Automated Transformation of Probabilistic Knowledge for a Medical Diagnostic System

Yu-Chuan Li, MD, Peter J. Haug, MD, Homer R. Warner, MD, Ph.D.
Department of Medical Informatics, University of Utah, Salt Lake City, Utah

Iliad is a large medical diagnostic system that covers more than 2000 diagnoses and 9000 findings. Due to the size and the complexity of this system, a robust knowledge representation is essential to consistently and efficiently model the medical knowledge involved. In this paper, we describe the knowledge representation currently used in Iliad and a probabilistic representation based on the Bayesian network formalism which can be derived using the information that the Iliad knowledge base contains.

1. INTRODUCTION

Iliad is a medical diagnostic support system developed at the University of Utah by Warner et al. [1]. Though it began as a system designed to help diagnose diseases in the Internal medicine area, it has grown to cover knowledge domains including Obstetrics/Gynecology, Dermatology, and Psychiatry. Currently, its knowledge base (KB) comprises of 2300 diseases and intermediate diagnoses and 9000 relevant findings. These findings include sociodemographic data, medical history, medications, physical examinations, laboratory test results, and pathological and radiological findings.

Iliad operates in three different modes: consultation, critiquing and simulation [2]. It has been proven to be a useful tool in teaching medical students diagnostic skills [1, 3, 4] and is currently used for that purpose in several medical schools in the United States. The expansion and refinement of the KB has been one of the most important aspects of this project. Due to the size and complexity of knowledge encompassed by this system, a robust knowledge representation (KR) is essential.

The KR that is currently used in Iliad will be referred to as the Iliad-KR in the following text. In addition to the Iliad-KR, we are currently exploring an alternate probabilistic representation based on Bayesian networks. In this paper, we discuss the weakness and strength of both KRs and similarities that have led to the development of a computer program that can automatically transform any KB expressed in the Iliad-KR into a Bayesian network.

2. THE ILIAD KNOWLEDGE REPRESENTATION

Instead of using heuristic scores like earlier diagnostic expert systems [5, 6], Iliad was designed around a multi-membership Bayesian model. This choice was based on the potential benefits associated with probabilistic formulations and on local experience with Bayesian medical decision systems [7, 8, 9]. Diagnostic medical knowledge is encoded in self-contained modules called frames. Each frame contains a list of findings associated with the disease that the frame represents. The TPR (true positive rate) and FPR(false positive rate) of each finding for a given disease is also included to facilitate the Bayesian calculation. During the early effort at knowledge engineering, the researchers discovered that multi-membership Bayes alone was not able to model all of the medical knowledge they intended to capture for two reasons: (1) The assumption of conditional independence of findings seldom held true for all findings associated with a disease, and (2) some medical diagnosis are routinely described using deterministic decision logic. Three mechanisms were devised to amend the multimembership model. These are referred to as "clustering", "OR sets" and "deterministic frames". "Clustering" and "OR sets" were used to address the problem of conditionally dependent findings, while "deterministic frames" were used to capture the Boolean logic sometimes used by physicians to describe a diagnosis.

2.1 The Deterministic Frames

As opposed to the frames that contain TPR/FPR, which we call probabilistic frames, a deterministic frame contains a list of relevant findings, Boolean decision logic, and the expected base frequencies for each of the findings in an inpatient population. The value of the frame is then determined by a heuristic algorithm that combines the truth status of the Boolean logic and the frequencies of the findings [10].

2.2 The Clusters and the OR sets

Clusters are frames (either deterministic or probabilistic) which contain a set of related findings that represent intermediate pathophysiologic states or syndromes [11]. These states or syndromes are then used as findings in the probabilistic frames that represent diseases or higher-level concepts (which can also be clusters). A deterministic frame does not contain clusters in the Iliad-KR.

Findings that are considered conditionally dependent but do not constitute an intact intermediate concept can be assigned as an OR set . Findings in an OR set are treated as mutually exclusive. When more than one finding in an OR set are instantiated (known to be true or false), only the finding with the most information will be active. A multi-membership, Bayesian formalism, together with these complimentary mechanisms make up the current Iliad knowledge representation.

2.3 The Algorithms Used to Reason in the Iliad-KR

In order to reason within this complex KR and obtain posterior probabilities for likely diagnoses, several algorithms were implemented in addition to the standard multi-membership Bayes' calculation. These algorithms

0195-4210/94/$5.00

include (1) a heuristic algorithm that propagates probabilities from clusters to higher-level frames, and (2) an algorithm that evaluates deterministic frames and returns probabilistic interpretations. Detailed descriptions of these algorithms can be found in [10].

3. BAYESIAN NETWORKS AND THE ILIAD-KR

3.1 Similarity Between the Two Knowledge Representations

Bayesian networks have been vigorously studied in the last few years as a normative knowledge representation in domains involving probabilistic dependencies. Possible applications in medical decision-support systems have also been enthusiastically explored [12-14]. We found the Bayesian network appealing because of its consistent representation of probabilistic dependencies among variables (also referred to as "nodes" in the following text) and the mathematical characteristics it embodies [15].

Further examination of the Iliad-KR and Bayesian networks provided us insight into the similarity between these two KRs. We found that the causal relations which need to be specified (as arcs) in Bayesian networks can be identified in the Iliad-KR: (1) The Iliad-KR uses frames to represent the relation between the diseases (or intermediate states) and the findings. These relations are, in most cases, direct causal relations. For example, variant angina is implemented as a finding in the frame "coronary artery spasm", this can be easily translated into a causal link from the "coronary artery spasm" node to the node that represents variant angina. (2) The clusters in the Iliad-KR can be treated as intermediate nodes caused directly by the diseases nodes that are their parents. Each cluster links, in turn, to a set of finding nodes. (3) The OR sets in the Iliad-KR imply hidden intermediate causes that can be added as an intermediate node in Bayesian network terms.

The need for exponential number of conditional probabilities has been criticized as one of Bayesian networks weakness. This requirement becomes problematic when a node has multiple parents or predecessors. Then, the number of probabilities required is $2^{(\text{number of parents})}$. A common approach to managing this requirement is through the "noisy OR gate model." Using this model reduces the probabilities needed for any node to one conditional probability of that node given each of its predecessors. The noisy OR gate model is based on the assumptions that (1) an event is presumed false if all of its listed causes are false (accountability) and (2) each exception to a normal causal relation between variables acts independently (exceptional independence) [15]. If these assumptions hold true, a complex conditional probability can be decomposed into simple ones using formula(6) below.

In the Iliad-KR, these simple conditional probabilities are explicitly assigned in the probabilistic frames. In the deterministic frames, the probabilistic relationship is actually implied in the Boolean logic of each frame. As a highly simplified example, if C causes A and B, conditional probabilities P(A|C) and P(B|C) are both 1. If C causes A or B, the conditional probabilities P(A|C) and P(B|C) are both 0.5. Formula (3) and (4) show the general form of this derivation.

3.2 Weakness and Strength of the Two Knowledge Representations

The Iliad-KR has given us an efficient and workable system capable of providing useful probabilities. Domain experts and knowledge engineers have created thousands of frames using this KR. The simplicity of the calculations associated with this KR, has made it possible to do any diagnostic inference in seconds on a personal computer. However, the heuristic components in the Iliad-KR calculations make the probabilities generated by the system mathematically unsound. In addition, the multi-membership basis of Iliad-KR also relies on the false assumption that all diseases are completely independent. This has resulted in the generation of less discriminative and overly confident posterior probabilities by the system [16].

On the other hand, Bayesian networks provide a mathematically sound representation with extensive expressiveness. Our experience with a renal mass diagnostic KB suggests that Bayesian network models demonstrated better reliability and discriminating ability than the current Iliad model [16]. Nonetheless, the price for this theoretically sound solution is the demand for combinatorially increasing numbers of probabilities and the use of NP-hard inference algorithms. Although the number of probabilities needed can be dramatically decreased by using a noisy OR gate model, this is only appropriate when the assumptions of the noisy OR gate model are not violated.

The most serious drawback to a Bayesian network model may be its inference algorithms: Both exact and approximate probabilistic inference in general Bayesian networks have been proven to be NP-hard [17, 18]. This means that for some Bayesian networks, the computation time needed to reach a solution will grow exponentially with the size of the network. Researchers in this area have not clearly characterized the classes of Bayesian networks that require exponential running time. Yet there has been significant work to restrict the topology and the conditional probabilities of the network to guarantee polynomial running time for approximation algorithms [19, 20]. For Bayesian networks with arbitrary topology, empirical evaluation is often useful to gain insight into the computation time needed [21].

4. THE TRANSFORMATION ALGORITHMS

Based on the similarity between Bayesian networks and the Iliad-KR discussed above, we have developed a set of algorithms to facilitate a transformation from the Iliad-KR to a Bayesian network.

Most of the findings in a frame can be transformed directly into the successors of the node that represents the frame itself. Some exceptions are nodes that represent age, sex and risk factors. Although they are used in the Iliad KB as findings of a disease, age and sex are obviously not caused by the diseases. In another example, cigarettes smoking is considered a finding under the frame "lung cancer", yet it is not caused by "lung cancer". Under these conditions, we have chosen to reverse the causal relationship by applying a Bayesian calculation

$$P(d^+|r^+)=\frac{P(r^+|d^+)P(d^+)}{P(r^+|d^+)P(d^+)+P(r^+|d^-)(1-P(d^+))}$$

..... (1)

where D is the disease node and R is the node to be reversed. Because TPR, FPR and prior probabilities for disease are readily available in the Iliad KB, $P(r^+|d^+)$, $P(r^+|d^-)$ and $P(d^+)$ can be obtained to calculate the conditional probability of D given R, i.e., the right hand side of equation (1). Since R now becomes a root node, we estimated its prior probability $P(r^+)$ by using formula (2) where n is the number of the original parents of R.

$$P(r^+)=\frac{\sum_i^n P(r^+|d_i^+)P(d_i^+)+P(r^+|d_i^-)(1-P(d_i^+))}{n}$$

..... (2)

In deterministic frames, a joint probability distribution is implied by the Boolean logic embedded in each frame. Let $f_1 \ldots f_n$ be the findings in a deterministic frame X, this joint probability distribution can be represented by formula (3). However, the causal relationship implied by this formula (X is dependent on $f_1 \ldots f_n$) is inconsistent with the causal semantics that are modeled throughout the Iliad-KR (X causes $f_1 \ldots f_n$). In order to have consistent causal semantics, we asserted that $f_1 \ldots f_n$ should be children of X and derive the conditional probabilities $P(f_i^+|x^+)$ by using formula (4) where f_i is the ith finding in X.

$$P(x,f_1,\ldots,f_n)=P(x|f_1,\ldots,f_n)\prod_i P(f_i)\ldots..(3)$$

$$P(f_i^+|x^+)=\frac{P(x^+,f_1,\ldots,f_i^+,\ldots,f_n)}{P(x^+,f_1,\ldots,f_n)}\ldots\ldots\ldots\ldots(4)$$

In the right hand side of formula (3), $P(x|f_1,\ldots,f_n)$ can be easily obtained from the Boolean logic and $P(f_i)$ is the frequency of the finding f_i embedded in Iliad's deterministic frames.

To accommodate the OR set heuristic in the Iliad-KR, we insert a synthetic intermediate node between the frame and the findings in an OR set. The TPR for the most important finding in the OR set is used as the conditional probability for the link between the synthetic node and the node that represents the frame. The conditional probability between this finding and the synthetic node will then be assigned 1 and the conditional probabilities for the rest of the findings are normalized accordingly. This implementation is intended more to maintain the correct semantics of the resulting Bayesian network than to simulate the OR set heuristic, although it does act similarly whenever the most important finding in the set is instantiated.

All the transformation algorithms were developed under the noisy OR gate model. As we describe in Section 3.1, this model is only valid when the assumptions of accountability and exceptional independence hold true. Although the exceptional independence assumption does hold true in the Iliad-KR most of the time, the accountability assumption is violated if the list of causes for a node is not exhaustive. To accommodate this assumption, we have added a parent node labeled as "Other causes" to each non-cluster findings [12]. Each of the "Other-causes" node was assigned a prior probability of 1 and a conditional probability derived from calculating its lower bound:

$$P(f^+)\le\sum_i P(f^+|d_i^+)P(d_i^+)+P(f^+|c^+)P(c^+)$$

$$P(f^+|c^+)\ge\frac{P(f^+)-\sum_i P(f^+|d_i^+)P(d_i^+)}{P(c^+)}\ldots\ldots(5)$$

where d_is are the listed causes of finding f, and c is the "Other-causes" node. Thus, $P(f|c)$ is the conditional probability for the link between c and f. $P(c^+)$ is by definition 1. The equal sign only holds when the d_is and c are mutually exclusive and exhaustive. All the information on the right hand side of formula (5) is either stored in the Iliad KB or derivable from it.

Using the noisy OR gate model, we can decompose a complex conditional probability into simple ones. For example, assuming d_1 and d_2 are parents of f, the probability of f conditioned on both d_1^+ and d_2^+ can be derived from formula (6).

$$P(f^-|d_1^+,d_2^+)=P(f^-|d_1^+)P(f^-|d_2^+)\ldots\ldots\ldots\ldots(6)$$

Thus all the complex conditional probabilities needed for Bayesian networks can be derived from the probability of f conditioned on each of its parents.

5. THE RESULTING BAYESIAN NETWORK

By utilizing the algorithms described above, we have developed a computer program that can read directly the Iliad KB and transform it into a Bayesian network. The result is a multiply connected Bayesian network

consisting of 11,406 nodes which has a multi-level structure as deep as 36 levels. The nodes are heavily interconnected and common findings are shared by as many as 62 parents. The size and complexity of this Bayesian network makes exact algorithms impractical for its inference. Among the existing approximation algorithms, we have chosen to use the likelihood weighting algorithm as our first inference algorithm because of its simplicity in implementation [22]. We are also investigating other weighting algorithms as well as various Markov sampling techniques [23, 24].

The initial results based on the forward simulation algorithms have been encouraging. We observed a trend of convergence when we increase the number of iterations for the simulation. On a synthesized case with 12 pieces of evidence, this Bayesian network consistently generated reasonable results after 40,000 iterations. However, under the likelihood weighting algorithm that we are using, the rate of convergence could degrade if the evidence occurred on nodes with extreme conditional probabilities (close to 0 or 1). We have found that the reversal of age, sex and risk factor nodes actually contribute to a better convergence rate. Many of the reversed nodes, are root nodes and are frequently present as part of the evidence.

6. DISCUSSION

The restricted assumptions associated with earlier probabilistic models like simple Bayes and multi-membership Bayes have limited their applications to narrow medical domains [9]. Bayesian networks eliminate many of these restrictions. In addition, the expressiveness of Bayesian networks can be especially advantageous in building large, broad-spectrum diagnostic systems where numerous intermediate pathophysiologic states may be present and shared by many diseases. However, it is a tremendous effort to build a comprehensive diagnostic system from scratch. One way to prevent the duplication of a knowledge engineering effort is to convert an existing system to a Bayesian network. Shwe and Middleton et al. have demonstrated this approach by reformulating Internist-I into a Bayesian network [12]. Their new system called QMR-DT showed a diagnostic accuracy comparable to the original Internist-I, despite the many approximations that were used in the conversion processes. However, because of the inherent two-level structure of the Internist-I KB, the resulting Bayesian network did not take advantage of this formalism's ability to express the multi-level structure of diagnostic reasoning.

Iliad was constructed using a KR that accommodates intermediate pathophysiologic states and conditionally dependent findings. We have found that, by utilizing the set of algorithms described above, most of the parameters needed by a Bayesian network can be derived from this KR. A program has been developed to automatically transform the Iliad-KR into a Bayesian network. Several advantages are associated with this approach: (1) Whenever the Iliad KB is updated, we only need to rerun this program once to include the updated knowledge into the Bayesian network. (2) This program allows us to explore different options in the transformation processes. For example, the conditional probability between a finding and the "Other-causes" node was derived by calculating the lower bound of its value. We can run this program multiple times with different values of this probability to generate Bayesian networks with different emphasis on "Other-causes".

We have also found that not all the probabilities in the Iliad KB are consistent with the clinical setting we are trying to model. This is partly because many of the probabilities have been tested and adjusted under the system's own inference heuristics. Converting to a well defined, mathematically consistent model has highlighted some of these questionable probabilities. Where appropriate, we are replacing them with estimates derived from data in the HELP clinical database [25].

Computation time has always been an issue in large Bayesian networks. Given the size and complexity of the Bayesian network derived from the Iliad KB, no existing inference algorithm can guarantee a response time suitable for interactive consultation although many are highly parallelizable. A more likely scenario is to integrate the inference engine into a health information system, process the data in the background once they are available, and provide the results when needed. We have potential applications which we are exploring include quality assurance for medical care [26], and medical free text processing [27].

* This publication is supported in part by grant number 5 R01 LM05323 from the National Library of Medicine.

References

[1]. Warner HR, Haug PJ, Bouhaddou O, Lincoln MJ, Warner HRJ, Sorenson D, Williamson JW, Fan C. Iliad as an expert consultant to teach differential diagnosis. In: Proceedings of the 12th Symposium on Computer Applications in Medical Care (SCAMC), IEEE Computer Society Press 1988:371-376.

[2]. Warner HR Jr. Iliad - Moving medical decision-making into new frontiers. In: Proceedings of International Symposium of Medical Informatics and Education. Salamon R, Protti D, Moehr J. eds. University of Victoria, B.C., Canada, 1989:267-70.

[3]. Cundick R, Turner CW, Lincoln MJ, Buchanan JP, Anderson C, Warner HRJ, and Bouhaddou O. Iliad as a patient case simulator to teach medical problem solving. In: Proceedings of the 13th Symposium on Computer Applications in Medical Care (SCAMC), IEEE Computer Society Press. 1989:902-6.

[4]. Turner CW, Williamson JW, Lincoln MJ, Haug PJ, Buchanan JP, Anderson C, Grant M, Cundick R, Warner HR. The effects of Iliad on medical student problem solving. In: Proceedings of the 14th Symposium on Computer Applications in Medical

Care (SCAMC), IEEE Computer Society Press. 1990:478-81.
[5]. Barnett GO, Cimino JJ, Hupp JA, Hoffer EP. DXplain: An evolving diagnostic decision-support system. JAMA 1987;258:67-74.
[6]. Miller RA. Pople HEJ, Myers JD. Internist-I: An experimental computer-based diagnostic consultation for general internal medicine. N Engl J Med 1982;307:468-76.
[7]. Ben-Bassat M, Carlson RW, Puri VK, Davenport MD, Schriver JA, Latif M, Smith R, Portigal LD, Lipnick EH, Weil MH. Pattern-Based Interactive Diagnosis of Multiple Disorders: The Medas System. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.PAMI-2, No.2, March 1980.
[8]. Warner HR, Olmsted CM, Rutherford BD. HELP - A program for medical decision making. Comp Biomed Res 1972; 5:65-74.
[9]. Warner HR, Toronto AF, Veasy LG. Experience with Bayes' theorem for computer diagnosis of congenital heart disease. Ann N.Y. Acad Sci 1964; 115:558-67.
[10]. Sorenson DK, Cundick RM, Fan C, Warner HR. Passing Partial Information among Bayesian and Boolean Frames. In: Proceedings of the 13th Symposium on Computer Applications in Medical Care (SCAMC). Kingsland LC, ed. Los Alamitos, CA: IEEE Computer Society Press. 1989:50-54.
[11]. Turner CW, Lincoln MJ, Haug PJ, Warner HR, Williamson JW, Whitman N. Clustered disease findings: aspects of expert systems. In: Proceedings of International Symposium of Medical Informatics and Education. Salamon R, Protti D, Moehr J. eds. University of Victoria, B.C., Canada, 1989:259-63.
[12]. Shwe MA, Middleton B, Heckerman DE, Henrion M, Horvitz EJ, Lehmann HP, Cooper GF. Probabilistic Diagnosis Using a Reformulation of the INTERNIST-1/QMR Knowledge Base I and II. Meth Inf Med 1991; 30: 241-67.
[13]. Olesen KG, Kjaerulff U, Jensen E, Jensen FV, Falck B. Andreassen S, Andersen SK. A MUNIN network for the median nerve - a case study on loops. Appl Artif Intell 1989; 3:384-404.
[14]. Heckerman DE, Horvitz, Nathwani BN. Update on the Pathfinder project. In: Proceedings of the 13th Symposium on Computer Applications in Medical Care (SCAMC). Kingsland LC, ed. Los Alamitos, CA: IEEE Computer Society Press. 1989:203-7.
[15]. Pearl J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufman, San Mateo, CA, 1988.
[16]. Li YC, Haug PJ. Evaluating the quality of a probabilistic diagnostic system using different inferencing strategies. In: Proceedings of the 17th Symposium on Computer Applications in Medical Care (SCAMC), Washington DC, 1993:471-477.
[17]. Cooper GF. The Computational Complexity of Probabilistic inference using Bayesian Belief Networks. Artificial Intelligence 1990; 42:393-405.
[18]. Dagum P, Luby M. Approximating probabilistic inference in Bayesian belief networks is NP-hard. Artificial Intelligence 1993; 60:141-153.
[19]. Heckerman DE. A tractable inference algorithm for diagnosing multiple diseases. In: Proceedings Fifth Conference on Uncertainty in Artificial Intelligence, Windsor, Ont, 1989:174-81.
[20]. Henrion M. Search-based methods to bound diagnostic probabilities in very large belief nets. In: Proceedings Seventh Workshop on Uncertainty in Artificial Intelligence, Los Angeles, CA, 1991.
[21]. Shwe M, Cooper GF. An empirical analysis of likelihood-weighting simulation on a large, multiply connected belief network. In: Proceedings Sixth Conference on Uncertainty in Artificial Intelligence, Boston, MA, 1990:498-508.
[22]. Fung R, Chang KC. Weighting and Integrating Evidence for Stochastic Simulation in Bayesian Networks. In: Machine Intelligence and Pattern Recognition: Uncertainty in Artificial Intelligence 5,Vol 10. Henrion M, Shachter R, Kanal LN, Lemmer JF. eds. Elsevier, Amsterdam, 1990:209-20.
[23]. Henrion M. An introduction to algorithms for inference in belief nets. In: Machine Intelligence and Pattern Recognition: Uncertainty in Artificial Intelligence 5, Vol 10. Henrion M, Shachter R, Kanal LN, Lemmer JF .eds. Elsevier, Amsterdam, 1990:129-38.
[24]. Shachter RD, Peot M. Simulation Approaches to General Probabilistic Inference on Belief Networks. In: Machine Intelligence and Pattern Recognition: Uncertainty in Artificial Intelligence 5, Vol 10. Henrion M, Shachter R, Kanal LN, Lemmer JF .eds. Elsevier, Amsterdam, 1990:221-31.
[25]. Pryor TA, Gardner RM, Clayton PD, Warner HR. The HELP system. J Medical System 1983; 7:87-102.
[26]. Lau LM, Warner HR. Performance of a Diagnostic System (Iliad) as a Tool for Quality Assurance. In: Proceedings of the 15th Symposium on Computer Applications in Medical Care (SCAMC), IEEE Computer Society Press. 1991:1005-10.
[27]. Haug PJ, Ranum DL, Frederick PR. Computerized Extraction of Coded Findings from Free-Text Radiologic Reports. Radiology 1990; 174:543-48.