Evidential Networks from a Different Perspective

Jiřina Vejnarová

Abstract. Bayesian networks are, at present, probably the most popular representative of so-called graphical Markov models. Naturally, several attempts to construct an analogy of Bayesian networks have also been made in other frameworks as e.g. in possibility theory, evidence theory or in more general frameworks of valuation-based systems and credal sets. We collect previously obtained results concerning conditioning, conditional independence and irrelevance allowing to define a new type of evidential networks, based on conditional basic assignments. These networks can be seen as a generalization of Bayesian networks, however, they are less powerful than e.g. so-called compositional models, as we demonstrate by a simple example.

Keywords: Conditional independence, conditioning, evidence theory, evidential networks, multidimensional models.

1 Introduction

Bayesian networks are, at present, probably the most popular representative of so-called graphical Markov models. Naturally, several attempts to construct an analogy of Bayesian networks have also been made in other frameworks as e.g. in possibility theory [5], evidence theory [4] or in the more general frameworks of valuation-based systems [11] and credal sets [7].

In this paper we bring an alternative to [4], which does not seem to us to be satisfactory, as graphical tools well-known from Bayesian networks are used in different sense. An attempt, using the technique of the operator of

Jiřina Vejnarová
Institute of Information Theory and Automation of the ASCR,
182 08 Prague, Czech Republic
e-mail: vejnar@utia.cas.cz

R. Kruse et al. (Eds.): Synergies of Soft Computing and Statistics, AISC 190, pp. 429–436.
springerlink.com

composition [9] was already presented in [13], but in that paper we concentrated ourselves only on structural properties of the network, the problem of definition of conditional basic assignments was not solved there. After solving this problem [15], in this paper we present a new concept of evidential networks, which can be seen as a generalization of Bayesian networks. However, simultaneously we show, that these evidential networks are less powerful than e.g. so-called compositional models.

The paper is organized as follows. After a brief summary of basic notions from evidence theory (Section 2), in Section 3 we recall recent concepts important for introduction of evidential networks, such as conditioning, conditional independence and irrelevance. In Section 4 we present a theorem allowing a direct generalization of Bayesian networks to evidential framework as well as a simple example demonstrating the potential weakness of these networks.

2 Basic Notions

In this section we will briefly recall basic concepts from evidence theory [10] concerning sets and set functions.

2.1 Set Projections and Joins

For an index set $N = \{1, 2, \ldots, n\}$ let $\{X_i\}_{i \in N}$ be a system of variables, each X_i having its values in a finite set $\mathbf{X}_i$. In this paper we will deal with *multidimensional frame of discernment* $\mathbf{X}_N = \mathbf{X}_1 \times \mathbf{X}_2 \times \ldots \times \mathbf{X}_n$, and its *subframes* (for $K \subseteq N$) $\mathbf{X}_K = \times_{i \in K} \mathbf{X}_i$. When dealing with groups of variables on these subframes, X_K will denote a group of variables $\{X_i\}_{i \in K}$ throughout the paper.

For $M \subset K \subseteq N$ and $A \subset \mathbf{X}_K$, $A^{\downarrow M}$ will denote a *projection* of A into $\mathbf{X}_M$:

$$A^{\downarrow M} = \{y \in \mathbf{X}_M \mid \exists x \in A : y = x^{\downarrow M}\},$$

where, for $M = \{i_1, i_2, \ldots, i_m\}$,

$$x^{\downarrow M} = (x_{i_1}, x_{i_2}, \ldots, x_{i_m}) \in \mathbf{X}_M.$$

In addition to the projection, in this text we will also need an opposite operation, which will be called a join. By a *join*[1] of two sets $A \subseteq \mathbf{X}_K$ and $B \subseteq \mathbf{X}_L$ $(K, L \subseteq N)$ we will understand a set

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

Let us note that for any $C \subseteq \mathbf{X}_{K \cup L}$ naturally $C \subseteq C^{\downarrow K} \bowtie C^{\downarrow L}$, but generally $C \neq C^{\downarrow K} \bowtie C^{\downarrow L}$.

[1] This term and notation are taken from the theory of relational databases [1].

2.2 Set Functions

In evidence theory [10] (or Dempster-Shafer theory) two dual measures are used to model the uncertainty: belief and plausibility measures. Both of them can be defined with the help of another set function called a *basic (probability or belief) assignment* m on $\mathbf{X}_N$, i.e.,

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

where $\mathcal{P}(\mathbf{X}_N)$ is the power set of $\mathbf{X}_N$, and $\sum_{A\subseteq \mathbf{X}_N} m(A) = 1$. Furthermore, we assume that $m(\emptyset) = 0$. A set $A \in \mathcal{P}(\mathbf{X}_N)$ is a *focal element* if $m(A) > 0$.

Belief and *plausibility measures* are defined for any $A \subseteq \mathbf{X}_N$ by the equalities

$$Bel(A) = \sum_{B\subseteq A} m(B), \qquad Pl(A) = \sum_{B\cap A\neq\emptyset} m(B),$$

respectively. It is well-known (and evident from these formulae) that for any $A \in \mathcal{P}(\mathbf{X}_N)$

$$Bel(A) \leq Pl(A), \qquad Pl(A) = 1 - Bel(A^C), \tag{1}$$

where A^C is the set complement of $A \in \mathcal{P}(\mathbf{X}_N)$. Furthermore, basic assignment can be computed from belief function via Möbius inverse:

$$m(A) = \sum_{B\subseteq A} (-1)^{|A\setminus B|} Bel(B), \tag{2}$$

i.e. any of these three functions is sufficient to define values of the remaining two.

For a basic assignment m on $\mathbf{X}_K$ and $M \subset K$, a *marginal basic assignment* of m on $\mathbf{X}_M$ is defined (for each $A \subseteq \mathbf{X}_M$):

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

3 Conditioning, Independence and Irrelevance

Conditioning and independence belong to the most important topics of any theory dealing with uncertainty. They are cornerstones of Bayesian-like multidimensional models.

3.1 Conditioning

In evidence theory the "classical" conditioning rule is so-called Dempster's rule of conditioning, nevertheless a lot of alternative conditioning rules for events have been proposed [8].

However, from the viewpoint of evidential networks conditioning of variables is of primary interest. In [14] we presented two definitions of conditioning by variables, based on Dempster conditioning rule and focusing, we proved that these definitions are correct, nevertheless, their usefulness for multidimensional models is rather questionable, as thoroughly discussed in the above-mentioned paper.

Therefore, in [15] we proposed a new conditioning rule defined as follows.

Definition 1. Let X_K and X_L ($K \cap L = \emptyset$) be two groups of variables with values in $\mathbf{X}_K$ and $\mathbf{X}_L$, respectively. Then the *conditional basic assignment* of X_K given $X_L \in B \subseteq \mathbf{X}_L$ (for B such that $m^{\downarrow L}(B) > 0$) is defined as follows:

$$m_{X_K|_P X_L}(A|_P B) = \frac{\displaystyle\sum_{\substack{C \subseteq \mathbf{X}_{K \cup L}: \\ C^{\downarrow K} = A \& C^{\downarrow L} = B}} m(C)}{m^{\downarrow L}(B)} \tag{3}$$

for any $A \subseteq \mathbf{X}_K$.

It is evident that the conditioning is defined only for focal elements of the marginal basic assignment, but we do not consider it a substantial disadvantage, because all the information about a basic assignment is concentrated in focal elements. Its correctness is expressed by Theorem 1, proven in [15].

Theorem 1. *Set function $m_{X_K|_P X_L}$ defined for any fixed $B \subseteq \mathbf{X}_L$, such that $m^{\downarrow L}(B) > 0$ by Definition 1 is a basic assignment on $\mathbf{X}_K$.*

3.2 Independence and Irrelevance

In evidence theory the most common notion of independence is that of random set independence [6]. It has already been proven [12] that it is also the only sensible one.

This notion can be generalized in various ways [3, 11, 12]; the concept of conditional non-interactivity from [3], based on conjunctive combination rule, is used for construction of directed evidential networks in [4]. In this paper we will use the concept introduced in [9, 12], as we consider it more suitable (the arguments can be found in [12]).

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 the equality

$$m^{\downarrow K \cup L \cup M}(A) \cdot m^{\downarrow M}(A^{\downarrow M}) = m^{\downarrow K \cup M}(A^{\downarrow K \cup M}) \cdot m^{\downarrow L \cup M}(A^{\downarrow L \cup M}) \quad (4)$$

holds for any $A \subseteq \mathbf{X}_{K \cup L \cup M}$ such that $A = A^{\downarrow K \cup M} \bowtie A^{\downarrow L \cup M}$, and $m(A) = 0$ otherwise.

It has been proven in [12] that this conditional independence concept satisfies so-called semi-graphoid properties taken as reasonable to be valid for any conditional independence concept and it has been shown in which sense this conditional independence concept is superior to previously introduced ones [3, 11].

Irrelevance is usually considered to be a weaker notion than independence [6]. It expresses the fact that a new piece of evidence concerning one variable cannot influence the evidence concerning the other variable.

More formally: group of variables X_L is *irrelevant* to X_K ($K \cap L = \emptyset$) if for any $B \subseteq \mathbf{X}_L$ such that $Pl^{\downarrow L}(B) > 0$ (or $Bel^{\downarrow L}(B) > 0$ or $m^{\downarrow L}(B) > 0$)

$$m_{X_K|X_L}(A|B) = m^{\downarrow K}(A) \quad (5)$$

for any $A \subseteq \mathbf{X}_K$.[2]

Generalization of this notion to conditional irrelevance may be done as follows. Group of variables X_L is *conditionally irrelevant* to X_K given X_M (K, L, M disjoint, $K \neq \emptyset \neq L$) if

$$m_{X_K|X_{L \cup M}}(A|B) = m_{X_K|X_M}(A|B^{\downarrow M}) \quad (6)$$

is satisfied for any $A \subseteq \mathbf{X}_K$ and $B \subseteq \mathbf{X}_{L \cup M}$ (whenever both sides are defined).

Let us note that the conditioning in equalities (5) and (6) stands for an abstract conditioning rule [8]. However, the validity of (5) and (6) may depend on the choice of conditioning rule, as we showed in [14] — more precisely irrelevance with respect to one conditioning rule need not imply irrelevance with respect to the other. Nevertheless, when studying the relationship between (conditional) independence and irrelevance based on Dempster conditioning rule and focusing we realized that they do not differ too much from each other [14].

However, the new conditioning rule introduced by Definition 1 exhibits much suitable properties as expressed by the following theorem proven in [15].

---

[2] Let us note that somewhat weaker definition of irrelevance can be found in [2], where equality is substituted by proportionality. This notion has been later generalized using conjunctive combination rule [3].

Theorem 2. *Let K, L, M be disjoint subsets of N such that $K, L \neq \emptyset$. If X_K and X_L are independent given X_M (with respect to a joint basic assignment m defined on $\mathbf{X}_{K\cup L\cup M}$), then X_L is irrelevant to X_K given X_M under the conditioning rule given by Definition 1.*

4 Evidential Networks

However, in Bayesian networks also the reverse implication plays an important role, as for the inference, the network is usually transformed into a decomposable model. Unfortunately, in the framework of evidence theory the reverse implication is not valid, in general, as was shown in [15]. Nevertheless, the following assertion holds true.

Theorem 3. *Let K, L, M be disjoint subsets of N such that $K, L \neq \emptyset$ and $m_{X_K|_P X_{L\cup M}}$ be a (given) conditional basic assignment of X_K given $X_{L\cup M}$ and $m_{X_{L\cup M}}$ be a basic assignment of $X_{L\cup M}$. If X_L is irrelevant to X_K given X_M under the conditioning rule given by Definition 1, then X_K and X_L are independent given X_M (with respect to a joint basic assignment $m = m_{X_K|_P X_{L\cup M}} \cdot m_{X_{L\cup M}}$*[3] *defined on $\mathbf{X}_{K\cup L\cup M}$).*

Proof. Irrelevance of X_L to X_K given X_M means that for any $A \subseteq \mathbf{X}_K$ and any $B \subseteq \mathbf{X}_{L\cup M}$ such that $m^{\downarrow L\cup M}(B) > 0$

$$m_{X_K|_P X_{L\cup M}}(A|B) = m_{X_K|_P X_M}(A|B^{\downarrow M}).$$

Multiplying both sides of this equality by $m^{\downarrow L\cup M}(B) \cdot m^{\downarrow M}(B^{\downarrow M})$ one obtains

$$\begin{aligned} m_{X_K|_P X_{L\cup M}}(A|B) \cdot m^{\downarrow L\cup M}(B) \cdot m^{\downarrow M}(B^{\downarrow M}) \\ = m_{X_K|_P X_M}(A|B^{\downarrow M}) \cdot m^{\downarrow L\cup M}(B) \cdot m^{\downarrow M}(B^{\downarrow M}), \end{aligned}$$

which is equivalent to

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

Therefore, the equality (4) is satisfied for $C \subseteq \mathbf{X}_{K\cup L\cup M}$ such that $C = A \times B$, where $A \subseteq \mathbf{X}_K$ and $B \subseteq \mathbf{X}_{L\cup M}$. Due to Theorem 1 it is evident, that

$$\sum_{A\subseteq \mathbf{X}_K, B\subseteq \mathbf{X}_{L\cup M}} m(A \times B) = 1,$$

and therefore equality (4) is trivially for satisfied also for any other $C = C^{\downarrow K\cup M} \bowtie C^{\downarrow L\cup M}$, and $m(C) = 0$ otherwise as well. Therefore, X_K and X_L

[3] Let us note that due to Theorem 1 $m_{X_{L\cup M}}$ is marginal to m and $m_{X_K|_P X_{L\cup M}}$ can be re-obtained from m via Definition 1.

are independent given X_M with respect to a joint basic assignment $m = m_{X_K|X_M} \cdot m_{X_{L\cup M}}$. □

This theorem makes possible to define evidential networks in a way analogous to Bayesian networks, but simultaneously brings a question: are these networks advantageous in comparison with other multidimensional models in this framework? The following example brings, at least partial, answer to this question.

Example 1. Let X_1, X_2 and X_3 be three binary variables with values in $\mathbf{X}_i = \{a_i, \bar{a}_i\}, i = 1, 2, 3$, and m be a basic assignment on $\mathbf{X}_1 \times \mathbf{X}_2 \times \mathbf{X}_3$ defined as follows

$$m(\mathbf{X}_1 \times \mathbf{X}_2 \times \{\bar{a}_3\}) = .5,$$
$$m(\{(a_1, a_2, \bar{a}_3), (\bar{a}_1, \bar{a}_2, a_3)\}) = .5.$$

Variables X_1 and X_2 are conditionally independent given X_3 with respect to m. Therefore also X_2 is irrelevant to X_1 given X_3, i.e.

$$m_{X_1|X_{23}}(A|B) = m_{X_1|X_3}(A|B^{\downarrow\{3\}}),$$

for any focal element B of $m^{\downarrow\{23\}}$. As both $m^{\downarrow\{23\}}$ and $m^{\downarrow\{3\}}$ have only two focal elements, namely $\mathbf{X}_2 \times \{\bar{a}_3\}$ and $\{(a_2, \bar{a}_3), (\bar{a}_2, a_3)\}$ and $\{\bar{a}_3\}$ and $\mathbf{X}_3$, respectively, we have

$$m_{X_1|_P X_{23}}(\mathbf{X}_1|\mathbf{X}_2 \times \{\bar{a}_3\}) = m_{X_1|_P X_3}(\mathbf{X}_1|\{\bar{a}_3\}) = 1,$$
$$m_{X_1|_P X_{23}}(\mathbf{X}_1|\{(a_2, \bar{a}_3), (\bar{a}_2, a_3)\}) = m_{X_1|_P X_3}(\mathbf{X}_1|\mathbf{X}_3) = 1.$$

Using these conditionals and the marginal basic assignment $m^{\downarrow\{23\}}$ we get a basic assignment $\tilde{m}$ different from the original one, namely

$$\tilde{m}(\mathbf{X}_1 \times \mathbf{X}_2 \times \{\bar{a}_3\}) = .5,$$
$$\tilde{m}(\mathbf{X}_1 \times \{(a_2, \bar{a}_3), (\bar{a}_2, a_3)\}) = .5.$$

Furthermore, if we interchange X_1 and X_2 we get yet another model, namely

$$\hat{m}(\mathbf{X}_1 \times \mathbf{X}_2 \times \{\bar{a}_3\}) = .5,$$
$$\hat{m}(\mathbf{X}_2 \times \{(a_1, \bar{a}_3), (\bar{a}_1, a_3)\}) = .5.$$ ◇

From this example it is evident, that evidential networks are less powerful than e.g. compositional models [9], as any of these threedimensional basic assignments can be obtained from its marginals using the operator of composition (cf. e.g. [9]).

5 Conclusions

We presented a conditioning rule for variables which is compatible with our notion of conditional independence — in other words, if we use this conditioning rule, we obtain conditional irrelevance concept, which is implied by this conditional independence. We also proved a theorem showing that under some specific conditions conditional irrelevance implies conditional independence. However, by a simple example we revealed the weakness of conditional basic assignments in comparison with the joint ones and therefore also the fact that evidential networks are less powerful in comparison with e.g. compositional models.

Acknowledgements. The support of Grant GAČR P402/11/0378 is gratefully acknowledged.

References

1. Beeri, C., Fagin, R., Maier, D., Yannakakis, M.: On the desirability of acyclic database schemes. J. Assoc. Comput. Mach. 30, 479–513 (1983)
2. Ben Yaghlane, B., Smets, P., Mellouli, K.: Belief functions independence: I. the marginal case. Int. J. Approx. Reason. 29, 47–70 (2002)
3. Ben Yaghlane, B., Smets, P., Mellouli, K.: Belief functions independence: II. the conditional case. Int. J. Approx. Reason. 31, 31–75 (2002)
4. Yaghlane, B., Smets, P., Mellouli, K.: Directed Evidential Networks with Conditional Belief Functions. In: Nielsen, T.D., Zhang, N.L. (eds.) ECSQARU 2003. LNCS (LNAI), vol. 2711, pp. 291–305. Springer, Heidelberg (2003)
5. Benferhat, S., Dubois, D., Gracia, L., Prade, H.: Directed possibilistic graphs and possibilistic logic. In: Proc. of the 7th Int. Conf. IPMU 1998, pp. 1470–1477 (1998)
6. Couso, I., Moral, S., Walley, P.: Examples of independence for imprecise probabilities. In: Proc. of ISIPTA 1999, pp. 121–130 (1999)
7. Cozman, F.G.: Credal networks. Artif. Intel. 120, 199–233 (2000)
8. Daniel, M.: Belief conditioning rules for classic belief functions. In: Proc. of WUPES 2009, pp. 46–56 (2009)
9. Jiroušek, R., Vejnarová, J.: Compositional models and conditional independence in evidence theory. Int. J. Approx. Reason. 52, 316–334 (2011)
10. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, Princeton (1976)
11. Shenoy, P.P.: Conditional independence in valuation-based systems. Int. J. Approx. Reason. 10, 203–234 (1994)
12. Vejnarová, J.: On conditional independence in evidence theory. In: Proc. of ISIPTA 2009, Durham, UK, pp. 431-440 (2009)
13. Vejnarová, J.: An Alternative Approach to Evidential Network Construction. In: Borgelt, C., González-Rodríguez, G., Trutschnig, W., Lubiano, M.A., Gil, M.Á., Grzegorzewski, P., Hryniewicz, O. (eds.) Combining Soft Computing and Statistical Methods in Data Analysis. AISC, vol. 77, pp. 619–626. Springer, Heidelberg (2010)
14. Vejnarová, J.: Conditioning, conditional independence and irrelevance in evidence theory. In: Proc. of ISIPTA 2011, Innsbruck, Austria, pp. 381–390 (2011)
15. Vejnarová, J.: Conditioning in Evidence Theory from the Perspective of Multidimensional Models. In: Greco, S., Bouchon-Meunier, B., Coletti, G., Fedrizzi, M., Matarazzo, B., Yager, R.R. (eds.) IPMU 2012, Part III. CCIS, vol. 299, pp. 450–459. Springer, Heidelberg (2012)