REPRESENTATION TECHNIQUES FOR DISTRIBUTED KNOWLEDGE MODELS

Knowledge fusion with aggregation and sampling

Alexander Holland[a] , Madjid Fathi[b] and Stefan Berlik[c]

[a] Institute of Knowledge Based Systems
University of Siegen, 57068 Siegen, Germany
Tel : +49.271.740.2276, Fax : +49.271.740.2322, E-mail : alex@informatik.uni-siegen.de

[b] Institute of Knowledge Based Systems
University of Siegen, 57068 Siegen, Germany
Tel : +49.271.740.2311, Fax : +49.271.740.2322, E-mail : fathi@informatik.uni-siegen.de

[b] Institute of Practical Computer Science
University of Siegen, 57068 Siegen, Germany
Tel : +49.271.740.2617, Fax : +49.271.740.2610, E-mail : berlik@informatik.uni-siegen.de

Keywords: Knowledge management, knowledge fusion, graphical models, distributed artificial intelligence.

Abstract: We address the point of introducing fusion and representation techniques for distributed knowledge sources often available as graphical structures. The distributed artificial intelligence field established new aspects for distributed problem solving and planning or learning in distributed systems with coordination. Knowledge fusion of distributed knowledge sources requires techniques to come to a global broad model including all knowledge network experts. A practical approach in this scenario is the aggregation of medical practitioners in hospitals. Based on probability foundations an aggregation and sampling fusion technique is explained.

1 INTRODUCTION

Distributed artificial intelligence evolved and diversified rapidly during the last years. It is an established and promising research and application field which brings together and draws on results, concepts and ideas from disciplines like artificial intelligence and computer science. Their focus is the study, construction and application of multiagent systems, in which several intelligent agents pursue some set of goal or perform some set of tasks (Weiss, 1999). Agents acts autonomous and their behaviour (Abecker, Bernardi, Elst, 2003) at least partially depends on its own experience. As an intelligent entitiy, an agents operates flexible and rationally in a variety of environmental circumstances given its perceptual and effectual equipment (D'Inverno, Luck, 2003). Behavioral flexibility and rationality are achieved by an agent on the basis of key processes such as problem solving, decision making and learning. As an interacting entity, an agent can be affected in its activities by other agents and perhaps by humans (Wooldrigde, 2000). A key pattern of interaction is goal and task oriented coordination, both in cooperative and competitive situations (Padgham, Winikoff, 2004). Concerning aspects like distributed rational decision making different doctors $d_1,\ldots,d_n$ distributed located should cooperate and medicate diseases. Each doctor has his own knowledge base often compiled available in structured form. We can represent this knowledge as graphical model for each doctor. Fusion aspects are necessary to merge this models come to a global model structure representing the distributed knowledge base D_i. Often the output of classical knowledge management projects is a centralized knowledge management system (KMS), which solely can be accessed by users in a distributed, decentralized way (Bonifacio, Bouquet, Traverso, 2002). These inherently centralized approaches tend to ignore that knowledge is usually distributed in an among complex knowledge based organisations. Distributed KMSs try to overcome these limitations by splintering a centralized KMS into a network of cooperating knowledge nodes (Bonifacio, Bouquet, Cuel, 2002). Knowledge nodes are abstractions of formal or informal units which are parts of networks (Cuel, 2003). We can combine this aspect with our graphical representation aspect to describe models as bayesian network structures and then apply fusion techniques working with knowledge under uncertainty. Knowledge fusion in general is an important component of knowledge science and engineering, which can transform and integrate diversiform knowledge resources to generate new knowledge objects like aggregated knowledge representation structures. We next introduce formal foundations of graphical representations and then pass to the point of fusioning such models (Preece, Hui, 2000) with different fusion techniques like aggregation or sampling methods based on underlying cases stored in knowledge bases (e.g. medical expert systems or decision support systems).

2 GRAPHICAL REPRESENTATIONS AND BAYESIAN NETWORKS

Bayesian networks are graphical models to represent knowledge under conditions of uncertainty. The use of such probabilistic models is based on direct acyclic graphs (DAG) with a probability table for each node. The nodes $\mathcal{V}$ in a Bayesian network represent propositional variables in a domain, the edges $\mathcal{E}$ between the nodes represent the dependency relationship among the variables. Each node has a conditional probability table $P(X \mid X_1,\ldots,X_n)$ attached that quantifies the effects that the parents $X_1,\ldots,X_n$ have on the node. We could say that the conditional probabilities encode the strength of dependencies among the variables. For each a conditional probability distribution is defined that specifies the probabilities of $\mathcal{V}$ given the values of the parents of X (Borgelt, Kruse, 2000). For instance, the recruitment and evolution of qualified employees can represent using graphical models that facilitate a decision process consistent with the company's strategic planning (Holland, 2004). Based on the gathered skills of the employees a modern decision maker has to integrate this knowledge in a decision making process (Holland, Peitzsch, 2005). The decision maker makes decisions by combining his own knowledge, experience and intuition with that available from other sources (Gonzales, Dankel, 1993). Given a learned network structure like Bayesian networks the decision maker can implement additional information in applying an inference algorithm. We use the learned Bayesian network to calculate new probabilities when particular information is achieved. For instance let A have n states with $P(A) = (x_1, ..., x_n)$ and assume that we get the information e that A can only be in state i or j. This statement expresses that all states except i and j are impossible, so next we can illustrate the probability distribution as $P(A, e) = (0,...,0, x_i, 0,...,0, x_j ,0, ...,0)$ (Grimmett, Stirzaker, 2004). Assume a joint probability table $P(U)$ where $\underline{e}$ is the preceding finding (n-dimensional table of zeros and ones). Using the chain rule for Bayesian networks (Russel, Norvig, 2003) we can express the following

$$P(U, e) = \Pi\ \mathrm{A \epsilon U}\ \ P(A \mid \mathrm{parents}(A)) \cdot \Pi\ i\ \underline{e}_i \qquad (1).$$

On constructing Bayesian networks from data sources (e.g. skill data, medical data) we use nodes to represent database attributes. Different Bayesian network structure learning algorithms have been developed. A good overview demonstrating general approaches to graphical probabilistic model learning from data is introduced by (Borgelt, Kruse, 2002) and (Jensen, 2001). In general we can distinguish between the search and score methods and the dependency analysis approach. In the first case the algorithm views the skill learning problem as searching for a structure that best fits the data. The methods start as graphical representation without any edges, using some search method to add an edge to the representation. In the next step they can use score methods to compare the new with the older structure. The main problem to learn Bayesian networks using search and scoring methods is the NP-hard complexity (Kleiter, 1992). Representative algorithms belonging to the search and scoring method are polytree construction algorithms, the K2 algorithm applying a Bayesian scoring method or the Lam-Bacchus algorithm applying the minimal description length principle (Papoulis, Pillai, 2002). Using the second dependency analysis method is a different approach. The algorithms try to discover the dependencies from the data and next use these dependencies to infer the structure. Our approach introduced in the following section belongs to the dependency analysis method without node ordering as an extension. One representative of this second probabilistic model learning algorithms is the boundary DAG algorithm introduced by (Pearl, 1988). Based on the complexity examination we count the number of independent network parameters as

$$|\theta_m| = \sum_i (|X_i| - 1) \cdot |pa(X_i)| \qquad (2)$$

to obtain a comprehensive complexity boundary as model complexity.

3 KNOWLEDGE FUSION

The integration of distributed knowledge is an important concept in many fields like engineering and medical applications. Expert knowledge can be represented as graphical model and special as bayesian network structure as motivated in the preceding section. Constructing a medical expert database causes close collaboration of different experts and doctors. They are often distributed located and have not the ability to come together for the construction of a global complete medical network structure. It is necessary to apply a knowledge fusion technique for the generation of a global graphical structure representing all ambulant patients cases. The cooperativeness of the experts differentiates and tends to result in different representation techniques for distributed knowledge.

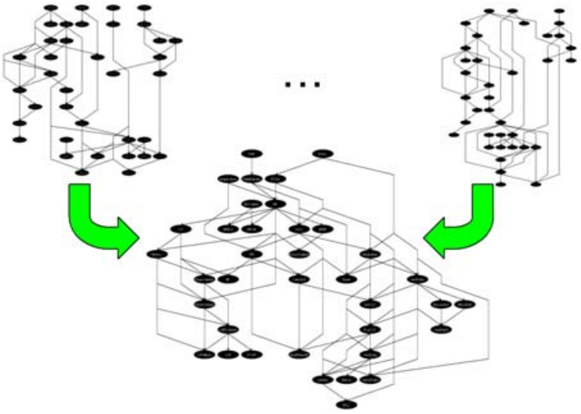

Figure 1: Fusion scenario of distributed knowledge stored in different graphical models

3.1 Concurrent Fusion with Distributed Knowledge

The development of knowledge-based systems involves knowledge acquisition from a diversity of sources often geographically distributed. It is obviously difficult to bring together information from different knowledge sources about a subject of common interest. The sources includes for instance written documents, interviews and application data stored in distributed knowledge bases inserted from different experts often specialized in a medical field like intensive care or a specialist in internal medicine. Intuitively merging medical knowledge bases is to find a knowledge base that has at least as much information as each component medical knowledge base and it is the smallest such medical knowledge base (Pradhan, 1997). Different experts working together are not in a position to generate the complete Bayesian network structure including conditional probability tables and all necessary random variables. An expert E_i (i=1,…,n) working in an specific hospital division has only access to specific medical data or medical knowledge to build a substructure of a complete Bayesian network structure. The main problem occurs when all included experts compose their individual medical knowledge to build up the Bayesian network structure displaying the domain knowledge. We can integrate knowledge stored in different Bayesian networks BN_i through knowledge fusion. Researchers differentiates the aspects concurrent knowledge fusion, complementary knowledge fusion and cooperative knowledge fusion. Complementary knowledge fusion integrates disjoint expert knowledge stored in disjoint network structures based on case databases. This approach can solve the problem description, if uncomplete knowledge appears in the underlying knowledge base. Cooperative knowledge fusion integrates expert knowledge (Holland, Fathi, 2006) depending on the bayesian network structure and domain knowledge from the other experts. This approach is independent on the network structure generation process (structure learning algorithm or manually proposed). We will focus in this paper on concurrent knowledge fusion describe in detail next.

To merge different Bayesian networks (example BN_1 and BN_2) we must first determine a measure for the approximation quality. A suitable measure is the maximal expected Kullback-Leibler divergence between Bayesian network structures like BN_1 and BN_2, where maximization ranges over all possible choices. The Kullback-Leibler divergence (van der Gaag, Renooij, 2001) express the difference or distance between two probability distributions. Given the probability distributions p and q define $KL(p,q)$ as follows: The Kullback-Leibler divergence is obviously not symmetric and can also verbalise using the Cross entropy $H(p,q)$ as follows:

$$KL(p,q) = \sum_{x} p(x) \log \frac{p(x)}{q(x)} \quad (3).$$

The Kullback-Leibler divergence is obviously not symmetric and can also verbalise using the Cross entropy $H(p,q)$ as follows:

$$KL(p,q) = -\sum_{x} p(x) \log q(x) + \sum_{x} p(x) \log p(x) = H(p,q) - H(p) \quad (4).$$

The KL divergence values are not negative with $KL(p,q) = 0$ if and only if $p=q$.

Concurrent Fusion includes combined expert knowledge from different fields like medical problem scenarios. Each expert can generate a case database with embedded and not obvious inferable conditional probability table settings and network structures. The Bayesian network learning algorithms introduced in section 3 delivers Bayesian networks for each expert to fusion via sampling or via LinOp aggregation.

Concurrent fusion via sampling:

1. Synthesize for each expert network a case database using a Monte Carlo technique.
2. Aggregate the expert case databases.
3. Learn the aggregated Bayesian network structure based on the case database determined in 2. using a machine learning algorithm.

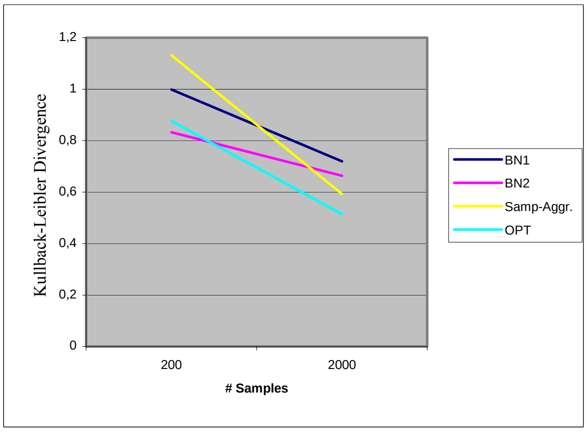

Figure 2: Sampling aggregation results with Kullback-Leibler divergence

We have splitted the ALARM network (Beinlich, Suermondt, Chavez, Chooper, 1989) in two disjoint case databases BN_1 and BN_2 based on the Monte Carlo technique named Probabilistic Logic Sampling. The aggregated, splitted and optimal network structure with varied samples between 200 and 2000 are illustrated in figure 2.

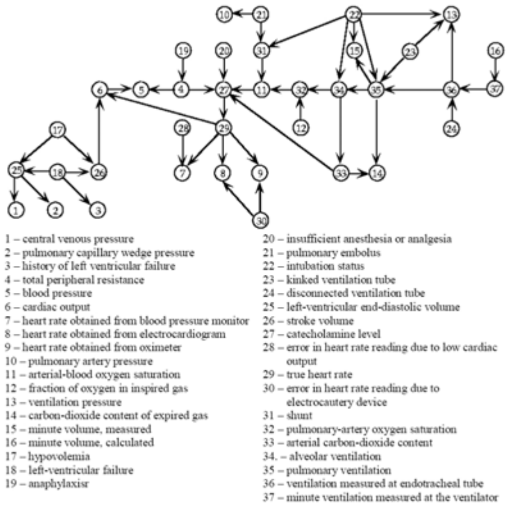

Figure 3: Logical alarm reduction mechanism network structure

We can avoid the aggregation of expert case databases in applying an aggregation operator for a common unified probability distribution. The aggregation of L expert probability distributions $p_1,\ldots,p_L$ using the LinOp operator (Pradhan,1997) leads to the following algorithm we have constructed:

1. Aggregate the probability distributions $p_1,\ldots,p_L$ of L Bayesian networks using the LinOp operator for a common probability distribution $p^* = \sum_{i=1}^{L} \frac{\alpha_i}{\sum_{j=1}^{L} \alpha_j} p_i$ with

   LinOp $LinOP(\alpha_1, p_1, \ldots, \alpha_L, p_L) = \sum_{i=1}^{L} \alpha_i p_i$.
2. Learn the aggregated Bayesian network based on a local score metric based learning algorithm using p^*.

The ALARM network results using concurrent fusion via LinOp aggregation demonstrates in figure 4. BN_1 and BN_2 represents bayesian network structured based on sampled case databases. The optimal bayesian network was founded on the aggregation of BN_1 and BN_2 related case databases and the linOP aggregation curve occurs in applying our introduced fusion technique for both bayesian networks BN_1 and BN_2 (Preece, Hui, 2000) without knowledge about the underlying case databases. We could increase the model accuracy via linOP aggregation in the case of small sample items in comparison with the source networks BN_1 and BN_2.

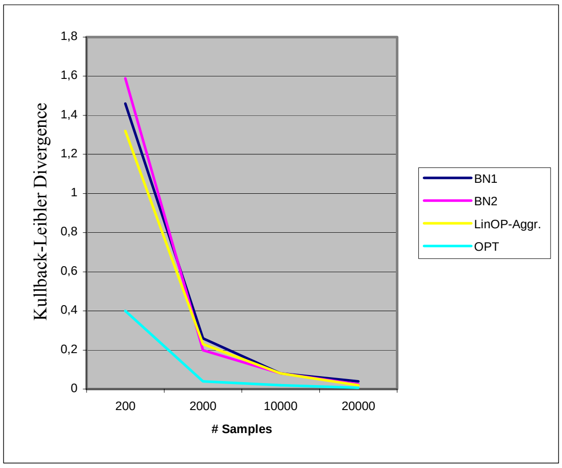

Figure 4: LinOP aggregation results with Kullback-Leibler divergence

4 SUMMARY AND FUTURE WORK

We have presented a new representation technique for global graphical network structures as fusion framework. The sampling and aggregation approach introduced in this paper improves the efficiency and reduces the computational complexity. Tests with medical data sets emphasize the model accuracy measurable with the KL divergence. In the future we will establish the merged network structure as foundation for distributed rational decision making approaches using influence diagrams with coalition formation structures to coordinate expert activities. This could help solving optimization problems of coalitions.

REFERENCES

Abecker, A.; Bernardi, A.; Elst, L., 2003. *Agent Technology for Distributed Organizational Memory*. ICEIIS 2003, pp. 2-10.

Beinlich, I.A.; Suermondt, H.J.; Chavez, R.M.; Chooper, G.F., 1989. *The ALARM Monitoring System. A Case Study with two Probabilistic Inference Techniques for Belief Networks.* Proceedings of the Second European Conference on Artificial Intelligence in Medical Care, pp. 247-256.

Bonifacio, M.; Bouquet, P.; Cuel, R., 2002. *Knowlede-Nodes : The Building Blocks of a Distributed Approach to Knowledge Management*. Proceedings of I-KNOW 2002, Graz, Austria, pp. 191-200.

Bonifacio, M.; Bouquet, P.; Traverso, P., 2002. Enabling Distributed Knowledge Management. Managerial and Technological Implications. Informatik, Vol. 1.

Borgelt, C.; Kruse, R., 2002. *Graphical Models. Methods for Data Analysis and Mining*. John Wiley & Sons, New York.

Cuel, R., 2003. *A New Methodology for Distributed Knowledge Management Analysis.* Proceedings of I-KNOW 2003, Graz, Austria, pp. 531-537.

D'Inverno, M.; Luck, M., 2003. *Understanding Agent Systems.* Software agents activities. Springer, Berlin. 2nd Edition.

Gonzales, A.J.; Dankel, D.D., 1993. *The Engineering of Knowledge-Based Systems in Theory and Practice.* Prentice Hall, Englewood Cliffs, New Jersey.

Grimmett, G.; Stirzaker, D., 2004. *Probability and Random Processes*. Third Edition, Oxford University Press.

Holland, A., 2004. *A Bayesian Approach to Model Uncertainty in Network Balanced Scorecards*. Proceedings of the 10th International Conference on Soft Computing (Mendel Conference), Brno, Czech Republic, pp. 134-138.

Holland, A.; Peitzsch, K.M., 2005. *Competence Management. Learning Skills from Data Based on XML Structured Qualification Profiles*. 5th International Conference on Knowledge Management, Graz, Austria, pp. 170-178.

Holland, A.; Fathi, M., 2006. *Concurrent Fusion via Sampling and LinOP-Aggregation*. Modeling Decisions for Artificial Intelligence, MDAI 2006, Tarragona, Spain, April 3-5, 2006.

Jensen, F.V., 2001. *Bayesian Networks and Decision Graphs.* Statistics for Engineering and Information Science, Springer, Berlin Heidelberg New York.

Kleiter, G.D., 1992. *Bayesian Diagnosis in Expert Systems*. Artificial Intelligence, Vol. 54, pp. 1-32.

Padgham, L.; Winikoff, M., 2004. *Developing Intelligent Agent Systems.* John Wiley & Sons, Wiley Series in Agent Technology.

Papoulis, A.; Pillai, S.U., 2002. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill Sercies in Electrical and Computer Engineering, 4th Edition.

Pearl, J., 1988. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Francisco.

Pradhan, M., 1997. *Focussing Attention in Anytime Decision Making*. PhD thesis, Section on Medical Informatics, Stanford University.

Preece, A.D.; Hui, K.Y., 2000. *The Kraft Architecture for Knowledge Fusion and Transformation*. Knowledge Based Systems, Vol. 13, pp. 113-120.

Russel, S.; Norvig, P., 2003. *Artificial Intelligence. A Modern Approach.* Prentice Hall, New Jersey.

van der Gaag, L.C.; Renooij, S., 2001. *On the evaluation of Probabilistic Networks.* Proceedings of the 8th Conference on Artificial Intelligence in Medicine. Lecture Notes in Computer Science, Vol. 2101. Springer, Berlin Heidelberg New York, pp. 457–461.

Weiss, G., 1999. *Multiagent Systems. A Modern Approach to Distributed Artificial Intelligence*. MIT Press Cambridge (Massachusetts), London (England).

Wooldridge, M., 2000. *Reasoning about Rational Agents. Intelligent Robotics and Autonomous Agents.* Bradford Book.