Approximation of Data by Decomposable Belief Models

Radim Jiroušek

Faculty of Management, University of Economics
and
Institute of Information Theory and Automation, Acad. of Sciences
radim@utia.cas.cz

Abstract. It is well known that among all probabilistic graphical Markov models the class of decomposable models is the most advantageous in the sense that the respective distributions can be expressed with the help of their marginals and that the most efficient computational procedures are designed for their processing (for example professional software does not perform computations with Bayesian networks but with decomposable models into which the original Bayesian network is transformed). This paper introduces a definition of the counterpart of these models within Dempster-Shafer theory of evidence, makes a survey of their most important properties and illustrates their efficiency on the problem of approximation of a "sample distribution" for a data file with missing values.

1 Introduction

For data analysis, data preprocessing and management of missing values form an important step substantially influencing the expected results. This concerns in particular the analysis performed with the help of "classical" statistical procedures based on probability. The situation is changing fundamentally when one starts considering models within Dempster-Shafer theory of evidence [13]. In this theory (and it is the main difference with probability theory) one can easily model ignorance and therefore missing data may remain missing - unknown. Unfortunately, nothing is free and this advantage is paid by an increase of computational complexity. This is due to the fact that a basic assignment, in contrast to a probability distribution, cannot be represented by a point function. Therefore any idea decreasing computational complexity of the necessary procedures is desirable.

The goal of this paper is to show that within the framework of Dempster-Shafer theory one can construct decomposable models and that their representation is much less space-demanding than general Dempster-Shafer models. Moreover, by an example of data approximation with the help of a decomposable model we show that we gain not only an efficient representation of basic assignments but also possibility to design efficient ("local") computational procedures.

E. Hüllermeier, R. Kruse, and F. Hoffmann (Eds.): IPMU 2010, Part I, CCIS 80, pp. 40–49, 2010.

After introducing the necessary notation we will define an operator of composition, which plays a leading role in the definition of decomposable models. When introducing these models in Section 3 we will also deal with the concepts of independence and bring reasons in favor of a (relatively) new definition of conditional independence in Dempster-Shafer theory.

2 Basic Notion

2.1 Set Notation

In the whole paper we shall deal with a finite number of variables $X_1, X_2, \ldots, X_n$ each of which is specified by a finite set $\mathbf{X}_i$ of its values. So, we will consider multidimensional space of discernment

$$\mathbf{X}_N = \mathbf{X}_1 \times \mathbf{X}_2 \times \ldots \times \mathbf{X}_n,$$

and its *subspaces*. For $K \subset N = \{1, 2, \ldots, n\}$, $\mathbf{X}_K$ denotes a Cartesian product of those $\mathbf{X}_i$, for which $i \in K$:

$$\mathbf{X}_K = \bigtimes_{i \in K} \mathbf{X}_i.$$

A *projection* of $x = (x_1, x_2, \ldots, x_n) \in \mathbf{X}_N$ into $\mathbf{X}_K$ will be denoted $x^{\downarrow K}$, i.e. for $K = \{i_1, i_2, \ldots, i_\ell\}$

$$x^{\downarrow K} = (x_{i_1}, x_{i_2}, \ldots, x_{i_\ell}) \in \mathbf{X}_K.$$

Analogously, for $K \subset L \subseteq N$ and $A \subset \mathbf{X}_L$, $A^{\downarrow K}$ will denote a *projection* of A into $\mathbf{X}_K$:

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

Let us remark that we do not exclude situations when $K = \emptyset$. In this case $A^{\downarrow \emptyset} = \emptyset$.

Set $A \subseteq \mathbf{X}_N$ is said to be a *point-cylinder* if it can be expressed as a Cartesian product of a singleton and a subspace $\mathbf{X}_L$. More precisely: a point-cylinder is a set $A \subseteq \mathbf{X}_N$ for which there exists an index set (possibly empty) $L \subseteq N$ such that $|C^{\downarrow L}| \leq 1$ and

$$C = C^{\downarrow L} \times \mathbf{X}_{N \setminus L}.$$

Let us stress that if $L = \emptyset$ then $C = \mathbf{X}_N$ (it is the only situation when $|C^{\downarrow L}| < 1$), and when $L = N$ then $|C| = 1$, C is a singleton.

In addition to the projection, in this text we will also need the opposite operation which is called a join. By a *join* of two sets $A \subseteq \mathbf{X}_K$ and $B \subseteq \mathbf{X}_L$ we understand a set

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

Notice that if K and L are disjoint then the join of the corresponding sets is just their Cartesian product $A \otimes B = A \times B$. For $K = L$, $A \otimes B = A \cap B$. If $K \cap L \neq \emptyset$ and $A^{\downarrow K \cap L} \cap B^{\downarrow K \cap L} = \emptyset$ then also $A \otimes B = \emptyset$.

In view of this paper it is important to realize that if $x \in C \subseteq \mathbf{X}_{K\cup L}$, then $x^{\downarrow K} \in C^{\downarrow K}$ and $x^{\downarrow L} \in C^{\downarrow L}$, which means that always

$$C \subseteq C^{\downarrow K} \otimes C^{\downarrow L}.$$

However, it does not mean that $C = C^{\downarrow K} \otimes C^{\downarrow L}$.

2.2 Assignment Notation

The role of a probability distribution from a probability theory is in Dempster-Shafer theory played by any of the set functions: belief function, plausibility function or basic (*probability or belief*) assignment. Knowing one of them, one can deduce the two remaining. In this paper we shall use exclusively basic assignments.

A *basic assignment* m on $\mathbf{X}_K$ $(K \subseteq N)$ is a function

$$m : \mathcal{P}(\mathbf{X}_K) \longrightarrow [0,1],$$

for which

$$\sum_{\emptyset \neq A \subseteq \mathbf{X}_K} m(A) = 1.$$

For the sake of this paper it is reasonable to consider only normalized basic assignments, for which $m(\emptyset)$ equals always 0. If $m(A) > 0$, then A is said to be a *focal element* of m.

Having a basic assignment m on $\mathbf{X}_K$ one can consider its *marginal assignment* on $\mathbf{X}_L$ (for $L \subseteq K$), which is defined (for each $\emptyset \neq B \subseteq \mathbf{X}_L$):

$$m^{\downarrow L}(B) = \sum_{A \subseteq \mathbf{X}_K : A^{\downarrow L} = B} m(A).$$

Basic assignment m is said to be *Bayesian* if all its focal elements are *singletons* (i.e. $m(A) > 0 \implies |A| = 1$). Basic assignment m is said to be *cylindrical* if all its focal elements are point-cylinders. Since each singleton is a point-cylinder, it is obvious that a Bayesian basic assignments is also cylindrical. An advantage of Bayesian and cylindrical basic assignments is that the number of possible focal elements does not grow up superexponentially (as it is for general basic assignments) with the number of dimensions but only exponentially.

2.3 Operator of Composition

Definition 1. 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 $(m_1 \triangleright m_2)(C) = 0$.

First of all we want to stress that the operator of composition is something else than the famous Dempster's rule of combination [4]. For example it is (in contrary to Dempster's rule) neither commutative nor associative. In [9,8] we proved a number of properties concerning the operator of composition. In view of the forthcoming text the following ones are the most important(m_1 and m_2 are basic assignments defined on $\mathbf{X}_K, \mathbf{X}_L$, respectively):

(i) $m_1 \triangleright m_2$ is a basic assignment on $\mathbf{X}_{K\cup L}$;
(ii) $(m_1 \triangleright m_2)^{\downarrow K} = m_1$;
(iii) $m_1 \triangleright m_2 = m_2 \triangleright m_1 \quad \Longleftrightarrow \quad m_1^{\downarrow K\cap L} = m_2^{\downarrow K\cap L}$;
(iv) If $A \subseteq \mathbf{X}_{K\cup L}$ is a focal element of $m_1 \triangleright m_2$ then $A = A^{\downarrow K} \otimes A^{\downarrow L}$;
(v) If m_1 and m_2 are cylindrical then $m_1 \triangleright m_2$ is also cylindrical.

3 Decomposable Models

3.1 Conditional Independence

First, let us present a generally accepted notion of unconditional independence[1] ([1,14,16]).

Definition 2. Let m be a basic assignment on $\mathbf{X}_N$ and $K, L \subset N$ be nonempty disjoint. We say that groups of variables X_K and X_L are *independent*[2] with respect to basic assignment m (in notation $K \perp\!\!\!\perp L\ [m]$) if for all $A \subseteq \mathbf{X}_{K\cup L}$

$$m^{\downarrow K\cup L}(A) = (m^{\downarrow K} \circledcap m^{\downarrow L})(A^{\downarrow K\cup L}).$$

Symbol $\circledcap$ denotes the famous *conjunctive combination rule* (non-normalized Dempster's rule of combination). It was proved in [8] that Definition 2 is equivalent to the following Definition 2a.

Definition 2a. Let m be a basic assignment on $\mathbf{X}_N$ and $K, L \subset N$ be nonempty disjoint. We say that groups of variables X_K and X_L are *independent* with respect to basic assignment m if for all $A \subseteq \mathbf{X}_{K\cup L}$

$$m^{\downarrow K\cup L}(A) = \begin{cases} m^{\downarrow K}(A^{\downarrow K}) \cdot m^{\downarrow L}(A^{\downarrow L}) & \text{if } A = A^{\downarrow K} \times A^{\downarrow L}, \\ 0 & \text{otherwise.} \end{cases}$$

[1] Some authors call it *marginal* independence.

[2] Couso et al. [3] call this independence *independence in random sets*, Klir [11] (*non-interactivity)).*

Though it it not obvious these definitions are equivalent to each other. However, when considering a generalization of these definitions to the conditional case we can get different notions. Most of the authors use the generalization based on Definition 2 (see for example papers [2,3,11,14,15,16]). In this text we will use a simple and straightforward generalization of Definition 2a, which was introduced in [6] and [8], and which can hardly be expressed with the help Dempster's rule of combination (or with the help of its non-normalized version: conjunctive combination rule). The resulting notion differs from the notion of conditional independence used, for example, by Shenoy [14] and Studený [16] (their notion of conditional independence is the same as the *conditional non-interactivity* used by Ben Yaghlane *et al.* in [2]).

Definition 3. 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 $A = A^{\downarrow K\cup M} \otimes A^{\downarrow L\cup M}$ the equality

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

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

Our definition (in the same way as the definition used in [2,14,16]) meets the following important properties:

- for $M = \emptyset$ the concept coincides with Definition 2;
- the notion meets all the properties required from the notion of conditional independence, so-called *semigraphoid properties* ([12,16,17]):

  (A1) $K \perp\!\!\!\perp L \,|\, M\,[m] \implies L \perp\!\!\!\perp K \,|\, M\,[m]$;
  (A2) $K \perp\!\!\!\perp L \cup M \,|\, J\,[m] \implies K \perp\!\!\!\perp M \,|\, J\,[m]$;
  (A3) $K \perp\!\!\!\perp L \cup M \,|\, J\,[m] \implies K \perp\!\!\!\perp L \,|\, M \cup J\,[m]$;
  (A4) $(K \perp\!\!\!\perp L \,|\, M \cup J\,[m])$ & $(K \perp\!\!\!\perp M \,|\, J\,[m]) \implies K \perp\!\!\!\perp L \cup M \,|\, J\,[m]$.

The main differences between our definition and that used in [2,14,16]) are the following

- our definition does not suffer from the *inconsistency with marginalization*[3];
- for our notion, the Dempster-Shafer counterpart to the probabilistic factorization lemma has been proved in [7].

[3] As it was showed by Studený, when the definition used in [2,14,16] is accepted, then it can happen that for two consistent overlapping basic assignments there does not exist their common extension with the required conditional independence property (for the Studený's example see [2,8]).

3.2 Decomposition

Consider a sequence $K_1, K_2, \ldots, K_r$ meeting the *running intersection property* (RIP), i.e. the sequence for which for all $i = 2, \ldots, r$ there exists $j, 1 \leq j < i$, such that

$$K_i \cap (K_1 \cup \ldots \cup K_{i-1}) \subseteq K_j.$$

Without a loss of generality we will assume that $K_1 \cup \ldots \cup K_r = N$.

Definition 4. We say that a basic assignment m is decomposable (with respect to a sequence $K_1, K_2, \ldots, K_r$ meeting RIP) if

$$m = (\ldots((m_1 \triangleright m_2) \triangleright m_3) \triangleright \ldots \triangleright m_{r-1}) \triangleright m_r.$$

In [7] we showed that, analogously to probabilistic decomposable models, also Dempster-Shafer decomposable models possess special independence structures described in the following assertion.

Theorem 1. *If a basic assignment m is decomposable with respect to a sequence $K_1, K_2, \ldots, K_r$ (meeting RIP) then for all $i = 2, \ldots, r$*

$$(K_i \setminus (K_1 \cup \ldots \cup K_{i-1})) \perp\!\!\!\perp ((K_1 \cup \ldots \cup K_{i-1}) \setminus K_i) \mid (K_i \cap (K_1 \cup \ldots \cup K_{i-1})) \; [m].$$

As showed in the following example, the dependence structure of decomposable models allows for their very efficient representation.

Example 1. Consider a 4-dimensional basic assignment on $\mathbf{X}_{\{1,2,3,4\}} = \mathbf{X}_1 \times \mathbf{X}_2 \times \mathbf{X}_3 \times \mathbf{X}_4$, where $|\mathbf{X}_i| = 2$ for all $i = 1, 2, 3, 4$. Since there are

$$2^{(2^4)} - 1 = 2^{16} - 1 = 65\ 535$$

nonempty subsets of the considered frame of discernment, this number expresses also the maximum number of focal elements of a general basic assignment. However, the situation drastically simplifies when one considers a basic assignment decomposable with respect (let us say) $\{1, 2\}, \{2, 3\}, \{3, 4\}$ (it is obvious that this sequence meets RIP). The simplification follows immediately from the fact that, due to Theorem 1, there is a system of conditional independence relations valid for basic assignment m. From this one can deduce that for all the focal elements A of m

$$A = A^{\downarrow\{1,2\}} \otimes A^{\downarrow\{2,3\}} \otimes A^{\downarrow\{3,4\}}$$

holds true. This equality holds only for 657 out of 65 535 nonempty subsets of $\mathbf{X}_{\{1,2,3,4\}}$. Nevertheless, thanks to the fact that

$$m = m^{\downarrow\{1,2\}} \triangleright m^{\downarrow\{2,3\}} \triangleright m^{\downarrow\{3,4\}},$$

we do not need to store basic assignment m but only its three marginals $m^{\downarrow\{1,2\}}$, $m^{\downarrow\{2,3\}}$ and $m^{\downarrow\{3,4\}}$. Each of them has at most 15 focal elements and therefore one needs only 45 numbers to represent this 4-dimensional decomposable basic assignment.

4 Approximation of a Primitive Sample Assignment

In this section we will show that application of decomposable models may result not only in possibility to store multidimensional[4] basic assignments but also in possibility to compute with them using "local" computational procedures. Let us illustrate this possibility by the way of example of approximation of a data file with the help of a decomposable basic assignment. The reader is asked to keep in mind that it is just an illustration. We do not propose to realize the following primitive procedure for practical applications. Because of lack of space we cannot present here any more sophisticated process based on more complex ideas like the procedures studied in [5].

Having a data file with missing values one can quite naturally assign to each data record a point-cylinder $C = C^{\downarrow L} \times \mathbf{X}_{N \setminus L}$ from $\mathbf{X}_N$ expressing that the record contains $|L|$ specific data values corresponding to $C^{\downarrow L}$ and $|N \setminus L|$ missing values. By a *primitive sample assignment* we will understand a basic assignment m, where value $m(C)$ is computed as a relative frequency (within the data file) of records assigned with point cylinder C. It means that any primitive sample assignment is cylindrical.

The approximation task is to find a decomposable basic assignment m, which is in a sense best approximation of the primitive sample assignment for a given data file. For this one has to specify a criterion according to which a "goodness" of the approximation is evaluated. To do so one can consider a number of possible divergences proposed in literature (for a nice survey see [10]). However, not all of them are such that they make the "local" computations possible. As the simplest example of a suitable distance let us consider a "relative entropy" type of divergence defined

$$Div(m; \bar{m}) = \sum_{A \subseteq \mathcal{F}(m)} m(A) \log \frac{m(A)}{\bar{m}(A)},$$

where m is the primitive basic assignment to be approximated, $\bar{m}$ is an approximating decomposable basic assignment and $\mathcal{F}(m) \subset \mathbf{X}_N$ is the set of focal elements of m. It is well known that this divergence is always nonnegative, equals 0 if and only if $m = \bar{m}$ and if

$$m(A) > 0 \quad \Longrightarrow \quad \bar{m}(A) > 0 \tag{1}$$

then it is also finite.

Consider a sequence $K_1, K_2, \ldots, K_r$ meeting RIP such that $K_1 \cup \ldots \cup K_r = N$, and an arbitrary basic assignment $\bar{\bar{m}}$ decomposable with respect to $K_1, \ldots, K_r$. Further define a decomposable basic assignment constructed from the marginals of m

$$\bar{m} = (\ldots((m^{\downarrow K_1} \triangleright m^{\downarrow K_2}) \triangleright m^{\downarrow K_3}) \triangleright \ldots \triangleright m^{\downarrow K_{r-1}}) \triangleright m^{\downarrow K_r}. \tag{2}$$

[4] When speaking about multidimensionality in connection with Dempster-Shafer theory we have in mind several tens rather than hundreds of dimensions.

It is not difficult to show that for $\bar{m}$ defined by formula (2) implication (1) is valid (this is because we assume that m is cylindrical and for a cylinder A, $A = A^{\downarrow K} \otimes A^{\downarrow L}$ always holds true) and so we get that $Div(m;\bar{m})$ is finite. Moreover

$$Div(m;\bar{m}) \leq Div(m;\bar{\bar{m}}).$$

This is why it is enough to look for an approximation of m in the form of a *compositional model* (2)[5].

Let us now show that the search for the best approximation of a basic assignment m (i.e. for the most advantageous sequence $K_1, K_2, \ldots, K_r$ meeting RIP) can be based on "local" computations only, i.e. that the procedure stores only and computes with marginal basic assignments $m^{\downarrow K_i}$.

To make our consideration more lucid, consider first $r = 2$. For this

$$Div(m;\bar{m}) = \sum_{A \subseteq \mathcal{F}(m)} m(A) \log \frac{m(A)}{(m^{\downarrow K_1} \triangleright m^{\downarrow K_2})(A)}.$$

The following modifications are correct because for $A \subseteq \mathcal{F}(m)$

$$m^{\downarrow K_1 \cap K_2}(A^{\downarrow K_1 \cap K_2}) > 0$$

and therefore value of $(m^{\downarrow K_1} \triangleright m^{\downarrow K_2})(A)$ is positive and computed according to case [**a**] of Definition 1.

$$\begin{aligned}
Div(m;\bar{m}) &= \sum_{A \subseteq \mathcal{F}(m)} m(A) \log \frac{m(A)}{(m^{\downarrow K_1} \triangleright m^{\downarrow K_2})(A)} \\
&= \sum_{A \subseteq \mathcal{F}(m)} m(A) \log \frac{m(A) \cdot m^{\downarrow K_1 \cap K_2}(A^{\downarrow K_1 \cap K_2})}{m^{\downarrow K_1}(A^{\downarrow K_1}) \cdot m^{\downarrow K_2}(A^{\downarrow K_2})} \\
&= \sum_{A \subseteq \mathcal{F}(m)} m(A) \log m(A) + \sum_{A \subseteq \mathcal{F}(m)} m(A) \log m^{\downarrow K_1 \cap K_2}(A^{\downarrow K_1 \cap K_2}) \\
&\quad - \sum_{A \subseteq \mathcal{F}(m)} m(A) \log m^{\downarrow K_1}(A^{\downarrow K_1}) - \sum_{A \subseteq \mathcal{F}(m)} m(A) \log m^{\downarrow K_2}(A^{\downarrow K_2}).
\end{aligned}$$

The second term of the last expression can be simplified in the following way

$$\begin{aligned}
\sum_{A \subseteq \mathcal{F}(m)} m(A) \log m^{\downarrow K_1}(A^{\downarrow K_1}) &= \sum_{B \subseteq \mathcal{F}(m^{\downarrow K_1})} \sum_{\substack{A \subseteq \mathcal{F}(m) \\ A^{\downarrow K_1} = B}} m(A) \log m^{\downarrow K_1}(A^{\downarrow K_1}) \\
&= \sum_{B \subseteq \mathcal{F}(m^{\downarrow K_1})} \log m^{\downarrow K_1}(B) \sum_{\substack{A \subseteq \mathcal{F}(m) \\ A^{\downarrow K_1} = B}} m(A) \\
&= \sum_{B \subseteq \mathcal{F}(m^{\downarrow K_1})} m^{\downarrow K_1}(B) \log m^{\downarrow K_1}(B).
\end{aligned}$$

[5] Notice that in spite of the fact that the described approximation does not decrease the number of focal elements, it can be very efficiently represented.

Denoting

$$\mathfrak{H}(m^{\downarrow K_1}) = - \sum_{B \subseteq \mathcal{F}(m^{\downarrow K_1})} m^{\downarrow K_1}(B) \log m^{\downarrow K_1}(B),$$

and using analogous symbols also for the other marginals of m we get

$$Div(m; \bar{m}) = \mathfrak{H}(m^{\downarrow K_1}) + \mathfrak{H}(m^{\downarrow K_2}) - \mathfrak{H}(m^{\downarrow K_1 \cap K_2}) - \mathfrak{H}(m).$$

Repeating the above computations for a general r one gets

$$Div(m; \bar{m}) = \mathfrak{H}(m^{\downarrow K_1}) + \left(\sum_{i=2}^{r} \mathfrak{H}(m^{\downarrow K_i}) - \mathfrak{H}(m^{\downarrow K_i \cap (K_1 \cup \ldots \cup K_{i-1})}) \right) - \mathfrak{H}(m).$$

This formula shows that when searching for a suitable sequence $K_1, K_2, \ldots, K_r$ meeting RIP one can omit the term $\mathfrak{H}(m)$ because it appears in all compared expressions. Moreover, when modifying the sequence $K_1, K_2, \ldots, K_r$ only slightly one usually does not need to recompute all the terms

$$\left(\mathfrak{H}(m^{\downarrow K_i}) - \mathfrak{H}(m^{\downarrow K_i \cap (K_1 \cup \ldots \cup K_{i-1})}) \right)$$

but only some of them. These properties indicate that a quite efficient method searching for a suboptimal approximation can easily be designed.

5 Conclusions

In the paper we supported a relatively new notion of conditional independence for Dempster-Shafer theory of evidence. This notion was first introduced in [6] (in that paper under the name of *conditional irrelevance*, though) and later also in [8] and [7], where its theoretical properties were studied. It appears that our notion (in comparison with the notion usually used by other authors [2,3,11,14,15,17,16]) possesses more properties of the probabilistic notion of conditional independence: here we have in mind especially that it does not suffer from the *inconsistency with marginalization* [2] and that it enables us to prove the *factorization lemma*. And it is these very properties that enables us to define decomposable models within Dempster-Shafer theory. Perhaps we do not need to stress that we believe that the introduced decomposable models, just as the probabilistic decomposable models, will allow us to design efficient computational procedures for computation in Dempster-Shafer theory.

Acknowledgement

The research was financially supported by the Czech National Science Foundation under the grants no. ICC/08/E010, and 201/09/1891, and by the Ministry of Education of the Czech Republic under grants 1M0572 and 2C06019.

References

1. Ben Yaghlane, B., Smets, P., Mellouli, K.: Belief Function Independence: I. The Marginal Case. Int. J. of Approximate Reasoning 29(1), 47–70 (2002)
2. Ben Yaghlane, B., Smets, P., Mellouli, K.: Belief Function Independence: II. The Conditional Case. Int. J. of Approximate Reasoning 31(1-2), 31–75 (2002)
3. Couso, I., Moral, S., Walley, P.: Examples of independence for imprecise probabilities. In: de Cooman, G., Cozman, F.G., Moral, S., Walley, P. (eds.) Proceedings of ISIPTA 1999, pp. 121–130 (1999)
4. Dempster, A.: Upper and lower probabilities induced by a multi-valued mapping. Annals of Mathematical Statistics 38, 325–339 (1967)
5. Hutter, M., Zaffalon, M.: Bayesian treatment of incomplete discrete data applied to mutual information and feature selection. In: Günter, A., Kruse, R., Neumann, B. (eds.) KI 2003. LNCS (LNAI), vol. 2821, pp. 396–406. Springer, Heidelberg (2003)
6. Jiroušek, R.: On a conditional irrelevance relation for belief functions based on the operator of composition. In: Beierle, C., Kern-Isberner, G. (eds.) Dynamics of Knowledge and Belief, Proceedings of the Workshop at the 30th Annual German Conference on Artificial Intelligence, Fern Universität in Hagen, Osnabrück, pp. 28–41 (2007)
7. Jiroušek, R.: Factorization and Decomposable Models in Dempster-Shafer Theory of Evidence. In: Workshop on the Theory of Belief Functions, Brest, France (2010)
8. Jiroušek, R., Vejnarová, J.: Compositional models and conditional independence in evidence theory. Accepted to Int. J. Approx. Reasoning, doi: 10.1016/j.ijar.2010.02.005
9. Jiroušek, R., Vejnarová, J., Daniel, M.: Compositional models of belief functions. In: de Cooman, G., Vejnarová, J., Zaffalon, M. (eds.) Proc. of the 5th Symposium on Imprecise Probabilities and Their Applications, Praha, pp. 243–252 (2007)
10. Jousselme, A.L., Maupin, P.: On some properties of distances in evidence theory. In: Workshop on the Theory of Belief Functions, Brest, France (2010)
11. Klir, G.J.: Uncertainty and Information. Foundations of Generalized Information Theory. Wiley, Hoboken (2006)
12. Lauritzen, S.L.: Graphical models. Oxford University Press, Oxford (1996)
13. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, Princeton (1976)
14. Shenoy, P.P.: Conditional independence in valuation-based systems. Int. J. of Approximate Reasoning 10(3), 203–234 (1994)
15. Shenoy, P.P.: Binary join trees for computing marginals in the Shenoy-Shafer architecture. Int. J. of Approximate Reasoning 17(2-3), 239–263 (1997)
16. Studený, M.: Formal properties of conditional independence in different calculi of AI. In: Moral, S., Kruse, R., Clarke, E. (eds.) ECSQARU 1993. LNCS, vol. 747, pp. 341–351. Springer, Heidelberg (1993)
17. Studený, M.: On stochastic conditional independence: the problems of characterization and description. Annals of Mathematics and Artificial Intelligence 35, 323–341 (2002)