquad && \mu(w_3) = 0.185 \\
&ii(b \wedge c) = \{w_4\} && \mu(w_4) = 0.06 \\
&ii(a) = \{w_5\} && \mu(w_5) = 0.05 \\
&ii(b) = \{w_6\} && \mu(w_6) = 0.055 \\
&ii(c) = \{w_7\} && \mu(w_7) = 0.07
\end{aligned}$$

Step 3. As $\Sigma_j p'_*(\phi_j) = 0.165 + 0.36 + 0.185 + 0.06 + 0.05 + 0.055 + 0.07 = 0.945$, we define $\mu(ii(true)) = 1 - \Sigma_j \mu(ii(\phi_j)) = 1 - \mu(\{w_1, ..., w_7\}) = 0.055$.

Step 4. Eventually, let $ii(true) = \{w_8\}$ and $\mathcal{A}$ be $\mathcal{A}_0$, then we obtain $\mathcal{W} = \{w_1, ..., w_8\}$ with probability distribution μ on it. The incidence function derived from $i(\phi) = \cup_{\phi_j \models \phi} ii(\phi_j)$ is as shown below.

$$i(a \wedge b \wedge c) = \{w_1\}$$

$$i(a \wedge b) = \{w_1, w_2\}$$

$$i(a \wedge c) = \{w_1, w_3\}$$

$$i(b \wedge c) = \{w_1, w_4\}$$

$$i(a) = \{w_1, w_2, w_3, w_4\}$$

$$i(b) = \{w_1, w_2, w_4, w_6\}$$

$$i(c) = \{w_1, w_3, w_4, w_7\}$$

For any other formula ψ, if $\mu(ii(\psi)) = 0$, we explain this in two ways: there is no possible world making this formula true or the probability of the subset which makes ψ true is 0. In any case, it doesn't matter whether we add $ii(\psi)$ to the whole set of possible worlds or not. The incidence calculus theory which can produce the lower bound of a probability distribution p_* on $\mathcal{A}$ is $< \mathcal{W}, \mu, P, \mathcal{A}, i >$ as well as $i^*(\phi)$ and $p^*(\phi)$.

For any formula $\phi \in \mathcal{L}(P) \setminus \mathcal{A}$, we can calculate both $i_*(\phi)$ and $p_*(\phi)$.

When we apply Algorithm B, there may be more than one formula satisfying the conditions in Step 2, but the order of choosing these formulae has no effect on the final result. For this example, after we choose $a \wedge b \wedge c$ and come to Step 2, it doesn't matter to choose $a \wedge b$ first or $a \wedge c$ first. The final result remains the same.

Theorem 1 *Applying Algorithm B on $(\mathcal{A}, \mu_*)$ produces the same result regardless the order of selecting formulae in Step 2.*

PROOF

Assume that after Step 1, the set of $\mathcal{A}_0$ is $\mathcal{A}_0 = \{\phi_1, ..., \phi_k\}$ and that at Step 2, there are two formulae ψ_1, ψ_2 satisfying the condition specified in that step.

In Step 1, for every $\phi_j \in \mathcal{A}_0$, we have

$$p'_*(\phi_j) := p_*(\phi_j)$$

$$\mu(w_j) := p'_*(\phi_j)$$

$$ii(\phi_j) = \{w_j\}$$

Assume that we first choose ψ_1 in Step 2, then we have

$$p'_*(\psi_1) := p_*(\psi_1) - \Sigma_{\phi_j \in \mathcal{A}_0, \psi_j \models \psi_1} p'_*(\phi_j)$$

Now we choose ψ_2, we obtain

$p'_*(\psi_2) := p_*(\psi_1) - \Sigma_{\phi_j \in \mathcal{A}_0 \cup \psi_1, \psi_j \models \psi_1} p'_*(\phi_j)$ (because ψ_1 is a smallest formula now)

$:= p_*(\psi_1) - \Sigma_{\phi_j \in \mathcal{A}_0, \psi_j \models \psi_1} p'_*(\phi_j)$ (because $\psi_1 \not\models \psi_2$)

which indicates that adding ψ_1 into set $\mathcal{A}_0$ has no effect on the outcome of ψ_2.

In the same way we could prove that choosing ψ_2 first then ψ_1 gives exactly the same result as above. That is, p'_* on set $\mathcal{A}_0 \cup \{\psi_1, \psi_2\}$ is the same no matter which formula is chosen first. Similarly, we could prove the theorem for any set of formulae $\{\psi_1, ..., \psi_n\}$ in Step 2.

QED

The Prolog execution time for this example is 0.759 (seconds) in the axiom order $\{a \wedge b \wedge c, a \wedge c, b \wedge c, a, b, a \wedge b, c\}$. If the order of the axioms is reversed as $\{a, b, c, b \wedge c, a \wedge c, a \wedge b, a \wedge b \wedge c\}$, there is not much difference. The runtime for the latter case is 1.189 (seconds). The algorithm creates a set of possible worlds with 8 elements.

3.2.3 Generating multiple consistent incidence assignments

In contrast to the methods for incidence assignments in the original incidence calculus, the new incidence assignment algorithm in generalized incidence calculus doesn't assume the size of a set of possible worlds in advance. The disadvantage of the methods in the original incidence calculus is that it takes too long a time to find a consistent incidence assignment. At the moment, the new algorithm in generalized incidence calculus constructs only one basic incidence assignment. In this section, we will examine how to generate multiple consistent incidence assignments from this basic assignment.

In fact, what is created in this algorithm can be regarded as a *model* or an *abstract*, denoted as S, for an unspecified set of possible worlds. There are three different possibilities to generate multiple consistent incidence assignments from S. Before giving these three situations, I need to introduce the definition of probability spaces first.

Definition 3.4: *Probability spaces*

A probability space (X, χ, μ) *has:*

X*: a* sample space *usually containing all the possible worlds;*

χ*: a* σ*-algebra containing some subsets of* X*, which is defined as containing* X *and closed under complementation and countable union.*

μ*: a probability measure* $\mu : \chi \to [0, 1]$ *with the following features:*

P1. $\mu(X_i) \geq 0$ *for all* $X_i \in \chi$*;*

P2. $\mu(X) = 1$*;*

P3. $\mu(\cup_{j=1}^{\infty} X_j) = \Sigma_{j=1}^{\infty} \mu(X_j)$*, if the* X_j*'s are pairwise disjoint members of* χ*.*

A subset χ' of χ is called a *basis* of χ if it contains non-empty and disjoint elements, and if χ consists precisely of countable unions of members of χ'. For any finite χ there is a unique basis χ' of χ and it follows that

$$\Sigma_{X_i \in \mathcal{X}'} \mu(X_i) = 1$$

In the following, we only consider finite probability spaces.

Situation 1 when we know nothing about the set of possible worlds, we could either take this model S as the set or take it as the basis of a set $\mathcal{W}$ which is going to be defined. This $\mathcal{W}$ is allowed to contain as many possible worlds as necessary. The elements in this set are then divided into $|S|$ groups. The jth group W_j matches to an element s_j in the basis. For the elements w_{jl} in W_j, we define $\mu'(w_{jl}) = \mu(s_j)/|W_j|$. For a formula ϕ_j in $\mathcal{A}$, if $i(\phi_j) = \{s_j\}$ then we define $i(\phi_j) = W_j$ when extending the incidence function i from S to $\mathcal{W}$.

For instance, in Example 3.1, it is possible to assume that the set of possible worlds has 8 elements, or assume that the set contains 80 (or any other number) possible worlds. For simplicity, we assume that these 80 possible worlds are divided into 8 groups and each group has 10 possible worlds. Suppose that the first group, containing the first 10 elements, matches to the first possible world w_1 in the model, then the probability for each element in the group is 0.165/10. Similarly, it is possible to calculate the probability for every other element.

Situation 2 when we know the set of possible worlds $\mathcal{W}$ and the probability distribution μ' on the set, we take S as the basis of $\mathcal{W}$. In this case, it is assumed that for each $s_j \in S$, there exists at least one subset W_j of $\mathcal{W}$ which guarantees $\mu(s_j) = \Sigma_{w_{jl}} \mu'(w_{jl})$ for $w_{jl} \in W_j$. Otherwise, the given lower bound of a probability distribution could not be generated from this set of possible worlds.

As there may be more than one method to divide the set $\mathcal{W}$ to form the basis S, it is possible to generate more than one consistent incidence assignment.

Situation 3 when we know the set of possible worlds and some relation among possible worlds but not the probability distribution, we take S as the basis of

$\mathcal{W}$. This basis can certainly be used to generate many consistent incidence assignments. This is particularly useful in constraint satisfaction problems. For example, if it is known that the possible worlds s_j and s_l are very likely to support same formulae, then w_j and w_l should possibly be in the same group. In this case, elements in $\mathcal{W}$ are grouped mainly based on their properties.

In summary, the method of model creation for incidence assignments is more flexible and can be used to generate multiple consistent assignments when required.

In Chapter 2, we mentioned that in [McLean, 1992], [McLean, Bundy and Liu 1994] McLean develops two algorithms for incidence assignments. The idea in McLean's work is that given a numerical assignment on a set of axioms $\mathcal{A}$, a set of possible worlds $\mathcal{W}$ is fixed and then an algorithm is used to divide elements of $\mathcal{W}$ into different groups and then each group is mapped onto each axiom in $\mathcal{A}$. Executing an algorithm each time may give different grouping result. These different results give different but consistent incidence assignments based on an initial numerical assignment. However, in our incidence assignment approach, we try to find a 'model' or a 'basis' first, and then we generate multiple consistent assignments based on this model. A model is usually easy to be constructed, so the run time is much faster than McLean's algorithms (see Example 3.1). The actual methods about how to generate different incidence assignments are largely dependent on real situations as we specified above. Therefore, our algorithm is more flexible.

3.2.4 Estimating lower bounds of probabilities

A critical condition is placed on the new incidence assignment algorithm proposed above, that is a set of axioms must be closed under the operator $\wedge$. When this condition doesn't hold, we can only apply Algorithm B under the assumption that the lower bounds of probabilities of those axioms, for which that are not specified initially, are 0. However, if a piece of evidence shows that such assumption does not exist, for instance, if $p_*(a \wedge b \wedge c) = 0.2$ then assuming $p_*(a \wedge b) = 0$ is unrealistic, we then have to estimate the lower pounds for those axioms before it is possible for us to apply the algorithm.

In this section, we briefly discuss how to estimate the lower bounds of probabilities for those axioms which are not defined initially.

Considering Example 2.3, an initial numerical assignment on a set of axioms $(a, b, c, a \wedge b \wedge c, a \wedge c, b \wedge c)$ is

$$
\begin{aligned}
p_*(a) &= 0.8 \\
p_*(b) &= 0.6 \\
p_*(c) &= 0.5 \\
p_*(a \wedge c) &= 0.4 \\
p_*(b \wedge c) &= 0.2 \\
p_*(a \wedge b \wedge c) &= 0.2
\end{aligned}
$$

$\mathcal{A}$ is not closed under $\wedge$, as $a \wedge b$ is not in $\mathcal{A}$ and the lower bound of probability on axiom $a \wedge b$ is not known. Algorithm B cannot be used immediately before $p_*(a \wedge b)$ is supplied. From the assignment on other axioms, we can calculate that $p_*(a \wedge b)$ might be between 0.2 to 0.6.

Case 1 We assign $p_*(a \wedge b) = 0.4$ and then execute Algorithm B on $\mathcal{A} \cup \{a \wedge b\}$, the result is wrong as we have $\Sigma_w \mu'(w) > 1$ for the constructed set of possible worlds. So $p_*(a \wedge b) \leq 0.4$ is impossible.

Case 2 We assign $p_*(a \wedge b) = 0.5$ and then execute Algorithm B on $\mathcal{A} \cup \{a \wedge b\}$, the result is quite good as shown below, because it gives a much lower numerical distribution on *true*. This means that there is almost no ignorance.

$$
\begin{aligned}
\mu(w_6) &= 0.1 & \qquad ii(a) &= \{w_6\} \\
\mu(w_5) &= 0.1 & ii(b) &= \{w_5\} \\
\mu(w_4) &= 0.1 & ii(c) &= \{w_4\} \\
\mu(w_3) &= 0.3 & ii(a \wedge b) &= \{w_3\} \\
\mu(w_2) &= 0.2 & ii(a \wedge c) &= \{w_2\} \\
\mu(w_1) &= 0.2 & ii(a \wedge b \wedge c) &= \{w_1\}
\end{aligned}
$$

$$\mu(w_7) = 0.1^{-17} \qquad ii(true) = \{w_7\}$$

Case 3 We assign $p_*(a \wedge b) = 0.6$ and then execute Algorithm B on $\mathcal{A} \cup \{a \wedge b\}$, the result is not wrong but not good. Because we have

$$ii(a) = \{w_6\}, \mu(w_6) = 0.1^{-15}$$

and

$$ii(true) = \{w_7\}, \mu(w_7) = 0.1$$

This means that there is a large amount of ignorance (0.1), but $p_*(a)$ carries almost nothing significant.

In general, the range of the lower bound of probability on an axiom ϕ can be roughly estimated as

$$min\{p_*(\psi) \mid \phi \rightarrow \psi\} \geq p_*(\phi) \geq max\{p_*(\psi) \mid \psi \rightarrow \phi\} \tag{3.6}$$

This estimation gives looser bounds than that in the real case. For instance, if we use this mathematical formula to guess the bounds for formula $a \wedge b$, we have $0.6 \geq p_*(a \wedge b) \geq 0.2$. However, the lower bounds of $a \wedge b$ should somehow be bigger than 0.4.

3.2.5 Implementation of incidence assignments

Converting an incidence function from a numerical assignment is one of the most important issues in applying incidence calculus. The algorithm of deriving incidence functions from numerical assignments, through basic incidence assignments, shows the possibility of constructing incidence functions and then generalized incidence calculus theories in most circumstances.

When a set of axioms is not closed under $\wedge$, most of situations, it is necessary to estimate the lower bounds of probabilities on those axioms which are not given initially. This topic has been briefly discussed in this section and it shows that the best estimation values of lower bounds on axioms can be found based on the feedback of generating incidence functions.

Algorithm B has been implemented using Sicstus Prolog. This program is used and tested both in recovering incidence functions when a set of axioms is closed under $\wedge$ and in evaluating an estimation when a set of axioms is not closed under $\wedge$. As far as for the examples we tested, the program gives out the result immediately, see Table 3.2.

No. of axioms	runtime(seconds)
7	0.759
12	10.678
15	12.08

Table 3.2 Test result of Algorithm B

3.3 Combining Evidence

What we have considered previously is limited to only one generalized incidence calculus theory. If a generalized incidence calculus carries the message provided by one source of evidence, then as new evidence comes in, more and more generalized incidence calculus theories will be constructed. In this circumstances, it is necessary to pull out the common effects of all the evidence and represent it using a single generalized incidence calculus theory. This procedure is normally called *the combination of evidence.* In this section, we discuss our approach for combining multiple pieces of evidence.

3.3.1 Effects of new information

Suppose we have already had an incidence calculus theory, $< \mathcal{W}, \mu, P, \mathcal{A}_1, i_1 >$, for a given problem, if a new piece of information regarding this problem is known, then it may have one of the following effects.

Effect 1: This source of information gives a new probability distribution on the set of possible worlds to replace the old probability distribution, then a new generalized incidence calculus theory will be created to substitute the old one and further inference will be made upon the new generalized incidence calculus theory only.

This can be seen through an example in [Fagin and Halpern, 1989a]. The example is stated as: *Suppose that we have 100 agents, each holding a lottery ticket, numbered 00 to 99. Suppose that agent a_1 holds ticket number 17. Assume that the lottery is fair, so, a priori, the probability that a given agent will win is 1/100. We are then told that the first digit of the winning ticket is 1, the problem is to determine the probability that agent a_1 will win.*

Considering this problem in incidence calculus, we can first form a generalized incidence calculus theory as $< \mathcal{W}, \mu_1, P, \mathcal{A}_1, i_1 >$ where $\mathcal{W} = \{00, ..., 99\}$, $\mu(w) = 1/100$, $P = \{a_1, ..., a_{100}\}$, $\mathcal{A}_1 = P$ and $i_1(a_i) = \{w\}$ when a_i's number is w. Here a_i stands for the proposition a_i *will win*. When we are told that *the first digit of the winning ticket is 1 later*, the probability distribution on $\mathcal{W}$ will be changed as $\mu_2(w) = 1/10$ when w is in $\{10, ..., 19\}$ and $\mu_2(w) = 0$ otherwise. Therefore the new generalized incidence calculus theory is $< \mathcal{W}, \mu_2, P, \mathcal{A}_1, i_1 >$. In this case, because the old probability distribution is replaced by the new one, the new generalized incidence calculus theory disables the effect of the old one. It is then easy to know that the probability that a_1 will win is $1/10$.

It would be interesting to notice that this new piece of evidence doesn't change the supporting relations between set $\mathcal{L}(P)$ and $\mathcal{W}$. It only changes the μ on $\mathcal{W}$. It results in one generalized incidence calculus theory being evoked only. This is entirely different from the situations below.

Effect 2: This source of evidence specifies a new generalized incidence function from sets $\mathcal{L}(P)$ to $\mathcal{W}$ without changing the set of possible worlds and its probability distribution. Then a new generalized incidence calculus theory is formed. Both the new and old incidence calculus theories have impacts on $\mathcal{L}(P)$. So it is necessary to consider how to obtain their joint impact.

Considering the weather example again, we have, first of all, a generalized incidence calculus theory as $< \mathcal{W}, \mu, P, \mathcal{A}_1, i_1 >$. If a new piece of information tells us that $i_2(rainy) = \{fri, sat, sun\}$ and $i_2(windy) = \{wed, fri\}$, then another generalized incidence calculus theory $< \mathcal{W}, \mu, P, \mathcal{A}_2, i_2 >$ is formed which gives an alternative interrelation among the elements of the two sets (without changing the probability distribution on set $\mathcal{W}$). We need to consider the joint impact of both the old and new information on the formula set. That is we must combine the two pieces of information. For a particular formula in $\mathcal{L}(P)$, if we have $i_1(\phi) = \mathcal{W}_1$ and $i_2(\phi) = \mathcal{W}_2$ from the two sources respectively, then the common impact of the two sources will produce $i_1 \odot i_2(\phi) = \mathcal{W}_1 \cap \mathcal{W}_2$. More generally if $i_1(\phi) = \mathcal{W}_1$ and $i_2(\psi) = \mathcal{W}_2$ then $i_1 \odot i_2(\phi \wedge \psi) = \mathcal{W}_1 \cap \mathcal{W}_2$. This is the basic idea of giving a combination mechanism in incidence calculus which will be further discussed in greater detail in the next section. Here $\odot$ indicates that a kind of combination mechanism, going to be defined later, is applied on i_1 and i_2.

Effect 3 This source of information defines a new incidence calculus theory different from the above two cases in the sense that the new information gives different sets of possible worlds and its probability distribution. Like situation 2) both the new and old generalized incidence calculus theories will make impacts on $\mathcal{L}(P)$, so it is necessary to consider how to obtain their joint impact. If the old incidence calculus theory is $< \mathcal{W}_1, \mu_1, P, \mathcal{A}_1, i_1 >$ and the new one is $< \mathcal{W}_2, \mu_2, P, \mathcal{A}_2, i_2 >$, then we form two probability spaces $(\mathcal{W}_1, \mathcal{W}_1, \mu_1)$ and $(\mathcal{W}_2, \mathcal{W}_2, \mu_2)$. However in contrast to situation 2), these two probability spaces are not the same.

In summary, apart from some very simple cases shown in situation 1), usually when the new pieces of information are obtained it is necessary to combine them with the existing information. The corresponding combination mechanism is essential to play such a role in producing the final effect of all the information. Currently incidence calculus doesn't have such a facility to cope with this problem. So it is important to propose a combination mechanism in incidence calculus to combine multiple pieces of evidence.

3.3.2 The Combination Rule in incidence calculus

Definition 3.5: *Combination Rule*

Suppose there are two generalized incidence calculus theories $< \mathcal{W}, \mu, P, \mathcal{A}_1, i_1 >$*,* $< \mathcal{W}, \mu, P, \mathcal{A}_2, i_2 >$*, then the joint impact of information carried by the two theories is represented by a quintuple:* $< \mathcal{W} \setminus \mathcal{W}_0, \mu', P, \mathcal{A}, i >$ *where*

$$\mathcal{W}_0 = \bigcup\{i_1(\phi) \cap i_2(\psi) \mid (\phi \wedge \psi = \perp), \phi \in \mathcal{A}_1, \psi \in \mathcal{A}_2\}$$

$$\mathcal{A} = \{\varphi \mid \varphi = \phi \wedge \psi \mid \phi \in \mathcal{A}_1, \psi \in \mathcal{A}_2, \varphi \neq \perp\}$$

$$i(\varphi) = \bigcup\{i_1(\phi) \cap i_2(\psi) \mid (\phi \wedge \psi \models \varphi), \varphi \in \mathcal{A}, \phi \in \mathcal{A}_1, \psi \in \mathcal{A}_2, \phi \wedge \psi \neq \perp\}$$

for any $w \in \mathcal{W} \setminus \mathcal{W}_0$

$$\mu'(w) = \frac{\mu(w)}{1 - \Sigma_{w' \in \mathcal{W}_0} \mu(w')}$$

and let

$$i(false) = \{\} \qquad i(true) = \mathcal{W} \setminus \mathcal{W}_0$$

Where $\perp$ means false and $\phi \wedge \psi \neq \perp$ means $\phi \wedge \psi$ is not contradictory.

$\mathcal{W}_0$ is a subset of $\mathcal{W}$ reflecting the conflict of two pieces of information and the conflict weight is $\Sigma_{w' \in \mathcal{W}_0} \mu(w')$. If the conflict weight is 1 then these two pieces of information are completely contradictory with each other and they cannot be combined using the rule.

When $\mathcal{A} = \{\}$, these two observations are irrelevant to each other and their combined result tells us nothing.

When $\mathcal{A} \neq \{\}, \forall \phi \in \mathcal{A}$, $i(\phi) = \{\}$, these two observations repel each other. In other words, only one of them can be held at each time.

Theorem 2 *Given two generalized incidence calculus theories* $< \mathcal{W}, \mu, P, \mathcal{A}_1, i_1 >$, *and* $< \mathcal{W}, \mu, P, \mathcal{A}_2, i_2 >$, *the combined structure* $< \mathcal{W} \setminus \mathcal{W}_0, \mu', P, \mathcal{A}, i >$ *is a generalized incidence calculus theory.*

PROOF

According to Definition 3.1, we only need to prove that $\mathcal{A}$ is closed under $\wedge$.

Assume that ϕ, ψ are two distinct formulae ($\phi \not\equiv \psi$) in $\mathcal{A}$ and they are derived from formulae in $\mathcal{A}_1$ and $\mathcal{A}_2$ as

$$\phi = \phi_1 \wedge \psi_1$$

$$\psi = \phi_2 \wedge \psi_2$$

where $\phi_1, \phi_2 \in \mathcal{A}_1$ and $\psi_1, \psi_2 \in \mathcal{A}_2$, so $\phi \wedge \psi = (\phi_1 \wedge \phi_2) \wedge (\psi_1 \wedge \psi_2)$. There are three possibilities whether $\phi \wedge \psi$ will be in $\mathcal{A}$.

(i) if $\phi_1 \wedge \phi_2 = \perp$ or $\psi_1 \wedge \psi_2 = \perp$, then $\phi \wedge \psi = \perp$. $\phi \wedge \psi$ doesn't need to be in $\mathcal{A}$.

(ii) if $i_1(\phi_1 \wedge \phi_2) = \{\}$ or $i_2(\psi_1 \wedge \psi_2) = \{\}$, then $i(\phi \wedge \psi) = \{\}$. $\phi \wedge \psi$ doesn't need to be in $\mathcal{A}$.

(iii) Otherwise, $\phi_1 \wedge \phi_2 \in \mathcal{A}_1$ and $\psi_1 \wedge \psi_2 \in \mathcal{A}_2$, so $(\phi_1 \wedge \phi_2) \wedge (\psi_1 \wedge \psi_2) = (\phi_1 \wedge \psi_1) \wedge (\phi_2 \wedge \psi_2)$ is in the combined set $\mathcal{A}$. That is, $\phi \wedge \psi$ is in $\mathcal{A}$.

To summarize these three situations, we have that $\mathcal{A}$ is closed under $\wedge$. Next we prove that $i(\phi \wedge \psi) = i(\phi) \cap i(\psi)$ when $\phi, \psi, \phi \wedge \psi$ are all in $\mathcal{A}$. Suppose that $i(\varphi_1 \wedge \varphi_2) \neq \{\}$, for any $w \in \mathcal{W} \setminus \mathcal{W}_0$, if $w \in i(\varphi_1 \wedge \varphi_2)$, then we have

$$\begin{aligned}
& w \in i(\varphi_1 \wedge \varphi_2) \\
& \Longrightarrow \exists \varphi_0(\varphi_0 \models \varphi_1 \wedge \varphi_2) \wedge (w \in ii(\varphi_0)) \\
& \Longrightarrow \exists \varphi_0(\varphi_0 \models \varphi_1) \wedge (\varphi_0 \models \varphi_2) \wedge (w \in ii(\varphi_0)) \\
& \Longrightarrow \exists \varphi_0(i(\varphi_0) \subseteq i(\varphi_1)) \wedge (i(\varphi_0) \subseteq i(\varphi_2)) \wedge (w \in ii(\varphi_0)) \\
& \Longrightarrow \exists \varphi_0(i(\varphi_0) \subseteq i(\varphi_1) \cap i(\varphi_2)) \wedge (w \in ii(\varphi_0)) \\
& \Longrightarrow w \in i(\varphi_1) \cap i(\varphi_2)
\end{aligned}$$

The other way around,

$$\begin{aligned}
& w \in i(\varphi_1) \cap i(\varphi_2) \\
& \Longrightarrow w \in i(\varphi_1) \wedge w \in i(\varphi_2) \\
& \Longrightarrow (\exists \varphi_0(\varphi_0 \models \varphi_1) \wedge w \in ii(\varphi_0)) \wedge (\exists \varphi_0'(\varphi_0' \models \varphi_2) w \in ii(\varphi_0')) \\
& \Longrightarrow \exists \varphi_0(\varphi_0 \models \varphi_1) \wedge (\varphi_0 \models \varphi_2) \wedge (w \in ii(\varphi_0)) \\
& \qquad\qquad\qquad \text{(as } \varphi_0 \text{ and } \varphi_0' \text{ must be equivalent)} \\
& \Longrightarrow \exists \varphi_0(\varphi_0 \models \varphi_1 \wedge \varphi_2) \wedge (w \in ii(\varphi_0)) \\
& \Longrightarrow \exists \varphi_0(i(\varphi_0) \subseteq i(\varphi_1 \wedge \varphi_2) \wedge (w \in ii(\varphi_0)) \\
& \Longrightarrow w \in i(\varphi_1 \wedge \varphi_2)
\end{aligned}$$

So

$$i(\varphi_1 \wedge \varphi_2) = i(\varphi_1) \cap i(\varphi_2)$$

As we have defined that $i(true) = \mathcal{W} \setminus \mathcal{W}_0$ and $i(false) = \{\}$, function i is a generalized incidence function. It is also easy to prove that $\Sigma_{w \in \mathcal{W} \setminus \mathcal{W}_0} \mu'(w) = 1$. So $< \mathcal{W} \setminus \mathcal{W}_0, \mu', P, \mathcal{A}, i >$ is a generalized incidence calculus theory.

QED

The explanation of this combination rule is that if observation X says that $\mathcal{W}_{1i}$ makes statement ϕ true, and observation Y says that $\mathcal{W}_{2j}$ makes statement ψ true, then $\mathcal{W}_{1i} \cap \mathcal{W}_{2j}$ should make statement $(\phi \wedge \psi)$ true when we know that both X and Y hold.

3.3.3 DS-independent information

The crucial issue in applying the rule to two generalized incidence calculus theories is that these two theories are based on the same set of possible worlds, but possibly based on different sets of axioms and incidence functions. The combination procedure unifies two sets of axioms into one set and two incidence functions into one incidence function. In this way, generalized incidence calculus is expected to be used to combine dependent evidence directly.

In general, the relations between two generalized incidence calculus theories (provided by two pieces of evidence) can be divided into the following three categories.

1). The two sets of possible worlds in the two generalized incidence calculus theories are the same. In this case, the Combination Rule above is applied to combine the two generalized incidence calculus theories.

2). The two sets of possible worlds in the two generalized incidence calculus theories are different and they are DS-independent[3]. In this case, it is possible to transform the two generalized incidence calculus theories into new forms so that two new generalized incidence calculus theories are based on the same set. Then the Combination Rule is applied on them. This is described in Theorem 2 below.

3). The two sets of possible worlds in the two generalized incidence calculus theories are different but not DS-independent. At the moment, we don't have a framework to deal with this in general. It has to be done individually. For a case in this category, if it is possible to find a common set of possible worlds in some way to replace the two existing sets of possible worlds, then the Combination Rule is applicable. However, when it is not possible to find a common set of possible worlds to replace the two existing sets of possible worlds, generalized incidence calculus cannot cope with the case. Example 3.2 below demonstrates this situation.

---

[3]See definition 3.6 below.

It needs to be pointed out that if two sets of possible worlds in a case are different but they are both derived from a well-defined set[4], this case is put into the first category as shown in Example 5.4.

As cases in category 2 can be transformed into cases in category 1, category 2 is regarded as an extension of category 1.

Definition 3.6: *DS-independence*

Two probability spaces, (X_1, χ_1, μ_1) and (X_2, χ_2, μ_2), are said to be DS-independent, if they satisfy the conditions

$$\mu_{\chi}(C_i \mid D_j) = \mu_1(C_i) \quad \mu_{\chi}(D_j \mid C_i) = \mu_2(D_j)$$

for all subsets C_i and D_j, where $\chi'_1 = \{C_1, ..., C_n\}$ and $\chi'_2 = \{D_1, ..., D_m\}$ are bases for χ_1 and χ_2 respectively. μ_{χ} is the (a priori) probability measure on χ which is the σ-algebra of the joint space X. Spaces X_1 and X_2 are constructed from the joint space X. If the joint space is $X_1 \otimes X_2$, then $\mu_{\chi}(C_i)$ is an abbreviation for $\mu_{\chi}(\{(C_i, D_j) \mid 1 \leq j \leq m\})$.

It is easy to see that if two probability spaces are DS-independent, then they must be probabilistically independent. If two probability spaces are DS-independent, then their common probability space is their set product, that is (X, χ, μ) is $(X_1 \otimes X_2, \chi_1 \otimes \chi_2, \mu_1 \otimes \mu_2)$. Based on this, we have the following theorem.

Theorem 3 *Suppose we have two generalized incidence calculus theories, $< \mathcal{W}_1, \mu_1, P, \mathcal{A}_1, i_1 >$ and $< \mathcal{W}_2, \mu_2, P, \mathcal{A}_2, i_2 >$, where $(\mathcal{W}_1, \mathcal{W}_1, \mu_1)$ and $(\mathcal{W}_2, \mathcal{W}_2, \mu_2)$ are DS-independent. Applying the Combination Rule to them we get*

$$< \mathcal{W}_3, \mu_3, P, \mathcal{A}_3, i_3 >$$

---

[4]See Example 5.4. Suppose that S is a set with probability distribution p, S_1 and S_2 are subsets of S and p_1 on S_1 and p_2 on S_2 are defined from p through statistical methods, then S_1 and S_2 are said to be derived from well-defined set S.

which is a generalized incidence calculus theory, where

$$\mathcal{W}_0 = \bigcup\{i_1(\phi) \otimes i_2(\psi) \mid (\phi \wedge \psi = \perp), \phi \in \mathcal{A}_1, \psi \in \mathcal{A}_2\}$$

$$\mathcal{W}_3 = \mathcal{W}_1 \otimes \mathcal{W}_2 \setminus \mathcal{W}_0$$

$$\mathcal{A}_3 = \{\varphi \mid \varphi = \phi \wedge \psi, \phi \in \mathcal{A}_1, \psi \in \mathcal{A}_2, \ \varphi \neq \perp\}$$

$$i_3(\varphi) = \bigcup\{i_1(\phi) \otimes i_2(\psi) \mid (\phi \wedge \psi \models \varphi), \phi \wedge \psi \neq \perp\}$$

the new probability distribution on $\mathcal{W}_3$ is

$$\mu_3(< w_{1l}, w_{2j} >) = \frac{\mu_1(w_{1l})\mu_2(w_{2j})}{1 - \Sigma_{< w_{1t}, w_{2m} > \in \mathcal{W}_0}\mu_1(w_{1t})\mu_2(w_{2m})}$$

Where $w_{1l}, w_{1t} \in \mathcal{W}_1$ and $w_{2j}, w_{2m} \in \mathcal{W}_2$

$\mathcal{W}_0$ is a subset of $\mathcal{W}_1 \otimes \mathcal{W}_2$ which supports contradictory.

For any formula φ in $\mathcal{L}(P)$, our belief in φ is

$$p_*(\varphi) = \Sigma_{w \in i_3(\varphi)}\mu_3(w)$$

PROOF

For two generalized incidence calculus theories $< \mathcal{W}_1, \mu_1, P, \mathcal{A}_1, i_1 >$ and $< \mathcal{W}_2, \mu_2, P, \mathcal{A}_2, i_2 >$, when their probability spaces $(\mathcal{W}_1, \mathcal{W}_1, \mu_1)$ and $(\mathcal{W}_2, \mathcal{W}_2, \mu_2)$ are DS-independent, then it is possible to generate incidence functions from $\mathcal{W}_1$ and $\mathcal{W}_2$ to P as two new incidence functions from $\mathcal{W}_1 \otimes \mathcal{W}_2$ to P. So two new generalized incidence calculus theories can be constructed from them as:

$$< \mathcal{W}_3, \mu_3, P, \mathcal{A}_1, i'_1 >$$

and

$$< \mathcal{W}_3, \mu_3, P, \mathcal{A}_2, i'_2 >$$

where

$$\mathcal{W}_0 = \bigcup_{\phi \wedge \psi \models \perp} i_1(\phi) \otimes i_2(\psi)$$

$$\mathcal{W}_3 = \mathcal{W}_1 \otimes \mathcal{W}_2 \setminus \mathcal{W}_0$$

$$i'_1(\phi) = (i_1(\phi) \otimes \mathcal{W}_2) \setminus \mathcal{W}_0 \qquad where \;\; \phi \in \mathcal{A}_1$$

$$i'_2(\psi) = (\mathcal{W}_1 \otimes i_2(\psi)) \setminus \mathcal{W}_0 \qquad where \;\; \psi \in \mathcal{A}_2$$

the new probability distribution on $\mathcal{W}_3$ is:

$$\mu_3(< w_{1i}, w_{2j} >) = \frac{\mu_1(w_{1l})\mu_2(w_{2j})}{1 - \Sigma_{<w_{1t}, w_{2m}> \in \mathcal{W}_0} \mu_1(w_{1t})\mu_2(w_{2m})} \tag{3.7}$$

Applying the Combination Rule on these two new generalized incidence calculus theories, we have the combined generalized incidence calculus theory $< \mathcal{W}_3, \mu_3, P, \mathcal{A}_3, i_3 >$.

QED

$\Sigma_{<w_{1t}, w_{2m}> \in \mathcal{W}_0} \mu_1(w_{1t})\mu_2(w_{2m})$ is the weight of the conflict between two theories. If the conflict part is 1 then these two pieces of information completely conflict with each other and they cannot be combined.

For the joint product of spaces $\mathcal{W}_1$ and $\mathcal{W}_2$, an element $< w_{1i}, w_{2j} >$ in $\mathcal{W}_1 \otimes \mathcal{W}_2 \setminus \mathcal{W}_0$ tells us that possible worlds w_{1i} and w_{2j} may support a formula at the same time. An element $< w_{2j}, w_{1i} >$ in $\mathcal{W}_2 \otimes \mathcal{W}_1$ implies the same meaning as $< w_{1i}, w_{2j} >$. Therefore we treat $\mathcal{W}_1 \otimes \mathcal{W}_2$ and $\mathcal{W}_2 \otimes \mathcal{W}_1$ as the same set. So the Combination Rule is both commutative and associative because the result of combining several incidence calculus theories is unique irrespective of the sequence in which they are combined.

In the following, we say that two generalized incidence calculus theories are DS-independent if their sets of possible worlds (together with their probability distributions) are DS-independent.

Example 3.2

We now use an example adopted from ([Pearl, 1988] pp.58) to show the situation in which two generalized incidence calculus theories are based on different sets of possible worlds but these two sets are not DS-independent. The example is as follows.

> There are three prisoners, A, B and C, have been tried for murder, and their verdicts will be read tomorrow. They know only that one of them will be declared guilty and the other two will be set free. The identity of the condemned prisoner is revealed to the very reliable prison guard, but not to the prisoners themselves. In the middle of the night, Prisoner A calls the guard over and makes the following request: 'Please give this letter to one of my friends -to one who is to be released. You and I know that at least one of them will be freed'. Later Prisoner A calls the guard again and asks who received the letter. The guard answers, 'I gave the letter to Prisoner B, he will be released tomorrow'. After this Prisoner A feels that his chance to be guilty has been increased from 1/3 to 1/2. What did he do wrong?

Assume that I_B stands for the proposition 'Prisoner B will be declared innocent' and G_A stands for the proposition 'Prisoner A will be declared guilty'. The task is to compute the probability of G_A given all the information obtained from the Guard.

Solving this problems in formal probability theory, Pearl gets

$$Pr(G_A \mid I_B) = \frac{Pr(I_B \mid G_A)Pr(G_A)}{Pr(I_B)} = \frac{Pr(G_A)}{Pr(I_B)} = \frac{1/3}{2/3} = 1/2 \tag{3.8}$$

where $Pr(I_B \mid G_A) = 1$ since $G_A \supset I_B$ and $Pr(G_A) = Pr(G_B) = Pr(G_C) = 1/3$ from the prior probability distribution.

Pearl argues that this is a wrong result and the wrong result arises from omitting the *full context* in which the answer was obtained by Prisoner A. He further explains that 'By context we mean the entire range of answers one could possibly obtain, not just the answer actually obtained'. Therefore, Pearl introduces another proposition I'_B, stands for 'The guard said that B will be declared innocent', and he gives that

$$Pr(G_A \mid I'_B) = \frac{Pr(I'_B \mid G_A)Pr(G_A)}{Pr(I'_B)} = \frac{1/2.1/3}{1/2} = 1/3 \tag{3.9}$$

which he believes is the correct result.

Using incidence calculus to solve this problem, we let $P = \{G_A, G_B, G_C\}$ and G_A stand for the proposition 'Prisoner A is guilty'. Then it is possible to form a set of possible worlds $\mathcal{W}_1 = \{w_1, w_2, w_3\}$ with $\mu_1(w_j) = 1/3$ from the prior probability distribution. w_1 implies A is guilty.

From this information, a generalized incidence calculus theory is formed as $< \mathcal{W}_1, \mu_1, P, P, i_1 >$ where $i_1(G_A) = \{w_1\}$, $i_1(G_B) = \{w_2\}$ and $i_1(G_C) = \{w_3\}$.

After the guard passed the letter to a prisoner, it is possible to form another set of possible worlds $\mathcal{W}_2 = \{L_B, L_C\}$ where L_B means Prisoner B received the letter. $\mu_2(L_B) = \mu_2(L_C) = 1/2$.

So the second generalized incidence calculus theory is constructed as

$$< \mathcal{W}_2, \mu_2, P, \mathcal{A}_2, i_2 >$$

where $i_2(G_A \vee G_C) = \{L_B\}$, $i_2(G_A \vee G_B) = \{L_C\}$ and $\mathcal{A}_2 = \{G_A \vee G_C, G_A \vee G_B\}$.

These two theories are based on different sets of possible worlds and they are not DS-independent. If we attempt to solve this example using Theorem 1, we can only get the result as shown in equation (3.8).

However whether it is possible to construct different generalized incidence calculus theories in order to reflect the *full context* of answers (the meaning of I'_B not I_B) remains open.

3.3.4 Implementation of combination rule

The combination procedure of two or more generalized incidence calculus theories involves multiple searches through two (or more) sets of axioms. Therefore, the combination procedure, like other similar combination techniques, has a computational problem, especially when sets of axioms are large.

The combination rule we proposed in this chapter has been implemented in Sicstus Prolog without considering computational complexity problem.

For example, assume that the initial incidence assignments on sets of axioms $\{a, b, a \wedge b, c, d, c \wedge d\}$ and $\{a, c\}$ are

$$i_1(a) = \{1, 2, 3, 4\}$$

$$i_1(b) = \{1, 2, 3\}$$

$$i_1(a \wedge b) = \{1, 2, 3\}$$

$$i_1(c) = \{4, 5, 6, 7\}$$

$$i_1(d) = \{6, 7, 8, 9\}$$

$$i_1(c \wedge d) = \{6, 7\}$$

and

$$i_2(a) = \{1, 2, 3\}$$

$$i_2(c) = \{4, 5, 6\}$$

then the result of combining these two sets is

$$i_3(a) = \{1, 2, 3\}$$

$$i_3(b \wedge a) = \{1, 2, 3\}$$

$$i_3(a \wedge c) = \{4\}$$

$$i_3(c) = \{4, 5, 6\}$$

$$i_3(c \wedge d) = \{6\}$$

The corresponding generalized incidence calculus is

$$< \mathcal{W}, \mu, P, \mathcal{A}, i_3 >$$

This program has been tested on several examples. When two sets of axioms are large (with more than 5 axioms), the run time is a bit slow, see Table 3.3.

When two generalized incidence calculus theories are DS-independent, we will have to redefine them to obtain two generalized incidence calculus theories which are based on the same set of possible worlds before we execute the program.

Size of $\mathcal{A}_1$	Size of $\mathcal{A}_2$	runtime
3	2	10.728
7	2	21.457
7	3	727.377

Table 3.3 Test result of combination rule

3.4 Summary of Extended Incidence Calculus

The generalized incidence calculus is no longer *truth functional.* So a Legal Assignment Finder doesn't exist in the generalized theory. We can use generalized incidence calculus to describe a wider range of information after we drop some conditions on the incidence function i.

A new notion, *basic incidence assignment*, is defined. Although this function is derived from a generalized incidence function, it can also be separated out from an original incidence function. So Algorithm B for incidence assignment can be used to recover an original incidence function as long as a set of axioms $\mathcal{A}$ is closed under $\wedge$. This function tells the difference between the incidence set of a compound formula and the union of incidence sets of all its parts. The meaning behind this difference is that the current information cannot fully allocate the incidence set of this formula to its parts.

We have also discussed an approach to giving an incidence assignment based on the lower bound measure of probabilities on the set of axioms.

Given a set of axioms $\mathcal{A} = \{\phi_1, \phi_2, ..., \phi_n\}$, the main step in the algorithm is to select the 'smallest' elements each time to process. The complexity of the selection procedure is $n + (n-1) + (n-2) + ... + 1$. So the actual complexity of the new algorithm should be $\mid \mathcal{A} \mid^2 = n^2$.

There are two algorithms in [McLean, 1992] for assigning incidences on axioms based on a probability assignment. The common feature of the two algorithms is that a set of possible worlds has to be fixed first. Given a set of axioms, both algorithms try to divide possible worlds into groups and assign each group to an axiom. Therefore, not only the number of axioms but also the interrelationship of these axioms affect the division procedure. For instance, assume that there are only two axioms a, b in $\mathcal{A}$, then the assignment procedure could be simply done by choosing two subsets of the set of possible worlds which can produce $p(a)$ and $p(b)$ respectively. However if $p(a \wedge b)$ is known as well in addition to $p(a)$ and $p(b)$, then we need not only two subsets of possible worlds W_1, W_2 to match $p(a)$ and $p(b)$, but also another subset W_3 which matches $p(a \wedge b)$ with the condition that $W_3 = W_1 \cap W_2$. That is, the complexity of these two algorithms increases along with the interrelationship of axioms considerably. Because of this, the order of the axioms also affects the efficiency of the algorithms as pointed out in [McLean, 1992].

In summary, there are three factors associated with the complexity of each algorithm in [McLean, 1992]. They are the numbers of axioms, the relations among axioms and the order of axioms in addition to the requirement of the fixed number of possible worlds. The complexity for the best case of the algorithms could be $\mid \mathcal{A} \mid$ when all the axioms in $\mathcal{A}$ have no interrelation at all. The complexity for the worst case is exponential and most of cases have this complexity. However, in our algorithm in extended incidence calculus, there is only one factor affecting the complexity, that is, the number of axioms. Besides, the new algorithm does not require a set of possible worlds to be predefined.

Example 3.1 has also been tested by McLean using the two algorithms. Although there are only 8 axioms in $\mathcal{A}$, both algorithms take a long time to find a consistent assignment of incidences (with runtime 374.520 seconds and 355.060 seconds respectively). Our algorithm only needs 1.189 seconds runtime to find a consistent assignment. This example, in some sense, shows us that the relations among axioms could slow the algorithms down enormously, much worse than the size of set of axioms. In real world cases, axioms always have some interrelations. Therefore, we concluded that on average the new incidence assignment algorithm we designed is much faster than McLean's methods.

This assignment algorithm can be used to restore a mass function from a belief function in Dempster-Shafer theory and to restore the probability space in general. We will mention this later in this thesis.

Chapter 4

The Dempster-Shafer theory of evidence

The Dempster-Shafer (DS) theory of evidence (sometimes called evidential reasoning or belief function theory) was first introduced by Dempster [Dempster, 1967] and later developed by Shafer [Shafer76]. The transferable belief model [Smets and Kennes, 1994] is developed based on this theory. DS theory has been popular since the early 1980 and a number of applications of the theory have been reported. Its relations with related theories have also been intensively discussed [Yager, Fadrizzi, Kacprzyk, 1994].

There are two main reasons why DS theory has attracted a lot of attention. It has the ability to model information flexibly and provides a convenient and simple mechanism (Dempster's combination rule) to combine two or more pieces of evidence which satisfy certain conditions. The former allows a user to describe ignorance because of lacking information and the latter allows a user to narrow the possible answer space as more evidence is accumulated.

Even though DS theory has been widely used, it has been found that Dempster's combination rule gives counterintuitive results in many cases. The conditions under which the rule should be used are crucial to the successful applications of the theory but they were not fully defined when Shafer gave the rule in the first instance [Shafer76]. Various discussions and criticisms of the rule have appeared in the literature. A mathematical description of the conditions on applying Dempster's combination rule appears in [Voorbraak, 1991], in which the rule condition is called DS-independence.

In this chapter I first introduce the basics of the theory and then study the probabilistic basis of basic functions defined by Shafer in this theory to see how we can derive a mass function from a probability distribution through a multivalued mapping in Dempster's paper [Dempster, 1967]. The discussion shows that DS theory is closely related to probability theory and provides a convenient way to describe the conditions of using Dempster's combination rule. Some other aspects of DS theory will also be discussed briefly at the end of the chapter.

4.1 Basic Concepts in Dempster-Shafer Theory

A piece of information is usually described as a mass function on a frame of discernment. We first give some definitions of the theory [Shafer76].

Definition 4.1: *Frame of discernment*

A set is called a frame of discernment (or simply a frame) if this set contains mutually exclusive and exhaustive possible answers to a question. It is usually denoted as Θ. It is required that at any time, one and only one element in the set is true.

For instance, if we assume that Emma lives in one of the cities, $city_1, ..., city_6$, then $\Theta = \{city_1, city_2, city_3, city_4, city_5, city_6\}$ is a frame of discernment for the question 'In which city does Emma live?'.

Definition 4.2: *Mass function*

A function $m : [0,1] \rightarrow 2^{\Theta}$ is called a mass function on frame Θ if it satisfies the following two conditions:

$$1.m(\emptyset) = 0 \qquad 2.\Sigma_A m(A) = 1$$

where $\emptyset$ is an empty set and A is a subset of Θ.

A mass function is also called *a basic probability assignment*, denoted as *bpa*.

For instance, if we know that Emma lives in the area covering these six cities, but we have no knowledge about in which city she lives, then we can only give a mass function $m(\Theta) = 1$. Alternatively, if we know that Emma lived in $city_3$ two years ago and she intended to move to other cities and tried to find a job somewhere within these six cities, but we have no definite information about where she lives now. A mass function could be defined as $m(city_3) = p, m(\Theta) = 1 - p$ for the situation where p stands for the degree of our belief that she is still in $city_3$.

A subset A with $m(A) > 0$ is called a focal element of this mass function. If all focal elements of a mass function are the singletons of Θ then this mass function is exactly a probability distribution on Θ. So mass functions are generalized probability distributions.

Definition 4.3: *Belief function*

A function $bel : [0,1] \to \Theta$ is called a belief function if bel satisfies:

$$1. bel(\Theta) = 1$$
$$2. bel(\cup_1^n A_i) \geq \Sigma_i bel(A_i) - \Sigma_{i>j} bel(A_i \cap A_j) + ... + (-1)^{-n} bel(\cap_i A_i)$$

It is easy to see that $bel(\emptyset) = 0$ for any belief function. A belief function is also called *a support function*. The difference between $m(A)$ and $bel(A)$ is that $m(A)$ is our belief committed to the subset A excluding any of its subsets while $bel(A)$ is our degree of belief in both A and all its subsets.

If there is only one focal element for a belief function and this focal element is the whole frame Θ, this belief function is called a *vacuous belief function*. It represents the total ignorance (because of lack of knowledge). In the following, we call (Θ, bel) a DS structure.

In general, if m is a mass function on a frame of Θ then bel defined in (4.1) is a belief function on Θ.

$$bel(B) = \Sigma_{A \subseteq B} m(A) \tag{4.1}$$

Definition 4.4: *Plausibility function*

A function pls defined below is called a plausibility function.

$$pls(A) = 1 - bel(\neg A)$$

$pls(A)$ represents the degree to which the evidence fails to refute A. From a mass function, we can get its plausibility function as [Shafer, 1990]

$$pls(B) = \Sigma_{A \cap B \neq \emptyset} m(A) \tag{4.2}$$

Recovering a mass function from a belief function is as follows [Shafer, 1990].

$$m(A) = \Sigma_{B \subseteq A} (-1)^{|B|} bel(B)$$

For any finite frame, it is always possible to get the corresponding mass function from a belief function and the mass function is unique.

In a system using evidential reasoning, knowledge or inference results are usually represented by the interval of bel and pls. There are several special features of this interval [Wesley, 1983].

$[bel(A), pls(A)] = [1, 1]$ subset A completely true;
$[bel(A), pls(A)] = [0, 0]$ subset A completely false;
$[bel(A), pls(A)] = [0, 1]$ subset A completely ignorant;
$[bel(A), pls(A)] = [bel, 1]$, $0 < bel < 1$ tends to support A;
$[bel(A), pls(A)] = [0, pls]$, $0 < pls < 1$ tends to refute A;
$[bel(A), pls(A)] = [bel, pls]$, $0 < bel < pls < 1$ may support or refute A.

When more than one mass function is given on the same frame of discernment, the combined impact of these pieces of evidence is obtained using a mathematical formula called Dempster's combination rule. If m_1 and m_2 are two mass functions on frame Θ, then $m = m_1 \oplus m_2$ is the mass function after combining the two mass functions.

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

$\oplus$ means that Dempster's combination rule is applied on two (or more) mass functions. The condition of using the rule is stated as "two or more pieces of evidence are based on distinct bodies of evidence" [Shafer76]. This description is a little confusing and causes a lot of misapplications and counterintuitive results [Voorbraak, 1991]. We will have a more detailed discussion on the conditions of the rule in the later part of this chapter.

4.2 Probability Background of Mass Functions

Even though Shafer does not agree with the idea that belief function theory is generalized probability theory and regards it as a new terminology to represent evidence and knowledge, some people argue that the theory has strong links with probability theory [Fagin and Halpern, 1989a]. Here we explore the motivation underlying Shafer's definition of mass functions under Dempster's assumptions. We argue that in Dempster's paper [Dempster, 1967], Dempster gave the prototype of mass functions implicitly. Shafer's contribution is to make it clear and use it to represent evidence directly.

4.2.1 Dempster's probability prototype of mass functions

Definition 4.5: *Dempster's probability space*

A structure (X, τ, μ) is called a Dempster probability space where

- *X is a sample space usually containing all the possible worlds;*
- *τ is a class of subsets of X;*

- *μ is a probability measure which gives $\mu : \tau \to [0, 1]$.*

Definition 4.6: *Multivalued mapping*

A function Γ is a multivalued mapping from Dempster probability space (X, τ, μ) to another space S if Γ assigns a subset $\Gamma x \subset S$ to every $x \in X$.

From a multivalued mapping Γ, the probability measure μ can be propagated to space S in such a way that for any subset T of S, the lower and upper bounds of probabilities of T are defined as

$$P_*(T) = \mu(T_*)/\mu(S_*) \tag{4.3}$$

$$P^*(T) = \mu(T^*)/\mu(S^*) \tag{4.4}$$

where

$$T_* = \{x \in X, \Gamma x \neq \emptyset, \Gamma x \subset T\}$$

$$T^* = \{x \in X, \Gamma x \cap T \neq \emptyset\}$$

The equations (4.3) and (4.4) are defined only when $\mu(S^*) \neq 0$. The denominator $\mu(S^*)$ is a renormalizing factor necessitated by the fact that the model permits, in general, outcomes in X which do not map into a meaningful subset of S. That is, there may exist a x, such that $\Gamma x = \emptyset$. Dempster argued that the subset $\{x, \Gamma x = \emptyset\}$ should be removed from X and the measure of the remaining set S^* renormalized to unity.

A multivalued mapping Γ from a space X to another S says that if the possible answer to a question described in the first space X is x, then the possible answer to a question described in the second space S is in Γx.

For the case that $S = \{s_1, s_2, ..., s_m\}$ is finite, the propagation procedure can be done as follows. Suppose that $S_{\gamma_1\gamma_2...\gamma_m}$ denotes the subset of S which contains s_i if $\gamma_i = 1$ and excludes s_i if $\gamma_i = 0$ for $i = 1, 2, ..., m$. If for each $S_{\gamma_1\gamma_2...\gamma_m}$ we define $X_{\gamma_1\gamma_2...\gamma_m}$ as

$$X_{\gamma_1\gamma_2...\gamma_m} = \{x \in X, \Gamma x = S_{\gamma_1\gamma_2...\gamma_m}\} \tag{4.5}$$

then all the subsets of X defined in (4.5) form a partition[1] of X into

$$X = \cup_{\gamma_1\gamma_2\dots\gamma_m} X_{\gamma_1\gamma_2\dots\gamma_m} \tag{4.6}$$

Certainly some of these subsets may be empty. The idea of forming $X_{\gamma_1\gamma_2\dots\gamma_m}$ is that each $X_{\gamma_1\gamma_2\dots\gamma_m}$ contains those elements in X which have the same mapping environments in S.

In order to calculate $P_*(T)$ and $P^*(T)$, Dempster assumed that each $X_{\gamma_1\gamma_2\dots\gamma_m}$ is in τ, then for any $T \subset S$, $P_*(T)$ and $P^*(T)$ are uniquely determined by the 2^m quantities $p_{\gamma_1\gamma_2\dots\gamma_m}$.

$$p_{\gamma_1\gamma_2\dots\gamma_m} = \mu(X_{\gamma_1\gamma_2\dots\gamma_m}) \tag{4.7}$$

Here we use an example to demonstrate the idea.

Example 4.1

Assume that $S = \{s_1, s_2, s_3\}$. Using $p_{\gamma_1\gamma_2\dots\gamma_m}$, all the possible lower and upper probabilities on S are given in Table 4.1.

T	$P^*(T)$	$P_*(T)$
$\emptyset$	0	0
$\{s_1\}$	$(p_{100} + p_{110} + p_{101} + p_{111})/k$	p_{100}/k
$\{s_2\}$	$(p_{010} + p_{110} + p_{011} + p_{111})/k$	p_{010}/k
$\{s_3\}$	$(p_{001} + p_{101} + p_{011} + p_{111})/k$	p_{001}/k
$\{s_1, s_2\}$	$(p_{100} + p_{010} + p_{110} + p_{101} + p_{011} + p_{111})/k$	$(p_{100} + p_{010} + p_{110})/k$
$\{s_1, s_3\}$	$(p_{100} + p_{001} + p_{110} + p_{101} + p_{011} + p_{111})/k$	$(p_{100} + p_{001} + p_{101})/k$
$\{s_2, s_3\}$	$(p_{010} + p_{001} + p_{110} + p_{101} + p_{011} + p_{111})/k$	$(p_{010} + p_{001} + p_{011})/k$
S	1	1

Table 4.1 Upper and lower probabilities on all subsets of S.

[1]A list of subsets $X_1, X_2, ..., X_n$ of X is called a partition of X if $X_i \cap X_j = \emptyset$ and $\cup_i X_i = X$.

Here we use k to denote $1 - p_{000}$. Given a subset T of S, the corresponding lower and upper subsets in X are known. For instance if $T = S_{110} = \{s_1, s_2\}$, then $T_* = X_{100} \cup X_{010} \cup X_{110}$ and $T^* = X_{100} \cup X_{110} \cup X_{101} \cup X_{011} \cup X_{111}$. In particular $\mu(S^*) = 1 - p_{000}$.

If we define a function m on S as $m(S_{\gamma_1\gamma_2\cdots\gamma_m}) = m_{\gamma_1\gamma_2\cdots\gamma_m} = p_{\gamma_1\gamma_2\cdots\gamma_m}/(1 - p_{00\ldots0})$, then Table 4.1 is replace by Table 4.2 below.

T	$P^*(T) = pls$	$P_*(T) = bel$	m
$\emptyset$	0	0	0
$\{s_1\}$	$m_{100} + m_{110} + m_{101} + m_{111}$	m_{100}	m_{100}
$\{s_2\}$	$m_{010} + m_{110} + m_{011} + m_{111}$	m_{010}	m_{010}
$\{s_3\}$	$m_{001} + m_{101} + m_{011} + m_{111}$	m_{001}	m_{001}
$\{s_1, s_2\}$	$m_{100} + m_{010} + m_{110} + m_{101} + m_{011} + m_{111}$	$m_{100} + m_{010} + m_{110}$	m_{110}
$\{s_1, s_3\}$	$m_{100} + m_{001} + m_{110} + m_{101} + m_{011} + m_{111}$	$m_{100} + m_{001} + m_{101}$	m_{101}
$\{s_2, s_3\}$	$m_{010} + m_{001} + m_{110} + m_{101} + m_{011} + m_{111}$	$m_{010} + m_{001} + m_{011}$	m_{011}
S	1	1	m_{111}

Table 4.2 Upper and lower probabilities on all subsets of S using function m.

Some of $m_{\gamma_1\gamma_2\cdots\gamma_m}$ may be 0. If we compare this table with equations (4.1) and (4.2), it is easy to see that the function m is exactly a mass function if S is a frame. P_* and P^* define a belief function and a plausibility function on S respectively. I assume that this is the model for defining mass functions in the style of Shafer. S being a frame of discernment is a special case of S being any space in Dempster's paper.

The vital step in calculating probability bounds in Dempster's prototype is that for any subset $X_{\gamma_1\gamma_2\cdots\gamma_m}$, this subset should be in τ. If τ, a collection of subsets of X, does not suit this requirement, then the rest of the calculations in Dempster's paper could not be carried out.

4.2.2 Deriving mass functions from probability spaces

In Chapter 3, we have introduced probability spaces in Definition 3.3. Here we recall it again to refresh our mind.

A structure (X, χ, μ) is called a probability space if (X, χ, μ) is a Dempster's probability space and χ is a $\sigma-$algebra of X (a set of subsets of X containing X and closed under complementation and countable union, but not necessarily containing all subsets of X)

A subset χ' of χ is called a *basis* of χ if it contains non-empty and disjoint elements, and if χ consists precisely of countable unions of members of χ'. For any finite χ there is a basis of χ and it follows that

$$\Sigma_{X_i \in \chi'} \mu(X_i) = 1$$

If χ is finite, it must have a basis and the basis is unique.

For any subset X_i of X, if X_i is not in χ, then we can get two bounded probability measures of X_i, usually called the *inner measure* and the *outer measure*. For any $X_i \subseteq X$, we define

$$\mu_*(X_i) = sup\{\mu(X_j) \mid X_j \subseteq X_i, X_j \in \chi\} \tag{4.8}$$

$$\mu^*(X_i) = inf\{\mu(X_j) \mid X_j \supseteq X_i, X_j \in \chi\} \tag{4.9}$$

It is proved in [Fagin and Halpern, 1989b] that μ_* is a belief function on X when X is a frame.

Given a probability space (X, χ, μ), assume that there is a multivalued mapping function Γ from X to another frame S. For a subset S' of S, we define $bel(S') = \mu(\{x \mid \Gamma x \subseteq S'\})$ as our degree of belief in S'. When $\{x \mid \Gamma x \subseteq S'\}$ is not measurable, that is when $\{x \mid \Gamma x \subseteq S'\}$ is not in χ, we define

$$bel(S') = \mu_*(\{x \mid \Gamma x \subseteq S'\}) \tag{4.10}$$

bel is also a belief function on S.

In a Dempster probability structure, because those nonempty subsets $X_{\gamma_1\gamma_2...\gamma_m}$ of X form a partition of X, the equation $\Sigma_{\gamma_1\gamma_2...\gamma_m}\mu(X_{\gamma_1\gamma_2...\gamma_m}) = 1$ holds. So these

$X_{\gamma_1\gamma_2\ldots\gamma_m}$ are members of the 'basis' of τ. If we use χ' to denote this partition, then we can form a σ−algebra of X using this basis. So a Dempster probability space can be represented by a normal probability space. Therefore belief function theory is closely related to probability theory and comes out of probability theory.

In the following, we use probability spaces to stand for both normal and Dempster probability spaces.

4.3 Problems with Dempster's Combination Rule

Based on Dempster's paper [Dempster, 1967] we can simply state his idea about the combination procedure as follows: suppose there are two pieces of evidence which are given in the form of two probability spaces (X_1, χ_1, μ_1) and (X_2, χ_2, μ_2). Further, suppose there is another space S and some mapping relationships from spaces X_1 and X_2 to S. The relation between one space and another says that the truth of some elements in the former space suggests the possibility of truth of some elements in the latter space. Given the probability of truth of some elements in spaces X_1 and X_2, we are interested in knowing the impact of the evidence on the space S (we may think that S contains answers to our questions or the possible values of a variable). In order to get the impact of evidence on the space S, Dempster suggested that we can get the joint probability space $(X, \chi, \mu) = (X_1 \otimes X_2, \chi_1 \otimes \chi_2, \mu_1 \otimes \mu_2)$ out of the two original spaces as well as the joint mapping relation from X to S first and then propagate the effect of probability distribution μ to S. Dempster further suggested that it is also possible to obtain the impact of several pieces of evidence by propagating the probability distributions from original probability spaces to S first and then combining them on S. Therefore Dempster gave two alternative approaches to calculate the impact of multiple pieces of evidence carried by different probability spaces. In order to refer to them easily we name these two methods as *Approach 1* and *Approach*

2[2]. The whole framework containing these two combination approaches is called *Dempster's combination framework.* Figures 4.1 and 4.2 show the meaning of these two approaches intuitively. In *Approach 1*, Dempster addressed combining several sources first and then propagating the unified source to the target space. In *Approach 2*, he addressed propagating the sources to the target space separately first and then combining them.

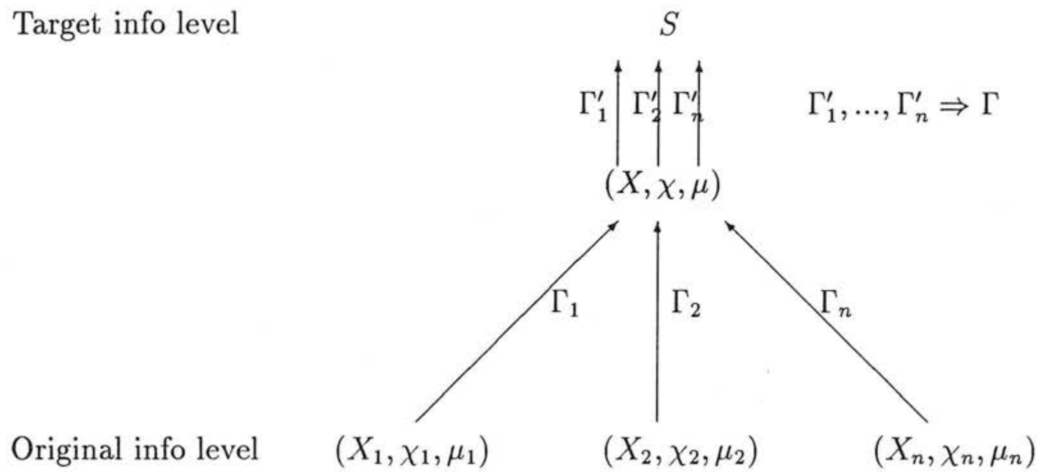

Figure 4.1. Combining evidence first and then propagating probabilities

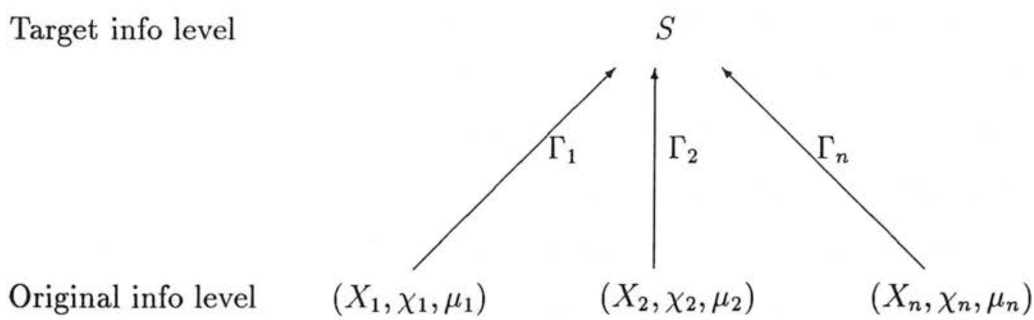

Figure 4.2. Propagating evidence first and then combining them

These two alternative approaches of combination require the same condition and it was implicitly claimed by Dempster that the results obtained from these two methods are the same under that condition. The condition is "*the sources (if we*

[2]See detailed descriptions of these approaches later

treat a space and its probability distribution as a source) are assumed independent. ... Opinions of different people based on overlapping experiences could not be regarded as independent sources. Different measurements by different observations on different equipments would often be regarded as independent ... the sources are statistical independent" [Dempster, 1967]. If we refer to the levels containing spaces (X_1, χ_1, μ_1) and (X_2, χ_2, μ_2) as the *original information level*, and space S as the *target information level*, then Dempster's condition of independence is assumed at the original information level. This requirement is called *DS-Independent* in [Voorbraak, 1991].

When gaving Dempster's combination rule in his book [Shafer76], Shafer followed *Approach 2* suggested by Dempster. After propagating two probability distributions from X_1, X_2 to S, the information accumulated on S needs to be combined. On S, these two pieces of evidence are in the form of belief functions, so Shafer proposed a mathematical formula named Dempster's combination rule to combine two (or more) belief functions. If we follow the idea that the rule proposed by Shafer was abstracted out from Dempster's combination framework *i.e., Approach 2*, then Dempster's combination rule should obey the condition defined by Dempster given above. Later, in some of his papers, Shafer began to address the importance of independence among original sources and make the condition more clear. For example in [Shafer, 1986], [Shafer, 1987a], Shafer stated that the condition of using Dempster's combination rule is 'two or more belief functions on the same frame but based on independent arguments or items of evidence' and in [Shafer, 1982] he used random encoded messages to describe the condition in using Dempster's combination rule. These explanations are much closer to the definition given by Dempster in his combination framework. However, as Dempster's combination rule does not require or reflect any information about the sources which support the corresponding belief functions and it only needs belief functions (or mass functions) in order to carry out the combination, it is, therefore, difficult to describe independent conditions precisely using only belief functions.

In contrast to the two views on belief functions in [Halpern and Fagin, 1992], we argue that the main cause of giving counterintuitive results in most cases[3] in using Dempster's combination rule is *overlooking (or ignorance of)* the condition of combination given in Dempster's original paper. Describing independent conditions on the target information level solely is inadequate. In other words, Dempster's combination rule is too simple (compared to Dempster's combination framework) to show (or carry) enough information and provide a precise mathematical description about the dependency relations of multiple evidence.

Even though two approaches in Dempster's combination framework aim at coping with independent sources of evidence equivalently, the extension of *Approach 1* in Dempster's combination framework can also be used to deal with dependent sources of evidence as discussed in [Shafer, 1986], [Shafer, 1987a], [Lingras and Wong, 1990]. But they didn't provide a unique rule for general cases.

4.3.1 The condition for using Dempster's combination rule

Suppose n pieces of information are known, *i.e.* (X_i, χ_i, μ_i) for $i = 1, ..., n$, which all have mapping relations (Γ_i) with another set S, and they are independent, Dempster [Dempster, 1967] suggested that the combined source (X, χ, μ) and Γ are defined in Equation (4.11).

$$\begin{aligned} X &= X_1 \otimes X_2 \otimes ... \otimes X_n \\ \chi &= \chi_1 \otimes \chi_2 \otimes ... \otimes \chi_n \\ \mu &= \mu_1 \otimes \mu_2 \otimes ... \otimes \mu_n \\ \Gamma(x) &= \Gamma_1(x) \cap \Gamma_2(x) \cap ... \cap \Gamma_n(x) \end{aligned} \tag{4.11}$$

The fourth formula can also be stated and explained as

$$\Gamma(x) = \Gamma'_1(x) \cap \Gamma'_2(x) \cap ... \cap \Gamma'_n(x)$$

[3] These don't include the situations such as the example 6.12 in [Voorbraak, 1991] and the murderer case in [Smets, 1988].

where $\Gamma'_i(x) = \Gamma_i(x_i)$ when $x \in X_1 \otimes ... \otimes X_{i-1} \otimes \{x_i\} \otimes ... \otimes X_n$.

The meaning behind this set of mathematical equations is that from n independent sources we can get the joint source which denotes the message carried by all separate sources and establish different mapping relations from the joint source to the target space S. Different mapping relations are further unified to get the joint mapping function Γ and using Γ the joint probability distribution μ is propagated to S. The definition of Γ reflects that $x = (x_1, x_2, ..., x_n) \in X$ is consistent with $s_i \in S$ if and only if s_i belongs to all $\Gamma_i(x_i)$ simultaneously.

Therefore Dempster suggested two approaches for performing the combination:

Approach 1: Combining them at the original information level by producing a joint space and a single probability distribution on the space. This should consider the different mappings from the joint space to the target information space, unify these mappings into one mapping and propagate the joint probability distribution to the target information level.

Approach 2: Propagating different pieces of evidence at the original information level to the target information level and then combining them.

Dempster assumed implicitly that the results obtained in the above two ways are the same under the condition that the n sources are *statistically independent.* As we discussed in the introduction, Shafer followed Approach 2 in getting Dempster's combination rule. But in this simplified combination rule (*i.e.* Dempster's combination rule) the original sources are hidden. *The invisibility of the original sources in the simplified combination rule makes it difficult to judge the dependent relations among the belief functions which in turn causes counterintuitive results in many cases.*

Proposition 6 *Two belief functions on a frame can be combined using Dempster's combination rule if the two sources, (X_1, χ_1, μ_1) and (X_2, χ_2, μ_2) from which the two belief functions are derived, are DS-independent.*

The idea of describing and judging dependent relations among the original probability spaces has also been mentioned implicitly by Shafer [Shafer, 1982], Shafer and Tversky [Shafer and Tversky, 1985] and Voorbraak [Voorbraak, 1991] but not explicitly defined at the original information level.

4.3.2 Examples

We now examine two examples. The first two examples come out from Shafer's paper [Shafer, 1986]. The first example shows that when two pieces of evidence are DS-independent, there are two alternative ways to combine them while the second one shows that when they are not independent, only the first method works.

Example 4.2:

Suppose that Shafer wants to know whether the street outside is slippery, instead of observing this himself, he asks another person Fred. Fred tells him that 'it is slippery'. However Shafer knows that Fred is careless in answering questions sometimes. Based on his knowledge about Fred, Shafer estimates that 80% of the time Fred reports what he knows and he is careless at 20% of the time. So Shafer believes that there is only a 80% chance that the street is slippery. In fact Shafer forms two frames X_1 and S in his mind for this problem where X_1 is related to Fred's truthfulness and S is related to the possible answers of slippery outside.

$$X_1 = \{truthful, careless\}$$

$$S = \{yes, no\}$$

A probability measure p_1 on X_1 is defined as

$$p_1\{truthful\} = .8 \qquad p_1\{careless\} = .2$$

He obtains the answer when this probability measure is propagated from the frame X_1 to the frame S. The principle of propagation is based on a multivalied mapping function between X_1 and S. Assume that a multivalued mapping function Γ_1 is defined as $\Gamma_1 truthful = \{yes\}$ and $\Gamma_1 careless = \{yes, no\}$, Shafer obtains that

$$bel_1(\{yes\}) = .8 \qquad bel_1(\{no\}) = 0 \tag{4.12}$$

which is a belief function based on the equation (4.10).

Further more, suppose Shafer has some other evidence about whether the street is slippery: his trusty indoor-outdoor thermometer says that the temperature is 31 degrees Fahrenheit, and he knows that because of the traffic ice could not form on the streets at this temperature. However he knows that the thermometer could be wrong even though it has been very accurate in the past. Suppose that there is a 99% chance that the equipment is working properly, so he could form another frame X_2 with its probability distribution as

$$X_2 = \{working, not\_working\}$$

$$p_2\{working\} = .99 \qquad p_2\{not\_working\} = .01$$

and a mapping function Γ_2 as $\Gamma_2 working = \{no\}$ and $\Gamma_2 not\_working = \{yes, no\}$. Therefore another belief function on S is calculated using (4.10) as

$$bel_2(\{yes\}) = 0 \qquad bel_2(\{no\}) = .99 \tag{4.13}$$

Now the problem is that there are two pieces of evidence available regarding the same question 'slippery or not?' and Shafer wants to know what the joint impact of the evidence on S is. In his paper, Shafer adopts two alternative approaches to do this.

Using Approach 1 in Dempster's combination framework:

First of all, since he believes that Fred's answer is independent of the equipment, *i.e. the two original pieces of evidence are DS-independent*, Shafer gets a joint space X out of X_1 and X_2 as $X = X_1 \otimes X_2$ with probability distribution p as $p((x_1, x_2)) = p_1(x_1) \times p_2(x_2)$. This is shown in the first two columns in Table 4.3. Then the new compatibility relation between X and S is given in the last column of Table 4.3 which takes into account what Fred and the thermometer have said. Because the joint element (truthful, working) does not match to any elements in S, he rules this element out and renormalizes the three others by eliminating the probability for (truthful, working). Finally, the posterior probability on X is given in the third column in the Table 4.3. Eventually applying (4.10) to the posterior probability on X, a belief function on S is defined as

$$bel(\{yes\}) = .04 \qquad bel(\{no\}) = .95$$

X	Probability of a		subset of S
	Initial	Posterior	mapped with x
(truthful, working)	.792	0.0	–
(truthful, not)	.008	.04	{yes}
(careless, working)	.198	.95	{no}
(careless, not)	.002	.01	{yes, no}

Table 4.3.

Here 'not' means 'not_working'.

For instance, for the element $(truthful, not)$, its initial probability is

$$p((truthful, not)) = p_1(truthful) \times p_2(not\_working) = 0.8 \times 0.01 = 0.008$$

As element $(truthful, working)$ matches to the empty set of S, the probabilities of other elements in X should be renormalized in order to assign zero probability to $(truthful, working)$. The posterior probability of $(truthful, not)$, therefore, is $p'((truthful, not)) = 0.008/0.208 = 0.04$. The multivalued mapping Γ assigns $(truthful, not)$ to element yes in S and assigns no more elements to yes, so $bel(\{yes\}) = p'((truthful, not)) = 0.04$.

Using Approach 2 in Dempster's combination framework:

In the second approach Shafer propagates the two pieces of evidence from sources X_1 and X_2 to space S first and then gets two belief functions defined as in (4.12) and (4.13) on S. Because he believes that the two sources are DS-independent, he uses Dempster's combination rule directly to combine them. This procedure is shown in Table 4.4.

m		{ yes }	.8	{ yes, no }	.2
{ no }	.99	{}	.792	{no }	.198
{yes, no }	.01	{ yes }	.008	{yes, no }	.002

Table 4.4.

where the first row and the first column stand for the two mass functions derived from the two belief functions defined in (4.12) and (4.13) respectively. After the normalization, the result is the same as he obtains by using *Approach 1.*

For instance, from this table, it is possible to calculate the combined mass function on *yes* as $m(\{yes\}) = 0.008/0.208 = 0.04$. So that $bel(\{yes\}) = 0.04$.

From this example we can see that when the two sources are statistically independent, regardless of which approach is used the results are the same. We should also notice that in the first approach, before the propagation procedure, Shafer has to calculate a posterior probability distribution in order to reflect the compatibility relation between the joint frame X and the target frame S. So it is the posterior probability distribution rather than the initial one that is actually propagated.

However if the original pieces of information are not independent, the second way may not work properly while the first one may still work as we can see in the next example.

Example 4.3:

Continuing the first example, assume that Shafer believes that the Fred's answer relates to the thermometer as Fred accesses to the thermometer regularly to see whether it is working. If it is not working properly then Fred would be careless in answering questions. Assume that Fred has a 90% chance of being careless if the thermometer is not working, then Fred's answer is somehow affected when the thermometer is not working.

Using Approach 1 in Dempster's combination framework:

Because of the relationship between the two sources, it is impossible to get the joint space from X_1 and X_2 by set product. Using *Approach 1*, Shafer gives the second column in Table 4.5 to replace the second column in Table 4.3 in order to take into account the dependency of Fred and the thermometer. Once again this initial probability distribution is renormalized in order to assign 0 probability to the element (truthful, working). Based on the compatibility relation between X and S, applying (4.10) another belief function on S is obtained.

$$bel'(\{yes\}) = .005 \qquad bel'(\{no\}) = .95$$

X	Probability of a		Subset of S
	Initial	Posterior	mapped with x
(truthful, working)	.799	0.0	–
(truthful, not)	.001	.005	{yes}
(careless, working)	.191	.950	{no}
(careless, not)	.009	.045	{yes, no}

Table 4.5.

This result is obviously different from the result given in Example 4.2 because of the relationship between the two pieces of evidence.

Using Approach 2 in Dempster's combination rule:

Using *Approach 2* in Dempster's combination framework requires that the sources must be independent, *i.e., irrelevant in any form or in any manner*, in this case, the two sources are no longer independent, so *Approach 2* in his framework cannot be used. That means Dempster's combination rule cannot be applied. If we apply Dempster's combination on this case, we will find that the combined result gives

$$bel(\{yes\}) = 0.04$$

which is the same as obtained in the previous example and this result, under new assumptions, is wrong. The correct answer should be $bel(\{yes\}) = 0.005$. Therefore only *Approach 1* can be used to deal with it.

The summary of this analysis is shown in Table 4.6.

	Dempster's combination framework		Dempster's combination rule
	Approach 1	Approach 2	bel_1 bel_2
Example 4.2	Applicable and correct	Applicable and correct	Applicable and correct
Example 4.3	Applicable and correct	Inapplicable	Applicable or inapplicable?

Table 4.6.

In fact, the two probability spaces (X_1, X_1, p_1) and (X_2, X_2, p_2) are not DS-independent (they are even not probabilistically independent). So Dempster's combination rule cannot be used.

However if we purely consider the two belief functions bel_1 and bel_2 on S, it is difficult to tell whether Dempster's combination rule is applicable or not (see Table 4.6.). The rule is applicable in the case shown in the first example but inapplicable in the case shown in the second example. Shafer's explanation of Dempster's combination rule is right in [Shafer, 1986] and some of his other papers, but in practice Dempster's rule does not carry (or ask for) enough information to make the correct judgement about the dependency of several pieces of information. This causes some confusion when one describes the condition of applying Dempster's combination rule.

If one uses Dempster's combination rule under the condition that the rule comes out from Dempster's combination framework, then one can normally use the rule correctly. If one doesn't have Dempster's combination framework in one's mind and only has Dempster's combination rule to use, then it is sometime very difficult to judge whether the rule is applicable given two belief functions.

The importance of considering relations among the original information sources has been discussed above. The result tells us that it is more natural to consider the combination at both the original information level and the target information level than only at the target information level. However neither Dempster's combination framework nor Dempster's combination rule provides such combination facilities conveniently.

Several authors have pointed out the problems of applying Dempster's combination rule and argued that the application of this rule could give wrong result [Black,1987], [Hunter,1987], [Lingras and Wong, 1990], [Nguyen and Smets, 1991], [Pearl, 1988] and [Voorbraak, 1991].

4.4 Some Other Aspects of DS Theory

In this section, we will discuss some other opinions on DS theory.

4.4.1 Computational complexity problems in DS theory

Soon after DS theory was used in practice, it was pointed out that using Dempster's combination rule has high computational complexity. A few algorithms have been developed to reduce the computational complexity in the theory.

Barnett [Barnett, 1981] first considered reducing the computational complexity of Dempster's rule of combination. He gave an algorithm for a special case in which every piece of evidence can be represented as a special kind of mass function, a mass function which at most has one focal element apart from the whole frame and the focal element is either a singleton or the complement of a singleton.

Mathematically a mass function m which can be considered in Barnett's algorithm is defined as

$$m(A) \neq 0, \ iff \ A = \{a\} \ or \ A = \Theta \setminus \{a\}$$

and

$$m(\Theta) = 1 - m(A)$$

Barnett made the combination of l mass functions on a frame with n elements in a time linear in n ($i.e.$ $o(n)$) when all the mass functions are given in the above form. For more details, see [Barnett, 1981].

Barnett's algorithm is limited when a mass function is not in the form he required. Gordon and Shortliffe [Gordon and Shortliffe, 1985] tried to reduce the complexity using a tree structure. They tried to reduce the computational complexity of combination when different pieces of evidence are relevant to different levels of specificity in a hierarchy of diseases and they suggested an approach to approximating Dempster's rule in this case. Later on this approach was strengthened by Shafer and Logan [Shafer and Logan, 1987] by an exact implementation of Dempster's rule in the case of hierarchical evidence shown in Figure 4.1. The mass functions which can be combined in such a structure are also limited to special cases where for each mass function, its focal element can only be a node in such a tree or the complement of a node or the whole frame. Their algorithm can be carried out in time linear in $\mid \Theta \mid$.

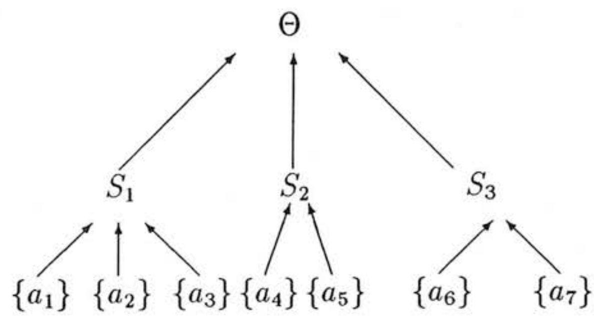

Figure 4.3 Hierarchical evidence space

Computational complexity in DS theory has been discussed in some special cases by Barnett, Shafer and Logan. Their work can be further generalized to a class of trees of partitions or variables which are called Markov Trees. The problem of propagating belief functions in such Markov trees was discussed in [Shafer, Shenoy and Mellouli, 1987]. Propagating belief functions in networks of variables has also been studied by Kong [Kong, 1987] and Mellouli [Mellouli, 1987]. A different approach for reducing the complexity of Dempster's combination rule has also been studied using Mobius transform of a graph in [Kennes, 1992] and [Kennes and Smets, 1990].

Voorbraak in [Voorbraak, 1988] argued that the computational problem in DS theory also arises when calculating belief functions from mass functions. He defines a Bayesian approximation of a belief function and shows that combining Bayesian approximations of belief functions is computationally less complex than combining belief functions. An approximation method for belief functions has also been studied in [Dubois and Prade 1990]. Recently Tessem [Tessem, 1993] provided an evaluation of their approximations.

4.4.2 Heuristic knowledge representation in DS theory

Another limitation of DS theory is that it lacks the ability to represent heuristic knowledge when the theory is applied to expert systems. Several attempts have also been made to extend the theory to tackle this difficulty [Bonissone and Tong 1985], [Ginsberg, 1984], [Liu, 1986], [Yen, 1988], [Liu, Hughes and McTear, 1992] and [Liu, Hughes and McTear, 1994].

Consider the following piece of heuristic knowledge: if X *is* X_1, then Y *is* Y_1 with a degree of belief r_1. If we get a piece of evidence which says that X *is* X_1 with a degree of belief a_1, by invoking this rule we should be able to obtain the corresponding degree y_1 for Y *is* Y_1. Certainly the value of y_1 must be a function F of a_1 and r_1 (i.e. $y_1 = F(a_1, r_1)$).

Generally, we suppose that a set of heuristic rules R includes:

R_1: if E_1 then H_{11} with a degree of belief r_{11};
H_{12} with a degree of belief r_{12};

...

R_2: if E_2 then H_{21} with a degree of belief r_{21};

H_{22} with a degree of belief r_{22};

...

...

R_n: if E_n then H_{n1} with a degree of belief r_{n1};

H_{n2} with a degree of belief r_{n2};

...

where $E_1, E_2, ..., E_n$ are values (or propositions) of the variable E, and E_i is called an antecedent of rule R_i. H_{ij} in rule R_i is a subset of the values (or propositions) of the variable H and it is called one of the conclusions of rule R_i. r_{ij} is called a rule strength.

Assume we have a piece of evidence which says that E_1 is confirmed with degree a_1, E_2 is confirmed with degree a_2, ..., E_n is confirmed with degree a_n, how can we solve the following problems:

1. what conditions should $\sum_i a_i$ satisfy?

2. what conditions should $\sum_j r_{ij}$ satisfy?

3. what is the function F to determine h_{ij}(the degree of belief on H_{ij}) from those a_i and r_{ij}?

4. if more than one set of rules is invoked and the same conclusion H_{ij} is obtained, what will be the final degree of belief on H_{ij} from those $h_{ij}, ..., h_{kl}$?

Generally, if the variable E is a Cartesian product of variables $A, B, ..., C$, that is, each E_i is in a form of $(A_i \, and \, B_j and ... \, and \, C_k)$, assuming we know the evidence for $A, B, ..., C$, then

5. what is the function F' to determine the degree of belief on the premise $(A_i \; and B_j \; and ... \; and \; C_k)$?

These problems have been modelled in fuzzy theory using a fuzzy extension of modal logic, based on Zadeh's concepts of necessity and possibility (Prade 1981). They were also solved in Mycin's certainty factor model (Shortliffe and Buchanan 1976). Can these problems be solved in Dempster-Shafer theory?

In [Yen, 1988], Yen used probabilistic mappings between two sets to replace multivalued mappings in DS theory in order to represent a heuristic rule as stated above. Later on in [Liu et al, 1993], [Liu, Hughes and McTear, 1992] and [Liu, Hughes and McTear, 1994], this approach is extended to more general cases by using evidential mappings to replace multivalued mappings. The advantage of this method and its comparison with Bayesian conditional probabilities [Pearl, 1988], with the approaches used in [Ginsberg, 1984] and [Hau and Kashyap, 1990] have been fully discussed in [Liu, Hughes and McTear, 1994].

Such extended theory is able to describe any kind of knowledge and to make inference in practice.

4.4.3 Open world assumptions by Smets

DS theory is ofter criticized for combining two almost conflicting pieces of information (provided they are DS-Independent). We can illustrate this problem with the following example:

Example 4.4

Suppose we have a murder case with three suspects: Peter, Paul and Mary and two witnesses [Smets, 1988]. Table 4.7 presents the degree of belief of each witness about who might be the murderer and the combined result of these two pieces of evidence.

suspect	witness1	witness2	combined result	unnormalized result
	m_1	m_2	m_{12}	m'_{12}
Peter	0.99	0.00	0.00	0.00
Paul	0.01	0.01	1.00	0.0001
Mary	0.00	0.99	0.00	0.00
{}	0.00	0.00	0.00	0.9999

Table 4.7 Result of combination

Zadeh [Zadeh, 1984] does not accept this solution, as it gives full certainty to a solution (Paul) that is hardly supported at all from the two witnesses. Looking at the column Unnormalized Result m'_{12}, this indicates that the 0.9999 portion of this belief has been committed to the empty set.

Smets [Smets, 1988] argued that in such a situation the meaning of the empty set should be reconsidered. Generally when one considers a problem, one constructs three sets, the *Known as Possible (KP)* set including those propositions that are known to be possible, the *Unknown Proposition (UP)* set including those propositions for which one has no idea whether they are possible or impossible, and the *Known as Impossible (KI)* set including those propositions known as impossible. In DS theory the *Unknown Proposition (UP)* set is always empty, and a frame of discernment is the *Known as Possible (KP)* set. In the above example, one solution to the conflicting result is to accept the *Unknown Proposition* set and believe that the real murderer must be a fourth person. Another method for handling the present inconsistency is that a meta-level belief should be used to consider the reliability of the witnesses. Smets [Smets, 1988] addressed that certainly discounting is one way to take into account this meta-level belief. If a further piece of evidence says *The murderer is necessarily one of the group Peter, Paul and Mary*, we have to accept the *Unknown Proposition* set is empty and $m(\{\})$ should be 0. Therefore $m(Peter) = m(Mary) = 0$ and $m(Paul) = 1$.

4.5 Summary of DS theory

4.5.1 Our argument in this chapter

In this chapter, we have reviewed DS theory from the perspective of probability theory and tried to clarify the independence requirement in DS theory by defining the original information level and the target information level. We argue that any independent judgement in DS theory should be made explicitly at the original level. Purely considering Dempster's combination rule without examining the original information will cause problems. However Dempster's combination rule does not give us (or require from us) any information about what the original sources are. So the conclusion we get from the above analysis is that those counterintuitive examples given in some articles [Black,1987], [Hunter,1987], [Lingras and Wong, 1990], [Nguyen and Smets, 1991] and [Pearl, 1988], some of examples in [Voorbraak, 1991][4] (like examples 2.4 and 3.3 in his paper) are caused by ignoring the independent requirement defined by Dempster in his combination framework. In the sense of DS-Independence required by Dempster's combination framework, those examples don't satisfy this requirement, so Dempster's combination framework is not applicable. However if we purely consider Dempster's combination rule and believe that those examples satisfy the independent requirement needed by Dempster's combination rule then Dempster's combination rule is applicable, but the combined results are counterintuitive. From the former point of view, they are caused by the misapplication of the framework, from the latter

[4]Some of the examples in [Voorbraak, 1991] show the weakness of the definition of frames. The author argues that the accuracy of reasoning result depends on at which level the frame is constructed such as example 4.1. Some other examples explain that even though Dempster's combination rule can be used in some situations, the results are still counterintuitive (violate with common sense) like example 6.12. This problem is also discussed in [Smets, 1988], [Zadeh, 1984] and [Zadeh, 1986]. Readers can refer to those papers if interested in more details.

point of view they are caused by the weakness of the combination formula. Neither of them is able to deal with those cases. Based on such a discussion, those belief functions, which can only be viewed as generalized probabilities, are precisely the cases which fail to satisfy the requirement of DS-Independence. So Dempster's combination rule is not suitable to cope with them.

4.5.2 Summary

The basics of DS theory can be summaried as follows.

DS theory has two advantages:

- Representing ignorance due to lack of information
- Combining multiple pieces of evidence using Dempster's combination rule

DS theory also has the following limitations:

- The computational complexity in using Dempster's combination rule is very high
- The application areas of Dempster's combination rule are rather limited
- There is a difficulty in representing heuristic knowledge

Chapter 5

A comprehensive comparison between generalized incidence calculus and the Dempster-Shafer theory of evidence

In Chapter 3 we have extended the original incidence calculus. Extended incidence calculus has the advantages to represent ignorance and combine evidence. These advantages are also possessed by DS theory. In this chapter, we provide a comprehensive comparison between extended incidence calculus and DS theory. We will compare these two theories in both representing and combining evidence. We will prove that 1) they have the same ability in representing evidence. 2) any two pieces of evidence which can be combined using Dempster's combination rule, can also be combined in incidence calculus by applying Theorem 2 and they obtain the same results. So Dempster's combination rule and Theorem 2 are totally equivalent. 3) those dependent cases which can be combined by the new Combination Rule (but not Theorem 2) cannot be combined by Dempster's combination rule [Liu and Bundy, 1994].

A few examples are used to demonstrate the similarities and differences between these two theories.

5.1 Comparison I: Representing Evidence

In DS theory, a mass function is defined on a frame Θ. Given a set of propositions P, P may not be a frame. However the basic elements set of P, $\mathcal{A}t$, is a frame. In [Fagin and Halpern, 1989a], any arbitrarily defined frame Θ is taken to be a subset of some $\mathcal{A}t$ (in fact, given a Θ, it is always possible to define $\mathcal{A}t = \Theta$ for a proper P). In this chapter, we follow the same idea and use $\mathcal{A}t$ to denote any frame of discernment.

Given a DS structure $(\mathcal{A}t, bel)$ and a generalized incidence calculus theory $< \mathcal{W}, \mu, P, \mathcal{A}, i >$ (where the frame $\mathcal{A}t$ is the basic element set of P), we say that this DS structure is equivalent to the generalized incidence calculus theory if for any $A \subseteq \mathcal{A}t$, $bel(A) = p_*(\phi_A)$. Here ϕ_A is defined as

$$\phi_A = \vee \delta_j \ where \ \delta_j \in A$$

That is if we use DS theory to describe the degree of belief on $\mathcal{A}t$, then we consider $\mathcal{A}t$ to be a frame, but if we use incidence calculus to describe the degree of belief on $\mathcal{A}t$, then we consider $\mathcal{A}t$ to be a collection of the basic elements formed from P. Therefore, a subset A of $\mathcal{A}t$ in $2^{\mathcal{A}t}$ is treated as equivalent to the formula $\vee \delta_j$ (where $\delta_j \in A$) in $\mathcal{L}(\mathcal{A}t)$.

Theorem 4 *For any DS structure $(\mathcal{A}t, bel)$, there is an equivalent generalized incidence calculus theory $< \mathcal{W}, \mu, \mathcal{A}t, \mathcal{A}, i >$.*

PROOF

Given a DS structure $(\mathcal{A}t, bel)$, suppose $A_{DS} = \{A_1, ...A_n\}$ is the focal element set of belief function bel and m is its mass function, then $\Sigma m(A_j) = 1$.

1) create a set of possible worlds $\mathcal{W} = \{w_1, ...w_n\}$ and let $\mu(w_j) = m(A_j)$.

2) let a subset $\mathcal{A}$ of $\mathcal{A}t$ be $\{\phi_{A_j} \mid A_j \in A_{DS}\}$;

3) define the basic incidence assignment ii as $ii(\phi_{A_j}) = \{w_j\}$;

4) define the incidence function i from ii as $i(\phi_A) = \cup_{\phi_{A_j} \in \mathcal{A}, \phi_{A_j} \models \phi_A} ii(\phi_{A_j})$

Then $< \mathcal{W}, \mu, \mathcal{A}t, \mathcal{A}, i >$ is a generalized incidence calculus theory (it is easy to prove that i has the features of Definition 3.1 in Chapter 3).

For any formula ϕ_A in $\mathcal{L}(\mathcal{A}t)$ and its related subset A of $\mathcal{A}t$, we have

$$\begin{aligned} p_*(\phi_A) &= \mu(i_*(\phi_A)) \\ &= \mu(\bigcup_{\phi_{A_j} \models \phi_A} i(\phi_{A_j})) \\ &= \mu(\bigcup_{\phi_{A_j} \models \phi_A} (\bigcup_{\phi_{A_l} \models \phi_{A_j}} ii(\phi_{A_l}))) \\ &= \mu(\bigcup_{\phi_{A_l} \models \phi_A} ii(\phi_{A_l})) \\ &= \Sigma_{\phi_{A_l} \models \phi_A} \mu(ii(\phi_{A_l})) \\ &= \Sigma_{\phi_{A_l} \models \phi_A} \mu(\{w_l\}) \\ &= \Sigma_{A_l \subseteq A} m(A_l) \\ &= bel(A) \end{aligned}$$

Then the belief function $bel(A)$ is exactly the same as $p_*(\phi_A)$.

□

So $pls(A) = 1 - bel(\neg A) = 1 - p_*(\neg \phi_A) = \mu(\mathcal{W} \setminus i_*(\neg \phi_A)) = p^*(A)$.

This theorem tells us that the belief function on frame $\mathcal{A}t$ given by a DS structure is the same as the lower bound of the probabilities on the formulae if we think of $\mathcal{A}t$ as a basic element set. Therefore, any belief function can be obtained as a lower bound from a generalized incidence calculus theory.

Example 5.1

The example used here is originally from [Fagin and Halpern, 1989b] and simplified by [Correa da Silva and Bundy, 1990b] as follows.

> A person has four coats: two are blue and single-breasted, one is grey and double-breasted and one is grey and single-breasted. To choose

which colour of coat this person is going to wear, one tosses a (fair) coin. Once the colour is chosen, to choose which specific coat to wear the person uses a mysterious nondeterministic procedure which we don't know anything about. What is the probability of the person wearing a single-breasted coat?

We solve this problem by using DS theory first and then deal with it in generalized incidence calculus.

DS structure: Let $P = \{g, d\}$ where g stands for "the coat is grey" and d stands for "the coat is double-breasted", then we have

$$\mathcal{A}t = \{g \wedge d, \neg g \wedge d, g \wedge \neg d, \neg g \wedge \neg d\}$$

which is a frame. The element $\neg g \wedge d$ in this frame is false because there is no coat which is not grey but double-breasted. So the real frame of discernment is reduced to;

$$\mathcal{A}t = \{g \wedge d, g \wedge \neg d, \neg g \wedge \neg d\}$$

According to the story that one tosses a (fair) coin to decide which colour to choose, we can define a mass function on the frame $\mathcal{A}t$ as

$$m(\{\neg g \wedge \neg d\}) = 0.5 \qquad m(\{g \wedge \neg d, g \wedge d\}) = 0.5$$

with the focal element set A_{DS} as

$$A_{DS} = \{\{\neg g \wedge \neg d\}, \{g \wedge \neg d, g \wedge d\}\}$$

Therefore, we have a DS structure $(\mathcal{A}t, bel)$. The degree of belief on $\neg d$ is $bel(\neg d) = m(\neg g \wedge \neg d) = 0.5$ and the degree of plausibility is 1.

The degrees of belief and plausibility say that the probability of the person wearing a single-breasted coat lies somewhere between 0.5 to 1 which cannot be measured in a single number.

Generalized incidence calculus theory: Based on the story we could have two possible worlds: w_1 for blue and single-breasted coats and w_2 for grey coats. The probability of each of the possible worlds is 0.5.

Given a set of propositions P and its basic element set $\mathcal{A}t$ as defined in DS structure, we know that w_1 supports formula $\neg g \wedge \neg d$ and w_2 makes the formula $(g \wedge \neg d) \vee (g \wedge d)$ true. So we define $i(\neg g \wedge \neg d) = \{w_1\}$ and $i(g) = \{w_2\}$. Then $< \mathcal{W}, \mu, P, \mathcal{A}, i >$ is a generalized incidence calculus theory.

From this generalized incidence calculus theory, we have that

$$i_*(\neg d) = i(\neg g \wedge \neg d)$$

$$i^*(\neg d) = \mathcal{W} \setminus i_*(d) = \mathcal{W}$$

so

$$p_*(\neg d) = 0.5 \qquad p^*(\neg d) = 1$$

which is identical to the result from DS theory.

Theorem 5 *For any generalized incidence calculus theory* $< \mathcal{W}, \mu, P, \mathcal{A}, i >$, *there is an equivalent DS structure* $(\mathcal{A}t, bel)$.

PROOF

Suppose $< \mathcal{W}, \mu, P, \mathcal{A}, i >$ is a generalized incidence calculus theory and ii is the corresponding basic incidence assignment,

1) define a subset $\mathcal{A}_{DS}$ of $\mathcal{A}t$ as $A_{DS} = \{A \mid \phi_A \in \mathcal{A}\}$.

2) if $\bigcup_{\phi_A} ii(\phi_A) \neq \mathcal{W}$, then $A_{DS} := A_{DS} \cup \{\mathcal{A}t\}$ where $ii(\mathcal{A}t) := \mathcal{W} \setminus \bigcup_{\phi_A} ii(\phi_A)$.

3) define $m(A_j) = \mu(ii(\phi_{A_j}))$ where $A_j \in A_{DS}$. Then $\Sigma_{A_j} m(A_j) = 1$.

So *bel*: $bel(A) = \Sigma_{B \subseteq A} m(B)$ gives a belief function on $\mathcal{A}t$ and we obtain a DS structure $(\mathcal{A}t, bel)$.

For any formula ϕ_A in $\mathcal{L}(\mathcal{A}t)$ and its related subset A of $\mathcal{A}t$, we have

$$
\begin{aligned}
p_*(\phi_A) &= \mu(i_*(\phi_A)) \\
&= \mu(\bigcup ii(\phi_B) \mid \phi_B \models \phi_A, \phi_B \in \mathcal{A}) \\
&= \Sigma_{\phi_B}(\mu(ii(\phi_B)) \mid \phi_B \models \phi_A, \phi_B \in \mathcal{A}) \\
&= \Sigma(m(B) \mid B \subseteq A, B \in A_{DS}) \\
&= bel(A)
\end{aligned}
$$

Therefore, $< \mathcal{W}, \mu, P, \mathcal{A}, i >$ and $(\mathcal{A}t, bel)$ are equivalent.

□

Example 5.2

This example demonstrates the procedure of producing a DS structure based on a given generalized incidence calculus theory as indicated in Theorem 5. This weather forecasting example continues the story in Chapter 3.

> Assume we know that on *fri, sat, sun, mon* it will rain and on *mon, wed, fri* it will be windy. The question we are interested in is on which days it will not rain.

Generalized incidence calculus theory: Let a set of possible worlds $\mathcal{W}$ be $\{sun, mon, tues, wed, thus, fri, sat\}$ and they have equal probability *i.e.* $\mu(w_i) = 1/7$ and let $P = \{rainy, windy\}$. The incidence function defined out of the above story is

$$i(rainy) = \{fri, sat, sun, mon\}$$

$$i(windy) = \{mon, wed, fri\}$$

the basic incidence assignment ii is

$$ii(rainy \wedge windy) = \{fri, mon\}$$

$$ii(rainy) = \{sat, sun\}$$

$$ii(windy) = \{wed\}$$

$$ii(\mathcal{A}t) = \{tues, thur\}$$

and the basic element set $\mathcal{A}t$ is

$$\mathcal{A}t = \{rainy \wedge windy, rainy \wedge \neg windy, \neg rainy \wedge windy, \neg rainy \wedge \neg windy\}$$

Therefore the generalized incidence calculus theory is

$$< \mathcal{W}, \mu, P, \mathcal{A}, i >$$

where $\mathcal{A} = \{rainy, windy, rainy \wedge windy\}$.

From this theory, we have

$$i_*(\neg rainy) = \{\}$$

$$i^*(\neg rainy) = \{tues, wed, thus\}$$

so

$$p_*(\neg rainy) = 0 \qquad p^*(\neg rainy) = 3/7$$

That is we cannot be sure on which day it will not rain but possibly on Tuesday, Wednesday or Thursday.

DS structure: For frame $\mathcal{A}t$ as defined above, we can derive a mass function m on it based on Theorem 4 as

$$m(rainy \wedge windy) = 2/7$$

$$m(rainy) = 2/7$$

$$m(windy) = 1/7$$

$$m(\mathcal{A}t) = 2/7$$

So we have $bel(\neg rainy) = 0$ and $pls(rainy) = 3/7$. The DS structure $(\mathcal{A}t, bel)$ gives the same result as incidence calculus.

A similar result has also been achieved in [Correa da Silva and Bundy, 1990b]. In their paper, it is proved that any original incidence calculus theory is equivalent to a Total Dempster-Shafer probability structure, and any Total Dempster-Shafer probability structure is equivalent to an original incidence calculus theory. In this paper, we have generalized incidence calculus theories and shown generalized incidence calculus theories are totally equivalent to DS structures.

5.2 Comparison II: Combining DS-Independent evidence

For any two DS structures $(\mathcal{A}t, bel_1)$ and $(\mathcal{A}t, bel_2)$, if we assume that the two belief functions are derived from two DS-independent pieces of evidence, then these two belief functions can be combined using Dempster's combination rule. From these two DS structures, two generalized incidence calculus theories can also be produced, and their combination leads to the third generalized incidence calculus theory using Theorem 3 in Chapter 3. What we need to prove in such a situation is that the combined result of the two DS structures turns out to be equivalent to the combined generalized incidence calculus theory.

Theorem 6 *Suppose $(\mathcal{A}t, bel_1)$ and $(\mathcal{A}t, bel_2)$ are two DS structures and bel_1 and bel_2 are obtained from the two DS-independent pieces of evidence and assume that the combined DS structure is $(\mathcal{A}t, bel)$. Further let $< \mathcal{W}_1, \mu_1, \mathcal{A}t, \mathcal{A}_1, i_1 >$ and $< \mathcal{W}_2, \mu_2, \mathcal{A}t, \mathcal{A}_2, i_2 >$ be the two generalized incidence calculus theories produced from $(\mathcal{A}t, bel_1)$ and $(\mathcal{A}t, bel_2)$, and $< \mathcal{W}, \mu, \mathcal{A}t, \mathcal{A}, i >$ be the combined generalized incidence calculus theory, then $(\mathcal{A}t, bel)$ is equivalent to $< \mathcal{W}, \mu, \mathcal{A}t, \mathcal{A}, i >$. That is, for any subset A of $\mathcal{A}t$,*

$$bel(A) = p_*(\phi_A)$$

Our proof is divided into two parts. In part one we need to prove that the conflict weight k in the combined DS structure is equal to $\mu(\mathcal{W}_0)$ in the combined generalized incidence calculus theory. In part two we need to prove that $bel(A) = p_*(\phi_A)$ for any $A \subseteq \mathcal{A}t$.

Because bel_1 and bel_2 are derived from two DS-independent pieces of evidence, $(\mathcal{W}_1, \mu_1)$ and $(\mathcal{W}_2, \mu_2)$ are DS-independent. So Theorem 2 is used to combine these two derived generalized incidence calculus theories.

PROOF

Suppose the two focal element sets in these two DS structures $(\mathcal{A}t, bel_1)$ and $(\mathcal{A}t, bel_2)$ are

$$A_{DS} = \{A_1, A_2, ..., A_n\} \ \Sigma m_1(A_l) = 1$$

$$B_{DS} = \{B_1, B_2, ..., B_m\} \ \Sigma m_2(B_j) = 1$$

The combined DS structure is $(\mathcal{A}t, bel)$ with bel defined as $bel_1 \oplus bel_2$.

Furthermore the two sets of axioms in the corresponding two generalized incidence calculus theories $< \mathcal{W}_1, \mu_1, P, \mathcal{A}_1, i_1 > < \mathcal{W}_2, \mu_2, P, \mathcal{A}_2, i_2 >$ are:

$$\mathcal{A}_1 = \{\phi_{A_1}, \phi_{A_2}, ..., \phi_{A_n}\}, ii_1(\phi_{A_l}) = \{w_{1l}\}, \mu_1(w_{1l}) = m_1(A_l)$$

$$\mathcal{A}_2 = \{\psi_{B_1}, \psi_{B_2}, ..., \psi_{B_m}\}, ii_2(\phi_{B_j}) = \{w_{2j}\}, \mu_2(w_{2j}) = m_2(B_j)$$

Part One

Part one proves $k = \mu(\mathcal{W}_0)$ where k is the weight of the conflict between these two DS structures, and $\mathcal{W}_0$, which is defined in Chapter 3, is the conflict set in the combined generalized incidence calculus theory.

Suppose $m = m_1 \oplus m_2$, for any $A_l \cap B_j = \{\}$, $(A_l \in A_{DS}, B_j \in B_{DS})$ $m_1(A_l)m_2(B_j)$ will be part of k. That is $k = k' + m_1(A_l)m_2(B_j)$.

For ϕ_{A_l} from A_l ($\phi_{A_l} \in \mathcal{A}_1$) and ψ_{B_j} from B_j ($\phi_{B_j} \in \mathcal{A}_2$), we have $\phi_{A_l} \wedge \psi_{B_j} \models \perp$. So

$$
\begin{aligned}
\mu(\mathcal{W}_0) &= \mu(\textstyle\bigcup_{\phi_{A_l}\wedge\psi_{B_j}\models\perp} i_1(\phi_{A_l}) \otimes i_2(\psi_{B_j}))\\
&= \mu(\textstyle\bigcup_{\phi_{A_l}\wedge\psi_{B_j}\models\perp}(\bigcup_{\phi_{A'_l}\models\phi_{A_l}} ii_1(\phi_{A'_l})) \otimes (\bigcup_{\psi_{B'_j}\models\psi_{B_j}} ii_2(\psi_{B'_j})))\\
&= \mu(\textstyle\bigcup_{\phi_{A_l}\wedge\psi_{B_j}\models\perp}(\bigcup_{\phi_{A'_l}\wedge\psi_{B'_j}\models\phi_{A_l}\wedge\psi_{B_j}} ii_1(\phi_{A'_l}) \otimes ii_2(\psi_{B'_j})))\\
&= \mu(\textstyle\bigcup_{\phi_{A'_l}\wedge\psi_{B'_j}\models\perp} ii_1(\phi_{A'_l}) \otimes ii_2(\psi_{B'_j}))\\
&= \Sigma(\mu_1(ii_1(\phi_{A'_l}))\mu_2(ii_2(\psi_{B'_j})) \mid \phi_{A'_l} \wedge \psi_{B'_j} \models\perp)\\
&= \Sigma(\mu_1(w_{1l'})\mu_2(w_{2j'}) \mid \phi_{A'_l} \wedge \psi_{B'_j} \models\perp)\\
&= \Sigma(m_1(A'_l)m_2(B'_j) \mid A'_l \cap B'_j = \{\})\\
&= k
\end{aligned}
$$

Part Two

For any subset C of $\mathcal{A}t$, and its corresponding formula φ_C, we need to prove that $bel(C) = p_*(\varphi_C)$.

For $A_l \in A_{DS}$ and $B_j \in B_{DS}$, if $A_l \cap B_j \subseteq C$, then $m_1(A_i)m_2(B_j)$ is a part of $bel(C)$.

For ϕ_{A_l} from A_l and ψ_{B_j} from B_j, we have $\phi_{A_l} \wedge \psi_{B_j} \models \varphi_C$. So

$$
\begin{aligned}
p_*(\varphi_C) &= \mu(i_*(\varphi_C))\\
&= \mu(\textstyle\bigcup_{\phi_{A_l}\wedge\psi_{B_j}\models\varphi_C} i_1(\phi_{A_l}) \otimes i_2(\psi_{B_j}))\\
&= \Sigma_{\phi_{A_l}\wedge\psi_{B_j}\models\varphi_C}(\mu_1(i_1(\phi_{A_l}))\mu_2(i_2(\psi_{B_j})))/(1-k)\\
&= \Sigma_{\phi_{A_l}\wedge\psi_{B_j}\models\varphi_C}(\mu_1(\textstyle\bigcup_{\phi_{A'_l}\models\phi_{A_l}} ii_1(\phi_{A'_l}))\mu_2(\bigcup_{\psi_{B'_j}\models\psi_{B_j}} ii_2(\psi_{B'_j})))/(1-k)\\
&= \Sigma_{\phi_{A_l}\wedge\psi_{B_j}\models\varphi_C}(\mu_1(\textstyle\bigcup_{\phi_{A'_l}\models\phi_{A_l}} \{w_{1l'}\})\mu_2(\bigcup_{\psi_{B'_j}\models\psi_{B_j}} \{w_{2j'}\}))/(1-k)\\
&= \Sigma_{\phi_{A'_l}\wedge\psi_{B'_j}=\varphi_{C'},\varphi_{C'}\models\varphi_C}(\mu_1(\{w_{1l'}\})\mu_2(\{w_{2j'}\}))/(1-k)\\
&= \Sigma_{C'\subseteq C, A'_l\cap B'_j=C'}(m_1(A'_l)m_2(B'_j))/(1-k)\\
&= \Sigma_{C'\subseteq C} m(C')\\
&= bel(C)
\end{aligned}
$$

□

Example 5.3 Combining two DS-independent pieces of evidence using both Dempster's combination rule and Theorem 2 in generalized incidence calculus.

Using Dempster's combination rule

Assume that we have two DS structures $(\mathcal{A}t, bel_1)$ and $(\mathcal{A}t, bel_2)$ with the following additional information.

$$\mathcal{A}t = \{a, b, c, d\}$$

$$A_{DS} = \{\{a, b, c\}, \mathcal{A}t\}$$

$$B_{DS} = \{\{c, d\}, \mathcal{A}t\}$$

$$m_1(\{a, b, c\}) = 0.7, m_1(\mathcal{A}t) = 0.3$$

$$m_2(\{c, d\}) = 0.6, m_2(\mathcal{A}t) = 0.4$$

Here A_{DS} and B_{DS} are the focal element sets for belief functions bel_1 and bel_2 respectively. Combining these two belief functions derived from m_1 and m_2, we get a joint belief function as shown in Table 5.1.

A / m	$\{a, b, c\}$ / 0.7	$\mathcal{A}t$ / 0.3
$\{c, d\}$ / 0.6	$\{c\}$ / 0.42	$\{c, d\}$ / 0.18
$\mathcal{A}t$ / 0.4	$\{a, b, c\}$ / 0.28	$\mathcal{A}t$ / 0.12

Table 5.1 Combination of two DS-independent pieces of evidence

From this table, it is possible to calculate degrees of belief on any subsets of $\mathcal{A}t$. For instance, for subset $\{a, b, c\}$, we have $bel(\{a, b, c\}) = 0.42 + 0.28 = 0.7$ and $pls(\{a, b, c\}) = 1$.

Using the incidence calculus combination rule

From the two DS structures given above, we are able to form two generalized incidence calculus theories from them as

$$< \mathcal{W}_1, \mu_1, P, \mathcal{A}_1, i_1 >$$

$$< \mathcal{W}_2, \mu_2, P, \mathcal{A}_2, i_2 >$$

with the following additional information

$$\mathcal{W}_1 = \{w_{11}, w_{12}\}, \mu_1(w_{11}) = 0.7, \mu_1(w_{12}) = 0.3$$

$$P := \mathcal{A}t$$

$$\mathcal{A}_1 = \{a \vee b \vee c, \mathcal{A}t\}$$

$$i_1(a \vee b \vee c) = \{w_{11}\}, i_1(\mathcal{A}t) = \mathcal{W}_1$$

$$\mathcal{W}_2 = \{w_{21}, w_{22}\}, \mu_2(w_{21}) = 0.6, \mu_2(w_{22}) = 0.4$$

$$P := \mathcal{A}t$$

$$\mathcal{A}_2 = \{c \vee d, \mathcal{A}t\}$$

$$i_2(c \vee d) = \{w_{21}\}, i_2(\mathcal{A}t) = \mathcal{W}_2$$

As $(\mathcal{W}_1, \mu_1)$ and $(\mathcal{W}_2, \mu_2)$ are DS-independent, Theorem 2 in Chapter 3 is used to combine these two incidence calculus theories as given in Table 5.2.

ϕ_A $i(\phi_A)$	$a \vee b \vee c$ $\{w_{11}\}$	$true$ $\mathcal{W}_1$
$c \vee d$ $\{w_{21}\}$	c $\{w_{11}\} \otimes \{w_{21}\}$ 0.42	$c \vee d$ $\mathcal{W}_1 \otimes \{w_{21}\}$ 0.6
$true$ $\mathcal{W}_2$	$a \vee b \vee c$ $\{w_{11}\} \otimes \mathcal{W}_2$ 0.7	$true$ $\mathcal{W}_1 \otimes \mathcal{W}_2$ 1

Table 5.2 Combination of two DS-independent generalized incidence calculus theories

The combined generalized incidence calculus theory is $< \mathcal{W}_1 \otimes \mathcal{W}_2, \mu, P, \mathcal{A}, i >$ where $\mu(< w_{1l}, w_{2j} >) = \mu_1(w_{1l})\mu_2(w_{2j})$. From this theory, we are also able to obtain the degree of our belief in any formula. For example, $p_*(a \vee b \vee c) = \mu(i_*(a \vee b \vee c)) = 0.7$ and $p^*(a \vee b \vee c) = 1$ which are the same as we got in DS theory.

Comparing Table 5.1 with Table 5.2 we will find that these two structures give the same result (numerically) on any subset (or formula). We will also find that whenever a numerical value (mass value) appears in Table 5.1, a corresponding incidence set replaces its position in Table 5.2. The combination procedure in generalized incidence calculus combines possible worlds instead of numbers. The degree of belief in a formula is calculated based on the incidence set.

Now it has been proved that what we can combine using Dempster's combination rule can also be combined in incidence calculus and they obtain the same result. Moreover in the next section we are going to show that we can handle a wider range of information in incidence calculus by applying the new combination rule.

5.3 Comparison III: Combining Dependent Evidence

In this section, we first show an example which can be dealt with using the combination rule in incidence calculus but cannot be dealt with using Dempster's combination rule. We then explore the theoretical difference between these two theories and explain why DS theory fails to deal with dependent evidence while incidence calculus succeeds.

Example 5.4

This example is from [Voorbraak, 1991]. There are 100 labelled balls in an urn as given in Table 5.3.

Label	Number of Balls	Subset Name in $\mathcal{W}$
axy	4	W_1
ax	4	W_2
ay	16	W_3
a	16	W_4
bxy	10	W_5
bx	10	W_6
by	20	W_7
b	20	W_8

Table 5.3. 100 balls and their labels

Suppose X and Y are separate observations of drawing a ball from the urn. The information carried by them is:

X: the drawn ball has label x;

Y: the drawn ball has label y.

Based on these two pieces of evidence, we are interested in knowing the degree of our belief that the drawn ball also has label b.

Using Dempster's combination rule:

Let a set of propositions P be $\{a, b\}$. a stands for a proposition 'The drawn ball has label a' and b stands for the proposition 'The drawn ball has label b'. Then the basic element set $\mathcal{A}t$ is the same as P which is a frame. Two mass functions are defined on $\mathcal{A}t$ based on the information carried by the two observations X and Y as:

$$m_X(a) = 2/7, \;\; m_X(b) = 5/7$$

$$m_Y(a) = 2/5, \;\; m_Y(b) = 3/5$$

where $m_X(a)$ is the mass value on a given by observation X which represents the possibility of a ball having label a when the ball is observed having label x and $m_Y(a)$ is the mass value on a given by observation Y which represents the possibility of a ball having label a when the ball is observed having label y.

The result of applying Dempster's combination rule to the above two mass functions is $m(b) = m_X \oplus m_Y(b) = 15/19$. So $bel(b) = 15/19$.

While in probability theory, the probability that a ball has both label x and y is

$$p(x)p(y) = 0.28 \times 0.5 = 0.14 = p(x \wedge y)$$

Therefore, we have $p(b \mid x \wedge y) = 5/7$. Obviously the results obtained in DS theory and in probability theory are not the same and the result given in DS theory is wrong. See the detailed analysis of the example in [Voorbraak, 1991].

Using the incidence calculus combination rule:

Let us examine this example in incidence calculus theory. First of all, we suppose that the set of possible worlds $\mathcal{W}$ contains 100 labelled balls.

$$\mathcal{W} = W_1 \cup W_2 \cup W_3 \cup W_4 \cup W_5 \cup W_6 \cup W_7 \cup W_8$$

where W_1 contains 4 possible worlds each of which specifies a ball with labels xya, ..., W_8 contains 20 possible worlds each of which specifies a ball with label b. The probability distribution on $\mathcal{W}$ is $\mu(w) = 1/100$ for any $w \in \mathcal{W}$. We further suppose the set of propositions P contains $\{a, b, x, y\}$ where a means that the chosen ball has label a etc.

From observations X and Y, it is possible to construct two generalized incidence calculus theories

$$< \mathcal{W}, \mu, P, \mathcal{A}_1, i_1 >$$

$$< \mathcal{W}, \mu, P, \mathcal{A}_2, i_2 >$$

where

$$i_1(x) = W_1 \cup W_2 \cup W_5 \cup W_6$$

$$i_1(a \wedge x) = W_1 \cup W_2, \quad i_1(b \wedge x) = W_5 \cup W_6$$

and

$$i_2(y) = W_1 \cup W_3 \cup W_5 \cup W_7$$

$$i_2(a \wedge y) = W_1 \cup W_3, \quad i_2(b \wedge y) = W_5 \cup W_7$$

where $\mathcal{A}_1 = \{x, a \wedge x, b \wedge x\}$ and $\mathcal{A}_2 = \{y, a \wedge y, b \wedge y\}$.

Applying the Combination Rule proposed in incidence calculus to these two theories, we can get the third incidence calculus theory $< \mathcal{W}, \mu, P, \mathcal{A}, i >$ with $\mathcal{A} = \{x \wedge y, a \wedge x \wedge y, b \wedge x \wedge y, a \wedge b \wedge x \wedge y\}$

$$i(b \wedge x \wedge y) = W_5 \qquad i(x \wedge y) = W_1 \cup W_5$$

$$i(a \wedge x \wedge y) = W_1 \qquad i(a \wedge b \wedge x \wedge y) = \{\}$$

It is easy to prove that for any $\phi \in \mathcal{A}$, $i_*(\phi) = i^*(\phi) = i(\phi)$, so a function p, defined as $p(\phi) = p_*(\phi) = \mu(i(\phi))$, is a probability function on $\mathcal{A}$. Further because $p(b \wedge x \wedge y) = 10/100$ and $p(x \wedge y) = 14/100$, according to Equation (3.3) in Chapter 3, we have

$$p(b \mid x \wedge y) = \frac{p(b \wedge x \wedge y)}{p(x \wedge y)} = \frac{\mu(i(b \wedge x \wedge y))}{\mu(i(x \wedge y))} = 5/7$$

This result is consistent with what we could get in probability theory.

Now we try to explain theoretically why Dempster's combination rule cannot be used in this case. In fact, the two mass functions are derived from two probability spaces (S_1, μ_1) and (S_2, μ_2) where $S_1 = W_1 \cup W_2 \cup W_5 \cup W_6$, $\mu_1(s) = 1/28$ and $S_2 = W_1 \cup W_3 \cup W_5 \cup W_7$ and $\mu_2(s) = 1/50$. These two probability spaces are defined from the unique space $(\mathcal{W}, \mu)$ and they share the information carried by the subset $W_1 \cup W_5$. Therefore Dempster's combination rule cannot be used to combine the two mass functions derived from the two probability spaces.

In incidence calculus, instead of combining numbers on set $\mathcal{A}t$, we combine two pieces of evidence symbolically at the original information level, *i.e.*, at the probability space level. For the above example, since the two probability spaces are somehow related to the unique space $(\mathcal{W}, \mu)$, we establish two generalized incidence functions from $\mathcal{W}$ to P rather than from S_1 and S_2 to P respectively. Therefore it is possible to cancel the overlapped information carried by the two observations. Because DS theory is a purely numerical uncertainty reasoning mechanism, it is not possible to combine evidence symbolically. So it is not possible to represent and cancel the joint (or overlapped) part of the information provided by two pieces of evidence.

Therefore, we conclude that even though the two theories have the same ability in representing evidence and combining DS-independent information, their theoretical structures are rather different. The essence of incidence calculus, *indirect encoding of probabilities of formulae,* makes it possible to cancel the effect of overlapped information and provide an alternative combination mechanism which combines dependent information. Although trying to combine dependent information at the probability space level has been considered in [Shafer, 1982] and in [Lingras and Wong, 1990], no unique rule was provided in DS theory for general cases because of the theoretical limitation of the theory.

5.4 Analysing Examples

In this section we reanalyze Example 4.2 and 4.3 used in Chapter 4. Shafer used these two example to show the idea that Dempster's combination framework can be used to deal with these two cases while Dempster's combination rule can only be used in the first case. We will show that the combination rule we proposed can deal with both cases.

Example 5.5

Example 4.2 is about to combine two DS-independent pieces of evidence in DS theory. The main point in that example is that Frad's report is irrelevant to the indication of the indoor-outdoor thermometer. So the two mass functions derived from these two sources can be combined using Dempster's combination rule. In

Example 4.2, three spaces, X_1, X_2 and S are involved. Two of them carry the evidence we collect and one is the target domain we would like to reason about. The corresponding probability spaces for the first two spaces are (X_1, X_1, p_{X_1}) and (X_2, X_2, p_{X_2}). Because these two probability spaces satisfy the DS-Independence requirement, the joint probability space is $(X_1 \otimes X_2, X_1 \otimes X_2, p_{X_1} \otimes p_{X_2})$. In order to combine these two pieces of evidence in incidence calculus, we need to describe them in incidence calculus terminology. From the two evidential spaces and their multivalued mappings to S (in this case, S is a set of propositions), two generalized incidence calculus theories are formed as

$$< X_1, p_{X_1}, S, \mathcal{A}_1, i_1 >$$

$$< X_2, p_{X_2}, S, \mathcal{A}_2, i_2 >$$

where

$$\mathcal{A}_1 = \{yes, yes \vee no\}$$

$$i_1(yes) = \{truthful\},\ \ i_1(yes \vee no) = \{careless\}$$

$$\mathcal{A}_2 = \{no, yes \vee no\}$$

$$i_2(no) = \{working\},\ \ i_2(yes \vee no) = \{not\}$$

Applying Theorem 2 in Chapter 3 to these two theories, we get the third generalized incidence calculus theory $< X, \mu, S, \mathcal{A}, i >$, where

$$\mathcal{W}_0 = \{< truthful, working >\}$$

$$X = X_1 \otimes X_2 \setminus \mathcal{W}_0$$

$$\mu(< x_1, x_2 >) = \frac{p_{X_1}(x_1) \times p_{X_2}(x_2)}{1 - \Sigma_{< x_1', x_2' > \in \mathcal{W}_0} p_{X_1}(x_1') \times p_{X_2}(x_2')}$$

$$\mathcal{A} = \{yes, no, yes \vee no\}$$

$$i(yes) = i(yes \wedge (yes \vee no)) = i_1(yes) \otimes i_2(yes \vee no) \setminus \mathcal{W}_0 = \{< truthful, not >\}$$

$$i(no) = i((yes \vee no) \wedge no) = i_1(yes \vee no) \otimes i_2(no) \setminus \mathcal{W}_0 = \{< careless, working >\}$$

$i(yes \vee no) = (i_1(yes) \otimes i_2(yes \vee no) \cup i_1(yes \vee no) \otimes i_2(no) \cup i_1(yes \vee no) \otimes i_2(yes \vee no)) \setminus \mathcal{W}_0 = \{< truthful, not >, < careless, working >, < careless, not >\}$

So we have $p_*(yes) = \mu(i(yes)) = .04$, $p_*(no) = .95$ which are the degrees of our belief in the two answers. This result is identical with the result obtained in Example 4.2. The detailed combination procedure is shown in Table 5.4.

ϕ $i(\phi)$	yes $\{truthful\}$	$yes \vee no = true$ $\{careless\}$
no $\{working\}$	$\perp$ $\{truthful\} \otimes \{working\}$	no $\{careless\} \otimes \{working\}$
$yes \vee no = true$ $\{not\}$	yes $\{truthful\} \otimes \{not\}$	$true$ $\{careless\} \otimes \{not\}$

Table 5.4 Combination of two DS independent complete DS structures

Example 5.6

Example 4.3 differs from Example 4.2 in the way that it does not assume that Fred's report is irrelevant to the indication of the thermometer. Rather, these two sources are interrelated in some ways. Although it is similar to what we did in Example 5.5 that there are still three spaces X_1, X_2 and S, and the joint space of X_1 and X_2 is still $X_1 \otimes X_2$, however, the joint probability measure on X is no longer $p_{X_1} \otimes p_{X_2}$ because of some dependency relation between the two probability measures on X_1 and X_2. So Dempster's combination rule cannot be applied and we cannot use Theorem 2 on this case any more. The actual probability measure is replaced by a new measure, denoted as μ, which is given in the second column of Table 4.5 to take into account of Fred's careless 'if the thermometer is not working'. The new probability space will be (X, X, μ). From the compatibility relations between X_1, X_2 and S, two incidence functions are decidable from X to S as

$i_1(yes) = \{< truthful, working >, < truthful, not >\}$

$$i_1(yes \vee no) = \{< careless, working >, < careless, not >\}$$

$$\mathcal{A}_1 = \{yes, yes \vee no\}$$

and

$$i_2(no) = \{< truthful, not >, < careless, not >\}$$

$$i_2(yes \vee no) = \{< truthful, working >, < careless, working >\}$$

$$\mathcal{A}_2 = \{no, yes \vee no\}$$

So the two corresponding incidence calculus theories are

$$< X, \mu, S, \mathcal{A}_1, i_1 >$$

$$< X, \mu, S, \mathcal{A}_2, i_2 >$$

Combining them using the rule given above, we get the third incidence calculus theory $< X', \mu', S, \mathcal{A}, i >$ where $\mathcal{W}_0 = \{< truthful, working >\}$, $X' = X \setminus \mathcal{W}_0$, and μ' is shown as in the third column of the Table 4.5. From this combined incidence calculus theory, the degrees of our belief in 'yes' is .005 and in 'no' is .95. It is once again the same as Shafer got in Example 4.3 in DS theory but this result cannot be obtained by using the Dempster's combination rule.

If we examine Examples 4.2, 4.3, 5.5, 5.6 carefully, we will find that the conflict set $\mathcal{W}_0$ in incidence calculus is always equivalent to the set of elements Shafer rules out from the joint set, and μ' is exactly the posterior probability measure given by Shafer. Therefore the new combination rule in incidence calculus covers the idea proposed by Dempster in his combination framework. The calculation of conflict set $\mathcal{W}_0$ and the modification of probability distribution μ' are automatic. The new rule integrates all these procedures into one mechanism in a natural way.

5.5 Comparison IV: Some Other Aspects Between The Two Theories

Apart from the comparisons between these two theories in the above three aspects, in this section, I will reveal some minor relations among these two theories in the following two aspects.

5.5.1 Similar but different mapping relations in the two theories

In both the original incidence calculus and in extended incidence calculus, we keep mentioning a mapping function, incidence function i. This mapping function is the most important component in a generalized incidence calculus theory (and in an original incidence calculus theory). An incidence function between a set of possible worlds $\mathcal{W}$ and a set of formulae $\mathcal{A}$ says that for each formula ϕ in $\mathcal{A}$, there is a subset W_ϕ in $\mathcal{W}$ corresponding to it. A probability distribution on $\mathcal{W}$ is discrete. That is for each $w \in \mathcal{W}$, $\mu(w)$ is known.

In DS theory, if we know that a belief function *bel* on $\mathcal{A}t$ is derived from a probability space (S, χ, μ'), then there must be a mapping function Γ between S and $\mathcal{A}t$. For each s in S, Γs is a subset of $\mathcal{A}t$. The probability distribution μ' on S is not necessarily discrete.

DS theory and generalized incidence calculus are equivalent in generating bounds on formulae (or sets). Bundy in [Bundy, 1992] said: Both systems[1] permit only partial definition of the probabilities of some formulae. DS theory achieves this by defining the incidence of all formulae, but not defining the probabilities of all the possible worlds, *i.e.*, Γ is a total function, but μ' is a partial function.

[1]Means DS theory and incidence calculus.

Incidence calculus achieves a similar effect the other way round, *i.e.*, μ is total but i is partial.

It is possible to draw a diagram below to demonstrate the similarity and the difference between these two mapping functions.

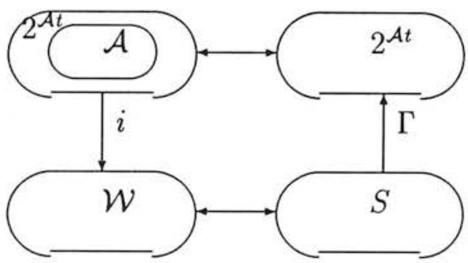

Figure 5.1. Γ in DS and i in incidence calculus

Because the direction of mapping function i goes down in incidence calculus while the direction of mapping function Γ goes up, it is possible to propose a new combination rule at the symbolic level. It seems that it is not possible to propose a similar combination mechanism in DS theory due to this difference.

5.5.2 Recovering mass functions

In DS theory, when a belief function *bel* is known, its mass functions can be recovered from it when the frame $\mathcal{A}t$ is finite. This is particular necessary when Dempster's combination rule is applied because this rule only applies on mass functions.

Since *bel* on a frame $\mathcal{A}t$ is equivalent to the lower bound of a probability distribution calculated from the lower bound of incidences on $\mathcal{A}t$, the application of Algorithm B in Chapter 3 on a DS structure $(\mathcal{A}t, bel)$ can recover its corresponding mass function [Liu, Bundy and Robertson, 1993a]. This is given in Algorithm C below.

Algorithm C

Given a function bel on the set $\mathcal{L}(P) = \mathcal{A}$, determine whether bel is a belief function on this language set [2] and obtain its mass function if it is.

Step 1: Delete all those elements in $\mathcal{A}$ in which $bel(*) = 0$. Then as in algorithm B, define a subset $\mathcal{A}_0$ out of $\mathcal{A}$. For any $\phi \in \mathcal{A}_0$, define $m(\phi) = bel(\phi)$. Assume that there are l elements in $\mathcal{A}_0$. Define $\mathcal{A}' = \mathcal{A} \setminus \mathcal{A}_0$.

Step 2: Chose a formula ψ from $\mathcal{A}'$ which satisfies the condition that $\forall \psi' \in \mathcal{A}'$, $\psi' \not\models \psi$.

For all $\phi_j \in \mathcal{A}_0$ repeat $bel(\psi) := bel(\psi) - bel(\phi_j)$ when $\phi_j \models \psi$.

If $bel(\psi) > 0$, define

$$
\begin{aligned}
l &:= l + 1 \\
\mathcal{A}_0 &:= \mathcal{A}_0 \cup \{\psi\} \\
\mathcal{A}' &:= \mathcal{A}' \setminus \{\psi\} \\
m(\psi) &:= bel(\psi)
\end{aligned}
$$

If $bel(\psi) = 0$ then ψ is not a focal element of this belief function.

If $bel(\phi) < 0$ then this assignment is not a belief function, stop the procedure.

Repeat this step until $\mathcal{A}'$ is empty.

Step 3: All the elements in $\mathcal{A}_0$ will be the focal elements of this belief function and the function m defined in Step 2 is the corresponding mass function. It is easy to prove that $\Sigma_A m(A) = 1$.

The algorithm tries to find the focal elements of a belief function one by one. Once all the focal elements are fixed and the uncertain values of these elements

---

[2] In fact, this language set can be any frame of discernment.

are defined, the corresponding mass function is known. The worst case of computational complexity of this algorithm is the same as the approach used in DS theory but it may be more efficient when the elements in $\mathcal{A}'$ are arranged in the decreasing sequence of their sizes.

Example 5.7

Assume that there are four elements in $\mathcal{A}t = \{a, b, c, d\}$ and $\mathcal{A} = \mathcal{L}(P)$ is $\mathcal{A} = \{a, b, c, d, a \vee b, a \vee c, a \vee d, b \vee c, b \vee d, c \vee d, a \vee b \vee c, a \vee c \vee d, a \vee b \vee d, b \vee c \vee d, a \vee b \vee c \vee d\}$ and the corresponding degrees of belief in elements of $\mathcal{A}$ are $bel(\mathcal{A}) = \{.5, 0, 0, .3, .7, .5, 8, 0, .3, .3, .7, 8, 1, .3, 1\}$.

By using the Algorithm C, the calculating procedure for a mass function is as follows.

Step 1. After deleting those elements with 0 degrees of belief, we have $\mathcal{A} = \{a \vee b \vee c \vee d, b \vee c \vee d, a \vee b \vee d, \ a \vee c \vee d, a \vee b \vee d, c \vee d, \ b \vee d, a \vee d, a \vee c, a \vee b, d, a\}$ and $\mathcal{A}_0 = \{a, d\}$. Define $m(a) = bel(a) = .5$, $m(d) = bel(d) = .3$, $l = 2$ and $\mathcal{A}' := \mathcal{A} \setminus \mathcal{A}_0$.

Step 2. Get $a \vee c$ from $\mathcal{A}'$. Because $a \models a \vee c$, we have $bel(a \vee c) := bel(a \vee c) - bel(a) = .5 - .5 = 0$. So $a \vee c$ is not a focal element. Repeat this procedure until we get $a \vee b$ and we have $bel(a \vee b) := .7 - .5 = .2$. Define

$$m(a \vee b) = bel(a \vee b) = .2$$
$$\mathcal{A}_0 := \mathcal{A}_0 \cup \{a \vee c\}$$
$$\mathcal{A}' := \mathcal{A}' \setminus \{a \vee c\}$$
$$l := l + 1$$

Repeat this procedure until $\mathcal{A}'$ is empty, we get $\mathcal{A}_0 = \{a, d, a \vee b\}$ and the mass function m is $m(a) = .5, m(d) = .3, m(a \vee c) = .2$.

If we take bel as an inner measure of a probability on $\mathcal{A}$ from an unknown probability space, this space can be recovered as $(\mathcal{W}, \chi, \mu)$ where the basis for χ is $\chi' = \{W_1, W_2, W_3\}$, $W_1 \cup W_2 \cup W_3 = \mathcal{W}$ and $\mu(W_1) = .5, \mu(W_2) = .3, \mu(W_3) = .2$.

5.6 Summary

In this chapter, we have made a comprehensive comparison between DS theory and incidence calculus on their abilities in the following three aspects: 1) representing evidence; 2) combining DS-independent evidence; 3) dealing with dependent evidence. We conclude that these two theories have the same ability in representing incomplete information and combining DS-independent evidence. However, incidence calculus is superior to DS theory in coping with overlapped information. This difference results from their different theoretical structures. DS theory is a pure numerical approach while incidence calculus possesses both symbolic and numerical features. That is incidence calculus can make an inference either at the symbolic level by producing incidence sets or on the numerical basis by calculating lower or upper bounds on probabilities of formulae. The new combination rule in incidence calculus is proposed based on the symbolic feature of the theory.

Trying to combine dependent pieces of information using Dempster's combination rule has been mentioned in [Dubois and Prade,1986], [Kennes, 1991], [Lingras and Wong, 1990], [Nguyen and Smets, 1991], [Shafer, 1986], [Smets, 1990], [Shafer, 1987a]. Some of their work focuses on how to improve Dempster's combination rule to deal with dependent situations as in [Dubois and Prade,1986], [Kennes, 1991], [Smets, 1990], [Nguyen and Smets, 1991]. In [Smets, 1990] and [Nguyen and Smets, 1991], the authors discussed the possible ways of combining dependent information in the transferable belief model. [Kennes, 1991] showed the way of solving this problem in the concept of *category*. [Dubois and Prade,1986] intended to model this problem in terms of set-theory in fuzzy logic. The approaches in [Lingras and Wong, 1990], [Shafer, 1986], [Shafer, 1987a] are closer to ours in this paper. In [Shafer, 1986], [Shafer, 1987a] Shafer showed that some dependent evidence can be combined within DS theory by following Dempster's framework, but cannot be combined by Dempster's combination rule. We have seen this in Chapter 4. Similar ideas also appeared in [Lingras and Wong, 1990] in the compatibility view defined by the authors. In [Lingras and Wong, 1990] two pieces of evidence are combined at the original information level and then propagated to the target frame. This combination relies on the compatibility relation among the two spaces which, they assume, is defined by the user and the probability distribution is defined based on whether the two spaces are probabilistically independent. Two original evidence spaces are assumed to be probabilistically independent if the spaces are logically independent. By logically independent, they mean that if every element in X is compatible with all elements in S then X and S are said to be logically independent. If two spaces are logically independent, the case can be dealt with using Dempster's rule. Otherwise, the authors use either a Bayesian approach or dependency functions to get a joint probability on the unified space. However the definition of logically independent on two spaces is not sufficient to guarantee that the two probability distributions are independent as we have seen in Example 4.3 given by Shafer.

In summary, all these papers we referred to above tried to either re-explain Dempster's combination rule in alternative terminologies, modify or complement the current Dempster combination rule from different angles. None of them tried to give a new combination mechanism which keeps the spirit of Dempster's combination idea but is distinct from Dempster's combination rule.

In contrast to the approaches above, we have proposed a new rule and tried to combine several pieces of evidence at the original information level. In our new combination rule, original sources are required in order to carry out the combination, so it is natural and convenient to make the independent judgement using the sources. Dempster's combination rule is a special case of this alternative rule when several sources are DS-independent. That is, when the several sources are DS-Independent the rule in incidence calculus can be simplified to be Dempster's combination rule. The main advantage of the rule is to unify the combination procedure and propagation procedure into one using incidence functions. Also, the definition of incidence functions makes incidence calculus possess some features of symbolic reasoning patterns which differ from other pure numerical mechanisms.

In general, independent relations among multiple sources of evidence can be considered as special cases of dependent situations. As Pearl indicated [Pearl, 1992],

> "If we have several items of evidence, each depending on the state of nature, these items of evidence should also depend on each other. This kind of dependency is not a nuisance but a necessary bliss; no evidential reasoning would otherwise be possible."

In our combination rule, we have indeed adopted the same idea and made some efforts towards combining dependent evidence. This result would be useful for further research work on either this topic or the relevant topics. It tells us that it is a promising way to cancel the overlapped and duplicated information from several pieces of evidence at the symbolic level rather than at the numerical level.

Chapter 6

Assumption-based truth maintenance systems

In this chapter, I am going to review assumption-based truth maintenance systems (ATMS). The main purpose of this chapter is to provide a basis for the discussion in the next chapter. I shall not review all of the work on the ATMS carried out so far. Rather I will only focus on two aspects of the ATMS: its basic reasoning mechanism and the possibility of associating it with numerical uncertainty mechanisms. After I have made these two aspects clear, we are then ready to walk through the next chapter which presents the third main contribution of this thesis.

6.1 The Reasoning Mechanism in The ATMS

The truth maintenance system (TMS) [Doyle, 1979] and later the ATMS [de Kleer, 1986] are both symbolic approaches to maintaining consistent sets of statements. The central issue in such a system is that for each statement, a set of supporting statements (called labels or environments generally in the ATMS) need to be produced. This set of supporting statements are obtained through a set of arguments attached to the given statement (called justifications). In an ATMS, a justification of a statement (or node) contains other statements (or nodes) from which the current statement can be derived. Justifications are specified by the system designer.

For instance, if we have two inference rules such as:

$$r_1 : q_1 \rightarrow q_2$$

$$r_2 : q_2 \rightarrow q_3$$

then logically we can infer that $r_3 : q_1 \rightarrow q_3$. In an ATMS, if r_1, r_2 and r_3 are represented by $node_1$, $node_2$ and $node_3$ respectively, then $node_3$ is derivable from the conjunction of $node_1$ and $node_2$ and we call (r_1, r_2) a justification of $node_3$. A rule may have several justifications. Furthermore, if r_1 and r_2 are valid under the conditions that A and B are true, respectively, then rule r_3 is valid under the condition that $A \wedge B$ is true, denoted as $\{A, B\}$. $\{A\}, \{B\}$ and $\{A, B\}$ are called sets of supporting statements (or environments) of r_1, r_2 and r_3, respectively. If we associate $node_3$ with the supporting statements such as $\{A, B\}$ and the dependent nodes such as (r_1, r_2) then $node_3$ is generally of the form of

$$r_3 : q_1 \rightarrow q_3, \{\{A, B\}...\}, \{(r_1, r_2)...\}$$

when $node_3$ has more than one justification. The collection of all the possible sets of supporting environments is called the label of a node. If we use $L(r_3)$ to denote the label of $node_3$, then $\{A, B\} \in L(r_3)$. If we assume that r_1, r_2 hold without requiring any dependent relations with other nodes, then $node_1$ and $node_2$ are represented as

$$r_1 : q_1 \rightarrow q_2, \{\{A\}\}, \{()\}$$

$$r_2 : q_2 \rightarrow q_3, \{\{B\}\}, \{()\}$$

Therefore, we can infer a label for any node as long as its justifications are known. For instance, if we know that the justifications of node r_n are

$$\{(r_{11}, ..., r_{t1}), ..., (r_{1l}, ..., r_{t'l})\}$$

and $L(r_{11})$,..., $L(r_{t'l})$ are known, then the label set of r_n can be obtained as

$$L(r_n) = \cup_k \{x \mid x = \cup_j x_j \ \ where \ \ x_j \in L(n_{jk})\}$$

and we denote this as

$$L(r_n) = \cup_k (L(r_{1k}) \otimes ... \otimes L(r_{jk}))$$

The advantage of this reasoning mechanism is that the dependent and supporting relations among nodes are explicitly specified, in particular, the supporting relations among assumptions and other nodes. This is obviously useful when we want to retrieve the reasoning path. It is also helpful for belief revision.

The limitation of this reasoning pattern is that we cannot infer those statements which are probably true rather than absolutely true. However, if we attach numerical degrees of belief to the elements in the supporting set of a node, we may be able to infer a statement with a degree of belief. For example, if we know that A is true with probability 0.8, B is true with probability 0.7 and both A and B are probabilistically independent, then the probability of $A \wedge B$ being true is 0.56 which is the product of 0.7 and 0.8. The belief in a node is considered as the probability of its label. So for $node_3$, our belief in it is 0.56 if we assume that $node_3$ has only environment $\{A, B\}$. Otherwise, 0.56 is our minimum degree of belief on $node_3$. Therefore, we are able to calculate the probability of a node through the probabilities on assumptions in its label set. In this way, a production rule $a \rightarrow b$ with rule strength m can be rewritten as [Pearl, 1988]

$$a \wedge C \rightarrow b$$

In this expression, C stands for how strong we believe in b is we know that a is true. C has different meaning in different systems. It is thought of the rule strength m in an expert system but it is called an assumption in the ATMS.

Some of research work towards this goal has been shown in [d'Ambrosio, 1988], [d'Ambrosio, 1990], [de Kleer and Williams, 1987], [Dubois, etal, 1990], [Fulvio Monai and Chehire, 1992], [Laskey and Lehner, 1989], [Pearl, 1988], [Proven, 1989] and [Liu, Bundy and Robertson, 1993b].

One common limitation in all these extensions of the ATMS is that the probabilities assigned to assumptions must be assumed probabilistically independent in order to calculate the degree of belief in a statement.

6.2 Non-Redundant Justification Sets and Environments

In this section, I first introduce some formal concepts used in the ATMS and then discuss sets of justifications and environments in detail.

node: a node (called a problem-solver's datum) in an ATMS represents any datum unit used in the system. This datum unit can be a proposition or any formula in the propositional language which the system uses. The dependencies among them are inferred during the system processing procedure.

assumptions: a set of distinguished nodes which are believed to be true without requiring any precondition are called assumptions.

justifications: justifications are supplied by the problem-solver. A justification for a node contains those nodes from which it can be derived. Usually, a node has several justifications representing multiple ways to infer the node.

label: a set of assumptions is called *an environment* of a node if the node holds in this environment. The label of a node contains all collections of such environments. Each environment in a label consists of non-redundant assumptions.

nogood: there is a nogood node in an ATMS system whose label consists of all environments in which falsity can be derived.

In an ATMS, each node is associated with a label and a set of justifications and the node is normally denoted as

$$< node_i, label, justifications >$$

The inference procedure in the ATMS propagates assumptions along justifications.

Both the label and the justifications for a node can be explained using material implication. Given a node c with label $\{\{A_1, A_2, ...\}\ \{B_1, B_2, ...\}...\}$ and with justifications $\{(z_1,\ z_2,\ ...)\ (y_1,\ y_2,\ ...)...\}$, the meaning of the label of c is that the conjunction of assumptions in each environment makes c true, such as $A_1 \wedge A_2...$ of environment $\{A_1, A_2...\}$ makes c true. So $L(c)$ is a set containing conjunctions of assumptions. $L(c) = \{(A_1 \wedge A_2 \wedge ...), (B_1 \wedge B_2 \wedge ...)...\}$. The label implies the node:

$$(A_1 \wedge A_2 \wedge ...) \vee (B_1 \wedge B_2 \wedge ...) \vee ... \to c$$

The relations between a justification and its node states that the conjunction of $z_i(y_j)$ logically supports the conclusion c. If we consider z_i and c as formulae in a propositional language, then $\wedge_i z_i$ is a formula in the language which implies c, that is, formula $\wedge_i z_i \to c$ is always true. In general if we let

$$j(c) = \{(z_1 \wedge z_2 \wedge ...), (y_1 \wedge y_2 \wedge ...)...\}$$

then every element in $j(c)$ semantically implies c, so $j(c) \models c$. Therefore there is the similar implication relation:

$$(z_1 \wedge z_2 \wedge ...) \vee (y_1 \wedge y_2 \wedge ...) \vee ... \to c$$

In general each environment is nonredundant. That is, deleting any element in an environment will destroy the implication relation between this environment and its node. For any two environments for one node, they don't imply each other. That is one environment is not a proper subset of another. If one environment is a subset of another, then the latter environment will be covered by the former one so the latter can be deleted. The same rules also apply to the justifications for a node. It is assumed in an ATMS that any justification is nonredundant and any two justifications don't imply each other. I will illustrate this in the following example [Laskey and Lehner, 1989].

Example 6.1

Assume that we have five inference rules such as:

$$r_1 : e \to d$$

$$r_2 : d \rightarrow b$$

$$r_3 : b \rightarrow a$$

$$r_4 : d \rightarrow c$$

$$r_5 : c \rightarrow a$$

and there are five assumptions Z, X, V, Y and W supporting them respectively. Then these five rules can be encoded into a set of ATMS nodes as[1]

$$\begin{array}{ll} node_1 : & < e \rightarrow d, \{\{Z\}\}, \{(Z)\} > \\ node_2 : & < d \rightarrow b, \{\{X\}\}, \{(X)\} > \\ node_3 : & < b \rightarrow a, \{\{V\}\}, \{(V)\} > \\ node_4 : & < d \rightarrow c, \{\{Y\}\}, \{(Y)\} > \\ node_5 : & < c \rightarrow a, \{\{W\}\}, \{(W)\} > \end{array}$$

Similarly we encode another two inference rules in this ATMS as

$$\begin{array}{ll} node_6 : & < d \rightarrow a, \{\{X, V\}, \{Y, W\}\}, \{(node_2, node_3), (node_4, node_5)\} > \\ node_7 : & < e \rightarrow a, \{\{Z, X, V\}, \{Z, Y, W\}\}, \{(node_1, node_6)\} > \end{array}$$

or in $node_7$'s justification, replacing $node_6$ by its justification set

$$\begin{array}{ll} node_7 : & < e \rightarrow a, \{\{Z, X, V\}, \{Z, Y, W\}\}, \\ & \qquad \{(node_1, node_2, node_3), (node_1, node_4, node_5)\} > \end{array}$$

We should notice that $(node_1, node_2, node_3)$ also implies $node_6$, but it is not in the justification set of $node_6$ as the effect of this justification has been covered by the justification $(node_2, node_3)$ and the justification set of a node should be non-redundant. The same situation applies to $node_7$ as well.

In fact there are in total seven conjunctions of nodes make $node_7$ true, but only two of them are included in the justification set of $node_7$.

---

[1]A node with only an assumption (or assumptions) in both its label and its justifications means that this node is supported and dependent on this assumption (or assumptions) only.

The justification set of a node in an ATMS contains those nodes (the conjunction of them) from which this node can be derived. If we require that a justification set of a node is non-redundant, then deleting any justification from the justification set of a node will result in losing a path which can derive the node.

6.3 Probabilistic Assumption Sets

In an ATMS, all nodes can be divided into four types: *assumptions, assumed nodes, premises,* and *derived nodes.* An assumption node is a node whose label contains a singleton environment mentioning itself, such as $< A, \{\{A\}\}, \{(A)\} >$.

An *assumed node* is a node which has justifications mentioning only assumptions[2]. For instance $< a, \{\{A\}\}, \{(A)\} >$ or $< b, \{\{A, B\}\}, \{(A, B)\} >$. All other nodes are either premises or derived nodes. A *premise* (or a fact) has an empty justification and empty label set, *i.e.*, it holds without any preconditions. A *derived node* usually doesn't include assumptions in its justifications, *e.g.*, $< c, \{\{A, B\}\}, \{(a, b)\} >$. In general, if we keep the restriction that non-assumptions cannot become assumptions, or assumptions cannot become another type of node [de Kleer, 1986], then it is possible to keep all assumptions in one set and the other nodes in another set, and these two sets are distinct.

The inference result of a node has one of three values: Believed, Disbelieved and Unknown. If one of the environments in the label c is believed, then c is believed. If one of the environments in the label $\neg c$ is believed, then c is disbelieved, otherwise c is unknown. When both c and $\neg c$ are believed, there is a conflict and falsity is derived. In this case, the label sets of some nodes should be revised, *e.g.*, delete *nogood* environments in which they appear. This kind of inference in an ATMS produce only three possible values. It cannot represent a plausible

[2]In [de Kleer, 1986], an assumed node has only one justification mentioning one assumption.

conclusion d with a degree of belief. Attempts to attach uncertain numbers with assumptions in the ATMS have appeared in [d'Ambrosio, 1988], [d'Ambrosio, 1990], [de Kleer and Williams, 1987], [Laskey and Lehner, 1989], [Dubois, etal, 1990] and [Fulvio Monai and Chehire, 1992]. The belief of a node is identified as the probability of its label $Bel(c) = Pr(L(c))$.

For example [Pearl, 1988], the rule *Turn the key* $\rightarrow$ *start the engine with* 0.8 can be represented in the ATMS as

$$< b \rightarrow a, \{\{B\}\}, \{(B)\} >$$

where B stands for an assumption (or a set of assumptions) which supports the implication relation $b \rightarrow a$ and assign 0.8 as the probability of B. a and b represent propositions *'start the engine'* and *'turn the key'* respectively.

Assume that for node b we have $< b, \{\{A\}\}, \{(A)\} >$, then the justification for node a is $b \wedge (b \rightarrow a)$ as $b \wedge (b \rightarrow a) \models a$. That is for node a we have

$$< a, \{\{A, B\}\}, \{(b, b \rightarrow a)\} >$$

a is a derived node.

Therefore $Bel(a) = Pr(L(a)) = Pr(A \wedge B) = 0.8$, if the probability distributions are probabilistically independent and the action *'turn the key'* is true, i.e., $p(A) = 1$.

In principle the ATMS has the ability to make plausible inferences with beliefs. For a simple case like the above, the calculation of probabilities on nodes is not difficult to carry out. However, in most cases labels of nodes are very complicated and probability distributions on assumptions may be somehow related. In those circumstances, calculating probabilities of labels of nodes is quite troublesome as shown in [Laskey and Lehner, 1989] and [Pearl, 1988]. We give the following two definitions to cope with this difficulty in general. The motivation of proposing the following two definitions is stimulated by the idea of managing possible worlds in incidence calculus. This part of the work is the extension to the original ATMS. It provides a theoretical basis for associating and managing probabilities in an ATMS. It covers the related work in [Laskey and Lehner, 1989] and [Pearl, 1988].

Definition 6.1: *Probabilistic assumption set*[3]

A set $\{A_1, ..., A_n\}$, denoted as $S_{A_1,...,A_n}$, is called a probabilistic assumption set for assumptions $A_1, ..., A_n$ if the probabilities on $A_1, ..., A_n$ are given by a probability distribution p from a piece of evidence and $\Sigma_{D\in\{A_1,...,A_n\}}p(D) = 1$. The simplest probabilistic assumption set has two elements A and $\neg A$, denoted as $S_{A,\neg A}$. For any two elements in a probabilistic assumption set, it is assumed that $A_i \wedge A_j \Rightarrow \perp$. For all elements in the set, we have $\vee_j A_j = true$ for $j = 1, ..., n$.

For two distinct probabilistic assumption sets $S_{A_1,...,A_n}$ and $S_{B_1,...B_m}$, the unified probabilistic assumption set is defined as $S_{A_1,...,A_n,B_1,...B_m} = S_{A_1,...,A_n} \otimes S_{B_1,...,B_m} = \{(A_i, B_j) \mid A_i \in S_{A_1,...,A_n}, B_j \in S_{B_1,...,B_m}\}$ where $\otimes$ means set product and $p(A_i, B_j) = p_1(A_i) \times p_2(B_j)$. p_1 and p_2 are the probability distributions on $S_{A_1,...,A_n}$ and $S_{B_1,...,B_m}$, respectively.

Example 6.2

Assume that the five assumptions in Example 6.1 are in different probabilistic assumption sets. An environment for $node_6$ derived from justification $\{(node_2, node_3)\}$ is $\{\{X, V\}\}$, then the joint probabilistic assumption set for this environment is $S_{X,\neg X} \otimes S_{V,\neg V}$. Similarly the joint probabilistic assumption set for environment $\{\{Y, W\}\}$ is $S_{Y,\neg Y} \otimes S_{W,\neg W}$.

Definition 6.2: *Full extension of a label*

Assume that an environment of a node n is $\{A, B, ..., C\}$ where $A, B, ..., C$ are in different probabilistic assumption sets $S_{A_1,...,A_x}$, $S_{B_1,...,B_y}$ and $S_{C_1,...,C_z}$. Because $A \wedge B \wedge ... \wedge C = A \wedge B \wedge ... \wedge C \wedge (\vee E_j \mid E_j \in S_{E_1,...,E_t})$, $A \wedge B \wedge ... \wedge C \rightarrow n$ and $A \wedge B \wedge ... \wedge C \wedge (\vee_j E_j \mid E_j \in S_{E_1,...E_t}) \rightarrow n$ are all true (where $S_{E_1,...,E_t}$

[3]A similar definition is given in [Laskey and Lehner, 1989] called an auxiliary hypothesis set.

is a probabilistic assumption set which is different from $S_{A_1,...A_x}, S_{B_1,...,B_y}$ *and* $S_{C_1,...,C_z}$*).* $\{A, B, ..., C\} \otimes S_{E_1,...,E_t}$ *is called a full extension of the environment to* $S_{E_1,...E_t}$*. If there are in total* m *probabilistic assumption sets in the ATMS, then* $\{A, B, ..., C\} \otimes S_{E_1,...,E_t} \otimes ... \otimes S_{F_1,...,F_j}$ *is called the full extension of the environment to all assumptions, or simply called the full extension of the environment. Similarly if every environment in a label has been fully extended to all assumptions, then we call the result the full extension of the label, denoted as* $FL(n)$*.*

To understand the idea behind this definition, we look at Example 6.1 again. There are 5 probabilistic assumption sets in this ATMS structure, $S_{Z,\neg Z}$, $S_{X,\neg X}$, $S_{V,\neg V}$, $S_{Y,\neg Y}$ and $S_{W,\neg W}$. One environment of $node_6$ is $\{X, V\}$ which contains assumptions in two probabilistic assumption sets $S_{X,\neg X}$ and $S_{Y,\neg Y}$. Based on Definition 6.2, the full extension of this environment is

$$\{X, V\} \otimes S_{Z,\neg Z} \otimes S_{Y,\neg Y} \otimes S_{W,\neg W}$$

and the full extension of label $L(node)$ is

$$\{X, V\} \otimes S_{Z,\neg Z} \otimes S_{Y,\neg Y} \otimes S_{W,\neg W} \cup \{Y, W\} \otimes S_{X,\neg X} \otimes S_{V,\neg V} \otimes S_{Z,\neg Z}$$

Similarly, we are able to calculate full extensions for all environments of nodes.

In particular, let $L(\perp)$ represent all inconsistent environments (*i.e. nogood*) and let $FL(\perp)$ represent the full extension of them. If a label of a node is $L(c) = \{\{A_1, A_2, ...\},\{B_1, B_2, ...\}, ...\}$, it means that $(A_1 \wedge A_2 \wedge ...) \vee (B_1 \wedge B_2 \wedge ...) \vee ... \rightarrow c$ is true. After we get the full extension of the label and represent it in disjunctive normal form (a disjunction of conjunctions), we have that $(A_1 \wedge A_2 \wedge ... \wedge B_1 \wedge ...C_1) \vee ... \vee (A_1 \wedge A_2 \wedge ...B_n \wedge ...C_1 \wedge ...) \vee ...(A_1 \wedge A_2 \wedge ... \wedge B_n \wedge ... \wedge C_m) \rightarrow c$ is true, each conjunction in the full extension contains the elements from different probabilistic assumption sets and any two such conjunctions are different. Such a full extension is convenient for calculating uncertainties related to assumptions.

Example 6.3

In Example 6.2, we have two different probabilistic assumption sets for two environments of $node_6$. However the probability of $node_6$ cannot be obtained by calculating them separately and then adding them together. Doing so may over count the joint part in these two sets. The solution to this is to apply Definition 6.2 to each of these environments and we have full extensions for these two environments as

$$S_{Z,\neg Z} \otimes \{X, V\} \otimes S_{Y,\neg Y} \otimes S_{W,\neg W}$$

$$S_{Z,\neg Z} \otimes S_{X,\neg X} \otimes S_{V,\neg V} \otimes \{Y, W\}$$

The full extension of the label of $node_6$ is the union of these two sets.

$$(S_{Z,\neg Z} \otimes \{X, V\} \otimes S_{Y,\neg Y} \otimes S_{W,\neg W}) \cup (S_{Z,\neg Z} \otimes S_{X,\neg X} \otimes S_{V,\neg V} \otimes \{Y, W\})$$

or

$$S_{Z,\neg Z} \otimes (\{X, V\} \otimes S_{Y,\neg Y} \otimes S_{W,\neg W} \cup S_{X,\neg X} \otimes S_{V,\neg V} \otimes \{Y, W\})$$

If we use p_Z to represent the probability distribution on probabilistic assumption set $S_{Z,\neg Z}$, then belief in this node is

$$\begin{aligned}
&Bel(node_6)\\
&= p_Z(S_{Z,\neg Z})(p_X(X)p_V(V)p_Y(S_{Y,\neg Y})p_W(S_{W,\neg W}) + p_X(S_{X,\neg X})p_V(S_{V,\neg V})p_Y(Y)p_W(W)\\
&-p_X(X)p_V(V)p_Y(Y)p_W(W))\\
&= p_Z(S_{Z,\neg Z})(p_X(X)p_V(V) + p_Y(Y)p_W(W) - p_X(X)p_V(V)p_Y(Y)p_W(W))
\end{aligned}$$

In general if the *nogood* environments are not empty, those non-empty environments should be deleted from the label of a node. The probability of a node is then changed to:

$$Bel(node) = Pr(FL(a) \setminus FL(\perp))$$

6.4 Summary

The main purposes of this chapter are to address the concept of non-redundant justification and label sets as well as propose a way of using them in the calculation of probabilities. The discussion in this chapter sets the scene for further discussion on the relations between incidence calculus in the next chapter.

Chapter 7

On the relations between extended incidence calculus and the ATMS

This chapter discusses the relations between extended incidence calculus and the ATMS. I first prove that managing labels for statements (nodes) in an ATMS is equivalent to producing incidence sets of these statements in extended incidence calculus. I then demonstrate that the justification set for a node is functionally equivalent to the implication relation set for the same node in extended incidence calculus. As a consequence, extended incidence calculus can provide justifications for an ATMS because implication relation sets are discovered by the system automatically. I also show that extended incidence calculus provides a theoretical basis for constructing a probabilistic ATMS by associating proper probability distributions on assumptions and the different probability distribution in extended incidence calculus don't necessarily need to be independent. In this way, we can not only produce labels for all nodes in the system, but also calculate the probability of any of such nodes in it. The nogood environments can also be obtained automatically. Therefore, incidence calculus and the ATMS are equivalent in carrying out inferences at both the symbolic level and the numerical level. This extends the result in [Laskey and Lehner, 1989].

7.1 Incidence Calculus Review

Incidence calculus was introduced in [Bundy, 1985], [Bundy, 1992] to deal with problems in purely numerical probabilistic reasoning. The special feature of this reasoning method is the indirect association of numerical uncertainty with formulae. In incidence calculus, probabilities are associated with the elements of a set of possible worlds (denoted as $\mathcal{W}$) and some formulae (called axioms) are associated with the subsets of the set of possible worlds. Each element in such a subset for a formula ϕ makes the formula true and this subset is normally called the incidence set of the formula, denoted as $i(\phi)$ ($i(\phi) \subseteq \mathcal{W}$). Our belief in a formula is regarded as the probability weight of the lower bound of its incidence set. Assume that the set of possible worlds is $\mathcal{W}$ and $q_1 \rightarrow q_2, q_2 \rightarrow q_3$ are two axioms in an incidence calculus theory and the incidence sets for $q_1 \rightarrow q_2$ and $q_2 \rightarrow q_3$ are $i(q_1 \rightarrow q_2) = W_1$ and $i(q_2 \rightarrow q_3) = W_2$, then the incidence set of $(q_1 \rightarrow q_2 \wedge q_2 \rightarrow q_3)$ is $W_1 \cap W_2$. As formula $q_1 \rightarrow q_3$ holds when formula $q_1 \rightarrow q_2 \wedge q_2 \rightarrow q_3$ holds, the incidence set of $q_1 \rightarrow q_2 \wedge q_2 \rightarrow q_3$ must be a subset of the incidence set of $q_1 \rightarrow q_3$. So $W_1 \cap W_2$ makes $q_1 \rightarrow q_3$ true and $W_1 \cap W_2 \subseteq i(q_1 \rightarrow q_3)$.

7.1.1 Essential semantic implication sets in incidence calculus

In this section, I give two more definitions in extended incidence calculus in order to carry out the analysis in the rest of the chapter.

Given a extended incidence calculus theory $< \mathcal{W}, \mu, P, \mathcal{A}, i >$, for any two formulae $\phi, \psi \in \mathcal{A}$, we have $i(\phi) \subseteq i(\psi)$ if $\phi \models \psi$. For any other formula $\phi \in \mathcal{L}(P) \setminus \mathcal{A}$, the lower bound of the incidence set for ϕ is defined as

$$i_*(\phi) = \bigcup_{\psi \in \mathcal{A}, \psi \models \phi} i(\psi) \tag{7.1}$$

The degree of our belief in a formula is defined as $p_*(\phi) = \mu(i_*(\phi))$ as we have seen in Chapter 3.

Definition 7.1: *Semantic implication set*

For any formula $\phi \in \mathcal{L}(P)$, if $\psi \models \phi$ then ϕ is said to be semantically implied by ψ. Let $SI(\phi) = \{\psi \mid \psi \models \phi, \psi \in \mathcal{A}\}$, set $SI(\phi)$ is called a semantical implication set of ϕ.

For instance, $(a \to b) \wedge (b \to c) \models (a \to c)$, if $(a \to b) \wedge (b \to c)$ is in $\mathcal{A}$, then it is in $SI(a \to c)$.

Given a set of axioms $\mathcal{A}$, the SI sets of some formulae may be empty. For instance, if $\mathcal{A} = \{a \to b, b \to c, a \to b \wedge b \to c\}$, then for $e \in \mathcal{L}(P)$, $SI(e)$ is empty.

Definition 7.2: *Essential semantic implication set*

Furthermore, let $ESI(\phi)$ be a subset of $SI(\phi)$ which satisfies the conditions (i) $\phi \in ESI(\phi)$ if $\phi \in SI(\phi)$ and (ii) a formula $\psi \in ESI(\phi)$ if for any $\psi' \neq \psi$ in $SI(\phi)$ then $\psi \not\models \psi'$, then $ESI(\phi)$ is called an essential semantical implication set of ϕ. This is denoted as $ESI(\phi) \models \phi$.

Given a $SI(\phi)$ set, $ESI(\phi)$ contains those 'biggest' formulae in $SI(\phi)$. This means that for any formula $\psi \in ESI(\phi)$, ψ does not imply any other formulae in $SI(\phi)$. For a formula ϕ, if $SI(\phi)$ set is empty, then $ESI(\phi)$ set must be empty. However, if $SI(\phi)$ is not empty, then $ESI(\phi)$ contains at least one element.

Proposition 7 *If $ESI(\phi)$ and $ESI'(\phi)$ are the two essential semantic implication sets for formula ϕ coming from the same incidence calculus theory, then $ESI(\phi) = ESI'(\phi)$.*

PROOF

Suppose that $ESI(\phi)$ and $ESI'(\phi)$ are different and further suppose that a formula ψ is in $ESI(\phi)$ but not in $ESI'(\phi)$. Since $\psi \in ESI(\phi)$ then for any formula $\psi' \in SI(\phi)$, we have that $\psi \not\models \psi'$.

However, as $\psi \notin ESI'(\phi)$, there is at least one formula ψ'' ($\psi'' \in SI(\phi)$) which makes the following equation true $\psi \models \psi''$. So according to Definition 7.2,

$\psi \notin ESI(\phi)$. Conflict. Therefore, $ESI(\phi) = ESI'(\phi)$ and the essential semantic implication set is unique.

QED

It will be proved later that the essential semantic implication set of a formula is exactly the same as the set of justifications of that formula in an ATMS.

Example 7.1

Suppose we have a generalized incidence calculus theory and we know that the following five inference rules are in the language set.

$$r_1 : e \rightarrow d$$

$$r_2 : d \rightarrow b$$

$$r_3 : b \rightarrow a$$

$$r_4 : d \rightarrow c$$

$$r_5 : c \rightarrow a$$

Further suppose that the set of axioms $\mathcal{A}$ contains these five rules and all the possible conjunctions of them, then the lower bounds of incidence set of other formulae can be inferred. For instance, for formula $e \rightarrow a$, the lower bound of its incidence set is

$$i_*(e \rightarrow a) = \bigcup_{\phi \models (e \rightarrow a)} i(\phi)$$

According to Definition 7.1, all the formulae ϕ in $\mathcal{A}$ satisfying the condition that $\phi \models (e \rightarrow a)$ are in the semantic implication set. So the calculation of lower bounds of incidence sets can be restated as:

$$i_*(\psi) = \bigcup_{\phi \in SI(\psi)} i(\phi)$$

In this example, there are in total seven axioms satisfying this requirement, so there are seven axioms in $SI(e \rightarrow a)$.

$$(e \to d) \wedge (d \to b) \wedge (b \to a)$$
$$(e \to d) \wedge (d \to c) \wedge (c \to a)$$
$$(e \to d) \wedge (d \to c) \wedge (c \to a) \wedge (d \to b)$$
$$(e \to d) \wedge (d \to c) \wedge (c \to a) \wedge (b \to a)$$
$$(e \to d) \wedge (d \to c) \wedge (c \to a) \wedge (d \to b) \wedge (b \to a)$$
$$(e \to d) \wedge (d \to b) \wedge (b \to a) \wedge (d \to c)$$
$$(e \to d) \wedge (d \to b) \wedge (b \to a) \wedge (c \to a)$$

However if we examine these seven axioms closely, we will find that only the first two axioms are necessary to be considered if we want to get $i_*(e \to a)$. The rest are unnecessary as their incidence sets are included into the incidence sets of the first two axioms. Based on Definition 7.2, these two axioms are in the essential semantic implication set of $e \to a$ and this set only has these two axioms. Therefore the following proposition is natural.

Proposition 8 *If $SI(\phi)$ and $ESI(\phi)$ are a semantic implication set and an essential semantic implication set of ϕ, then the following equation holds:*

$$i_*(\phi) = i_*(SI(\phi)) = i_*(ESI(\phi))$$

where $i_*(SI(\phi)) = \bigcup_{\phi_j \in SI(\phi)} i(\phi_j)$.

PROOF

Assume a set of axioms in a generalized incidence calculus theory is $\mathcal{A}$. For a formula ϕ, when $\phi \in \mathcal{A}$, we have

$$\phi \in SI(\phi), \phi \in ESI(\phi), ESI = \{\phi\}$$

so

$$i_*(\phi) = i(\phi) = i_*(SI(\phi)) = i_*(ESI(\phi))$$

When $\phi \notin \mathcal{A}$, we have a set of formulae $\phi_1, ..., \phi_n \in \mathcal{A}$ $(n \geq 0)$ each of which implies ϕ. So $SI(\phi) = \{\phi_1, ..., \phi_n\}$. Assume that the elements in $ESI(\phi)$ are

$\psi_1, ..., \psi_m$, then for ψ_j, there will be some formulae $\phi_{j'}$ (at least ψ_j itself) in $SI(\phi)$ which make the following equation hold

$$\phi_{j'} \models \psi_j$$

Let SI_{ψ_j} be a set containing these $\phi_{j'}$, *i.e.* $SI_{\psi_j} = \{\phi_{j'} \mid \phi_{j'} \models \psi_j\}$, then we have $i_*(\psi_j) = i_*(SI_{\psi_j})$ because $i(\phi_{j'}) \subseteq i(\psi_j)$. Repeating this procedure for each formula in $ESI(\psi)$, we obtain the following equation

$$i_*(ESI(\phi)) = \cup_{\psi_j} i_*(SI_{\psi_j})$$

To prove

$$i_*(SI(\phi)) = i_*(ESI(\phi))$$

we need to prove that

$$i_*(SI(\phi)) = \cup_{\psi_j} i_*(SI_{\psi_j})$$

Assume that $i_*(SI(\phi)) \setminus \cup_{\psi_j} i_*(SI_{\psi_j}) = S \neq \{\}$, we have

$S \neq \{\}$ and $w \in S \Rightarrow$
$w \in i_*(SI(\phi)) \setminus \cup_{\psi_j} i_*(SI_{\psi_j}) \Rightarrow$
$(\exists \varphi)\varphi \in SI(\phi), \varphi \notin ESI(\phi), w \in i(\varphi) \Rightarrow$
$(\exists \varphi')\varphi' \in SI(\phi), \varphi \models \varphi', \varphi' \notin ESI(\phi)$ (otherwise $\varphi \in SI_{\varphi'}$ and $\varphi \notin SI(\phi)) \Rightarrow$
$(\exists \varphi'')\varphi'' \in SI(\phi), \varphi' \models \varphi'', \varphi'' \notin ESI(\phi) \Rightarrow$
... (repeat this procedure until we find φ_t)
$(\exists \varphi_t)\varphi_t \in SI(\phi), \varphi_{t-1} \models \varphi_t, \varphi_t \notin ESI(\phi)$ and $\nexists \varphi'_t, \varphi_t \models \varphi'_t$ (as $\mathcal{A}$ is finite) $\Rightarrow$
$\varphi_t \notin ESI(\phi)$ and $\varphi_t \in ESI(\phi)$

Conflict, so S is empty. Therefore, $i_*(SI(\phi)) = i_*(ESI(\phi))$ and $i_*(\phi) = i_*(SI(\phi))$.

END

Based on a extended incidence calculus theory, the efficiency of calculating an incidence set for a formula is very much dependent on the efficiency of finding its semantic implication set as well as the essential semantic implication set.

7.1.2 Similarities of the reasoning models in extended incidence calculus and the ATMS

Abstractly, if we view the set of possible worlds in extended incidence calculus as the set of assumptions in an ATMS, and view the calculation of the incidence sets of formulae as the calculation of labels of nodes in the ATMS, then the two reasoning patterns are similar. Furthermore, as the probability weight of an incidence set can be calculated, extended incidence calculus has associated numerical uncertainty with symbolic reasoning into one mechanism. Extended incidence calculus has no such indications as justifications during its inference procedure. The implication relations are discovered automatically.

The apparent similarity of these two reasoning patterns motivated me to explore their relations more deeply. I focus my attention on the production of labels in the ATMS and calculations of incidence sets in extended incidence calculus. I will prove that the two reasoning mechanisms are equivalent in producing dependent relations among statements. As extended incidence calculus can draw a conclusion with a numerical degree of belief on it, extended incidence calculus actually possesses some features of both symbolic and numerical reasoning approaches. Therefore, extended incidence calculus can be used both as a theoretical basis for the implementation of a probabilistic ATMS by providing both labels and degrees of belief of statements and as an automatic reasoning model to provide justifications for an ATMS.

7.2 Constructing Labels and Calculating Beliefs in Nodes Using Extended Incidence Calculus

7.2.1 An example

Now I will use an example (from [Laskey and Lehner, 1989]) to show how to manage assumptions in the ATMS in the way we manage sets of possible worlds in extended incidence calculus. I will solve this problem using ATMS techniques and extended incidence calculus respectively. The result shows that both inference mechanisms can be used to solve the same problem and the results are the same. It also shows the procedure for transforming an ATMS into extended incidence calculus.

Example 7.2

Assume that we have five inference rules from Example 6.1 and the fact e is observed, we want to infer our belief in other statements, such as a. This is shown in Figure 7.1.

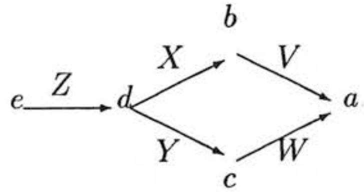

Figure 7.1. semantic network of inference rules

Approach 1: Solving this problem in an ATMS.

Assume that there are the following nodes in the ATMS shown in Figure 7.1 which are put into four categories.

assumed nodes:

$$n_1 :< e \rightarrow d, \{\{Z\}\}, \{(Z)\} >$$

$$n_2 :< d \rightarrow b, \{\{X\}\}, \{(X)\} >$$

$$n_3 :< b \rightarrow a, \{\{V\}\}, \{(V)\} >$$

$$n_4 :< d \rightarrow c, \{\{Y\}\}, \{(Y)\} >$$

$$n_5 :< c \rightarrow a, \{\{W\}\}, \{(W)\} >$$

premise node:

$$n_8 :< e, \{\{\}\}, \{()\} >$$

derived nodes:

$$n_6 :< d \rightarrow a, \{\{X, V\}, \{Y, W\}\}, \{(n_2, n_3), (n_4, n_5)\} >$$
$$n_7 :< e \rightarrow a, \{\{Z, X, V\}, \{Z, Y, W\}\}, \{(n_1, n_6)\} >$$

or replacing n_6 by its own justifications

$$n_7 :< e \rightarrow a, \{\{Z, X, V\}, \{Z, Y, W\}\}, \{(n_1, n_2, n_3), (n_1, n_4, n_5\} >$$
$$n_9 :< a, \{\{Z, X, V\}, \{Z, Y, W\}\}, \{(n_7, n_8)\} >$$

or

$$n_9 :< a, \{\{Z, X, V\}, \{Z, Y, W\}\}, \{(n_1, n_2, n_3, n_8), (n_1, n_4, n_5, n_8)\} >$$

assumption nodes: $< X, \{\{X\}\}, \{(X)\} >$ and so on.

If we are interested in calculating beliefs of nodes, having labels of nodes is not enough [Pearl, 1988], [Laskey and Lehner, 1989]. We would have to manipulate labels in some way in order to get the beliefs. In our approach, we need to obtain the full extension of a label first. In order to do so, probabilistic assumption sets are required and some new assumptions need to be created when necessary. For instance, for the premise node e, we need to associate it with a distinct assumption E, then node n'_8 can be rewritten as $n'_8 :< e, \{\{E\}\}, \{(E)\} >$. There are in total six probabilistic assumption sets. They are $S_{V,\neg V}$, $S_{W,\neg W}$, $S_{X,\neg X}$, $S_{Y,\neg Y}$, $S_{Z,\neg Z}$, $S_{E,\neg E}$.

The labels of derived nodes are obtained based on the justifications given by the problem solver, premise nodes and assumed nodes. The label of proposition a is $L(a) = \{\{Z, X, V\}\{Z, Y, W\}\}$ and its full extension is

$$FL(a) = S_{E,\neg E} \otimes \{Z\} \otimes (\{X, V\} \otimes S_{Y,\neg Y} \otimes S_{W,\neg W} \cup S_{X,\neg X} \otimes S_{V,\neg V} \otimes \{Y, W\})$$

If we assume that different probability distributions on different assumption sets give

$$p_V(V) = .7$$

$$p_W(W) = .8$$

$$p_X(X) = .6$$

$$p_Y(Y) = .75$$

$$p_Z(Z) = .8$$

$$p_E(E) = 1$$

and they are probabilistically independent, then the belief in node a is

$$Bel(a) = Pr(FL(a)) = 1 \times .8 \times (.6 \times .7 + .75 \times .8 - .6 \times .7 \times .75 \times .8) = 0.6144$$

A different calculation procedure can also be found in [Laskey and Lehner, 1989]. which produces the same result.

Approach 2: Using extended incidence calculus to solve the problem.

Now let us see how this problem can be solved in extended incidence calculus. Suppose that we have the following six generalized incidence calculus theories

$$< S_{V,\neg V}, \mu_1, P, \{b \rightarrow a, T\}, \{i_1(b \rightarrow a) = \{V\}, i_1(T) = S_{V,\neg V}\} >$$

$$< S_{W,\neg W}, \mu_2, P, \{c \rightarrow a, T\}, \{i_2(c \rightarrow a) = \{W\}, i_2(T) = S_{W,\neg W}\} >$$

$$< S_{X,\neg X}, \mu_3, P, \{d \rightarrow b, T\}, \{i_3(d \rightarrow b) = \{X\}, i_3(T) = S_{X,\neg X}\} >$$

$$< S_{Y,\neg Y}, \mu_4, P, \{d \rightarrow c, T\}, \{i_4(d \rightarrow c) = \{Y\}, i_4(T) = S_{Y,\neg Y}\} >$$

$$< S_{Z,\neg Z}, \mu_5, P, \{e \rightarrow d, T\}, \{i_5(e \rightarrow d) = \{Z\}, i_5(T) = S_{Z,\neg Z}\} >$$

$$< S_{E,\neg E}, \mu_6(E) = 1, P, \{e\}, \{i_6(e) = \{E\}, i_6(T) = S_{E,\neg E}\} >$$

where $S_V = \{V, \neg V\}$, ..., $S_Z = \{Z, \neg Z\}$, and $S_E = \{E, \neg E\}$ are probabilistic assumption sets.

As we assume that sets of $S_{X,\neg X}, ..., S_{Z,\neg Z}, S_{E,\neg E}$ are probabilistically independent, the combination of the first five theories produces a generalized incidence calculus theory $< S_7, \mu_7, P, \mathcal{A}_7, i_7 >$ in which the joint set is $S_7 = S_{Z,\neg Z} \otimes S_{X,\neg X} \otimes S_{V,\neg V} \otimes S_{Y,\neg Y} \otimes S_{W,\neg W}$.

$$\begin{aligned}
i_7(d \rightarrow b \wedge b \rightarrow a) &= S_{Z,\neg Z}\{X\}\{V\}S_{Y,\neg Y}S_{W,\neg W}{}^1 \\
i_7(d \rightarrow c \wedge c \rightarrow a) &= S_{Z,\neg Z}\{Y\}\{W\}S_{X,\neg X}S_{V,\neg V} \\
i_7(d \rightarrow b \wedge b \rightarrow a \wedge d \rightarrow c \wedge c \rightarrow a) &= S_{Z,\neg Z}\{X\}\{V\}\{Y\}\{W\} \\
i_7(e \rightarrow d \wedge d \rightarrow b \wedge b \rightarrow a) &= \{Z\}\{X\}\{V\}S_{Y,\neg Y}S_{W,\neg W} \\
i_7(e \rightarrow d \wedge d \rightarrow c \wedge c \rightarrow a) &= \{Z\}\{Y\}\{W\}S_{X,\neg X}S_{V,\neg V}
\end{aligned}$$

If we let $e \rightarrow d \wedge d \rightarrow b \wedge b \rightarrow a = \phi_1$ and $e \rightarrow d \wedge d \rightarrow c \wedge c \rightarrow a = \phi_2$, then

$$i_7(\phi_1 \wedge \phi_2) = \{Z\}\{X\}\{V\}\{Y\}\{W\}$$

Combining this theory with the sixth generalized incidence calculus theory we obtain

$$i(e \wedge \phi_1) = S_{E,\neg E}\{Z\}\{X\}\{V\}S_{Y,\neg Y}S_{W,\neg W}$$

$$i(e \wedge \phi_2) = S_{E,\neg E}\{Z\}\{Y\}\{W\}S_{X,\neg X}S_{V,\neg V}$$

---

[1] We use $S_{Z,\neg Z}\{X\}\{V\}S_{Y,\neg Y}S_{W,\neg W}$ to denote $S_{Z,\neg Z} \otimes \{X\} \otimes \{Y\} \otimes S_{Y,\neg Y} \otimes S_{W,\neg W}$. The sets $S_{Y,\neg Y} \otimes S_{W,\neg W}$ and $S_{W,\neg W} \otimes S_{Y,\neg Y}$ are considered to be the same.

$$i(e \wedge \phi_1 \wedge \phi_2) = S_{E,\neg E}\{Z\}\{X\}\{V\}\{Y\}\{W\}$$

Because $e \wedge \phi_1 \rightarrow a$, $e \wedge \phi_2 \rightarrow a$ and $e \wedge \phi_1 \wedge \phi_2 \rightarrow a$, the following equation holds:

$$i_*(a) = i(e \wedge \phi_1) \cup i(e \wedge \phi_2) \cup i(e \wedge \phi_1 \wedge \phi_2)$$

$$= S_{E,\neg E}\{Z\}\{X\}\{V\}S_{Y,\neg Y}S_{W,\neg W} \cup S_{E,\neg E}S_{X,\neg X}S_{V,\neg V}\{Z\}\{Y\}\{W\}$$

and

$$\begin{aligned}
p_*(a) &= \mu(i_*(a)) \\
&= \mu(S_{E,\neg E}\{Z\}\{X\}\{V\}S_{Y,\neg Y}S_{W,\neg W} \cup S_{E,\neg E}S_{X,\neg X}S_{V,\neg V}\{Z\}\{Y\}\{W\}) \\
&= \mu(S_{E,\neg E}) \times \mu(\{Z\}\{X\}\{V\}S_{Y,\neg Y}S_{W,\neg W} \cup S_{X,\neg X}S_{V,\neg V}\{Z\}\{Y\}\{W\}) \\
&= \mu(S_{E,\neg E}) \times \mu(\{Z\}) \times \mu(\{X\}\{V\}S_{Y,\neg Y}S_{W,\neg W} \cup S_{X,\neg X}S_{V,\neg V}\{Y\}\{W\}) \\
&= \mu(S_{E,\neg E}) \times \mu(\{Z\}) \times (\mu(\{X\}\{V\}S_{Y,\neg Y}S_{W,\neg W}) + \mu(S_{X,\neg X}S_{V,\neg V}\{Y\}\{W\}) \\
&\quad -\mu(\{X\}\{V\}\{Y\}\{W\})) \\
&= 1 \times 0.8 \times (.6 \times .7 \times 1 \times 1 + 1 \times 1 \times .75 \times .8 - .6 \times .7 \times .75 \times .8) \\
&= 0.6144
\end{aligned}$$

So our belief in a is also 0.6144.

Similarly we can obtain $i_*(d \rightarrow a)$, $i_*(e \rightarrow a)$ as:

$$i_*(d \rightarrow a) = S_{E,\neg E}S_{Z,\neg Z}\{X\}\{V\}S_{Y,\neg Y}S_{W,\neg W} \cup S_{E,\neg E}S_{Z,\neg Z}\{Y\}\{W\}S_{X,\neg X}S_{V,\neg V}$$

$$i_*(e \rightarrow a) = S_{E,\neg E}\{Z\}\{X\}\{V\}S_{Y,\neg Y}S_{W,\neg W} \cup S_{E,\neg E}\{Z\}\{Y\}\{W\}S_{X,\neg X}S_{V,\neg V}$$

These six generalized incidence calculus theories are in fact produced from assumed and premise nodes in the ATMS.

If we compare the full extensions of nodes in the ATMS and the lower bounds of incidence sets on formulae, we can find that the following equations hold:

$$i_*(d \rightarrow a) = FL(d \rightarrow a) \qquad i_*(e \rightarrow a) = FL(e \rightarrow a) \qquad i_*(a) = FL(a)$$

That is, the full extension of a node is the same as the lower bound of incidence set of the corresponding formula, *i.e.*, for an element $(a_1, a_2, .., a_k)$ in $i_*(\phi)$,

$(a_1, a_2, ..., a_k)$ is also in $FL(\phi)$. In the following I give the general procedure of encoding a list of ATMS nodes into the equivalent generalized incidence calculus theories.

7.2.2 The algorithm of equivalent transformation from an ATMS to extended incidence calculus

Algorithm D: *Transformation algorithm from an ATMS to extended incidence calculus*

Given a probabilistic ATMS we follow the following steps to convert it into an equivalent extended incidence calculus theory.

Step 1: divide the list of nodes into four sets: a set of assumption nodes, a set of assumed nodes, a set of derived nodes and a set of premises. The set of assumption nodes is called lower level nodes and the last three sets together are called higher level nodes. Based on the higher level nodes, a set of propositions P is established. A higher level node is either a proposition in P or a formula in $\mathcal{L}(P)$.

Step 2: from the set of assumption nodes, we can form a list of probabilistic assumption sets $S_{A_1,...,A_m}, S_{B_1,...,B_n}, ...$, based on Definition 6.1. It is also assumed that these sets are probabilistically independent. If they are not independent, a normally extended ATMS cannot solve them.

Step 3: divide those assumed nodes into groups under the following conditions: node n_t and n_j are in group k, when there exists an assumption A which is in an environment of $L(n_t)$ and also in an environment of $L(n_j)$ or an assumption in $L(n_t)$ and an assumption in $L(n_j)$ are in the same probabilistic assumption set. If both n_t and n_j are in the same group, and both n_j and n_l are in the same group, then n_t, n_j and n_l are in the same group.

Step 4: for any group k, create a corresponding structure $< \mathcal{W}_k, p_k, P, i_k, \mathcal{A} >$. The set of axioms $\mathcal{A}$ consists of assumed nodes in this group and all the possible conjunctions of them. The set $\mathcal{W}_k$ is either a probabilistic assumption set or the set product of several such sets if there is more than one probabilistic assumption set involved in the labels of these assumed nodes. For instance, if the label of node n_k is $\{\{A\},\{B\}\}$ and $S_{A,A_1,...}, S_{B,B_1,...}$ are different, then the set of possible worlds $\mathcal{W}_k$ should be $\mathcal{W}_k = S_{A,A_1,...} \otimes S_{B,B_1,...}$. The function i_k is defined as $i_k(n_t) = L(n_t)$ and $i_k(n_t \wedge n_j) = L(n_t) \otimes L(n_j)$. So i_k defined on $\mathcal{A}$ is closed under $\wedge$. We further define $i_k(false) = \{\}$ and $i_k(true) = \mathcal{W}_k$, then $< \mathcal{W}_k, p_k, P, i_k, \mathcal{A} >$ is a generalized incidence calculus theory. In the case that the set of possible worlds is a joint space of several probabilistic assumption sets, labels of nodes need to be reconstructed. Following the above case if $S_{A,A_1,...} = \{A, \neg A\}$ and $S_{B,B_1,...} = \{B, \neg B\}$, the label of node n_t can be changed into

$$\begin{aligned} L(n_t) &= \{\{A\} \otimes \{B, \neg B\}, \{A, \neg A\} \otimes \{B\}\} \\ &= \{\{\{A, B\}, \{A, \neg B\}\}, \{\{A, B\}, \{\neg A, B\}\}\} \\ &= \{\{A, B\}, \{A, \neg B\}, \{\neg A, B\}\} \end{aligned}$$

In general, $L(n_t) = \{\{A\} \otimes S_B, S_A \otimes \{B\}\}$.

Step 5: for each premise node, create a generalized incidence calculus theory and add the set of possible worlds to the list. For example, for premise e, a suitable generalized incidence calculus theory might be $< \{V\}, \mu(V) = 1, P, \{e\}, i_j(e) = \{V\} >$. The added probabilistic assumption set must be different from any set in the list.

Step 6: combining these generalized incidence calculus theories we have the result that for any derived node d_j, there is $i_*(d_j) = FL(d_j) \setminus FL(\perp)$. $FL(d_j) \setminus FL(\perp)$ means deleting those conjunctive parts which appear in both $FL(d_j)$ and $FL(\perp)$.

So both the label set and the degree of belief in a node can be obtained in this combined generalized incidence calculus theory correctly as proved below.

7.2.3 Formal proof

In this section I will give the formal proof about the equivalence between an ATMS and the transformed generalized incidence calculus theories.

Theorem 7 *Given an ATMS, there exists a set of generalized incidence calculus theories such that the reasoning result of the ATMS is equivalent to the result obtained from the combination of these theories. For any node d_l in an ATMS, $FL(d_l)\backslash FL(\perp)$ is equivalent to the lower bound of the incidence set of formula d_l in the combined generalized incidence calculus theory, that is $FL(d_l)\backslash FL(\perp) = i_*(d_l)$. The nogood environments are equivalent to a subset of the set of possible worlds which causes conflicts, that is $FL(\perp) = \mathcal{W}_0$.*

PROOF

The purpose of this proof is that, applying Algorithm D on a given ATMS, we get a list of generalized incidence calculus theories, the combined generalized incidence calculus theory of these theories generates the same label set and belief degree of a node as the ATMS does.

Assume that the nodes of an ATMS are divided into four sets, *e.g.*, a set of assumption nodes, a set of assumed nodes, a set of premise nodes and a set of derived nodes.

Step A: In order to carry out the proof below, we need to reconstruct the justifications of derived nodes to ensure that justifications of derived nodes contain only assumed nodes or premise nodes. This can be done as follows.

Given a derived node d_l, choose a node from its justifications. If the node is an assumption C, then create an assumed node c with single environment $\{C\}$ and single justification (C) and then replace C with c in any justifications where C appears. If the node is a derived node, then replace the node with the justifications of this node. For example if d_l is such a derived node with justifications $\{(z_1, z_2)(z_3, z_4)\}$ and d_l appears in a justification of node d_j as $\{(..., d_l, ...), ...\}$, then d_l is replaced with its justifications and the new justifications of d_j are $\{(..., z_1, z_2, ...), (..., z_3, z_4, ...), ...\}$.

Repeat this procedure until all nodes in the justifications of a derived node are either assumed nodes or premise nodes. As a consequence, an environment of a derived node contains only assumptions because labels of assumed and premise nodes contain only assumptions.

Step B: For any derived node d_l, suppose its justifications are

$$\{(a_1, a_2, ...), (b_1, b_2, ...), ...\}$$

then the conjunction of each justification of d_l implies d_l, such as $a_1 \wedge a_2 \wedge ... \rightarrow d_l$. If we denote this implication as $\models$, then we have $a_1 \wedge a_2 \wedge ... \models d_l$. If we let $j(d_l) = \{a_1 \wedge a_2 \wedge ..., b_1 \wedge b_2 \wedge ..., ...\}$ then $j(d_l) \models d_l$. The environments of d_l will be

$$(L(a_1) \otimes L(a_2) \otimes ...) \cup (L(b_1) \otimes L(b_2) \otimes ...) \cup ...$$

For example, if

$$L(a_1) = \{l_{i1}, l_{i2}, ...\}$$

and

$$L(a_2) = \{l_{j1}, l_{j2}, ...\}$$

then

$$L(a_1) \otimes L(a_2) = \cup_{t,k}\{l_{it} \cup l_{jk}\}$$

In general for a derived node d_l, assume that d_l has a justification $(n_1, n_2, ..., n_l)$, then

$$L(n_1) \otimes L(n_2) \otimes ... \otimes L(n_l) \setminus L(\perp)$$

is the label set of d_l.

Step C: After forming a language set from higher level nodes, a series of generalized incidence calculus theories (assume n theories in total) can be constructed from assumed nodes and premise nodes based on *steps 4 and 5* described in the equivalent transformation algorithm. Any two sets of possible worlds of such theories are required to be probabilistically independent and all of them can be combined using Theorem 2 in Chapter 3 and the subset of possible worlds which leads to contradictions is W_0.

Suppose $(n_1, n_2, ..., n_l)$ is a justification of a derived node d_i (we have ensured that these nodes are either assumed nodes or premise nodes) and they are arranged into t generalized incidence calculus theories. Combining them we will obtain the generalized incidence calculus theory

$$< \mathcal{W}_1, \mu'_1, P, \mathcal{A}'_1, i'_1 > \tag{7.2}$$

$$\begin{aligned} i'_1(n_1 \wedge n_2 \wedge ... \wedge n_l) &= i_1(n_{11} \wedge ... \wedge n_{1m_1}) \otimes ... \otimes i_t(n_{t1} \wedge ... \wedge n_{jtm_t}) \setminus W'_1 \\ &= (L(n_{11}) \otimes ... \otimes L(n_{1m_1}) \otimes ... \otimes (L(n_{t1}) \otimes ... \otimes L(n_{tm_t})) \setminus W'_1 \\ &= L(n_1) \otimes L(n_2) \otimes ... \otimes L(n_l) \setminus W'_1 \end{aligned}$$

where $\{n_1, ..., n_l\} = \{n_{11}, ..., n_{1m_1}, ...n_{t1}, ..., n_{tm_t}$ and $(n_{11} \wedge ... \wedge n_{1m_1})$, ..., $(n_{t1} \wedge ... \wedge n_{tm_t})$ are in these t different generalized incidence calculus theories, and W'_1 is the subset of possible worlds which leads to contradictions after combing these t generalized incidence calculus theories.

Assume that by combining the remaining $n - t$ generalized incidence calculus theories we have

$$< \mathcal{W}_2, \mu'_2, P, \mathcal{A}'_2, i'_2 > \tag{7.3}$$

where $\mathcal{A}'_2 = \{y_1, y_2, ..., y_n\}$ and the subset of possible worlds leading to contradictions is $\mathcal{W}'_2$. To combine the theories in (7.2) and (7.3), $\phi \wedge y_1, \phi \wedge y_2, ..., \phi \wedge y_n$ will be in the set of axioms of the new combined theory.

$$< \mathcal{W}_3, \mu'_3, P, \mathcal{A}'_3, i > \tag{7.4}$$

Here ϕ denotes $n_1 \wedge n_2 \wedge ... \wedge n_l$. Because $\phi \wedge y_j \models \phi$ and for any $\psi \wedge y_j \models \phi \wedge y_j$, $\psi \models \phi$, the following equation holds.

$$
\begin{aligned}
i_*(\phi) &= \bigcup_j i(\phi \wedge y_j) \\
&= \bigcup_j i'_1(\phi) \otimes i'_2(y_j) \setminus W'_3 \\
&= i'_1(\phi) \otimes \bigcup_j i'_2(y_j) \setminus W'_3 \\
&= i'_1(\phi) \otimes (\mathcal{W}_2 \setminus W'_2) \setminus W'_3 \qquad \text{as } \bigcup_j i'_2(y_j) = \mathcal{W}_2 \setminus W'_2 \\
&= i'_1(\phi) \otimes \mathcal{W}_2 \setminus ((i'_1(\phi) \otimes W'_2) \cup W'_3) \\
&= (L(n_1) \otimes L(n_2) \otimes ... \otimes L(n_l) \setminus W'_1) \otimes \mathcal{W}_2 \setminus ((i'_1(\phi) \otimes W'_2) \cup W'_3) \\
&= (L(n_1) \otimes L(n_2) \otimes ... \otimes L(n_l)) \otimes \mathcal{W}_2 \setminus ((W'_1 \otimes \mathcal{W}_2) \cup (i'_1(\phi) \otimes W'_2) \cup W'_3) \\
&= ((L(n_1) \otimes L(n_2) \otimes ... \otimes L(n_l)) \otimes \mathcal{W}_2) \setminus W_0
\end{aligned}
$$

where W'_3 is the set of possible worlds which leads to contradictions after combining the generalized incidence calculus theories i'_1 and i'_2. The incidence function is i in the final generalized incidence calculus theory. W_0 is the total set of possible worlds causing conflict after combining all generalized incidence calculus theories.

Because of the relation $n_1 \wedge ... \wedge n_l \rightarrow d_l$ in the ATMS, we have the relation $n_1 \wedge ... \wedge n_l \rightarrow d_l$ in extended incidence calculus. So $i_*(\phi) \subseteq i_*(d_l)$. In general, if there are k justifications for node d_l, the environments obtained from k justifications are $(L(a_{11}) \otimes ... \otimes L(a_{1x})) \cup ... \cup (L(a_{k1} \otimes ... \otimes L(a_{ky})) \setminus L(\perp)$, then there are k corresponding formulae $\phi_1, \phi_2, ..., \phi_k$, where $i_{j*}(\phi_j) \subseteq i_*(d_l)$ for $j = 1, ..., k$. So $\bigcup_j i_{j*}(\phi_j) \subseteq i_*(d_l)$.

Step D: In the ATMS, a *nogood* environment is derived if $\perp$ is proved. When c and $\neg c$ are both derived, $L(c) \otimes L(\neg c)$ is a *nogood* environment. For any higher level node a, $(a, \neg a)$ is automatically recognized as a justification of node $\perp$ and $L(\perp) =$ *nogood*. Certainly for an assumption A, $(A, \neg A)$ is also a justification of node $\perp$, but adding such justifications does not affect the result in our discussion, so in the following we only consider justifications of $\perp$ as $(a, \neg a)$.

Choosing a justification of node $\perp$, such as $(c, \neg c)$, $L(c) \otimes L(\neg c)$ will be the environments of nogood. When c or $\neg c$ is a derived node, we replace c or $\neg c$ with its label. Suppose that the justifications of c are $\{(z_1, z_2, ...), (x_1, x_2, ...), ...\}$ and the justifications of $\neg c$ are $\{(y_1, y_2, ...), ...\}$, then $\{(z_1, z_2, ..., y_1, y_2, ...), (x_1, x_2, ..., y_1, y_2, ...), ...\}$ will be the justifications of $\perp$. Therefore $(L(z_1) \otimes L(z_2) \otimes ... \otimes L(y_1) \otimes L(y_2) \otimes$

$...) \cup (L(x_1) \otimes L(x_2) \otimes ... \otimes L(y_1) \otimes L(y_2) \otimes ...)$ are nogood environments. Because $z_1 \wedge z_2 \wedge ... \wedge y_1 \wedge ... = \perp$ and $x_1 \wedge x_2 \wedge ... \wedge y_1 \wedge ... = \perp$, we have $(L(z_1) \otimes L(z_2) \otimes ... \otimes L(y_1) \otimes ...) \cup (L(x_1) \otimes L(x_2) \otimes ... \otimes L(y_1) \otimes ...) \subseteq W_0$ based on **Step C** above. Therefore $FL(\perp) \subseteq W_0$.

The other way around, for any element $w \in W_0$, in the combined theory there exists a formula $\phi_1 \wedge \phi_2 \wedge ... \wedge \phi_n = \perp$ and $w \in L(\phi_1) \otimes ... \otimes L(\phi_n)$. Deleting those ϕ_j which will not destroy the equation $\wedge_i \phi_i = \perp$, we will have $\psi_1 \wedge ... \wedge \psi_m = \perp$. Therefore there exists a node z, the conjunction of some ψ_i implies z and the conjunctions of remaining ψ_j implies $\neg z$. So $z \wedge \neg z = \psi_1 \wedge ... \wedge \psi_m = \perp$ and therefore $L(\psi_1) \otimes ... \otimes L(\psi_m)$ are nogood environments. It is straightforward that w is in the full extension of $L(\psi_1) \otimes ... \otimes L(\psi_m)$, so w is a nogood environment, that is $FL(\perp) \supseteq W_0$, so $FL(\perp) = W_0$.

Step E: Using the result from **Step C** and **Step D**, because $\bigcup_j i_{j*}(\phi_j) \subseteq i_*(d_l)$, we have the following equations.

$$((L(a_{11}) \otimes ... \otimes L(a_{1x})) \otimes ... \otimes (L(a_{k1} \otimes ... \otimes L(a_{ky}))) \setminus W_0 \subseteq i_*(d_l)$$

$$FL(d_l) \setminus FL(\perp) \subseteq i_*(d_l)$$

The other way around, for any $w \in i_*(d_l)$, there exists a formula $\phi = \phi_1 \wedge ... \wedge \phi_n$ and $w \in i(\phi)$. There is also a formula $\psi \in FL(d_l)$ such that $\psi = \psi_1 \wedge ... \wedge \psi_m$, $\phi \to \psi$. So $w \in i_*(\psi) = L(\psi_1) \otimes ... \otimes L(\psi_m) \setminus W_0$. Based on the definition of $FL(d_l)$, $\psi_1 \wedge ... \wedge \psi_m$ should be a justification of node d_l, so $L(\psi_1) \otimes ... \otimes L(\psi_m) \setminus L(\perp)$ will be the environments of d_l. Therefore w is in the full extension of $FL(d_l) \setminus FL(\perp)$. That is $FL(d_l) \setminus FL(\perp) \supseteq i_*(d_l)$, so eventually $FL(d_l) \setminus FL(\perp) = i_*(d_l)$.

QED

Example 7.3

Example 7.3 shows the way of dealing with conflicting information. Following the story in Example 7.2, suppose we are told later that f is also observed and there is a rule $f \rightarrow \neg c$ with degree 0.8 in the knowledge base. That is, three more nodes in the ATMS are used.

assumed node: $< f \rightarrow \neg c, \{\{U\}\}, \{(U)\} >$

premise node: $< f, \{\{\}\}, \{()\} >$

assumption node: $< U, \{\{U\}\}, \{(U)\} >$

pas: $S_U = \{U, \neg U\}$, $S_F = \{F, \neg F\}$.

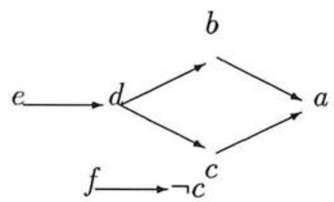

Figure 7.2. semantic network of inference rules

Here *pas* means probabilistic assumption set and $S_{F,\neg F}$ is created to support premise node f.

In the ATMS, we can infer that one environment of node c is $\{E, Z, Y\}$ and one environment of node $\neg c$ is $\{F, U\}$. So the *nogood* environment is $\{E, X, Y, F, U\}$. The belief in node a needs to be recalculated in order to re-distribute the weight of conflict on other nodes. The new belief in node a is 0.366 as given in [Laskey and Lehner, 1989].

In extended incidence calculus, similar to Example 7.1, two more generalized incidence calculus theories are constructed from the assumed node $f \rightarrow \neg c$ and the premise node f. Combining these two theories with the final one we obtained in Example 7.1, we have $W_0 = \{UZY\}$[2], $i_*(a) = \{ZXV \cup ZYW\} \setminus W_0$. Therefore $\mu(\{UZY\}) = 0.48$ which is the weight of conflict and $p'_*(a) = \mu(\{ZXV \cup ZYW\}) \setminus \{UZY\}) = 0.366$ which is our belief in a. Both of these results are the same as

[2] In order to state the problem simply, I use UZY instead of $\{U\}\{Z\}\{Y\}S_{X,\neg X}S_{W,\neg W}S_{V,\neg V}S_{E,\neg E}S_{F,\neg F}$.

those given in [Laskey and Lehner, 1989], but the calculation of belief in node a and the weight of conflict are based on generalized incidence calculus theory.

7.2.4 Comparison with Laskey and Lehner's work

The work carried out in this chapter has some similarity with Laskey and Lehner's work in [Laskey and Lehner, 1989]. The key idea in [Laskey and Lehner, 1989] is to create the medium level elements between a set of beliefs and numerical assignments and then associate the numerical assignments with the medium level elements. The medium level elements are exactly the set of possible worlds in extended incidence calculus and the set of assumptions in an ATMS. Both our and Laskey and Lehner's work try to group assumptions into different sets and each set is associated with a probability distribution. Both systems calculate labels and degrees of belief in nodes. Both systems have to normalize label sets after conflict is discovered and both of them obtain total conflict weight. I, however, have provided a formal proof of the connections between extended incidence calculus and the ATMS while Laskey and Lehner didn't. Moreover, the result obtained in this chapter also provides a theoretical basis for some results obtained in [Laskey and Lehner, 1989]. In this section, we will explain this point in detail.

Comparison 1). In [Laskey and Lehner, 1989] after the label of a node is obtained, in order to calculate the belief in this node, an algorithm is given to rewrite a label as a list of disjoint conjuncts of assumptions. For instance, in Example 7.2 the label of node a is rewritten as $L(a) = \beta_1 \vee \beta_2$ where $\beta_1 = W \wedge Y \wedge Z$ and $\beta_2 = (V \wedge X \wedge Z \wedge \neg W) \vee (V \wedge X \wedge Z \wedge W \wedge \neg Y)$.

If we simplify the elements in the full extension of a label (i.e. using Z to replace $(Z \wedge \neg W) \vee (Z \wedge W)$), we can get exactly those β lists required in [Laskey and Lehner, 1989].

Comparison 2). In [Laskey and Lehner, 1989] when *nogood* environments are produced, the beliefs in nodes are calculated in the following way

$$Bel(node) = \frac{Pr(label \cap \neg nogood)}{Pr(\neg nogood)} = \frac{Pr(label \cap \neg nogood)}{1 - Pr(nogood)}$$

It is suggested that all *nogood* environments can be divided into two groups $nogood_1$ and $nogood_2$, where $nogood_2$ has no overlap with environments in $nogood_1$ or *label*. So in the actual calculation *nogood* is replaced by $nogood_1$ and it is claimed that such replacement doesn't affect the whole result. They didn't provide a proof. I will prove below that this result is sound.

Theorem 8 *Assume that all nogood environments can be divided into two disjoint groups $nogood_1$ and $nogood_2$. For a node d_l, if $L(d_l)$ has no overlap with $nogood_2$, then the following equation holds.*

$$Bel(d_l) = \frac{Pr(L(d_l) \cap nogood)}{1 - Pr(nogood)} = \frac{Pr(L(d_l) \cap nogood_1)}{1 - Pr(nogood_1)}$$

PROOF

If all *nogood* environments can be divided into two disjoint groups, then it is possible to divide all the corresponding generalized incidence calculus theories into two groups based on **Step C** in section 7.2.3. The combination of generalized incidence calculus theories in two groups produces two conflict sets, referred to as $nogood_1$ and $nogood_2$ respectively. The final combination of these two generalized incidence calculus theories will not produce any conflict sets (if it does then the assumption that $nogood_1$ and $nogood_2$ are disjoint is wrong). Assume that the two generalized incidence calculus theories are i_1 and i_2 respectively after combining two groups of generalized incidence calculus theories, for a formula ϕ, if the list of axioms making ϕ true are $x_1, x_2, ..., x_n$, then

$$i_*(\phi) = \bigcup_j (i_1(x_j))$$

Assume that the list of all axioms for incidence function i_2 are $y_1, y_2, ..., y_m$, then combining i_1 and i_2 we have

$$\begin{aligned} i'_*(\phi) &= \cup_l(\cup_j i(x_l \wedge y_j)) \\ &= \cup_l(\cup_j i_1(x_l) \otimes i_2(y_j)) \\ &= \cup_l(i_1(x_l) \otimes \cup_j i_2(y_j)) \end{aligned}$$

$$= \cup_l (i_1(x_l) \otimes (\mathcal{W}_2 \setminus FL(nogood_2)))$$
$$= (\cup_l i_1(x_l)) \otimes (\mathcal{W}_2 \setminus FL(nogood_2))$$
$$= i_*(\phi) \otimes (\mathcal{W}_2 \setminus FL(nogood_2))$$

So $p_*(\phi) = \mu(i'_*(\phi)) = \mu(i_*(\phi)) \times \mu(\mathcal{W}_2 \setminus FL(nogood_2)) = \mu(i_*(\phi))$. That is

$$Bel(\phi) = \frac{Pr(L(\phi) \cap nogood_1)}{1 - Pr(nogood_1)}$$

Therefore, those nogood environments which don't have overlap with the label of a node don't affect the belief in this node.

END

Comparison 3). The major step in [Laskey and Lehner, 1989] is to create an auxiliary set for each belief function and let the auxiliary set carry the information provided by the belief function. So the probability distribution on an auxiliary set which in turn gives the belief function on another set can be thought as the source for this belief function. Therefore the two auxiliary sets defined in this way should be DS-Independent; otherwise these two belief functions cannot be combined by Dempster's Rule and the result obtained in an ATMS has no point of comparison with the result in DS theory.

However, in extended incidence calculus, we don't need to make such an assumption. For dependent probabilistic assumption sets, as long as we can find their joint probabilistic assumption set, we can still combine them using the Combination Rule in Chapter 3. If there are a number of probabilistic assumption sets and some of them are dependent, we combine dependent probabilistic assumption sets first and then carry out the combination for the rest.

Example 7.4

Example 7.4 demonstrates the point I discussed in comparison 2) above. Assume that the ATMS network is extended as in Figure 7.3 by adding more nodes in it. When the facts h and j are observed, both i and $\neg i$ will be derived, then there will be a conflict. So the total *nogood* environments are $\{UZY, HI\}$. If we let $nogood_1 = \{UZY\}$ and $nogood_2 = \{HI\}$, then $nogood_2$ has no overlap with $nogood_1$ and $L(a)$. So the belief in a shouldn't be changed even when h and j are observed.

assumed nodes:	$< h \rightarrow i, \{\{H\}\}, \{(H)\} >$
	$< j \rightarrow \neg i, \{\{I\}\}, \{(I)\} >$
premise nodes:	$< h, \{\{\}\}, \{()\} >$
	$< j, \{\{\}\}, \{()\} >$
assumption node:	$< H, \{\{H\}\}, \{(H)\} >$
	$< I, \{\{I\}\}, \{(I)\} >$
pas:	$S_H = \{H, \neg H\}$, $S_I = \{I, \neg I\}$
	$S_G = \{G, \neg G\}$, $S_L = \{L, \neg L\}$

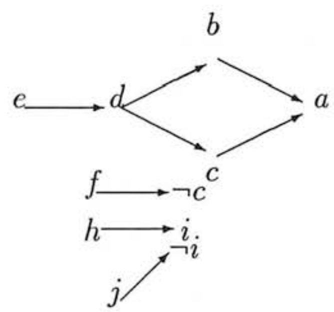

Figure 7.3. Extending the existing ATMS

If we wish to consider this problem in extended incidence calculus, after we have encoded the new assumed and premise nodes into generalized incidence calculus theories, the combination of these theories produces a conflict set $W'_0 = \{HI\}$. The further combination of this theory with the generalized incidence calculus theory obtained in Example 7.3 gives the final result of the impact of all evidence. In this final generalized incidence calculus theory, we have $p''_*(a) = p'_*(a) = 0.366$ while the whole weight of conflict is

$$\mu(FL(UZY \cup HI))$$
$$= p_U(U)p_Z(Z)p_Y(Y) + p_H(H)p_I(I) - p_U(U)p_Z(Z)p_Y(Y)p_H(H)p_I(I)$$
$$= 0.48 + p_H(H)p_I(I) - 0.48p_H(H)p_I(I)$$

Therefore in extended incidence calculus we don't need to divide *nogood* environments into different groups and the correct result can still be achieved.

7.3 Implementing Incidence Calculus Using an ATMS

In the previous section, I proved that given an ATMS system, it can be equivalently translated into incidence calculus terminology and the result achieved in extended incidence calculus is the same as what can be obtained in the ATMS. Furthermore, extended incidence calculus also provides a way of coping with numerical uncertainties and allows automatic calculation of beliefs in nodes (also including *nogood* environments).

In this section, I will show that a given list of generalized incidence calculus theories can be encoded into an ATMS and we can carry out the corresponding inference at the symbolic level in the ATMS. The transformation procedure allows us to use extended incidence calculus as a tool to provide justifications for an ATMS.

Remember that the major role of an ATMS is to create nodes and build links among nodes using justifications. The key step in transforming generalized incidence calculus theories into an ATMS lies in finding 'correct' formulae, creating nodes for these formulae and using these nodes as justifications for other formulae. For example, if a set of axioms $\mathcal{A}$ in a generalized incidence calculus theory is $\mathcal{A} = \{a \to b, b \to c, a \to b \wedge b \to c\}$, then we say $a \to b$ and $b \to c$ are the 'correct' formulae. The third axiom in $\mathcal{A}$ can be obtained through the conjunction of the other two axioms. In general our purpose is to find these 'correct' formulae and create ATMS nodes for them first and then use them to form justifications for any other nodes.

Definition 7.4: *Basic axiom set*

Given a generalized incidence calculus theory $< \mathcal{W}, \mu, P, \mathcal{A}, i >$, *a subset* $\mathcal{A}'$ *of* $\mathcal{A}$ *is called a basic axiom set if for any axiom* ψ *in* $\mathcal{A} \setminus \mathcal{A}'$, ψ *is the conjunction of some axioms in* $\mathcal{A}'$.

In Section 3.1.1, we have made the assumption that a set of axioms in a generalized incidence calculus theory is closed under $\wedge$. That is, for $\phi, \psi \in \mathcal{A}$, $\phi \wedge \psi$ is also in $\mathcal{A}$ holds. So we could, at least, delete those $\phi \wedge \psi$ from $\mathcal{A}$ in order to get $\mathcal{A}'$. It is easy to see that $\mathcal{A}'$ is unique for a given $\mathcal{A}$.

7.3.1 Examples

In order to see how to encode a list of generalized incidence calculus theories, in particular create nodes for the basic axioms, into an ATMS, we first examine two examples.

Example 7.5

Encoding a generalized incidence calculus theory into an ATMS

Assume that we have a set of propositions P as

$$\{storm, windy, snowy, fog, bad\_weather, traffic\_problem, ...\}$$

and $\mathcal{L}(P)$ as the language set of P. Further assume that we have a generalized incidence calculus theory $< \mathcal{W}_1, \mu_1, P, \mathcal{A}_1, i_1 >$ where the set of axioms contains $\mathcal{A}'$ and all the possible conjunctions of elements in $\mathcal{A}'$. $\mathcal{A}'$ is defined as $\{storm, windy \wedge snowy, storm \rightarrow bad\_weather, storm \rightarrow (windy \wedge snowy), (windy \wedge snowy) \rightarrow bad\_weather\}$. We will be able to infer the lower bounds of incidence sets of any formulae in $\mathcal{L}(P)$ based on this theory. For instance, for formula *bad_weather* (it is also a proposition), the lower bound of its incidence set is

$$i_*(bad\_weather) = \cup_{\phi \in SI(bad\_weather)} i(\phi) = \cup_{\psi \in ESI(bad\_weather)} i(\psi)$$

where $SI(bad\_weather)$ and $ESI(bad\_weather)$ are the semantic implication set and essential semantic implication set of $bad\_weather$ and $ESI(bad\_weather)$ is

$$ESI(bad\_weather) = \{storm \wedge (storm \rightarrow bad\_weather),$$
$$(windy \wedge snowy) \wedge ((windy \wedge snowy) \rightarrow bad\_weather)\}$$

If we assume the incidence sets of axioms in $\mathcal{A}'$ are:

$$i_1(storm) = A_1$$
$$i_1(windy \wedge snowy) = A_2$$
$$i_1(storm \rightarrow bad\_weather) = A_3$$
$$i_1((windy \wedge snowy) \rightarrow bad\_weather) = A_4$$
$$i_1(storm \rightarrow (windy \wedge snowy)) = A_5$$

then the lower bound of incidence set of $bad\_weather$ is

$$i_*(bad\_weather) = (A_1 \cap A_3) \cup (A_2 \cap A_4)$$

Encoding this generalized incidence calculus theory into an ATMS If we want to solve this problem in an ATMS, it is natural to create the following nodes in the system initially with appropriate assumptions supporting them.

$$n_1 :< storm, \{\{A_1\}\}, \{(A_1)\} >$$
$$n_2 :< windy \wedge snowy, \{\{A_2\}\}, \{(A_2)\} >$$
$$n_3 :< storm \rightarrow bad\_weather, \{\{A_3\}\}, \{(A_3)\} >$$
$$n_4 :< windy \wedge snowy \rightarrow bad\_weather, \{\{A_4\}\}, \{(A_4)\} >$$
$$n_5 :< storm \rightarrow windy \wedge snowy, \{\{A_5\}\}, \{(A_5)\} >$$

For node $bad\_weather$, the justification set is $\{(n_1, n_3), (n_2, n_4)\}$ and the corresponding label set is $\{\{A_1, A_3\}, \{A_2, A_4\}\}$. If we take $\{A_1, A_3\}$ as $A_1 \cap A_3$ and take the whole label set as $(A_1 \cap A_3) \cup (A_2 \cap A_4)$ (using $\cup$ to replace $\vee$), then this set is the same as $i_*(bad\_weather)$.

This example shows us the spirit that it is possible to encode a generalized incidence calculus theory into an ATMS if we can find a proper set of axioms (like the five axioms in $\mathcal{A}'$) and let them be the set of nodes including premise nodes and assumed nodes. For a formula in $\mathcal{L}(P)$ if its ESI set is not empty, then the rational reconstruction of this set can be the active justification set of this formula in the ATMS. If the ESI set of a formula is empty, it only tells us that from current assumed and premise nodes, we cannot infer any meaningful result for this formula, so the current active justification set is empty. The rational reconstruction of a formula in an $ESI(\phi)$ means that if this formula is $\phi_1 \wedge ... \wedge \phi_j$ and for each ϕ_i there is a node in the corresponding ATMS, then $(n_{\phi_i}, ..., n_{\phi_j})$ is a justification of n_ϕ. In a large ATMS, n_ϕ may have many justifications, but at one time based on the information available only some of them are used for inference while the rest are not. The used justifications are called active ones here. For node *bad_weather*, its justification set is $j(bad\_weather) = \{(node_1, node_3), (node_2, node_4)\}$ based on $ESI(bad\_weather)$ when we use node names to replace formula names such as $node_1$ for *storm*.

Example 7.6

Extending the ATMS by encoding more generalized incidence calculus theories

Apart from the generalized incidence calculus theory given in Example 7.5, if we further get another two generalized incidence calculus theories as

$$< \mathcal{W}_2, \mu_2, P, \mathcal{A}_2, i_2 >$$

$$< \mathcal{W}_3, \mu_3, P, \mathcal{A}_3, i_3 >$$

where

$\mathcal{A}_2 = \{bad\_weather \to traffic\_problem\}$
$i_2(bad\_weather \to traffic\_problem) = B_1$

and $\mathcal{A}_3 = \{fog, fog \rightarrow traffic\_problem, fog \wedge (fog \rightarrow traffic\_problem)\}$

$$i_3(fog) = C_1$$

$$i_3(fog \rightarrow traffic\_problem) = C_2$$

$$i_3(fog \wedge (fog \rightarrow traffic\_problem)) = C_1 \cap C_2$$

then after combining these three theories, the set $ESI(traffic\_problem)$ in the combined theory is

$$\begin{aligned} ESI(traffic\_problem) = \{ \; & fog \wedge (fog \rightarrow traffic\_problem), \\ & storm \wedge (storm \rightarrow bad\_weather) \\ & \wedge (bad\_weather \rightarrow traffic\_problem), \\ & (windy \wedge snowy) \wedge ((windy \wedge snowy) \rightarrow bad\_weather) \\ & \wedge (bad\_weather \rightarrow traffic\_problem)\} \end{aligned}$$

and the lower bound of incidence set of $traffic\_problem$ is

$$i_*(traffic\_problem) = (\mathcal{W}_1 \otimes \mathcal{W}_2 \otimes (C_1 \cap C_2)) \cup ((A_1 \cap A_3) \otimes B_1 \otimes \mathcal{W}_3) \cup ((A_2 \cap A_4) \otimes B_1 \otimes \mathcal{W}_3)$$

Encoding more generalized incidence calculus theories into the ATMS

If we extend the ATMS created above by adding more premise and assumed nodes into it, based on the second and the third generalized incidence calculus theories, we can do this by adding the following additional nodes to the system.

$$n_6 :< bad\_weather \rightarrow traffic\_problem, \{\{B_1\}\}, \{(B_1)\} >$$

$$n_7 :< fog, \{\{C_1\}\}, \{(C_1)\} >$$

$$n_8 :< fog \rightarrow traffic\_problem, \{\{C_2\}\}, \{(C_2)\} >$$

In such an ATMS, the justification set for node $traffic\_problem$ is $\{(n_7, n_8), (n_1, n_3, n_6), (n_2, n_4, n_6)\}$ and the label set is $\{\{C_1, C_2\}, \{A_1, A_3, B_1\}, \{A_2, A_4, B_1\}\}$. In general for an environment $\{X_1, X_2, ..., Y_1, ..., Z_1...\}$ of a node, we regard it as $(X_1 \cap X_2 \cap ...) \otimes (Y_1 \cap ...) \otimes (Z_1 \cap ...)$ if X_i, Y_j and Z_l are in different sets. So the environment $\{A_1, A_3, B_1\}$ is actually thought as $(A_1 \cap A_3) \otimes B_1$. Therefore, the label set of node *traffic_problem* is $(C_1 \cap C_2) \cup ((A_1 \cap A_3) \otimes B_1) \cup ((A_2 \cap A_4) \otimes B_1)$ and the full extension of the set if exactly the same as $i_*(traffic\_problem)$.

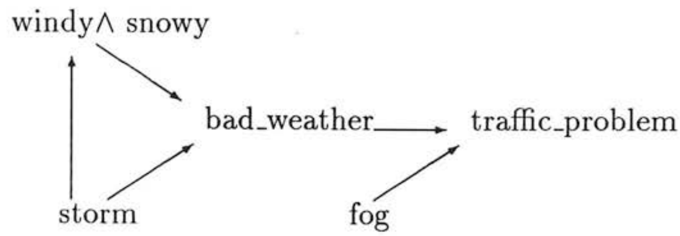

Fig. 7.4 Semantic network of created ATMS

Figure 7.4 shows the semantic relations among propositions after we gathered these three pieces of evidence and encoded them into terms of the ATMS.

Here we should notice the following four points.

First, even though there are three axioms in $\mathcal{A}_3$ we only use two of them to create new nodes in an ATMS as the third one is the conjunction of these two axioms. Secondly, in an ATMS the justification set and label set of a premise node or an assumed node only have one single assumption. Here we extend this to the general case that a justification or an environment for an assumed node or a premise node can have a set of assumptions and each of these assumptions supports the node. For instance, A_1 can be explained as a set of assumptions and each of its element A_{1i} has the same functions as an assumption possesses in a normal ATMS. Thirdly, when a generalized incidence calculus theory is encoded into nodes of an ATMS, we select only some formulae from the set of axioms as the basis and create a node for each such formula. The general principle of such selection is that any other axioms can be the conjunctions of selected formulae. The ideal situation should be that any selected axiom is either a proposition in P or a implication formula with only propositions at both left and right hand sides. For instance, *storm* or $(storm \rightarrow bad\_weather)$ in Example 7.5. However, given a generalized incidence calculus theory randomly, axioms selected may not always be that simple. Such as in Example 7.5, we have an axiom $(windy \wedge snowy)$ and we selected it as a node in order to build up an ATMS. The ideal case should be that we have both *windy* and *snowy* as axioms and both are selected. If we have enough information to specify all the incidence sets for these axioms, then Figure 7.4 will be changed into the following:

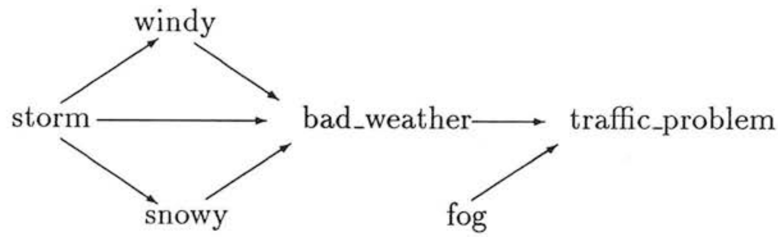

Figure 7.5. Standard ATMS semantic network

Given limited information it is only possible to create an ATMS in Figure 7.4 and the system in Figure 7.4 is in fact the compound system appearing in Figure 7.5.

Fourthly, it is always assumed that a premise node is either true or false in an ATMS. But in an ATMS created from a list of generalized incidence calculus theories, a premise node also has a degree of belief on it and it may be supported by a set of assumptions.

7.3.2 Transforming a set of generalized incidence calculus theories into an ATMS

When we encode a generalized incidence calculus theory into the ATMS terminology, we don't create a node for every axiom, rather we just select some of the axioms such as the subset $\mathcal{A}'$ of $\mathcal{A}$ in Example 7.5. Other axioms can be obtained by conjoining two or more of the selected axioms.

Algorithm E: *Transformation procedure from generalized incidence calculus theories to the ATMS*

Given a generalized incidence calculus theory $< \mathcal{W}, \mu, P, \mathcal{A}, i >$ *with* $\mathcal{A}'$ *as the basic axiom set, we transform it into an ATMS by the following steps.*

Step 1: For a basic axiom $\phi \in \mathcal{A}'$, if it is a proposition in P, then create a premise node $node_\phi$ for it. Otherwise create an assumed node for it. In this case we treat the set of possible worlds as a set of assumptions. So the justifications and environments for such a created node will be $\{(A_1), ..., (A_n)\}$ and $\{\{A_i\}, ..., \{A_n\}\}$ where $A_i \in i(\phi)$.

Step 2: For an axiom ψ in $\mathcal{A} \backslash \mathcal{A}'$, if $\psi = \wedge\phi_i$, then $\{(\phi_1, ..., \phi_j)\}$ is a justification of ψ. ψ can have several justifications. The label of ψ is $\{\{A_1\}, ...\{A_k\}\}$ where $A_j \in i(\psi)$. For any formula $\varphi \in \mathcal{L}(P)$, its justification set is $ESI(\varphi)$ by replacing non-basic axioms with their justifications.

If more generalized incidence calculus theories are introduced, we need to extend the existing ATMS structure. This extension means creating more premise and assumed nodes and assigning justifications to more formulae.

Step 3: Assume that the basic axiom set for another generalized incidence calculus theory is $\mathcal{A}'_1$, then in principle we can create premise and assumed nodes for each basic axiom in $\mathcal{A}'_1$ and assign justifications and labels for other axioms. But for an axiom ϕ both in $\mathcal{A}'$ and $\mathcal{A}'_1$, we only create one node by adding more justifications and environments for it. After finishing the creation of premise and assumed nodes, we combine these two generalized theories in extended incidence calculus. For an axiom φ in the combined theory, if $\varphi \notin (\cup\mathcal{A}'_j)$, then assign $\{(\phi_1, ..., \phi_k)\}$ as a justification of φ if $\wedge\phi_i = \varphi$. Here ϕ_i are the basic axioms in $\mathcal{A}' \cup \mathcal{A}'_1$. For any other formulae, we assign the essential semantic implication sets in the combined theory as their justification sets. The label sets can be obtained using justifications. In this procedure if $\perp$ is produced, that is $W_0 \neq \{\}$, the subset W_0 will be the *nogood* environments and the corresponding $ESI(\perp)$ will certainly be the justifications of it. We need to remember to replace non-basic axioms in $ESI(\phi)$ with basic axioms.

Step 4: In the similar way, we can add more generalized incidence calculus theories into the existing ATMS structure.

In such a designed structure, we have got premise and assumed nodes each of which has assumptions as its justifications and label. For other nodes, i.e., derived nodes, we assign them with justifications which contain only premise and assumed nodes. Finally, for each possible world, we create an assumption node for it, so the constructed structure is an ATMS.

It is not difficult to see that the reasoning results obtained from these generalized incidence calculus theories and the created ATMS are the same.

Theorem 9 *Given a list of generalized incidence calculus theories, after transforming them into an ATMS, the reasoning result in this constructed ATMS is the same as obtained in extended incidence calculus by combining these theories.*

As it is easy to prove this theorem, I will only give the outline rather than provide the whole proof here.

The main step in proving that the result in this transformed ATMS is equivalent to the result in extended incidence calculus is that the labels of derived nodes are the same as the lower bounds of incidence sets of these nodes. Based on *Step 3*, for a node ϕ, its justification set is the same as the semantic implication set of this formula in extended incidence calculus, so as a consequence, the label set of this node is the same as the lower bound of its incidence set.

The advantage of this procedure is the automatic assignment of justifications to nodes while in an ATMS justifications are assigned by the designer. The weakness is that when more generalized incidence calculus theories are introduced, we have to repeat the procedure of finding the essential semantic implication sets in order to assign justifications to nodes. This procedure is slow. I will further investigate this problem and try to find a fast way to do it in the future.

7.4 Extended Incidence calculus Can Provide Justifications For The ATMS

In the previous sections, I have discussed the formal relations between extended incidence calculus and the ATMS. The major similarity of the two reasoning mechanisms is that the justifications in an ATMS are equivalent to the essential semantic implication sets in extended incidence calculus. As a result, the labels of nodes are equivalent to the incidence sets of the corresponding nodes. However, a difference between these two reasoning patterns is that the justifications are assigned by the designers in an ATMS while essential semantic implication sets are discovered automatically in extended incidence calculus. Therefore, the whole reasoning procedure in extended incidence calculus is automatic while the one in an ATMS is semi-automatic. The procedure of discovering semantic implication sets in extended incidence calculus can be regarded as a tool to provide justifications for an ATMS. The application of this procedure into an ATMS can release a system designer from the task of assigning justifications and this procedure can guarantee those justifications are non-redundant. A problem with this procedure is that it is slow to find all essential semantic implication sets. If it is possible to have a fast algorithm for this procedure, then an ATMS can be established and extended automatically without a designer's involvement.

Example 7.7

Providing justifications automatically using extended incidence calculus

We examine Example 7.2 in [Laskey and Lehner, 1989] in a different way here. Assume that our objective in Example 7.2 is to calculate the impact on a when e is observed. Because there is no direct effect from e on a, a diagram shown as Figure 7.1 is created to build a link between e and a. In order to infer a, the justifications for node $e \rightarrow a$ are essential to be given in an ATMS. Assume that the information carried by this diagram is denoted as S_I and the information specifying justifications is denoted as S_J, then in an ATMS we have

$$S_I \cup S_J \Rightarrow L(e \rightarrow a) \tag{7.5}$$

Here notation $A \Rightarrow B$ means that from information carried by A, it is possible to infer information carried by B through some logical methods. $L(e \rightarrow a)$ stands for the label set of $e \rightarrow a$. S_J may either contain the justifications for node $e \rightarrow a$ only or consists of more justifications for the assisting nodes (such as $e \rightarrow b$). We say that S_J is the extra information for the system inference.

Given the same initial information carried by S_I, extended incidence calculus does inferences without requiring any more information. The inference procedure produces

$$S_I \Rightarrow i_*(e \rightarrow a) \cup ESI(e \rightarrow a)$$

The notation $\Rightarrow$ is explained as from the information on the left hand side, we can infer the information on the right hand side. So from the information in S_I, we can obtain both the lower bound of the incidence set and the inference paths of a node. The essential semantic implication set for a node contains exactly the justifications for the same node. Therefore the extra information required by the ATMS can be supplied by extended incidence calculus as an output in general and we are able to change (7.5) as follows in an ATMS

$$S_I \cup ESI(e \rightarrow a) \Rightarrow L(e \rightarrow a)$$

which takes the output from extended incidence calculus as an input in the ATMS.

So we can abstract out essential semantic implication sets for all necessary formulae and assign them on the corresponding nodes without considering assumptions on the initial nodes. In this way, a justification in an ATMS can be constructed.

So we can conclude that the inference result in extended incidence calculus provides justifications for an ATMS automatically.

7.5 Summary

The main contributions of this chapter are:

[1] It has been proved that extended incidence calculus and the ATMS are equivalent at both the symbolic reasoning level (if we view the set of possible worlds in extended incidence calculus as the set of assumptions in an ATMS) and numerical inference level if we associate proper probabilistic distributions on assumptions. They can be translated into each other's form.

[2] It has been shown that the integration of symbolic and numerical reasoning patterns is possible and extended incidence calculus itself is a typical example of this integration. Extended incidence calculus can therefore be regarded as an bridge between these two reasoning patterns.

[3] In [Liu and Bundy, 1994] it has been proved that generalized incidence calculus is equivalent to Dempster-Shafer theory of evidence in representing evidence and combining source-independent evidence. Therefore the result of investigating the relationship between extended incidence calculus and ATMS can provide a theoretical basis for some results in [Laskey and Lehner, 1989] which lacks theoretical explanations, namely the calculation of beliefs in nodes and the weight of conflict introduced by all evidence as well as its effect on individual nodes.

[4] It is assumed that justifications must be supplied by the problem solver if one uses the ATMS techniques. We have shown that extended incidence calculus can be used to provide justifications for nodes automatically without human's involvement. Therefore a complete automatic ATMS system is constructible.

[5] The calculation of probabilities in nodes is done under the assumption that all given probability distributions are probabilistically independent. When this

condition is not satisfied, the algorithm in [Laskey and Lehner, 1989] would not work.

A notable statement about the relations between the ATMS and extended incidence calculus has been given by Pearl [Pearl, 1988]. He said:"In the original presentation of incidence calculus, propositions were not assigned numerical degrees of belief but instead were given a list of labels called *incidences*, representing a set of situations in which the propositions are true. ... Thus, incidences are semantically equivalent to the ATMS notion of 'environments', and it is in this symbolic form that incidence calculus was first implemented by Bundy." In this chapter I have discussed the relations intensively. This discussion proves the equivalence between extended incidence calculus and the ATMS. The result tells us that extended incidence calculus itself is a unification of both symbolic and numerical approaches. It can therefore be regarded as a bridge between the two reasoning patterns. This result also gives theoretical support for research on the unification of the ATMS with numerical approaches. In extended incidence calculus structure, both symbolic supporting relations among statements and numerical calculation of degrees of belief in different statements are explicitly described. For a specific problem, extended incidence calculus can either be used as a support based symbolic reasoning system or be applied to deal with numerical uncertainties. This feature cannot be provided by pure symbolic or numerical approaches independently.

A advantage of using extended incidence calculus to make inferences is that it doesn't require the problem solver to provide justifications. The whole reasoning procedure is performed automatically. The inference result can be used to produce the ATMS related justifications. The calculation of degrees of beliefs in nodes is based on the probability distributions on assumption sets which can either be dependent or independent.

Chapter 8

Conclusions

8.1 Introduction

An intensive extension and discussion of incidence calculus has been carried out in this thesis. The discussion shows that incidence calculus, in particular, extended incidence calculus has the potential to deal with complicated uncertainty problems. The combination technique in extended incidence calculus provides an attractive beginning towards combining dependent pieces of evidence.

Incidence calculus shares the features of both numerical and symbolic reasoning mechanism. It can be taken as a bridge between the two reasoning patterns. This suggests that incidence calculus could be used where both numerical and symbolic reasoning techniques are required.

8.2 Contributions of The Thesis

In this chapter, I shall draw conclusions from the following three aspects which are the three contributions of this thesis.

- Extended incidence calculus
- A comprehensive comparison between extended incidence calculus and DS theory
- A study of the relations between extended incidence calculus and the ATMS

8.2.1 Extended incidence calculus

The first main contribution of the thesis is that the original incidence calculus is extended dramatically so that it possesses many attractive advantages.

The main difference between the original incidence calculus and extended incidence calculus is the conditions on incidence functions. The crucial point in this procedure is that we don't keep the assumption that if a possible world is not supporting a formula, it must support the negation of the formula. That is we don't consider that $i(\neg\phi) = \mathcal{W} \setminus i(\phi)$ captures all possible situations in practice. Once we have $i(\neg\phi) \cup i(\phi) \subset \mathcal{W}$, the condition $i(\phi \vee \psi) = i(\phi) \cup i(\psi)$ is not valid any more. Such generalized incidence calculus is not a nuisance but a requirement of real cases such as Example 2.5 on page 41.

More precisely, extended incidence calculus has the following main features.

- Extended incidence calculus has the ability to represent ignorance caused by incomplete information.

- Extended incidence calculus has the ability to combine DS-independent evidence.

- Extended incidence calculus has the ability to combine dependent information, which would create problems in DS theory.

- Extended incidence calculus provides a model-creation algorithm for incidence assignment. This algorithm can also be used to judge whether an numerical assignment is a belief function and obtain its mass function when it is.

8.2.2 Relations between extended incidence calculus and DS theory

Considering numerical aspect of extended incidence calculus, we made a comprehensive comparison between extended incidence calculus and DS theory.

The Dempster-Shafer theory of evidence is widely appreciated mainly because it has two advantages over other numerical uncertainty reasoning mechanisms. These two advantages are representing ignorance and combining evidence. The original incidence calculus doesn't have these two properties. However extended incidence calculus does. Therefore, it is necessary to investigate the similarities and differences of these two theories.

We proved in Chapter 5 that:

- Extended incidence calculus and the DS theory have the same ability to represent evidence.
- Extended incidence calculus and the DS theory have the same ability in combining DS-independent evidence.
- Extended incidence calculus can also combine some dependent pieces of evidence while Dempster's combination rule cannot. The combination mechanism in extended incidence calculus subsumes the combination rule, *(i.e.) Dempster's combination rule, in DS theory.*

Therefore, extended incidence calculus is an alternative of DS theory and it is expected to deal with some of the cases (dependent situations) which cannot be dealt with by DS theory.

8.2.3 Relations between extended incidence calculus and the ATMS

Since extended incidence calculus also possesses features of truth maintenance systems, we have investigated the relations between extended incidence calculus and the ATMS [de Kleer, 1986]. We proved that these two reasoning mechanisms also share some similarities. The conclusions we have got are:

- Extended incidence calculus and the ATMS are equivalent in calculating labels of nodes.

- Extended incidence calculus can be regarded as a basis for constructing probabilistic ATMS. So that an original ATMS could also have the ability to cope with uncertainty problems.

- Extended incidence calculus can be regarded as a basis for providing justifications for an ATMS. In this way, the reasoning procedure of an ATMS is expected to be completely automatic without requiring a system designer to supply justifications from time to time.

8.3 Issues of Implementation

There are several algorithms in the thesis. The implementation is done in Sicstus Prolog.

In Chapter 3, we define a new function, basic incidence assignment ii. The algorithm A for obtaining a basic incidence assignment ii from an incidence function i has been implemented.

The algorithm B in Chapter 3 for incidence assignment has also been implemented. The new algorithm is relatively faster than the methods used in the original incidence calculus. Using this algorithm, multiple consistent incidence assignments can be constructed. One of the applications of the algorithm is that when a lower bound of a probability distribution is assigned on every formula in the whole language set, the algorithm can be used to check whether the given lower bound is a belief function. When it is a belief function, the application of the algorithm will produce the corresponding mass function.

The combination rule in Chapter 3 has also been implemented for the cases where several generalized incidence calculus theories are based on the same set of possible worlds.

8.4 Limitations of Extended Incidence Calculus

Extended incidence calculus enriches the expressive and reasoning power of the original incidence calculus considerably. The advantages of the advanced theory have been fully discussed in the thesis. However, like any other reasoning mechanism, extended incidence calculus also has some limitations itself. Briefly, there are following weaknesses in extended incidence calculus.

1) Algorithm B requires a strict condition to be applied, that is, a set of axioms must be closed under operator $\wedge$. Although, it is possible to guess some missing probability values for some formulae, but the overall method still needs to be improved.

2). The combination rule in extended incidence calculus is exponential along with the sizes of sets of axioms in two generalized incidence calculus theories. Efficient algorithm is needed to improve the efficiency of combination.

3). We have proved that extended incidence calculus and the ATMS are equivalent in producing labels (incidences) for nodes (formulae) theoretically. What is missing from this part of work is that we need to design an efficient algorithm to obtain labels in practice.

4). The extension of incidence calculus to the first order logic is not discussed in the thesis.

5). Extended incidence calculus is no longer truth functional, so the inference mechanism in the legal assignment finder in the original incidence calculus cannot be applied any more.

8.5 Future Work

Here we describe some of the topics which need to be addressed in future in order to further improve extended incidence calculus.

Conditions on incidence assignment algorithm

At the moment the incidence assignment algorithm B is carried out under a very strict condition. That is, the set of axioms much be closed under $\wedge$. We have briefly discussed the technique on how to supplement the lower bounds on other relevant axioms if $\mathcal{A}$ is not closed initially. This technique should be further developed in order to give more precise estimation of lower bounds of probabilities.

Computational complexity

The combination mechanism in extended incidence calculus is quite slow, particularly when P has a large number of propositions. Working out an efficient algorithm to reduce the computational complexity problem is one of the main issues to be considered in the future. One possible method to do so is linked with the next topic.

How large would a set of propositions be

Another problem with the current development is that we expect a set of propositions containing all the descriptions we are interested in. That is, in either the original or extended incidence calculus, there is basicly only one set of propositions on which we describe evidence (information). The propagation of incidences on the set becomes increasingly inefficient when the size of the set increases. There are several factors affecting the inference, but the vital factor is the size of a set of axioms which is closed under $\wedge$. The structure of axioms (simple atomic propositions or compound formulae) has effects as well.

An empirical experiment is given below. When there are more than 100 axioms in $\mathcal{A}$ and the structures of these axioms are complex, extended incidence calculus has difficulties to deal with it.

Set	$\mid \mathcal{A} \mid$	Time (seconds)
P	8	160
P	14	290
P	20	470
P	41	63844
P	63	79776
P	*127*	*139552*

An alternative method is discussed in [Liu 1995] in which a set of propositions P is split into several small, but coherent sets. Incidences are propagated within and among these sets. The above example has also been tested using the new method. It takes only about 30 (seconds) to derive the bounds of incidence of a formula while it takes 139552 seconds to do so in the traditional method. The main step in this method is to split the big set into small sets based on some implication relations. Each small set obtained in this way will no longer have any implication relations, so a set of axioms should not be very large. Therefore, as long as it is possible to split up a big set into several small sets, the size of a problem can be very large.

Publications related to the Thesis

- Liu,W. and A.Bundy, The combination of different pieces of evidence using incidence calculus, Research Paper No. 599, Dept. of AI, Univ. of Edinburgh.

- Liu,W. and A.Bundy, A comprehensive comparison between generalized incidence calculus and DS theory, *The Int. J. of Human-Computer Studies* formerly *The Int. J. of Man-Machine Studies*, **40**:1009-1032.

- Liu,W., A.Bundy and D.Robertson, Recovering incidence functions. *2nd European conference on symbolic and quantitative approaches to reasoning and uncertainty.* Lecture Notes in Computer Science 747:241-248 Spinger. Long version is available as Department Research Paper 648.

- Liu,W., A.Bundy and D.Robertson, On the relations between incidence calculus and ATMS. *2nd European conference on symbolic and quantitative approaches to reasoning and uncertainty.* Lecture Notes in Computer Science 747:248-256 Spinger. Also presented at the IJCAI-93 workshop on Management of Uncertainty in AI. Available as Department Research Paper 611.

- Liu,W. and A.Bundy, Constructing probabilistic ATMS using incidence calculus. Submitted.

- Liu,W., Hong,J., McTear,M.F. and Hughes,J.G., An extended framework for evidential reasoning systems. *Int. J. of Pattern Recognition and Artificial Intelligence* 7(3):441-457.

- Liu,W., Hughes,J.G. and McTear,M.F., Representing heuristic knowledge in DS theory. *Uncertainty in Artificial Intelligence, Proc. of the 8th conference*:182-190, edited by Dubois, Wellman, D'Ambrosio and Smets. Morgan Kaufmann.

- Liu,W., Hughes,J.G. and McTear,M.F., Representing heuristic knowledge and propagating beliefs in the Dempster-Shafer theory of evidence. *Advances* *in the Dempster-Shafer Theory of Evidence*: pp441-472 edited by Fedrzzi, Kacprzyk and Yager. John Wiley and Sons, Inc., New York.

Bibliography

[Bacchus, 1988] Bacchus,F., *Representing and reasoning with probabilistic knowledge*, PhD thesis, The University of Alberta, 1988. Also abailable as University of Waterloo, Ontario, Canada, N2L 3G1, pp.1-135.

[Bacchus, 1990] Bacchus, F., *Representing and reasoning with probabilistic knowledge*, The MIT Press.

[Baldwin, 1987] Baldwin,J.F., Evidential Support Logic Programming, *Fuzzy Sets and Systems* 24:1-26.

[Barnett, 1981] Barnett,J.A. Computational methods for a mathematical theory of evidence. IJCAI-81: 868-875.

[Black,1987] Black,P., Is Shafer general Bayes? *Proc. of Third AAAI Uncertainty in Artificial Intelligence workshop.* 2-9.

[Bonissone and Tong 1985] Bonissone,P.P. and Tong,R.M., Reasoning with uncertainty in expert systems. *Int. J. Man-Machine Studies*, :241-250.

[Bundy, 1985] A. Bundy, Incidence calculus: a mechanism for probabilistic reasoning, *J. Automated Reasoning*, **1** 263-283.

[Bundy, 1986] A. Bundy, Correctness criteria of some algorithms for uncertain reasoning using incidence calculus, *J. Automated Reasoning*, **2** 109-126.

[Bundy, 1992] A. Bundy, Incidence calculus. *The Encyclopedia of Artificial Intelligence*.663-668. Also available as Research Paper 497, Dept. of Artificial Intelligence, Edinburgh.

[Cohen, 1985] Pohen, P.R., Heuristic Reasoning about Uncertainty: an artificial intelligence approach, PhD Disscerttion. Pitman Advanced Publishing Program, Boston, London, Melbourne.

[Corlett and Todd, 1985] Corlett, R.A. and Todd, S.J., A Monte-Carlo approach to uncertain inference. *Proceedings of AISB-85*, Ross ed, 28-34.

[Correa da Silva and Bundy, 1990a] Correa da Silva,F.S. and Bundy,A., A rational reconstruction of incidence calculus. Research Paper 517. Dept. of AI, Univ. of Edinburgh.

[Correa da Silva and Bundy, 1990b] Correa da Silva,F.S. and Bundy,A., On some equivalence relations between incidence calculus and Dempster-Shafer theory of evidence. *Proc. of the sixth conference on uncertainty in artificial intelligence*:378-383, G.E.Corp.

[d'Ambrosio, 1988] d'Ambrosio,B., A hybird approach to reasoning under uncertainty, *Int. J. Approx. Reasoning* 2:29-45.

[d'Ambrosio, 1990] d'Ambrosio,B., Incremental construction and evaluation of defeasible probabilistic models, *Int. J. Approx. Reasoning* 4:233-260.

[de Kleer, 1986] de Kleer,J., An assumption-based TMS, *Artificial Intelligence* 28:127-162.

[de Kleer, 1986] de Kleer,J., Extending the ATMS, *Artificial Intelligence* 28:163-196.

[de Kleer and Williams, 1987] de Kleer,J. and B.C.Williams, Diagnosing multiple faults, *Artificial Intelligence* 32:97-130.

[Dempster, 1967] Dempster,A.P., Upper and lower probabilities induced by a multivalued mapping, *Ann. Math. Stat.*. 38, 325-339.

[Doyle, 1979] Doyle,J., A truth maintenance stsrem, *Artificial Intelligence* 12:231-272.

[Duda etal, 1976] Duda,R.O., P.E.Hart and N.Nilsson, Subjective Bayesian methods for rule-base inference systems, in: *Proceedings 1976 National Computer Conference, AFIPS* 45: 1075-1082.

[Dubois and Prade,1986] Dubois,D. and H.Prade, On the unicity of Dempster rule of combination, *I.J. of Intelligence Systems.* 1,133-142.

[Dubois and Prade 1990] Dubois, D. and H. Prade, Consonant approximations of belief functions. *Int. J. Approx. Reasoning* 4:419-449.

[Dubois, etal, 1990] Dubois, D., J.Lang and H.Prade, Handling uncertain knowledge in an ATMS using possibilistic logic, *ECAI-90 workshop on Truth Maintenance Systems*, Stocckholm, Sweden.

[Dubois and Prade, 1992] Dubois,D. and H.Prade, Belief change and possibility theory. *Belief revision* P.Gardenfors Eds., Cambridge University Press:142-182.

[Dubois and Prade, 1994] Dubois, D. and H.Prade, Non-standard theories of uncertainty in knowledge representation and reasoning. *The knowledge Engineering Review*, Vol.9(4):399-416.

[Exploring AI] *Exploring Artificial Intelligence* eds by Howard E Shrobe and AAAI, survey talks from the National Conference on Artificial Intelligence.

[Gardenfors, 1992] Gardenfors,P., Belief revision: an introduction. In *Belief Revision* (Gardenfors ed.), Cambridge University Press, 1-28, 1992.

[Ginsberg, 1984] Ginsberg,M.L., Non-monotonic reasoning using Dempster-s rule, AAAI-84:126-129.

[Fagin and Halpern, 1989a] Fagin,R. and J. Halpern, Uncertainty, belief and probability, IJCAI-89:1161-1167.

[Fagin and Halpern, 1989b] Fagin,R. and J. Halpern, Uncertainty, belief and probability, Research Report of IBM, RJ 6191.

[Fulvio Monai and Chehire, 1992] Fulvio Monai,F. and T.Chehire, Possibilistic assumption based truth maintenance systems, validation in a data fusion application. *Proc. of the eighth conference on uncertainty in artificial intelligence*:83-91, Stanford.

[Ginsberg, 1984] Ginsberg,M.L., Non-monotonic reasoning using Dempster's rule, AAAI-84:126-129.

[Gordon and Shortliffe, 1985] Gordon,J. and Shortliffe,E.H., A method for managing evidential reasoning in a hierarchical hypothesis space. *Artificial Intelligence* 26:323-357.

[Halpern and Fagin, 1992] Halpern,J.Y., and R.Fagin, Two views of belief: belief as generalized probability and belief as evidence, *Artificial Intelligence.* 54: 275-317.

[Hau and Kashyap, 1990] Hau,H.Y. and Kashyap,R.L., Belief combination and propagation in a lattice-structured inference network. *IEEE-SMC*, 20(1):45-57.

[Hunter,1987] Hunter,D., Dempster-Shafer vs. probabilistic logic, *Proc. Third AAAI uncertainty in artificial intelligence workshop.* 22-29.

[Kennes, 1991] Kennes,R., Evidential reasoning in a categorical perspective: conjunction and disjunction of belief functions. *Proc. of 7th uncertainty in artificial intelligence*:174-181.

[Kennes and Smets, 1990] Kennes,R. and Smets,Ph., Computational aspects of the Mobius transform. *Procs. of the 6th conf. on uncertainty in AI*, edited by Bonissone, Herionn, Kanal and Lemmer, Uncertainty in AI 6:401-416. North Holland, Amsterdam.

[Kennes, 1991] Kennes,R., Evidential reasoning in a categorical perspective: Conjunction and disjunction of belief functions, *Proc. of 7th uncertainty in artificial intelligence.* 174-181.

[Kennes, 1992] Kennes,R., Computational aspects of the Mobius transform of a graph. *IEEE-SMC*, 22:201-223.

[Kruse et al, 1992] Kruse,R., E.Schwecke and J.Heinsohn, Uncertainty and Vagueness in Knowledge-Based Systems, Springer Verlag.

[Kong, 1987] Kong,A., Multivariate belief functions and graphical methods. PhD thesis, Dept. of Statistics, Harvard University.

[Kyburg, 1991] H.E. Kyburg Jr., Evidential Probability, Computer Science Technical Report number 376, University of Rochester.

[Laskey and Lehner, 1989] Laskey,K.B. and P.E.Lehner, Assumptions, beliefs and probabilities, *Artificial Intelligence*, 41:65-77.

[Lemmer, 1986] Lemmer,J.F., Confidence factors, empiricism and Dempster-Shafer theory of evidence, *Proc. of Uncertainty in Artificial Intelligence*. 117-125.

[Lingras and Wong, 1990] Lingras,P. and S.K.M.Wong, Two perspectives of the Dempster-Shafer theory of belief functions, *Int. J. Man-Machine Studies*. 33, 467-487.

[Liu and Bundy, 1992] Liu,W. and A.Bundy, The combination of different pieces of evidence using incidence calculus, Research Paper No. 599, Dept. of AI, Univ. of Edinburgh.

[Liu and Bundy, 1994] Liu,W. and A.Bundy, A comprehensive comparison between generalized incidence calculus and DS theory, *The Int. J. of Human-Computer Studies* formerly *The Int. J. of Man-Machine Studies*, **40**:1009-1032.

[Liu, Bundy and Robertson, 1993a] Liu,W., A.Bundy and D.Robertson, Recovering incidence functions. *2nd European conference on symbolic and quantitative approaches to reasoning and uncertainty*. Lecture Notes in Computer Science

747:241-248 Spinger. Long version is available as Department Research Paper 648.

[Liu, Bundy and Robertson, 1993b] Liu,W., A.Bundy and D.Robertson, On the relations between incidence calculus and ATMS. *2nd European conference on symbolic and quantitative approaches to reasoning and uncertainty.* Lecture Notes in Computer Science 747:248-256 Spinger. Also presented at the IJCAI-93 workshop on Management of Uncertainty in AI. Available as Department Research Paper 611.

[Liu and Bundy, 1993] Liu,W. and A.Bundy, Constructing probabilistic ATMS using incidence calculus. Submitted.

[Liu 1995] Liu,W., The incidence propagation method. In preparation.

[Liu et al, 1993] Liu,W., Hong,J., McTear,M.F. and Hughes,J.G., An extended framework for evidential reasoning systems. *Int. J. of Pattern Recognition and Artificial Intelligence* 7(3):441-457.

[Liu, Hughes and McTear, 1992] Liu,W., Hughes,J.G. and McTear,M.F., Representing heuristic knowledge in DS theory. *Uncertainty in Artificial Intelligence, Proc. of the 8th conference*:182-190, edited by Dubois, Wellman, D'Ambrosio and Smets. Morgan Kaufmann.

[Liu, Hughes and McTear, 1994] Liu,W., Hughes,J.G. and McTear,M.F., Representing heuristic knowledge and propagating beliefs in the Dempster-Shafer theory of evidence. *Advances in the Dempster-Shafer Theory of Evidence*: pp441-472 edited by Fedrzzi, Kacprzyk and Yager. John Wiley and Sons, Inc., New York.

[Liu, 1986] Liu,G.S.H., Causal and plausible reasoning in expert systems, AAAI-86:220-225.

[Lowrance, et al 1981] Lowrance, J.D., and T.D.Garvey, Evidential reasoning: an developing concept, *Proceedings IEEE international conference on cybernetics and society*:6-9.

[McLeith, 1988] McLeith,M., Probabilistic logic: some comments and possible use for nonmonotonic reasoning, in *Uncertainty in Artificial Intelligence 2*:55-62, eds by J.F.Lemmer and L.N.Kanal. Elsevier Science Publishers B.V. (North-Holland).

[McLean, 1992] McLean,D.R., *Testing and extending the incidence calculus*, M.Sc Dissertation, Dept. of artificial Intelligence, Univ. of Edinburgh.

[McLean, Bundy and Liu 1994] McLean,D.R., A.Bundy and W.Liu, Assignment methods for incidence calculus. To appear *the Journal of Approximate Reasoning.* Also available as Department research paper 649.

[Mellouli, 1987] Mellouli,K., On the propagation of beliefs in networks using the Dempster-Shafer theory of evidence. PhD thesis.

[Nilsson, 1986] Nilsson,N.L., Probabilistic logic, *Artificial Intelligence* 28:71-87.

[Nguyen and Smets, 1991] Nguyen,H.T. and Ph. Smets, On Dynamics of Cautious Belief and Conditional Objects, Technical Report No. TR/IRIDIA/91-13.2. *To appear in I. J. of Approx. Reasoning.*

[Pearl, 1988] Pearl,J., *Probabilistic Reasoning in Intelligence Systems: networks of plausible inference*, Morgan Kaufmann Publishers, Inc..

[Pearl, 1990] Pearl,J., Reasoning with belief functions: An analysis of compatibility, *International J. of Approximate Reasoning.* 4,363-389.

[Pearl, 1992] Pearl,J., Rejoinder to Comments on "Reasoning with belief function: An analysis of compatibility", *I.J.of Approx. Reasoning.* 6:425-443.

[Proven, 1989] Proven,G.M., AN analysis of ATMS-based techniques for computing Dempster-Shafer belief functions. *Proc. of 11th international joint conference on artificial intelligence*:1115-1120.

[Rich, 1983] Rich,E., Default reasoning as likelihood reasoning, *Proceedings of the IJCAI*:348-351.

[Ruspini et al, 1990] Ruspini,E.H., Lowrance,J.D. and Strat,T.M., Understanding Evidential Reasoning. SRI International Technical Note 501.

[Shafer76] Shafer,G., *A Mathematical theory of evidence*, Princeton university press.

[Shafer, 1982] Shafer,G., Belief functions and parametric models, *J. R. Stat. Soc. Ser.*. B 44: 322-352.

[Shafer, 1986] Shafer,G., Probability judgement in artificial intelligence, *Uncertainty in Artificial Intelligence 2*, (Kanal and Lemmer Eds.) 127-135.

[Shafer, 1987a] Shafer,G., Probability judgement in artificial intelligence and expert systems, *Stat. Sci.*. 2, 3-16.

[Shafer 1987b] Shafer,G., Belief function and possibility measures, *Analysis of Fussy Information*, 1, (J.C.Bezdek, Ed.) CRC Press.

[Shafer, 1990] Shafer,G., Perspectives on the theory and practice of belief functions, Working paper No.218. Univ. of Kansas, School of Business.

[Shafer and Tversky, 1985] Shafer,G. and A.Tversky, Languages and Designs for Probability Judgement, *Cognitive Science*. 9,309-339.

[Shafer and Logan, 1987] Shafer,G. and Logan,R., Implementing Dempster's rule for hierarchical evidence. *Artificial Intelligence*, 33:271-298.

[Shafer, Shenoy and Mellouli, 1987] Shafer,G., Shenoy,P.P., and Mellouli,K., Propagating belief functions in qualitative Markvo trees. *International journal of approximate reasoning*, 1:349-400.

[Shortliffe, 1976]
Shortliffe, E.H., Computer-based medical consultations: MYCIN. Elsevier, New York.

[Smets, 1988] Smets,Ph., Belief functions, *Non-Standard Logics for Automated Reasoning*, (Smets, Mamdani, Dubois and Prade Eds.), 253-286.

[Smets and Kennes, 1994] Smets,Ph. and R.Kennes, The transferable belief model, *Artificial Intelligence* 66(2):191-234.

[Smets, 1990] Smets,Ph. The Combination of Evidence in the Transferable Belief Model, *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 12(5), 447-458.

[Smets and Hsia, 1990] Smets,Ph. and Y.T.Hsia, Default reasoning and the transferable belief model. *Proc. of 6th uncertainty in artificial intelligence*:529-537.

[Strat, 1987] Strat, T.M., The generation of explanations within evidential reasoning systems, *AAAI-87*:308-313.

[Tessem, 1993] Tessem,B., Approximations for efficient computation in the theory of evidence. *Artificial Intelligence*, 61:315-329 (research note).

[Voorbraak, 1988] Voorbraak,F., A computational efficient approximation of Dempster-Shafer theory, *Int. J. Man-Mach. Stud.* 30:522-236.

[Voorbraak, 1991] Voorbraak,F., On the justification of Dempster's rule of combination, *Artificial Intelligence.* 48, 171-197.

[Wesley, 1983] Wesley,L.P. Reasoning about control: the investigation of an evidential approach. IJCAI-83:203-206.

[Yager, Fadrizzi, Kacprzyk, 1994] Yager, R.R., Fedrizzi, M. and Kacprzyk, J. (eds). *Advances in the Dempster-Shafer Theory of Evidence* John Wiley & Sons, Inc..

[Yen, 1988] Yen,J., Gertis: A Dempster-Shafer Approach to Diagnosing Hierarchical Hypotheses, *Communications of the ACM*:573-578.

[Zadeh, 1975] Zadeh,L.A., Fuzzy logic and approximate reasoning, *Synthese* 30:407-428.

[Zadeh, 1984] Zadeh,L.A., A mathematical theory of evidence (book review), *AI Mag.*. 5 (3): 81-83.

[Zadeh, 1986] Zadeh,L.A.,A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination, *AI Mag.*. 7, 85-90.