On the Conversion of Rule Bases into Belief Networks

Shijie Wang, Marco Valtorta

University of South Carolina

Abstract

Two types of belief network, namely, Dempster-Shafer belief networks and Bayesian networks, have emerged as an appealing alternative to rule bases in knowledge-based systems development. It is interesting to contrast some topics of knowledge representation for reasoning under uncertainty in rule-based systems and in belief network-based systems. This paper addresses three issues related to the conversion of rule bases into belief networks. We will first discuss a way for converting rule bases into the two different types of belief network, and then examine the relationship between rule chaining in rule bases and belief propagation in Dempster-Shafer belief networks. We will also show the difficulty involved in the conversion of rule bases into Dempster-Shafer belief networks, and point out that it is mainly caused by the exponential growth rate of belief functions, a phenomenon peculiar to Dempster-Shafer belief networks which has severe practical consequences but has long been overlooked.

Introduction

In most existing knowledge-based systems, MYCIN-type rule bases have been the dominating formalism for implementing knowledge bases involving uncertainties. In recent years, however, belief networks have become an appealing alternative. It is therefore interesting to study the relationship between these two formalisms of knowledge representation. For some application domains, it is also worthwhile to convert an existing rule base into a belief network [8].

Depending on the definition of belief functions, belief networks are categorized into two classes, namely, Dempster-Shafer networks (or simply belief networks) and Bayesian networks (sometimes also called belief networks, influence diagrams, causal probabilistic networks or simply causal networks). Several systems have been developed employing either type of belief network for various applications. For example, the MUNIN system designed to assist physicians in the diagnosis of muscle diseases using electromyography was built around a large Bayesian belief network of some 1000 nodes [1], whereas the expert system called Adaptive Reasoner for diagnosing problems in multi-stage air compressors was implemented using Dempster-Shafer networks [4].

In the next section, we will first give some basic definitions of belief network, and then briefly discuss two different ways for converting rule bases into the two types of belief network. In most rule-based systems, the inference procedure is realized through rule chaining (forward chaining and/or backward chaining). When rule bases involve uncertain knowledge, some uncertainty handling scheme must be incorporated into the inference procedure, such as MYCIN's certainty factor scheme. In belief network-based systems, the inference procedure is accomplished through belief propagation based on local combinations and projections. In Section 3, we will examine the relationship between the results of these two different types of inference procedure. Although the conversion of rule bases into either type of belief network is conceptually straightforward, we will show in Section 4 that the conversion into Dempster-Shafer networks could be computationally expensive due to the exponential growth rate of Dempster-Shafer belief functions and discuss the implications of this problem to the development of belief network-based systems.

Dempster-Shafer Networks vs. Bayesian Networks

A Dempster-Shafer network is normally defined as an undirected graph with associated belief functions (see, e.g. [12, 20, 21]). For any variable or node X in such a network, the frame of discernment $\Theta(X)$ is the set of exhaustive and mutually exclusive propositions or hypotheses about X (i.e. all possible values of the variable X). When X is a joint variable that consists of several individual variables, $\Theta(X)$ is the Cartesian product of the frames of all conjoining members. Let $2^{\Theta(X)}$ denote the power set of $\Theta(X)$. A *basic probability assignment (bpa)* function is defined over the frame of X as a mapping $\mathbf{m}_X: 2^{\Theta(X)} \rightarrow [0,1]$. For every subset of the frame $S \subseteq \Theta(X)$ (i.e. $S \in 2^{\Theta(X)}$), $\mathbf{m}_X(S) \geq 0$, $\mathbf{m}_X(\emptyset)=0$ and $\Sigma\{\mathbf{m}_X(S)|S \subseteq \Theta(X)\}=1$. $\mathbf{m}_X(S)$ (the *bpa* value or the *mass* of S) is interpreted as the amount of belief committed exactly to this subset S of the frame $\Theta(X)$ when a piece of evidence directly supports the proposition "The value of X is (in) S" but provides no further discriminating information about individual elements in S. The total belief in the subset S is measured by the *belief function* Bel_X defined as $\mathrm{Bel}_X(S)=\Sigma\{\mathbf{m}_X(R)|R \subseteq S\}$. It immediately follows from the definitions that for any singleton $S \in \Theta(X)$, $\mathrm{Bel}_X(S)=\mathbf{m}_X(S)$, while for the entire frame $\Theta(X)$, $\mathrm{Bel}_X(\Theta(X))=1$ and $\mathbf{m}_X(\Theta(X))=1-\Sigma\{\mathbf{m}_X(S)|S \subset \Theta(X)\}$. The latter in effect defines a measure of ignorance, that is, the amount of belief left in the entire frame due to lack of further evidence. We call a subset S of X's frame $\Theta(X)$ such that $\mathbf{m}_X(S)>0$ a *focal element* of the belief function Bel_X. Usually, a belief function is represented by a list of focal elements along with their associated *bpa* values.

Figure 1(a) shows a Dempster-Shafer network constructed from the widely-used fictitious example (with some changes) about diagnosis of dyspnoea [9]. Note that belief functions for joint variables specify the relations between their member variables, often in the convenient form of if-then rules. For example, heavy smoking is known to be a risk factor for bronchitis. This fact can be depicted by the belief function for the joint variable $R3$ = {*Smoking, Bronchitis*}. Suppose the frame for *Smoking* is {*long&heavy, medium, light*} and the frame for *Bronchitis* is simply {*yes, no*}. Let *rule1* be the statement "*if the patient has a long heavy smoking history then there is a 65% chance that s/he has bronchitis*". This statement can be expressed as $\mathbf{m}_{R3}(rule1)=0.65$ and $\mathbf{m}_{R3}(\Theta(R3))=0.35$.

Unlike Dempster-Shafer networks, a Bayesian network is usually defined as a directed acyclic graph with nodes representing variables and links the causal relations between nodes (see, e.g. [9], [16]). In such a network, belief functions associated with nodes are simply prior or posterior probabilities and belief functions associated with links are specified by conditional probability matrices. The Bayesian network shown in Figure 1(b) represents the joint probability distribution $p(A,S,T,C,B,X,D)$ =

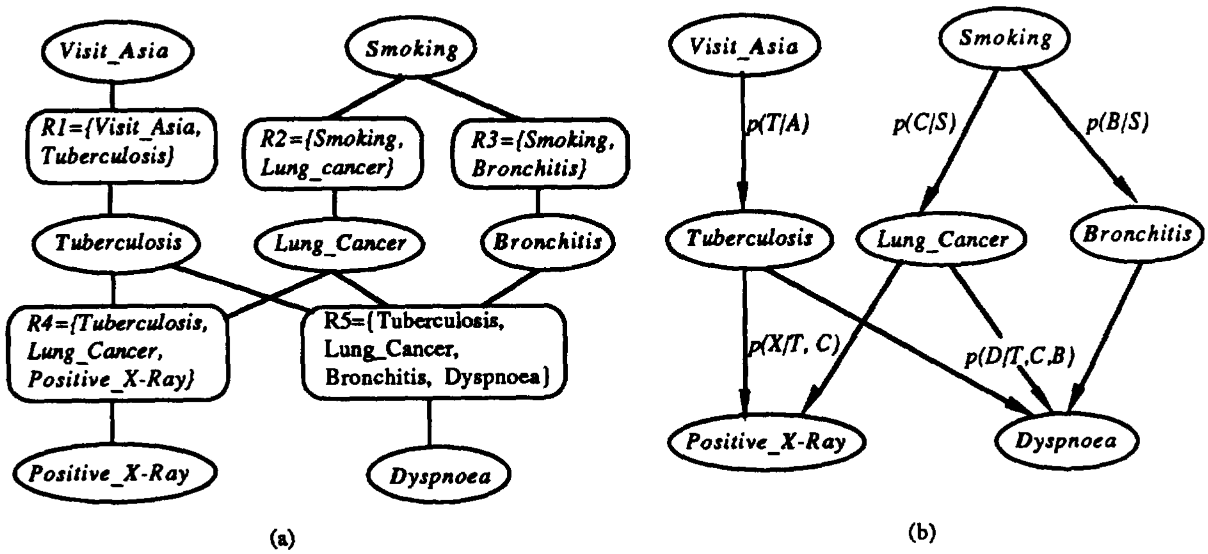

Fig. 1 The belief networks for the example about dyspnoea: (a) Dempster-Shafer network, (b) Bayesian network.

$p(A)p(T/A)p(C/S)p(B/S)p(X/T,C)p(D/T,C,B)$.

When a relation between variables is given in the form of if-then type rule, for instance, "*if* $A=a_i$ *then* $B=b_j$ (with certainty c)", the Bayesian school interprets such a rule as a conditional probability expression $p(B=b_j \mid A=a_i)=c$, while the Dempster-Shafer school views such a rule as an implication $(A=a_i) \Longrightarrow (B=b_j)$ (with certainty c), and converts it to a belief function with two focal elements: $\mathbf{m}_R(S)=c$ and $\mathbf{m}_R(\Theta(R))=1-c$, where $\Theta(R)$ is the joint frame $\Theta(A)\times\Theta(B)$ and $S=\{\neg(A=a_i) \vee (B=b_j)\}=(\Theta(A)-\{a_i\})\times\Theta(B)\cup\Theta(A)\times\{b_j\}$.

There has been much debate on the merits and drawbacks of these two types of networks (see, e.g. [6, 8, 11, 14, 18, 19, 22]). Also, it has been a controversy for a long time whether we should interpret in the first place a statement made by domain experts as a conditional probability or as an if-then rule (and thus an implication). Lewis [10] (as quoted by Goodman [7]) pointed out that "one could not identify implication with conditioning in the probability sense," and "no systematic approach exists for combination of evidence problems when individual inference rules are interpreted through conditional probabilities." Pearl [15] also argued that there is fundamental difference between the role of premise in logic and that of conditional events in probability calculus, and "the statement $p(B \mid A)=p$ denotes totally different operational semantics than the production rule "*if A then B* (with certainty p)." As a matter of fact, one can show that

$$p(A \Longrightarrow B)=p(B \mid A)+p(\neg A)\cdot p(\neg B \mid A), \qquad (1)$$

where $p(A \Longrightarrow B)\geq p(B \mid A)$ and the strict inequality holds unless $p(A)=1$ or $p(A \Longrightarrow B)=1$.

For practical applications, which type of belief network, and hence which interpretation, is more appropriate really depends upon the knowledge structure of the application domain. The Dempster-Shafer theory of belief functions allows beliefs to be assigned to subsets as well as singletons of a frame of discernment while Bayesian belief functions allow singletons only. Hence, one of the main advantages that many proponents have claimed for the Dempster-Shafer type of network is that it has more expressive power and it can represent hierarchical evidence conveniently. For rule bases that deal only with singletons, there is basically no advantage to convert it to a Dempster-Shafer network instead of a Bayesian network. However, sometimes Bayesian networks may be inadequate or inconvenient to model the domain knowledge.

Belief Propagation vs. Rule Chaining

In systems based on Dempster-Shafer belief networks, a belief function representing input evidence or hypothesis can be propagated throughout the network by performing a series of local combinations and projections, provided that the belief network possesses the qualitative Markov property (see, e.g. [12, 20]). There are similar procedures for systems based on Bayesian networks (see, e.g. [9, 16]).

Given two independent Dempster-Shafer belief functions Bel'_X and Bel''_X over the same frame $\Theta(X)$ with their respective *bpa* functions $\mathbf{m}'_X$ and $\mathbf{m}''_X$, which represent two independent pieces of evidence, Dempster's rule combines the two pieces of evidence using the direct sum operation and produces a new belief function denoted by $\mathrm{Bel}_X=\mathrm{Bel}'_X \oplus \mathrm{Bel}''_X$. For any focal element S_k of Bel_X, $S_k \subseteq \Theta(X)$, $1\leq k\leq 2^{|\Theta(X)|}$, the *bpa* function is computed by the following orthogonal products:

$$\mathbf{m}_X(S_k)=\kappa\sum\{\mathbf{m}'_X(S_i)\cdot\mathbf{m}''_X(S_j) \mid S_i, S_j\subseteq\Theta(X), S_i\cap S_j=S_k\} \quad (2)$$

where κ is a normalizing factor that ensures no belief is attributed to the null hypothesis,

$$\kappa^{-1}=1-\sum\{\mathbf{m}'_X(S_i)\cdot\mathbf{m}''_X(S_j) \mid S_i, S_j\subseteq\Theta(X), S_i\cap S_j=\varnothing\}, \quad (3)$$

In case of $\kappa^{-1}=0$, Bel_X is not defined and the two belief functions are said to be not combinable, indicating that the two pieces of evidence completely contradict each other.

In rule-based systems, belief propagation is realized through rule chaining, either forward-chaining or backward-chaining. In the XX system for oil exploration [2], expert knowledge of exploration geologists is represented in the form of rules with associated belief values, similar to the MYCIN rules except that belief values are assigned to subsets of a frame as well as singletons.

The XX system employs Dempster's rule to combine belief functions defined over the same frame. However, propagating belief functions from one frame to another is realized through combined forward-backward chaining, thus XX essentially is still a rule-based system.

In the XX rule base, most rules are in the following form except for control action rules:

if cond_1 & cond_2 & ... & cond_m BF **then**
(concl_1 bf_1) (concl_2 bf_2) ... (concl_n bf_n).

Both conditions and conclusions (hypotheses) are represented as (variable value) pairs. A variable corresponds to an attribute of an object in MYCIN [3]. however, its value can be any subset of the frame of this variable. Each bf_i is a belief value, i.e. basic probability assignment, assigned to the respective conclusion by domain experts, and $\sum_{i=1}^{n} bf_i \leq 1$. Note that at run time, each condition represents a piece of evidence with an associated belief value, obtained either from user supplied evidence or as the results of firing some other rules. BF is a function that computes the overall belief value for the condition part of the rule. In case of multiple conditions, the value of BF will be the minimum of the belief values of all individual conditions. Note that BF corresponds precisely to *TALLY* in a MYCIN rule, and a *bf* value corresponds to a Certainty Factor CF in MYCIN [3].

Suppose that three variables A, B, and C are involved in a set of if-then type rules with frames of $\Theta(A)=\{a_1,a_2,a_3\}$, $\Theta(B)=\{b_1,b_2\}$, and $\Theta(C)=\{c_1,c_2,c_3\}$, respectively. Let $R=\{A,B,C\}$ be the joint variable. Then, any rule involving these three variables can be defined as a focal element of the belief function Bel_R. Let us consider the simple rule R1 using the XX rule format as shown below:

R1: **if** $(A=a_1 \text{ and } B=b_2)$ BF *then* $((C=c_3) \text{ with } bf=bf_1)$.

Suppose now we have obtained two pieces of evidence, represented as belief functions defined on the frames $\Theta(A)$ and $\Theta(B)$, respectively:

$$e1:\quad m_A(\{a_1\})=BF_1,\ \ m_A(\Theta(A))=1-BF_1,$$
$$e2:\quad m_B(\{b_2\})=BF_2,\ \ m_B(\Theta(B))=1-BF_2.$$

In the XX system, by simply taking the minimum of belief values in the conditions as the value of BF and then multiplying it with the belief value in the conclusion $C=c_3$, firing this rule amounts to propagating beliefs (evidence $e1$, $e2$) from their home frames $\Theta(A)$ and $\Theta(B)$ via the composite frame $\Theta(R)$ to the target frame $\Theta(C)$:

$$m_C(\{c_3\})=\min(BF_1,BF_2)\times bf_1 \tag{4}$$

Computing Equation (4) needs only one comparison and one multiplication. This is a rough approximation of belief propagation within Dempster-Shafer's framework. In an exact implementation of the theory of belief functions and belief propagation based on projections, this simple process requires a series of projections and combinations in order to update the belief value of $m_C(\{c_3\})$. Let $\mathrm{Bel}_{S\to T}$ denote the projection of a belief function over the frame $\Theta(S)$ onto the frame $\Theta(T)$, where projections could be a vacuous extension or a marginalization or both depending on whether or not S is a subset or superset of T [12, 20]. Then, the new belief value of $m_C(\{c_3\})$ is obtained by taking

$$\mathrm{Bel}_R \leftarrow \mathrm{Bel}_R \oplus \mathrm{Bel}_{A\to R} \oplus \mathrm{Bel}_{B\to R} \tag{5a}$$

$$\mathrm{Bel}_C \leftarrow \mathrm{Bel}_C \oplus \mathrm{Bel}_{R\to C}. \tag{5b}$$

As the results of this propagation procedure, we have the resulting belief in the hypothesis $C=c_3$, i.e. $m_C(\{c_3\})$ computed by:

$$m_C(\{c_3\})=BF_1\times BF_2\times bf_1. \tag{6}$$

Note that belief values of all other focal elements are also updated in this process whether desired or not. It can be shown that the relation in Equation (6) can be generalized in two ways: more than two premises can be present in each rule, and rules may form a simple chain. Moreover, similar simple equations for the computation of masses in series-parallel belief networks are known to hold [15, Ch. 9]. As an obvious consequence, the more conditions we have in the premise of a rule, the greater the difference is between Equation (4) and Equation (6). Moreover, the longer the path between two nodes in a belief network or the longer the corresponding chain of rules, the greater the difference is between these two approaches. This implies that given a knowledge base, either in the form of a rule base or in the form of a belief network, the resulting belief distribution obtained through belief propagation based on projections tends to be more conservative than those obtained by rule chaining as in XX or MYCIN. Also, the amounts of computation needed differ drastically. However, Equation (4) is of *ad hoc* nature while Equation (6) represents a more coherent treatment of uncertainties with a sound theoretical basis.

Now, suppose XX or MYCIN is modified to take the product of BF_1 and BF_2 instead of MIN in Equation (4), then Equation (4) will be seemingly equivalent to Equation (6). A very natural question is: Is it always possible to get the same results by replacing the complex belief propagation procedure with the simple rule chaining procedure? It is simple to answer the question in the negative, when we are concerned with exact computation of masses or beliefs: since the computation of masses or beliefs using Dempster's rule is intractable [13, 17], while the computation of certainty factors using rules is simple, it is not possible to obtain the same results by substituting a Dempster-Shafer belief network with a MYCIN-style rule base of approximately the same size. It can also be shown that no matter whether we take the minimum or the product, different results are achieved even when we are not concerned with the exact value of the belief (or mass) in a proposition, but only with the relative ranking of a proposition relative to another. We will only sketch the proof of this result (concerning rankings) here.

We propose two formalizations of the notion of "relative ranking." In the first case, we are concerned with whether the belief (or mass) in a proposition is greater than or equal to a threshold (e.g., 0.5). In the second case, the question of interest is whether the belief (or mass) in a proposition is greater than or equal to the belief in a second proposition. (This is a special case of the situation in which we seek the proposition with the highest associated belief.)

Observe that the first case is a special case of the second, by the following construction: let T and P be the threshold and proposition of interest, respectively. Create an instance of the second case by adding a disjoint node with belief T to the belief network of interest. The second question has answer yes for the instance so constructed if and only if the first question has answer yes, since the threshold is now just the belief value of the (dummy) disjoint node.

By the observation in the previous paragraph, we need only show that Dempster's rule and MIN give different results (sometimes) for the first case. Rather than provide an example, we use a computational complexity argument and obtain a rather more general result for free: not only the use of MIN leads to different results than the use of Dempster's rule, but the use of any polynomial function of several beliefs to compute the belief in their combination is disallowed.

Orponen [13] and Provan [17] have independently shown that Dempster's rule is #P-Complete. It is easy to verify that

Orponen's Theorem 3.1 can be strengthened without any change in its proof as follows: Dempster's rule is #P-Complete when the possible values of the mass in a proposition are $1/2^n, 2/2^n, 3/2^n, \cdots, (2^n-1)/2^n, 1$. Note that there are 2^n such values.

If MIN (or any other polynomial-time function used to combine masses) were equivalent to Dempster's rule as far as rankings are concerned, we could solve Orponen's problem in polynomial time by applying binary search, where each "comparison" requires the solution of two ranking problems. (The outcome of each "comparison" is whether the actual value of the mass is less than or equal to or greater than some appropriate threshold.) This contradicts the intractability of Orponen's problem.

A similar argument also applies to Bayesian networks, because Cooper [5] proved that probabilistic inference using Bayesian networks is NP-complete, even when only rankings are desired.

Converting Rule Bases to Dempster-Shafer Belief Networks

We have designed and implemented in AKCL a belief network-based expert system shell BELFUN [24], and the procedure for converting a rule base to a Dempster-Shafer belief network is a part of the system. Currently, a project is under way to extend the system with a graphical user interface in OSF/Motif. The system can be run in three stages or modes. In the first stage, the knowledge engineer can build a new rule base or modify an existing one, and upon completion, invoke the conversion procedure to make a belief network. In the second stage, the knowledge engineer can construct a new belief network or modify an existing one, which could be the result of the first stage. When constructing a new belief network, it's quite natural that the knowledge engineer may as well prefer to specify relations between variables in the form of if-then type rule. The system employes the conversion procedure to represent such rules as belief functions. In this stage, s/he will also specify action rules for control purpose and prepare questions that the system will ask the user at run time (the third stage).

In principle, the conversion of an XX-type rule base to a belief function network is quite straightforward. However, since we have to combine all rules involving the same set of variables into a single belief function using Dempster's rule of combination which is exponential with respect to the size of the joint frame, the conversion could be computationally very expensive if there are large joint frames.

Moreover, even if the joint frames for the rule base are all of moderate sizes, the conversion can still be very expensive if there are large clusterings of rules. This is due to the fact that in general Dempster's rule of combination generates new focal elements at an exponential rate with respect to the number of rules (belief functions) to be combined [23]. Suppose k rules (belief functions) involving m variables are defined over their common frame $\Theta(R)$, the new belief function may have up to $n_1 \times n_2 \times \cdots \times n_k$ focal elements, or n^k focal elements if all n_i's are equal to n. Even if k is still in a moderate range, $n_1 \times n_2 \times \cdots \times n_k$ could have quickly approached $2^{|\Theta(R)|}$. This growth rate decreases when one or more of the belief functions being combined have already been close to $2^{|\Theta(R)|}$. Note that each focal element can be a large subset of m-tuples from the joint frame $\Theta(R)$. For example, in the partition "slope depositional settings" of the XX rule base, rules from *rule736* through *rule742* form a group of seven involving 5 variables. The combined belief function has as many as $3 \times 2^6 = 192$ focal elements. Each of the focal elements has 91 to 104 frame elements (5-tuples). The largest set of rules in this partition has nine rules: *rule703* and *rule768* through *rule775*. Even though the size of their joint frame is only 20 ($=1 \times 1 \times 5 \times 4$), the resulting belief function obtained by converting and combining this set of rules could have up to $2 \times 3^6 \times 4^2 = 729$ focal elements. The actual number of focal elements is 640, each having 2 to 18 frame elements. It took about 30 minutes CPU time on a SUN 3 workstation to convert this partition into a belief network, and about 6 minutes for converting and combining the seven rules *rule736* through *rule742* alone (with one additional rule which is made up to complete the frames for the two variables and thus not combined with the other rules). Yet, this is still far from the worst cases in the XX rule base. Some joint variables have more than 10 rules defined on each of them and some rules have as many as 8 conclusions (i.e. the belief function for such a rule alone has 8 focal elements).

Unfortunately, empirical evidence suggests that rule bases for some types of applications like classification tend to render very dense belief networks with high degrees of node linkage for joint variable nodes and especially with large clusterings of belief functions. In Figure 2, we show the belief network obtained from the partition "slope depositional settings" of the XX rule base. Note that at each joint variable node of networks, instead of showing the name of the joint variable, we list the set of rules that are defined on it. When the set has more than 2 rules, it is indicated by the range of rules in bold face. The joint variable nodes with the three largest set of rules are highlighted. This partition has 88 rules involving 18 individual variables. These variables are also used in two other partitions with 72 and 79 rules concerning "shelf depositional settings" and "basin depositional settings". The largest partition "carbonate facies" has 416 rules. Belief networks constructed from most of the 18 partitions in the XX rule base have very simple topological structures similar to the one shown in Figure 2.

MUNIN [1] and the belief network in [4] (called knowledge graph by Collins) also bear a similar topological structure but with multiple stages (Figure 2 can be viewed as a single stage belief network). And so does the MYCIN rule base if it is reconstructed as a belief network. If the XX rule base were not partitioned, the belief network constructed from it would be much more complex than MUNIN.

Concluding Remarks

The conversion of rule bases into belief networks involves combining sets of rules into a single belief function. However, Dempster-Shafer belief functions allow beliefs assigned to subsets as well as singletons of the frames. The problem of exponential growth rate of Dempster-Shafer belief functions is exactly the cost for this expressive power. Note that when two Bayesian belief functions are combined, the resulting belief function does not increase in size due to this combination because only singletons are involved. Therefore, for any practical application, a trade-off between expressive power (and other factors) and computational complexity must be considered when deciding whether we should build a rule-based system or a belief network-based system, or whether we should convert an existing rule base to a belief network, and whether we should employ a Bayesian network or a Dempster-Shafer network. In order to choose the most appropriate representation, it is very important to first investigate thoroughly the knowledge structure of the application domain under consideration. Dempster-Shafer networks would be an appropriate choice only if the knowledge structure of the application is a sparse belief network.

Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions.

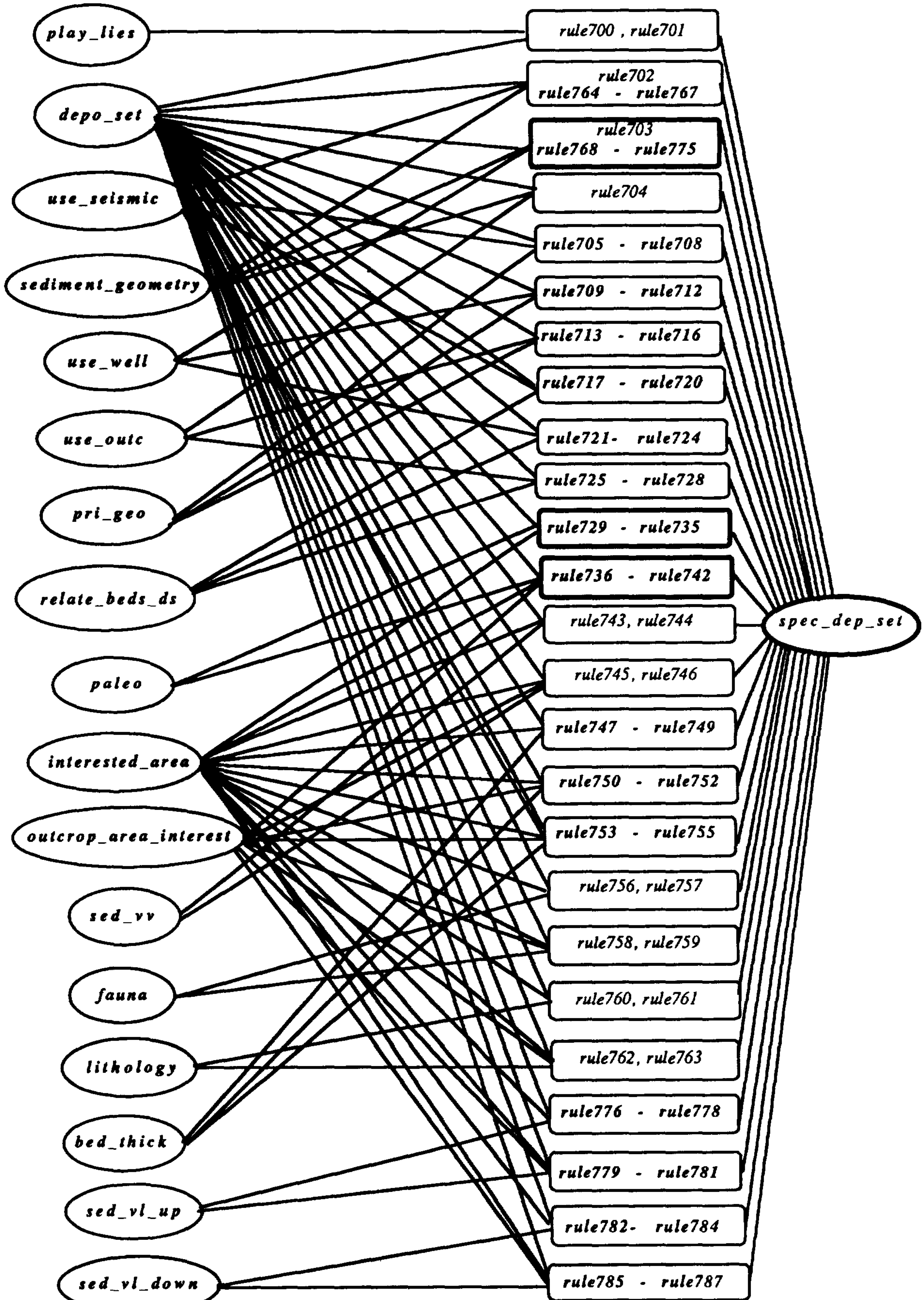

Fig. 2 The belief network constructed from the partition "slope depositional settings " of the XX rulebase.

References

1. Steen Andreassen, Marianne Woldbye, Bjorn Falck, and Stig K. Andersen "MUNIN - A Causal probabilistic network for interpretation of electromyographic findings," in *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, pp. 366-372, Milan, Italy, 1987.
2. G. Biswas and T. S. Anand, "Using the Dempster-Shafer scheme in a diagnostic expert system shell," in *Proceedings of the Third Workshop on Uncertainty in Artificial Intelligence*, pp. 98-105, Seattle, WA, 1987.
3. B. G. Buchanan and E. H. Shortliffe (eds.) *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, Addison-Wesley, Reading, MA, 1984.
4. Glen C. Collins, "Adaptive Heuristic Reasoning for Belief-Function-Based Diagnosis," Ph.D. Dissertation, Department of Electrical Engineering, Vanderbilt University, 1990.
5. Gregory F. Cooper, "Probabilistic Inference Using Belief Networks is NP-Hard," Memo KSL-87-27, Knowledge Systems Laboratory, Stanford University, 1987. (Long version of "The Computational Complexity of Probabilistic Inference Using Bayesian Belief Networks," *Artificial Intelligence*, 42, pp. 393-405, 1990.)
6. R. Fagin and J. Y. Halpern, "Uncertainty, belief and probability," in *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, pp. 1161-1167, Detroit, 1989.
7. I. R. Goodman, "A measure-free approach to conditioning," *Proceedings of the Third Workshop on Uncertainty in AI*, pp. 270-277, Seattle, WA, 1987.
8. J. Gordon and E. H. Shortliffe, "The Dempster-Shafer theory of evidence," in *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, B.G. Buchanan and E.H. Shortliffe (eds.), pp. 272-292, Addison-Wesley, Reading, MA, 1984.
9. S. L. Lauritzen and D. J. Spiegelhalter, "Local computations with probabilities on graphical structures and their application to expert systems (with discussion)," *Journal of the Royal Statistical Society, Series B*, 50, No. 2, pp. 157-224, 1988.
10. D. Lewis, "Probabilities of conditionals and conditional probabilities," *Phil. Rev.*, 85(3), 1976.
11. Dennis V. Lindley, "The Probability Approach to the Treatment of Uncertainty in Artificial Intelligence and Expert Systems (with discussion)," *Statistical Science*, Vol. 2, No. 1, pp. 17-24, 1987.
12. Khaled Mellouli, "On the propagation of beliefs in networks using the Dempster-Shafer theory of evidence," Doctoral dissertation, School of Business, University of Kansas, 1988.
13. Pekka Orponen, "Dempster's Rule of Combination is #P-complete," *Artificial Intelligence*, 44, pp. 245-253, 1990.
14. Judea Pearl, "Reasoning with belief functions: an analysis of compatibility," *International Journal of Approximate Reasoning*, No. 4, pp. 363-389, 1990.
15. J. Pearl, *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, San Mateo, CA, 1988.
16. Judea Pearl, "Fusion, propagation, and structuring in belief networks," *Artificial Intelligence*, 29, pp. 241-288, 1986.
17. Gregory M. Provan, "A logic-based analysis of Dempster Shafer theory," *International Journal of Approximate Reasoning*, No. 4, pp. 451-495, 1990.
18. Glenn Shafer, "Perspectives on the theory and practice of belief functions," *International Journal of Approximate Reasoning*, No. 4, pp. 323-362, 1990.
19. Glenn Shafer, "Probability judgement in artificial intelligence and expert systems (with discussion)," *Statistical Science*, Vol. 2, No. 1, pp. 3-16, 1987.
20. Glenn Shafer, Prakash P. Shenoy, and Khaled Mellouli, "Propagating belief functions in qualitative Markov trees," *International Journal of Approximate Reasoning*, No. 1, pp. 349-400, 1987.
21. Prakash P. Shenoy and Glenn Shafer, "Propagating belief functions with local computations," *IEEE Expert*, vol. 1, pp. 43-52, 1986.
22. P. Smets, "Belief functions (with discussion)," in *Non-Standard Logics for Automated Reasoning*, (eds.) P. Smets *et al.*, pp. 253-286, Academic Press, 1988.
23. S. Wang and M. Valtorta, "On the Exponential Growth Rate of Dempster-Shafer Belief Functions," to appear in *Proceedings of the SPIE Application of AI X Conference*, Orlando, FL, April 1992.
24. S. Wang and M. Valtorta, "A Prototype Belief Network-based Expert System Shell," *Proceedings of the Third International Conference on Industrial & Engineering Applications of Artificial Intelligence & Expert Systems*, pp.509-518, Charleston, SC, July 1990.