Conditional Independence and Compositional Models for Belief Functions

Radim Jiroušek*

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

Abstract: The main bottleneck of applications of belief function models to problems of practice lies in the fact that a belief measure, in contrast to a probability or possibility measure, cannot be represented by a density function. It is a set function and for its representation one needs an exponential number of parameters (exponential with the size of a finite space on which the belief function measure is defined). Therefore, one has to employ some approaches enabling reduction of necessary parameters. In the contribution we will discuss an approach utilizing properties of conditional independence relations that will enable us to assemble a multidimensional model from a system of its simple submodels. Unfortunately, concept of conditional independence has been for belief functions defined in several ways and none of them fully meets our requirements. Therefore we will also discuss a new approach how to define this basic concept, which is based on the notion of an *operator of composition.*

Keywords: Belief function, multidimensional model, operator of composition, conditional independence

1 Introduction

The main bottleneck of applications of belief function models to problems of practice lies in the fact that a belief measure, in contrast to a probability or possibility measure, cannot be represented by a density function. It is a set function and for its representation one needs an exponential number of parameters (exponential with the size of a finite space on which the belief function measure is defined). Therefore, one has to employ some approaches enabling reduction of necessary parameters. In this contribution we will discuss an approach utilizing properties of conditional independence relations that will enable us to assemble (*compose*) a multidimensional model from a system of its marginal submodels. This is also the reason why these models are called *compositional models.*

These models were originally proposed as an alternative to Bayesian networks for multidimensional probabilistic distributions representation. They were based on a simple idea: multidimensional distribution was composed from a system of low-dimensional distributions by repetitive application of a special operator of composition. In this paper, such an operator of composition will be introduced for belief functions, and it will be shown that it can be considered as a true generalization of the operator of composition for probability distributions.

*The research was financially supported by GAČR under the grant no. ICC/08/E010, and by Ministry of Education of the Czech Republic by grants no. 1M0572 and 2C06019.
The paper is a version of the paper "Compositional Belief Function Models" accepted for *Proceedings of the 4th International Conference on Soft Computing and Intelligent Systems*, Nagoya, Japan, September 17-21, 2008.

2 Belief Function Models

2.1 Set Projections and Extensions

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 (in the belief function setting it is usually called *frame 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 $i \in K$: $\mathbf{X}_K = \times_{i \in K} \mathbf{X}_i$, and $X_K = \{X_i\}_{i \in K}$ denotes the set of the respective variables.

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 $\mathbf{X}_K$:

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

In addition to the projection, in this text we will need also an opposite operation which will be called extension. By the *extension* of two sets $A \subseteq \mathbf{X}_{K_1}$ and $B \subseteq \mathbf{X}_{K_2}$ we will understand a set

$$A \otimes B = \{x \in \mathbf{X}_{K_1 \cup K_2} : x^{\downarrow K_1} \in A \ \ \& \ \ x^{\downarrow K_2} \in B\}.$$

In what follows, an important role will be played by special sets, which were in [1] called Z-layered rectangles. These are those sets $C \subseteq \mathbf{X}_{K_1 \cup K_2}$ for which

$$C = C^{\downarrow K_1} \otimes C^{\downarrow K_2}.$$

2.2 Composition of Basic Assignments

A belief function is defined with the help of a *basic (probability or belief) assignment* m on $\mathbf{X}_N$, which is a set function

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

with

$$\sum_{A \subseteq \mathbf{X}_N} m(A) = 1.$$

Therefore, for the sake of simplicity, we will not speak about belief functions but about basic assignments: We shall marginalize and compose basic assignments. For each $K \subset N$ *marginal basic assignment* of m is defined (for each $B \subseteq \mathbf{X}_K$):

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

We say that two basic assignments m_1 and m_2 defined on $\mathbf{X}_{K_1}$ and $\mathbf{X}_{K_2}$, respectively, are *projective* if $m_1^{\downarrow K_1 \cap K_2} = m_2^{\downarrow K_1 \cap K_2}$.

Now, we can define the most important notion of this paper, which was originally defined in [4].

Definition 1 *For arbitrary two basic assignments m_1 on $\mathbf{X}_{K_1}$ and m_2 on $\mathbf{X}_{K_2}$ $(K_1 \neq \emptyset \neq K_2)$ a* composition *$m_1 \triangleright m_2$ is defined for each $C \subseteq \mathbf{X}_{K_1 \cup K_2}$ by one of the following expressions:*

[a] *if $m_2^{\downarrow K_1 \cap K_2}(C^{\downarrow K_1 \cap K_2}) > 0$ and $C = C^{\downarrow K_1} \otimes C^{\downarrow K_2}$ then*

$$(m_1 \triangleright m_2)(C) = \frac{m_1(C^{\downarrow K_1}) \cdot m_2(C^{\downarrow K_2})}{m_2^{\downarrow K_1 \cap K_2}(C^{\downarrow K_1 \cap K_2})};$$

[b] *if $m_2^{\downarrow K_1 \cap K_2}(C^{\downarrow K_1 \cap K_2}) = 0$ and $C = C^{\downarrow K_1} \times \mathbf{X}_{K_2 \setminus K_1}$ then $(m_1 \triangleright m_2)(C) = m_1(C^{\downarrow K_1})$;*

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

2.3 Basic Properties of the Operator of Composition

Let us stress, for the reader familiar with the Dempster's rule of combination [6], that the introduced operator is something quite different.

First, Dempster's rule of combination was defined for two basic assignments defined on the same frame of discernment. In contrast to this, there is no restriction regarding frames of discernments of arguments connected with the introduced operator of composition. Nevertheless, composition of basic assignments defined on the same frame of discernment is uninteresting, because in this case the result is always the first argument - see property (2) of Lemma 2.

The reader should keep in mind that the operator of composition was designed for the situations when one has two basic assignments defined on different frames of discernment and wants to get a new basic assignments defined on a larger frame of discernment incorporating (as much as possible of) the information contained in the original basic assignments.

In this section we shall recollect most of the important properties of the operator of composition, most of which were proved in [4].

Lemma 2 *For arbitrary two basic assignments m_1 on $\mathbf{X}_{K_1}$ and m_2 on $\mathbf{X}_{K_2}$ the following properties hold true:*

1. $m_1 \triangleright m_2$ *is a basic assignment on* $\mathbf{X}_{K_1 \cup K_2}$;
2. $(m_1 \triangleright m_2)^{\downarrow K_1} = m_1$;
3. $m_1 \triangleright m_2 = m_2 \triangleright m_1 \iff m_1^{\downarrow K_1 \cap K_2} = m_2^{\downarrow K_1 \cap K_2}$;
4. *If* $L \subseteq K_1$ *then* $m_1^{\downarrow L} \triangleright m_1 = m_1$;
5. *If* $K_1 \supseteq (K_2 \cap K_3)$ *then* $(m_1 \triangleright m_2) \triangleright m_3 = (m_1 \triangleright m_3) \triangleright m_2$.

Realize that property (3) of the preceding Lemma says that the operator is commutative if and only if it is applied to two projective basic assignments. Generally, it is neither commutative nor associative.

3 Composition of Bayesian Basic Assignments

It is well known that if all focal elements (subsets of frame of discerment for which the basic assignment is positive) of m are *singletons*, i.e. if $m(A) > 0$ implies that $|A| = 1$, then this basic assignment corresponds to a probability distribution, and it is why some authors call it *Bayesian basic assignment.* Regarding the fact that operators of composition were originally defined for composition of probability distributions* a natural question arises: What is the relation of compositional models in these two theoretical frameworks? To answer this question we shall compare the properties of the corresponding operators of composition. But first, let us recollect how the operator of composition is defined in its probabilistic version.

Let us start considering probability distributions p_i defined on $\mathbf{X}_{K_i}$ (i.e. $p_i : \mathbf{X}_{K_i} \longrightarrow [0,1]$ ınd $\sum_{x \in \mathbf{X}_{K_i}} p_i(x) = 1$). Analogously to the notation used for basic assignments, their marginal listributions (for $L \subset K_i$) will be denoted $p_i^{\downarrow L}$. Realize that $p_i(\emptyset) = 0$, but $p_i^{\downarrow \emptyset}(\emptyset) = 1$.

*Probabilistic compositional models were designed as a non-graphical alternative to Bayesian networks and her Graphical Markov models in [3].

Definition 3 Consider arbitrary two probability distributions p_1 and p_2 defined on $\mathbf{X}_{K_1}, \mathbf{X}_{K_2}$, respectively ($K_1 \neq \emptyset \neq K_2$). If $p_1^{\downarrow K_1\cap K_2}$ is dominated by $p_2^{\downarrow K_1\cap K_2}$, i.e.

$$\forall z \in \mathbf{X}_{K_1\cap K_2} \quad p_2^{\downarrow K_1\cap K_2}(z) = 0 \Longrightarrow p_1^{\downarrow K_1\cap K_2}(z) = 0,$$

then $p_1 \triangleright p_2$ is for all $x \in \mathbf{X}_{K\cup L}$ defined by the expression

$$(p_1 \triangleright p_2)(x) = \frac{p_1(x^{\downarrow K_1}) \cdot p_2(x^{\downarrow K_2})}{p_2^{\downarrow K_1\cap K_2}(x^{\downarrow K_1\cap K_2})}.$$

(In case of necessity we define $\frac{0\cdot 0}{0} = 0$.) Otherwise the composition $p_1 \triangleright p_2$ remains undefined.

The reader certainly noticed the main difference between the definitions of operators of composition in the two considered theoretical settings: In contrast to composition of basic assignments, it may happen that the composition of probability distributions in not defined. It occurs when $p_2^{\downarrow K_1\cap K_2}$ does not dominate $p_1^{\downarrow K_1\cap K_2}$. In other words, it is undefined if there would be for some $x \in \mathbf{X}_{K\cup L}$ value $(p_1 \triangleright p_2)(x)$ defined by an indeterminate term

$$(p_1 \triangleright p_2)(x) = \frac{p_1(x^{\downarrow K_1}) \cdot 0}{0}$$

with $p_1(x^{\downarrow K_1}) > 0$.

In [4] we proved that if we compose by the operator† of composition two Bayesian basic assignments, such that the corresponding probability distributions may be composed by the probabilistic operator of composition (i.e. the probabilistic composition is defined) then the resulting distribution is again Bayesian. The assertion we are presenting here is a little bit stronger: It says that the resulting compositions coincide.

Lemma 4 *Let m_1 and m_2 be Bayesian basic assignments on $\mathbf{X}_{K_1}$ and $\mathbf{X}_{K_2}$, respectively, such that for all $A \subseteq \mathbf{X}_{K_1\cap K_2}$ it holds that $m_2{}^{\downarrow K_1\cap K_1}(A) = 0 \implies m_1{}^{\downarrow K_1\cap K_2}(A) = 0$. Let p_1 and p_2 be probabilistic distributions for which $m_1(\{x\}) = p_1(x)$ and $m_2(\{y\}) = p_2(y)$ hold true for all $x \in \mathbf{X}_{K_1}$ and $y \in \mathbf{X}_{K_2}$. Then $m_1 \triangleright m_2$ is a Bayesian basic assignment and*

$$\forall z \in \mathbf{X}_{K_1\cup K_2} (m_1 \triangleright m_2)(\{z\}) = (p_1 \triangleright p_2)(z).$$

4 Generalization of Probabilistic Models

In this section we shall make a couple of suggestions enabling us to understand multidimensional models of basic assignments as a real enrichment of probabilistic models. First let us have a look how the concept of conditional independence was introduced in these two theoretical settings.

Consider three disjoint sets $I, J, K \subset N$ ($I \neq \emptyset \neq J$) and a probability distribution p on $\mathbf{X}_N$. We say that for distribution p groups of variables X_I and X_J are *conditionally independent given variables* X_K if for all $x \in \mathbf{X}_{I\cup J\cup K}$ the following equality holds true

$$p^{\downarrow I\cup J\cup K}(x) \cdot p^{\downarrow K}(x^{\downarrow K}) = p^{\downarrow I\cup K}(x^{\downarrow I\cup K}) \cdot p^{\downarrow J\cup K}(x^{\downarrow J\cup K}).$$

It is well known that this is equivalent to the fact $p^{\downarrow I\cup J\cup K} = p^{\downarrow I\cup K} \triangleright p^{\downarrow J\cup K}$.

How is it for basic assignments? Answering the question is not so easy because of the fact that this notion for belief functions was introduced in several different ways. Perhaps the most

†Notice that by Definitions 1 and 3 we have introduced two operators of composition, both of which are denoted by the same symbol $\triangleright$. We believe that it is obvious that for composition of probability distributions one has to apply the probabilistic version, i.e. Definition 3, whilst for composition of basic assignments one has to apply operator from Definition 1.

quent (and maybe also with the greatest number of supporters) is the one, which can be easily ined with the help of *commonality function.* Using notation of Studený [8], commonality ıction Com_m is defined for basic assignment m (assuming that m is defined on $\mathbf{X}_N$) for each $\subset \mathbf{X}_N$ by a simple formula

$$Com_m(A) = \sum_{B \supseteq A} m(B).$$

n Yaghlane *et al.* [1, 2] define the concept of conditional non-interactivity (as well as Shenoy ines his concept of conditional independence [7]) in the way that variables X_I and variables are *conditionally non-interactive* given variables X_K if and only if for all $A \subseteq \mathbf{X}_N$

$$Com_{m^{\downarrow I\cup J\cup K}}(A^{\downarrow I\cup J\cup K}) \cdot Com_{m^{\downarrow K}}(A^{\downarrow K}) = Com_{m^{\downarrow I\cup K}}(A^{\downarrow I\cup K}) \cdot Com_{m^{\downarrow J\cup K}}(A^{\downarrow J\cup K}).$$

this paper we shall denote this property by $X_I \perp\!\!\!\perp_{[m]} X_J | X_K$.

Unfortunately, for basic assignments it does not hold true that $X_I \perp\!\!\!\perp_{[m]} X_J | X_K$ if and only he basic marginal assignment $m^{\downarrow I\cup J\cup K}$ *factorizes* in the following sense

$$m^{\downarrow I\cup J\cup K} = m^{\downarrow I\cup K} \triangleright m^{\downarrow J\cup K}. \qquad (\star)$$

evertheless, there are still properties indicating a similarity of these two notions, which are nmarized in the following simple assertion.

oposition 5 *Consider a basic assignment m on $\mathbf{X}_N$ and three disjoint subsets $I, J, K \subset N$ $\neq \emptyset \neq J$). If $A \subseteq \mathbf{X}_{I\cup J\cup K}$ is a focal element of $m^{\downarrow I\cup J\cup K}$ and $A \neq A^{\downarrow I\cup K} \otimes A^{\downarrow J\cup K}$ then neither the following two expressions holds true:* *1. $X_I \perp\!\!\!\perp_{[m]} X_J | X_K$, and*
2. $m^{\downarrow I\cup J\cup K} = m^{\downarrow I\cup K} \triangleright m^{\downarrow J\cup K}$.

So, the first property connecting the concepts of conditional non-interactivity and factoriza- n for basic assignments is that any of them guarantees that the focal elements of the respective ;ic assignment can be expressed as an extension of its corresponding projections (Z-layered tangles in the language of Ben Yaghlane *et al.*).

Another connecting property says that these notions coincide for Bayesian basic assignments. mely, in [8] Studený claims that for Bayesian basic assignments the concept of conditional ı-interactivity coincides with the concept of conditional independence of the corresponding ·bability distribution. Due to Lemma 4 the same holds also for the concept of factorization he sense of equality ($\star$).

Let us now pinpoint the difference between the studied concepts. In [1] the authors admit t their concept of conditional non-interactivity (as showed by Studený) is *not consistent with rginalization*[9, 10]. This means that it may happen that there are two basic assignments m_1 l m_2 defined on $\mathbf{X}_{I\cup K}$ and $\mathbf{X}_{J\cup K}$, respectively (I, J, K disjoint, $I \neq \emptyset \neq J$), for which there s not exist a basic assignment m on $\mathbf{X}_{I\cup J\cup K}$, such that m_1 and m_2 would be its marginal gnments and simultaneously $X_I \perp\!\!\!\perp_{[m]} X_J | X_K$. For an example see [1]. Such a situation, vever, cannot happen for the concept of factorization.

Taking into account also the fact that, as we showed in [5], factorization in the sense of ality ($\star$) meets all the semigraphoid axioms, we are making the following suggestion.

›posal 6 *Introduce the concept of conditional independence relation for basic assignments ı the help of factorization in the sense of equality ($\star$).*

Probabilistic compositional models have, from the point of view of practical applications, a ıdvantage that a necessary composition need not be defined. It is true that it may happen in situations when one composes probability distributions which are not consistent. But it easily occur when one constructs a model from data from different sources or when a source ı missing data is considered. To avoid this problem we propose the following solution.

Proposal 7 *Apply the operator of composition designed for basic assignments (Definition 1) even when handling probability distributions and consider in some cases sets of probability distributions.*

Surprisingly enough, realization of this proposal need not increase computational complexity of the used algorithms. This statement is based on the fact that space complexity of these models is not higher that space complexity of the corresponding probabilistic models.

5 Conclusion

In the paper we have introduced an operator of composition for basic assignments, which enables us to construct multidimensional models from a sequences of low-dimensional assignments, from so called generating sequences. Moreover, we showed these models are true generalization of probabilistic models and therefore we propose to use them whenever classical probabilistic model, due to incoherence of low-dimensional probability distributions, does not exist. To increase consistency of probabilistic models and a wider class of models constructed from basic assignments, we proposed also to introduce a new concept of conditional independence for basic assignments: the concept corresponding to factorization in the sense of equality ($\star$).

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] E. Ben Yaghlane, Ph. Smets, and K. Mellouli, "Directed Evidential Networks with Conditional Belief functions," *Proc. of ECSQARU 2003*, LNAI 2711, Springer, pp. 291–305, 2003.

[3] 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.

[4] R. Jiroušek, J. Vejnarová and M. Daniel, "Compositional models of belief functions," *Proc. of the 5th Symposium on Imprecise Probabilities and Their Applications* (G. de Cooman, J. Vejnarová and M. Zaffalon, eds.), Charles University Press, Praha, pp. 243–252, 2007.

[5] R. Jiroušek "Conditional Independence and Factorization of Multidimensional Models," accepted for publications in: *Proc. of the 2008 IEEE World Congress on Computational Intelligence (WCCI 2008)*, HongKong.

[6] G. Shafer, *A mathematical theory of evidence.*, Princeton University Press, 1976.

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

[8] M. Studený, "Formal properties of conditional independence in different calculi of AI," *Proceedings of ECSQARU'93*, (K. Clarke, R. Kruse and S. Moral, eds.). Springer-Verlag, 1993, pp. 341–351.

[9] M. Studený, "On stochastic conditional independence: the problems of characterization and description," *Annals of Mathematics and Artificial Intelligence,* vol. 35, p. 323-341, 2002.

[10] M. Studený, *Probabilistic Conditional Independence Structures.* Springer, London, 2005.