Proceedings of the 13th Czech-Japan Seminar on Data Analysis and Decision Making in Service Science

Editors: Takeshi ITOH and Kenichi SUZUKI

November 3-5, 2010

University Hall

(Otaru University of Commerce)

Otaru, Japan

On Graphical (Decomposable) Models and Belief Networks in Dempster-Shafer Theory of Evidence

Radim Jiroušek*

Faculty of Management, University of Economics
Institute of Information Theory and Automation, Academy of Sciences
Jindřichův Hradec, Czech Republic
radim@utia.cas.cz

Abstract: When applying AI models one has always to cope with what is often called "curse of multidimensionality" because problems of practice have to be described with a great number of characteristics/features. In probability theory the problem was successfully solved with the help of Graphical Markov Models (GMM), which made it possible to represent and process distributions with hundreds or even thousands of dimensions.

In this paper we show that the ideas on which GMM are based can be applied also within the framework of Dempster-Shafer theory of evidence. We introduce two types of factorization of basic assignments with the help of which we are able to define three types of GMM, namely the class of *graphical models* (in the sense the term was used by the authors of pioneer papers as e.g. Daroch, Lauritzen, Speed, Edwards, Havránek and others in early 80's of the last century), *Bayesian networks* (perhaps the most popular class within GMM), and *decomposable models* that allow for the most efficient computational procedures.

Keywords: discrete belief functions, Dempster-Shafer theory, conditional independence, graphical model, operator of composition

1 Introduction

It is well-known that problems of practice require knowledge bases comprising great number of attributes/properties. Therefore, when considering probabilistic or possibilistic models one has to cope with distributions of hundreds or even thousands of dimensions. The dimensionality explosion starts to be almost hopeless when considering models of Dempster-Shafer theory [3, 13], where the basic assignment is not a point function, like distributions in probability/possibility theories, but a set function. This is why any space-saving technique for model representation and/or processing [15] is in Dempster-Shafer theory of such a great importance.

Speaking about probability theory, a substantial decrease of computational complexity was achieved with the help of Graphical Markov Models (GMM), a technique developed in the last quarter of the last century. Studying properly probabilistic GMMs one can realize that it is the notion of *conditional independence* (which is closely connected with the notion of *factorization*) that makes it possible to represent multidimensional probability distributions efficiently. The goal of this paper is to show that GMMs can be introduced also in Dempster-Shafer theory. The paper is a brief survey summarizing results achieved in the contributions presented at Workshop on Theory of Belief Functions (held in Brest, France, April 2010) [6],

*This research was supported by GAČR under the grant no. ICC/08/E010, and 201/09/1891, and by MŠMT ČR under grants 1M0572 and 2C06019.

13th International Workshop on Non-Monotonic Reasoning (Toronto, Canada, May 2010) [7] and 5th International Conference on Soft Methods in Probability and Statistics (Oviedo and Mieres, Spain, September/October 2010) [8]. As a motivation let us start with recalling the notions of GMMs in probability theory.

2 Probabilistic Graphical Markov Models

Let us consider a probability measure π on a finite space $\mathbf{X}_N = \mathbf{X}_1 \times \mathbf{X}_2 \times \ldots \times \mathbf{X}_n$. For any $K \subseteq N$, symbol $\pi^{\downarrow K}$ will denote its respective marginal measure on subspace $\mathbf{X}_K = \times_{i \in K} \mathbf{X}_i$. Similarly, for a point $x \in \mathbf{X}_N$ its projection into subspace $\mathbf{X}_K$ will be denoted $x^{\downarrow K}$, and for $A \subseteq \mathbf{X}_N$

$$A^{\downarrow K} = \{y \in \mathbf{X}_K : \exists x \in A, x^{\downarrow K} = y\}.$$

By a *join* of two sets $A \subseteq \mathbf{X}_K$ and $B \subseteq \mathbf{X}_L$ we will understand a set

$$A \otimes B = \{x \in \mathbf{X}_{K \cup L} : x^{\downarrow K} \in A \;\;\&\;\; x^{\downarrow L} \in B\}.$$

Consider a probability measure π and three disjoint groups of variables $X_K = \{X_i\}_{i \in K}$, $X_L = \{X_i\}_{i \in L}$ and $X_M = \{X_i\}_{i \in M}$ $(K, L, M \subset N,\ K \neq \emptyset \neq L)$. We say that X_K and X_L are *conditionally independent given* X_M (in probability measure π) if for all $x \in \mathbf{X}_{K \cup L \cup M}$

$$\pi^{\downarrow K \cup L \cup M}(x) \cdot \pi^{\downarrow M}(x^{\downarrow M}) = \pi^{\downarrow K \cup M}(x^{\downarrow K \cup M}) \cdot \pi^{\downarrow L \cup M}(x^{\downarrow L \cup M}).$$

This property will be denoted by $K \perp\!\!\!\perp L \,|\, M \; [\pi]$.

In probability theory, all the graphical Markov models we will deal with in this paper can be defined as probability distributions (measures) factorizing with respect to a graph; either oriented (we call them digraphs) or undirected.

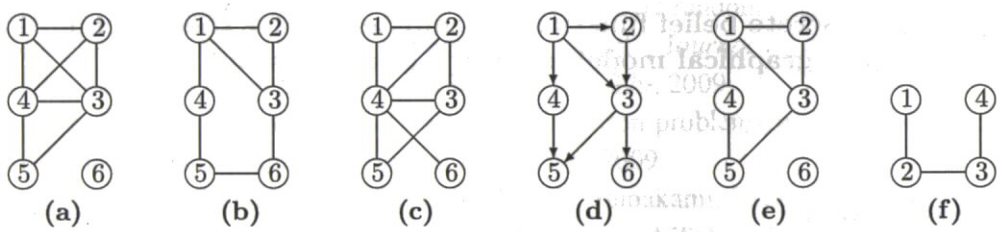

Figure 1: Graphs and a digraph

In addition to standard terms from graph theory, which can be found in any textbook, we will need a couple of advanced notions and nonstandard notation. If i is a node of a digraph then there is a set of its children ($ch(i) = \{j \in V : (i \to j) \in E\}$) and its parents ($pa(i) = \{j \in V : (j \to i) \in E\}$). By $fam(i)$ we understand $pa(i) \cup \{i\}$ (e.g. for the digraph in Figure 1(d) $ch(5) = \emptyset$, $pa(5) = \{3, 4\}$, and $fam(5) = \{3, 4, 5\}$). Let us recall a trivial result of theory of finite digraphs: nodes of any DAG (acyclic digraph) can be ordered so that parents are before their children. For example, for the graph in Figure 1(d) (notice that it is a DAG) the depicted enumeration possesses this property.

A graph (undirected) is *decomposable* if its cliques (maximal complete subsets of nodes) $K_1, K_2, \ldots, K_r$ can be ordered in the way that the sequence meets the so called *running intersection property* (RIP):

$$\forall i = 2, \ldots, r \; \exists j, 1 \leq j < i : K_i \cap (K_1 \cup \ldots \cup K_{i-1}) \subseteq K_j. \tag{1}$$

Notice that this property is met by any ordering of the cliques of the graph in Figure 1(a), and that the cliques of the graph in Figure 1(b) cannot be ordered to meet this property. It means that from the mentioned two graphs only the former is decomposable. The graph in Figure 1(c) is also decomposable, because the ordering of its cliques $\{1,2,4\},\{2,3,4\},\{4,6\},\{3,4,5\}$ meets RIP (in spite of the fact that, for example, $\{3,4,5\},\{1,2,4\},\{2,3,4\},\{4,6\}$ does not meet this property).

The last notions we will need are the notions of a *separation* and a *separating set*. We say that two different nodes $i,j \in N$ are *separated by a set* $K \subseteq N \setminus \{i,j\}$ if we cannot go along the graph edges from i to j without going through a node from K. So, if there is no path from i to j (as, for example there is no path from 1 to 6 in the graph in Figure 1(a)) then even the empty set may be a separating set. A set K is a *minimal separating set* if there exists a pair of nodes i and j, which is separated by K but no proper subset of K separates i and j. Notice that in the graph in Figure 1(c) both $\{2,4\}$ and $\{4\}$ are minimal separating sets; the former is a minimal separating set for 1 and 3, whereas the latter is a minimal separating set for 1 and 6.

If graph $G=(N,E)$ is not complete then it is always possible to find a couple of subsets $L,M \subset N$ (usually there are lot of such couples; the exception is a graph consisting of only two cliques, for which this couple is unique) such that

- $L \cup M = N$;
- $L \cap M$ is a minimal separating set;
- each pair of nodes $i \in L \setminus M$, $j \in M \setminus L$ is separated by $L \cap M$.

The set of all such couples will be denoted by symbol $\mathcal{S}(G)$. Now, we are ready to introduce a class of subsets of $\mathbf{X}_N$ whose structures *comply with graph* G (these sets will be used in the definition of graphical models in Section 4 – Definition 4):

$$\mathcal{R}(G)=\{A \subseteq \mathbf{X}_N : \forall (L,M) \in \mathcal{S}(G) \ \ (A = A^{\downarrow L} \otimes A^{\downarrow M})\}.$$

Probabilistic factorization - graphical model. Consider a graph $G=(N,E)$ with r cliques $K_1,K_2,\ldots,K_r$. We say that a probability distribution π *factorizes with respect to graph* G if there exist r functions $\phi_1,\phi_2,\ldots,\phi_r$,

$$\phi_i : \mathbf{X}_{K_i} \longrightarrow [0,+\infty),$$

such that for all $x \in \mathbf{X}_N$, $\pi(x)=\prod_{i=1}^{r} \phi_i(x^{\downarrow K_i})$. For more details see [12].

Decomposable models. We say that a probability distribution π is *decomposable* if it factorizes with respect to a decomposable graph $G=(N,E)$. It is a famous fact that decomposable distributions can be expressed in a "closed form" as a ration of two products of its marginal distributions.

Lemma 1 *π is decomposable with respect to a graph $G=(N,E)$ (with cliques $K_1,K_2,\ldots,K_r$ – assume they are ordered to meet RIP) if and only if*

$$\pi(x)=\prod_{i=1}^{r} \pi^{\downarrow K_i}(x^{\downarrow K_i \setminus (K_1\cup\ldots\cup K_{i-1})} | x^{\downarrow K_i \cap (K_1\cup\ldots\cup K_{i-1})}) = \frac{\prod_{i=1}^{r} \pi^{\downarrow K_i}(x^{\downarrow K_i})}{\prod_{i=2}^{r} \pi^{\downarrow K_i \cap (K_1\cup\ldots\cup K_{i-1})}(x^{\downarrow K_i \cap (K_1\cup\ldots\cup K_{i-1})})}.$$

Bayesian Networks. We say that a probability distribution π is a Bayesian network with a digraph $G=(N,E)$ if it factorizes with respect to this digraph, which means that

$$\pi(x)=\prod_{i\in N} \pi^{\downarrow fam(i)}(x^{\downarrow \{i\}} | x^{\downarrow pa(i)}) = \frac{\prod_{i\in N} \pi^{\downarrow fam(i)}(x^{\downarrow fam(i)})}{\prod_{i\in N} \pi^{\downarrow pa(i)}(x^{\downarrow pa(i)})}.$$

3 Dempster-Shafer theory

As in the previous section, we consider a finite *multidimensional frame of discernment* $\mathbf{X}_N$. In this paper we consider only normalized basic assignments, i.e. functions $m : \mathcal{P}(\mathbf{X}_K) \to [0,1]$, for which $\sum_{A \subseteq \mathbf{X}_K} m(A) = 1$ and $m(\emptyset) = 0$. Set $A \subseteq \mathbf{X}_K$ is said to be a *focal element* of m if $m(A) > 0$. In analogy to the probabilistic case, we denote marginal basic assignments by $m^{\downarrow K}$.

Conditional Independence. There are several ways how the notion of conditional independence is introduced (see for example papers [1, 2, 11, 14, 16]). In this text we will use the one, which was introduced in [5]* and [9], and which differs from the notion of conditional independence used, for example, by Shenoy [14] and Studený [16] (and which is the same as the *conditional non-interactivity* used by Ben Yaghlane *et al.* in [1]).

Definition 2 Let m be a basic assignment on $\mathbf{X}_N$ and $K, L, M \subset N$ be disjoint, $K \neq \emptyset \neq L$. We say that groups of variables X_K and X_L are *conditionally independent given* X_M *with respect to* m (and denote it by $K \perp\!\!\!\perp L|M\ [m]$), if for any $A \subseteq \mathbf{X}_{K\cup L\cup M}$ such that $A = A^{\downarrow K\cup M} \otimes A^{\downarrow L\cup M}$

$$m^{\downarrow K\cup L\cup M}(A) \cdot m^{\downarrow M}(A^{\downarrow M}) = m^{\downarrow K\cup M}(A^{\downarrow K\cup M}) \cdot m^{\downarrow L\cup M}(A^{\downarrow L\cup M})$$

holds true, and $m^{\downarrow K\cup L\cup M}(A) = 0$ for all $A \subseteq \mathbf{X}_{K\cup L\cup M}$, for which $A \neq A^{\downarrow K\cup M} \otimes A^{\downarrow L\cup M}$.

Operator of Composition. Operator of composition was introduced for probability theory in [4] and its Dempster-Shafer counterpart in [10].

Definition 3 For two arbitrary basic assignments m_1 on $\mathbf{X}_K$ and m_2 on $\mathbf{X}_L$ ($K \neq \emptyset \neq L$) a *composition* $m_1 \triangleright m_2$ is defined for each $C \subseteq \mathbf{X}_{K\cup L}$ by one of the following expressions:

[a] if $m_2^{\downarrow K\cap L}(C^{\downarrow K\cap L}) > 0$ and $C = C^{\downarrow K} \otimes C^{\downarrow L}$ then $(m_1 \triangleright m_2)(C) = \frac{m_1(C^{\downarrow K})\cdot m_2(C^{\downarrow L})}{m_2^{\downarrow K\cap L}(C^{\downarrow K\cap L})}$;

[b] if $m_2^{\downarrow K\cap L}(C^{\downarrow K\cap L}) = 0$ and $C = C^{\downarrow K} \times \mathbf{X}_{L\setminus K}$ then $(m_1 \triangleright m_2)(C) = m_1(C^{\downarrow K})$;

[c] in all other cases $(m_1 \triangleright m_2)(C) = 0$.

4 Graphical Markov Models in D-S Theory

Definition 4 (Factorization) *Let $G = (N, E)$ be a graph with r cliques $K_1, K_2, \ldots, K_r$. We say that basic assignment m* factorizes *with respect to graph G if there exist r functions $\psi_1, \psi_2, \ldots, \psi_r$, ($\psi_i : \mathcal{P}(\mathbf{X}_{K_i}) \longrightarrow [0, +\infty)$), such that for all $A \subseteq \mathbf{X}_N$*

$$m(A) = \begin{cases} \prod_{i=1}^{r} \psi_i(A^{\downarrow K_i}), & \text{if } A \in \mathcal{R}(G), \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Theorem 5 *If basic assignment m factorizes with respect to $G = (N, E)$ and $K \subset N$ separates nodes i, j in G, then $\{i\} \perp\!\!\!\perp \{j\} \,|\, K\,[m]$.*

Example. Consider a 6-dimensional basic assignment factorizing with respect to the graph in Figure 1(e). If all $\mathbf{X}_i$ are binary, then general basic assignment may have up to $2^{64} - 1$ focal elements. Nevertheless, since the considered graph consists of 5 cliques: $\{1,2,3\}$, $\{1,4\}$, $\{3,5\}$, $\{4,5\}$ and $\{6\}$, all the necessary factor functions are defined with by $2^8 + 3 \cdot 2^4 + 2^2 = 308$ numbers.

*In [5] the notion was called conditional irrelevance.

Decomposable Models. Analogously to probabilistic case, basic assignment m is said to be *decomposable* if it factorizes with respect to a decomposable graph. Since it is well-known that decomposable probability distributions had markedly preferential properties in comparison with general graphical distributions, one can expect something similar also in Demster-Shafer theory. And really, the following assertions express specificity concerning decomposable basic assignments.

Theorem 6 *If $K_1, K_2, \ldots, K_r$ are cliques of a decomposable graph $G = (N, E)$ ordered to meet RIP then*

$$\mathcal{R}(G) = \{A \subseteq \mathbf{X}_N : A = A^{\downarrow K_1} \otimes A^{\downarrow K_2} \otimes \ldots \otimes A^{\downarrow K_r}\}.$$

Theorem 7 *Let $K_1, K_2, \ldots, K_r$ be cliques of a decomposable graph $G = (N, E)$ ordered to meet RIP. Basic assignment m factorizes with respect to G if and only if*

$$m = m^{\downarrow K_1} \triangleright m^{\downarrow K_2} \triangleright \ldots \triangleright m^{\downarrow K_r}.$$

Example. Consider a 4-dimensional binary frame of discernment $\mathbf{X}_{\{1,2,3,4\}}$. A general basic assignment can have up to 65 535 focal elements. However, if m is decomposable with respect to the (decomposable) graph in Figure 1(f), then thanks to Theorem 6 the reader can deduce that (due to the requirements made by Definition 4 on m) this basic assignment can have only 658 focal elements. Nevertheless, we know that to define this assignment we do not need to assign values to all these potential focal elements but it is enough to define its three 2-dimensional marginals ($m^{\downarrow\{1,2\}}$, $m^{\downarrow\{2,3\}}$, $m^{\downarrow\{3,4\}}$), which means that it is enough to define only three times 15 numbers.

Belief Networks. To conclude considerations on graphical Markov models in Dempster-Shafer theory of evidence it has remained to say what are the models that are considered a Dempster-Shafer counterpart of Bayesian networks. In analogy with the probabilistic case we have to say what do we understand when saying that a basic assignment factorizes with respect to DAG. The answer is given in the following definition.

Definition 8 (Factorization with respect to DAG) Let $G = (N, E)$ be DAG, and $i_1, i_2, \ldots, i_{|N|}$ be its nodes ordered in the way that parents are before their children. We say that a basic assignment m *factorizes with respect to* G if

$$m = m^{\downarrow fam(i_1)} \triangleright m^{\downarrow fam(i_2)} \triangleright \ldots \triangleright m^{\downarrow fam(i_{|N|})}.$$

5 Conclusions

Inspired by GMM in probability theory, we showed that it is possible to introduce analogous models in Dempster-Shafer theory of evidence. We introduced "classical" graphical models (basic assignments factorizing with respect to undirected graphs), their special subclass of decomposable models, and belief networks which form a D-S theory counterpart to probabilistic Beyesian networks. It is evident that from the computational point of view belief networks and decomposable models deserve a special attention. Both these models can be defined with the help of their marginals, which makes their construction simple and tractable. Namely, the resulting multidimensional basic assignment is constructed from a given system of marginals by an iterative application of the operator of composition. The respective graph gives instructions which marginals should be used and in which order they are to be composed.

We fully agree with the anonymous reviewer of [6] who stated: "The idea of generalizing the fundamental concepts from probability theory to belief functions is very natural." It is simple and natural, and therefore it brings an additional supporting argument in favor of the definition of the conditional independence in Dempster-Shafer theory of evidence as used here in Definition 2.

References

[1] B. Ben Yaghlane, Ph. Smets, and K. Mellouli, "Belief Function Independence: II. The Conditional Case," *Int. J. of Approximate Reasoning*, vol. 31, no. (1-2), pp. 31–75, 2002.

[2] I. Couso, S. Moral and P. Walley, "Examples of independence for imprecise probabilities," in *Proceedings of ISIPTA'99*, G. de Cooman, F. G. Cozman, S. Moral, P. Walley, Eds., location and date, 1999, pp. 121–130.

[3] A. Dempster, "Upper and lower probabilities induced by a multi-valued mapping," *Annals of Mathematical Statistics* vol. 38, pp. 325–339, 1967.

[4] R. Jiroušek, "Composition of probability measures on finite spaces," *Proc. of the 13th Conf. Uncertainty in Artificial Intelligence UAI'97*, (D. Geiger and P. P. Shenoy, eds.). Morgan Kaufmann Publ., San Francisco, California, pp. 274–281, 1997.

[5] R. Jiroušek, "On a conditional irrelevance relation for belief functions based on the operator of composition," in *Dynamics of Knowledge and Belief, Proceedings of the Workshop at the 30th Annual German Conference on Artificial Intelligence,* Ch. Beierle, G. Kern-Isberner, Eds., Fern Universität in Hagen, Osnabrück, 2007, pp. 28-41.

[6] R. Jiroušek, "Factorization and Decomposable Models in Dempster-Shafer Theory of Evidence," in *Proceedings of the Workshop on Theory of Belief Functions,* Brest, 2010.

[7] R. Jiroušek, "Is It Possible to Define Graphical Models in Dempster-Shafer Theory of Evidence?" In: *Proceedings of the 13th Int. Workshop on Non-Monotonic Reasoning,* Toronto, 2010.

[8] R. Jiroušek, "An Attempt to Define Graphical Models in Dempster-Shafer Theory of Evidence," to appear in *Proceedings of the 5th International Conference on Soft Methods in Probability and Statistics.*

[9] R. Jiroušek and J. Vejnarová, "Compositional models and conditional independence in evidence theory." Accepted to *Int. J. Approx. Reasoning.*

[10] R. Jiroušek, J. Vejnarová and M. Daniel, "Compositional models of belief functions," in *Proc. of the 5th Symposium on Imprecise Probabilities and Their Applications,* G. de Cooman, J. Vejnarová, M. Zaffalon, Eds., Praha, date, 2007, pp. 243–252.

[11] G. J. Klir, *Uncertainty and Information. Foundations of Generalized Information Theory.* Wiley, Hoboken, 2006.

[12] S. L. Lauritzen, *Graphical models.* Oxford University Press, 1996.

[13] G. Shafer, *A Mathematical Theory of Evidence.* Princeton University Press, Princeton, New Jersey, 1976.

[14] P. P. Shenoy, "Conditional independence in valuation-based systems," *Int. J. of Approximate Reasoning,* vol. 10, no. 3, pp. 203–234, 1994.

[15] P. P. Shenoy, "Binary join trees for computing marginals in the Shenoy-Shafer architecture," *Int. J. of Approximate Reasoning,* vol. 17, no. (2-3), pp. 239–263, 1997.

[16] M. Studený, "Formal properties of conditional independence in different calculi of AI," in *Proceedings of European Conference on Symbolic and quantitative Approaches to Reasoning and Uncertainty ECSQARU'93,* K. Clarke, R. Kruse, S. Moral, Eds., location and date, Springer-Verlag, 1993, pp. 341–351.