

Learning Structure in Evidential Networks from Evidential DataBases

Narjes Ben Hariz^(✉) and Boutheina Ben Yaghlane

LARODEC Laboratory, Institut Supérieur de Gestion de Tunis,
Le Bardo, Tunisia
narjes.benhariz@gmail.com

Abstract. Evidential networks have gained a growing interest as a good tool fusing belief function theory and graph theory to analyze complex systems with uncertain data. The graphical structure of these models is not always clear, it can be fixed by experts or constructed from existing data. The main issue of this paper is how to extract the graphical structure of an evidential network from imperfect data stored in evidential databases.

1 Introduction

Data in real-world problems are generally characterized by different forms of imperfection: imprecision, uncertainty and/or inconsistency. Many theories have been proposed to deal with this problem of imperfection, one of the most popular is the belief function theory. It is a general framework that handles both partial and total ignorance and offers interesting rules for combining evidence.

Based on this theory, evidential networks are considered as a powerful and flexible tool for modeling complex systems by combining belief function theory and graph theory. Among the most popular evidential graphical models are the evidential networks with conditional belief functions proposed by Xu et al. [23] and the directed evidential networks with conditional belief functions proposed by Ben Yaghlane et al. [4].

As in Bayesian networks, evidential networks are based on two parts: the qualitative part represented by a directed acyclic graph and the quantitative part including a set of parameters modeled by conditional belief functions. The graphical structure of these networks is not always clear, specially in real complex systems. Therefore, a good way for constructing this structure is to estimate it from data.

We address in this paper the issue of learning structure in evidential networks from evidential databases, by extending the classical methods widely used for learning structure in Bayesian networks to the belief functions framework. More precisely, we are interested in generalizing learning methods based on tests of independency, including for example the algorithms proposed by Pearl and Verma [22] and the algorithms proposed by Spirtes et al. [20]. Our learning process is based on evidential chi-square test $E\chi_2$, a generalization of the statistical chi-square test in the belief functions framework.

The rest of the paper is organized as follows: we first recall some basic concepts of belief function theory and evidential databases, then we present briefly evidential networks and a short review of the most important algorithms used for learning structure in Bayesian networks. In the main part of the paper, we present our learning process and the evidential chi-square independency test. In the last part of the paper we more explain our approach by an illustrative example.

2 Belief Function Theory and Evidential DataBases

The belief function theory, evidence theory or also Dempster-Shafer theory [17,19], is a mathematical framework commonly used for handling imperfection in data. In the following, we present briefly some theoretical aspects of evidence theory and we introduce databases based on this theory.

2.1 Basic Concepts of Belief Function Theory

Let $N = \{N_1, \dots, N_n\}$ be a set of random variables.

Definition 1. Each variable N_i takes its values from a set of exclusive and exhaustive elements called the frame of discernment and denoted by Ω_{N_i} .

Definition 2. We denote by $2^{\Omega_{N_i}}$ the set of all subsets (propositions or events) from Ω_{N_i} .

Definition 3. The degree of belief accorded exactly to a proposition A , is called the basic belief assignment or a mass function (bba). It is a mapping from $2^{\Omega_{N_i}}$ to $[0, 1]$ such that:

$$\sum_{A \subseteq \Omega} m^{\Omega}(A) = 1 \tag{1}$$

Definition 4. Any event $A \in \Omega_{N_i}$ with $m^{\Omega_{N_i}}(A) > 0$ is called a focal element, and the set of all these elements is denoted by $F(m^{\Omega_{N_i}})$.

Definition 5. Let $m^{\Omega_{N_i}}[B](A)$ denote the conditional basic belief assignment of A given B , it is defined by Dempster's rule of conditioning:

$$m^{\Omega_{N_i}}[B](A) = \sum_{C \subseteq \bar{B}} m^{\Omega_{N_i}}(A \cup C), \tag{2}$$

where \bar{B} is the complement of the proposition B .

More details about the rules of conditioning in the belief function theory can be found in [18,21].

2.2 Evidential DataBases

An Evidential DataBase (EDB) is a database storing data with different forms of imperfection: certain, probabilistic, possibilistic, missing and/or evidential data, modeled in the belief functions framework. More details and examples about this notion can be found in [1,8].

Definition 6. Let $EDB(L,C)$ denote an evidential database with L lines (records) and C columns (variables), each variable N_i define its possible values in a frame of discernment denoted by Ω_{N_i} .

Definition 7. Let V_{lc} be the evidential value of cell in the l^{th} line and c^{th} column, V_{lc} is defined by a mass function m_{lc} from $2^{\Omega_{N_i}}$ to $[0,1]$ such as:

$$m_{lc}^{\Omega_{N_i}}(\emptyset) = 0 \text{ and } \sum_{A \subseteq \Omega_{N_i}} m_{lc}^{\Omega_{N_i}}(A) = 1 \quad (3)$$

3 Evidential Networks

Evidential networks or belief function networks are graphical models based on the belief functions framework for modeling uncertainty. These models are considered as a generalization of Bayesian networks for handling different types of uncertainty and taking into account both total and partial ignorance.

Among the most popular evidential graphical models Directed EVIDential Networks with conditional belief functions (DEVNs) proposed by Ben Yaghlane et al. [4] and Evidential Networks with Conditional belief functions (ENCs) proposed by Xu et al. [23]. DEVNs are developed to extend ENCs for handling n-ary relations between variables.

As in probabilistic networks, evidential graphical models are based essentially on two parts: the qualitative part describing the graphical structure of the network and the quantitative part describing the conditional dependencies between variables.

3.1 Qualitative Part

ENCs and DEVNs have the same graphical structure which is similar to the graphical structure of Bayesian networks (BNs). This structure is modeled by a Directed Acyclic Graph (DAG) $G = (N, E)$ characterized essentially by a set of nodes $N = \{N_1, \dots, N_x\}$ representing the different variables of the problem and a set of edges $E = \{E_1, \dots, E_y\}$ coding conditional dependencies between variables.

It is important to note that the graphical properties and concepts of a DAG are maintained in evidential networks including conditional independency criterions such as the d-separation, the converging connection (also called v-structure) and the CPDAG (Completed Partially Directed Acyclic Graph). More details about these notions can be found in [3,14].

3.2 Quantitative Part

The quantitative level is represented by a set of parameters θ modeled by conditional belief functions. Each node N_i in an evidential network is a representation of a random variable taking its values on a frame of discernment Ω_{N_i} . Each root node is associated with an a priori mass function, other nodes are associated with a conditional mass function. DEVNs are more flexible than ENCnS in modeling conditional beliefs. In DEVNs the conditional mass function can be defined in two manners: per edge or per child node.

In this paper we are interested in the qualitative part of evidential networks. As both ENCnS and DEVNs have the same qualitative part we adopt for the rest of the paper a general notation for evidential networks (ENs).

4 Learning Structure in Bayesian Networks

Learning the graphical structure of Bayesian networks from data remains an interesting topic of research. In this section we present a short overview of the literature on learning Bayesian network structure, more details can be found in [10, 12, 13].

The structure learning methods in BNs are grouped on three essential families:

Constrained based methods. These methods are based on the test of the conditional dependencies between variables in order to build the requested graph.

Score based methods. The main idea of these methods is to find the best structure from the search space of possible DAGs by maximizing a scoring function.

Hybrid methods. Combine constrained based methods and score based methods in order to deal with more complex problems.

In this work we will mainly interest on the methods based on testing the conditional dependencies between variables. The majority of these methods follow the same approach which is based on three steps:

- ① Build an undirected graph according to a statistical test.
- ② Detect the V-structures.
- ③ Get a CPDAG by propagating some edges orientation.

This family of methods includes two principal groups of algorithms:

- The algorithms proposed by Pearl and Verma [15, 22] such as IC and IC*. The main idea of these algorithms is to start from an empty graph and try to add edges between dependent variables according to the result of the statistical test.
- The algorithms proposed by Spirtes, Glymour and Scheines [20] such as SGS and PC. These algorithms are based also on a statistical test to delete edges between independent variables from a complete graph.

One of the most statistical tests commonly used for measuring the conditional independency between variables in learning structure algorithms is the chi-square (χ^2) test. This metric will be generalized in the next section for dealing with uncertain data.

5 Learning Structure in Evidential Networks

In this part, we focus on the main purpose of this paper which is how to build the graphical part of an evidential network from data stored in evidential databases.

As we mentioned previously, we are interested in this work in generalizing the constrained based methods, the extension of the other methods of structure learning will be the object of further works.

5.1 Evidential Independency Test

The idea of extending the statistical tests of independency (essentially the χ^2 test) to the belief function theory comes from the principle of the generalization of the maximum likelihood estimation to the evidence framework originally introduced in [7] and applied to learn parameters in evidential networks in [2].

As in probability theory these tests are based on a relation between the observed data and the expected one.

Let us consider X and Y two variables, and Z a set of variables from our $EDB(L, C)$.

Definition 8. *The evidential observed values corresponding respectively to X and Y and to X and Y given Z , denoted by EO_{xy} and $EO_{xy|z}$, are defined by the following equations:*

$$EO_{xy} = \sum_{l=1}^L m_{lc}^{\Omega_X}(X = x) * m_{lc}^{\Omega_Y}(Y = y) \tag{4}$$

$$EO_{xy|z} = \sum_{l=1}^L m_{lc}^{\Omega_X}(X = x) * m_{lc}^{\Omega_Y}(Y = y) * \prod_j m_{lcj}^{\Omega_{Z_j}}(Z_j = z_j) \tag{5}$$

Definition 9. *Let EE_{xy} and $EE_{xy|z}$ denote the evidential expected values such that:*

$$EE_{xy} = \frac{\sum_{l=1}^L m_{lc}(X = x) * \sum_{l=1}^L m_{lc}(Y = y)}{L} \tag{6}$$

$$EE_{xy|z} = \frac{\sum_{l=1}^L m_{lc}(X = x) * \prod_j m_{lcj}^{\Omega_{Z_j}}(Z_j = z_j) * \sum_{l=1}^L m_{lc}(Y = y) * \prod_j m_{lcj}^{\Omega_{Z_j}}(Z_j = z_j)}{\sum_{l=1}^L \prod_j m_{lcj}^{\Omega_{Z_j}}(Z_j = z_j)} \tag{7}$$

Definition 10. *The evidential test of independency between the two variables X and Y is defined as follows:*

$$E\chi^2_{XY} = \sum_{x=1}^{2^{\Omega_X}-1} \sum_{y=1}^{2^{\Omega_Y}-1} \frac{(EO_{xy} - EE_{xy})^2}{EE_{xy}} \tag{8}$$

with a degree of freedom $df = ((2^{\Omega_X} - 1) - 1)((2^{\Omega_Y} - 1) - 1)$.

The two variables X and Y are considered independent if the value of $E\chi^2_{XY}$ is less than the critical value in the chi-squared distribution.

Definition 11. *The conditional evidential test of independency between the two variables X and Y in the context of Z is defined as follows:*

$$E\chi^2_{XY|Z} = \sum_{x=1}^{2^{\Omega_X}-1} \sum_{y=1}^{2^{\Omega_Y}-1} \sum_{z=1}^{2^{\Omega_Z}-1} \frac{(EO_{xy|z} - EE_{xy|z})^2}{EE_{xy|z}} \tag{9}$$

with a degree of freedom $df = ((2^{\Omega_X} - 1) - 1)((2^{\Omega_Y} - 1) - 1) \prod_j (2^{\Omega_{z_j}} - 1)$.

The two variables X and Y are said independent in the context of Z if the value of $E\chi^2_{XY|Z}$ is less than the critical value in the chi-squared distribution.

The principle of the extension of the χ^2 test can be also used for generalizing other statistical tests such as the likelihood ratio (G^2) test or even the score functions based on the likelihood principle, (in order to generalize score based and hybrid algorithms).

It must be emphasized that, the $E\chi^2$ test has the same major limitation of the classical χ^2 test. This test become inappropriate when the number of variables is high and the amount of data is not sufficient. In probability theory Spirates et al. propose an heuristic to overcome this drawback: if the degree of freedom is higher than $\frac{L}{10}$ the two variables are considered dependent.

In the evidence theory this problem is even more serious, as we consider in the calculation process the power set of each variable. Thus we propose, in this case, to calculate the degree of freedom according to the focal elements of each variable as follows: $df' = (|F(m^{\Omega_X})| - 1)(|F(m^{\Omega_Y})| - 1)$.

If the problem persists, then the hypothesis of Spirates et al. can be applied.

5.2 Learning Approach

The generalization of the χ^2 test in the belief function framework will be the core of our learning process. In fact, our goal is to estimate the different dependency relations between variables from an $EDB(L, C)$ using the $E\chi^2$ test, in order to get the $DAG(N, E)$ that most closely matches the data. This approach is based on two main steps:

- ① Test the different independency relations between variables:
 - Calculate the evidential observed values using Eqs. (4) and (5).
 - Calculate the evidential expected values using Eqs. (6) and (7).
 - Test the dependency between variables according to the $E\chi^2$ test measured using Eqs. (8) and (9).

- ② Apply a learning structure algorithm based on independency tests:
 - Build an undirected graph according to $E\chi^2$: delete edges between independent variables if we started with a complete graph or add edges between dependent variables if we started with an empty graph.
 - Detect the V-structures according to the calculated $E\chi^2$ test.
 - Propagate edges to get a CPDAG representing the sought DAG.

Note that, a possible simplification in the first step of our approach is to consider only focal elements for each variable, in order to reduce the number of values that must be calculated in the $E\chi^2$ test. In this case we consider df' as a degree of freedom.

This approach generalizes any constrained based method classically used for estimating the graphical structure of a Bayesian network.

It will be also interesting to compare our learning approach using the $E\chi^2$ test and other independency tests dedicated to uncertain data such as the independency tests proposed in [16].

6 Illustrative Example

To further explain the details of our learning method, we introduce in this section an illustrative example based on the classical problem Asia Chest Clinic first described in [11]. Table 1 presents a part from an EDB corresponding to the latter problem.

Our problem includes eight variables $\{A, S, T, L, B, O, X, D\}$ having the power sets, respectively: $\{a, \bar{a}, a \cup \bar{a}\}$; $\{s, \bar{s}, s \cup \bar{s}\}$; $\{t, \bar{t}, t \cup \bar{t}\}$; $\{l, \bar{l}, l \cup \bar{l}\}$; $\{b, \bar{b}, b \cup \bar{b}\}$; $\{o, \bar{o}, o \cup \bar{o}\}$; $\{x, \bar{x}, x \cup \bar{x}\}$ and $\{d, \bar{d}, d \cup \bar{d}\}$.

Note that in the EDB a is denoted by 0, \bar{a} is denoted by 1 and $a \cup \bar{a}$ is denoted by $\{0, 1\}$. All other propositions are denoted by the same way.

The first step of our approach is to test dependency between variables. In the following we give some calculation details of the evidential chi-square test applied to the variables T and O using 20 instances from our EDB (represented in Table 1).

- $EO_{to} = \sum_{l=1}^{20} m_{lc}^{\Omega_T}(T=t) * m_{lc}^{\Omega_O}(O=o) = 1 * 1 + 1 * 1 + 1 * 0.3 + 1 * 1 + 1 * 0.4 + 1 * 1 = 4.7$
- By the same manner we get:
 $EO_{t\bar{o}} = 5.3$ $EO_{to\bar{o}} = 1$ $EO_{\bar{t}o} = 4.99$ $EO_{\bar{t}\bar{o}} = 2.41$ $EO_{\bar{t}o\bar{o}} = 1.6$ $EO_{t\bar{t}o} = 0$
 $EO_{t\bar{t}\bar{o}} = 0$ $EO_{t\bar{t}o\bar{o}} = 0$
- $EE_{to} = \frac{\sum_{l=1}^{20} m_{lc}(T=t) * \sum_{l=1}^{20} m_{lc}(O=o)}{20} = \frac{11 * (0.21 + 1 + 1 + 1 + 0.3 + 1 + 1 + 0.4 + 0.2 + 1)}{20} = 3.91$
- Applying the same formula:
 $EE_{t\bar{o}} = 5.65$ $EE_{to\bar{o}} = 1.43$ $EE_{\bar{t}o} = 3.19$ $EE_{\bar{t}\bar{o}} = 2.63$ $EE_{\bar{t}o\bar{o}} = 1.17$ $EE_{t\bar{t}o} = 0$
 $EE_{t\bar{t}\bar{o}} = 0$ $EE_{t\bar{t}o\bar{o}} = 0$
- $E\chi_{TO}^2 = \left[\frac{(EO_{to} - EE_{to})^2}{EE_{to}} \right] + \dots + \left[\frac{(EO_{t\bar{t}o\bar{o}} - EE_{t\bar{t}o\bar{o}})^2}{EE_{t\bar{t}o\bar{o}}} \right] = 2.52$
- $df = ((2^2 - 1) - 1) * ((2^2 - 1) - 1) = 4$

Table 1. EDB corresponding to the Asia Chest Clinic problem

A	S	T	L	B	O	X	D
0	0	1	0	0	0(0.21)1(0.79)	0	1
0	0	0	1	0	1	1	1
0(0.5)1(0.5)	0	0	{0,1}	0	0	1	0
1	0	1	0	0	1	1	0
0	0	1	0	0	1	1	1
0	0	1	0	0	0	0	0
{0,1}	0	0	1	0	0	1	1
0	0	0	0	0	0(0.3)1(0.7)	0	0
0(0.22){0,1}(0.78)	0	1	0	0	{0,1}	1	1
0	0	1	0	0	1(0.9){0,1}(0.1)	1	1
0(0.2){0,1}(0.8)	0	1	{0,1}	1	0	1	1
0	0	0	1	0	0	0	1
0	0	0	1	0	1	0	1
0(0.1)1(0.2){0,1}(0.7)	0	0	0	0	{0,1}	1	1
0	0	0	1	0	0(0.4)1(0.6)	0	0
1	0	1	0	0	0(0.2)1(0.3){0,1}(0.5)	0	0
1(0.45){0,1}(0.55)	0	0	0	0	0	1	1
1	0	1	1	0	1	1	1
0	0	0	0	0	1	0	1
0(0.36){0,1}(0.64)	0	0	0	0	1	1	1

Note that, when dealing with perfect data such as the case of variables S, T, B, X and D , the $E\chi^2$ give the same result as the classical χ^2 test.

Assuming that $\alpha = 5\%$, the critical value according to the chi-squared distribution is equal to 9.488 which is higher then the value of the calculated test $E\chi^2_{TO}$. According to this result, the two variables T and O are independent. However, we should note that the obtained value of $E\chi^2_{TO}$ is not significant, because the amount of data considered in this example is very small.

After finishing the different calculation steps and applying the PC algorithm, the result of the evidential learning process from the whole data set is presented in Fig. 1.

Phase 1, phase 2 and phase 3 represent the different iterations of the first step of the PC algorithm which is building an undirected graph by eliminating edges between independent variables from the complete graph, corresponding to the Asia network.

The next step in the learning process is to detect the different v-structures in the obtained undirected graph using the calculated $E\chi^2$ test in order to build a CPDAG modeling the required DAG.

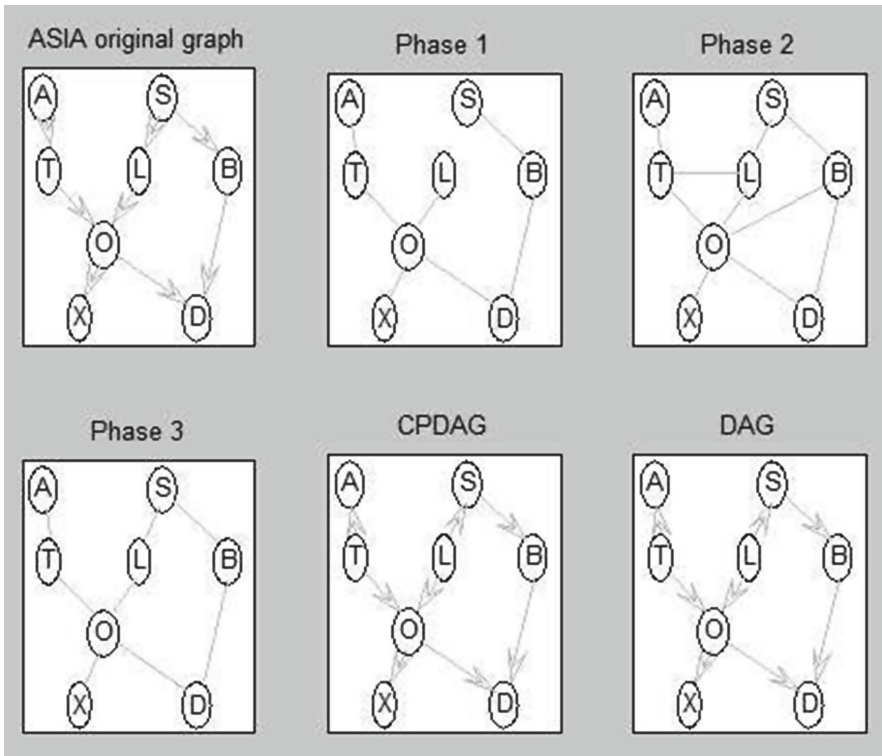


Fig. 1. Result of the evidential learning process

The last step will be then to pass from the CPDAG to the final network as shown in Fig. 1, the algorithm of the construction of a DAG from a CPDAG is detailed in [5].

7 Conclusion

A constrained based approach for learning evidential networks structure from evidential data has been proposed. This approach generalizes the classical constrained based methods usually used for learning structure in BNs by extending the statistical χ^2 test to the belief function framework in order to deal with different types of imperfection in data.

In the future works, we will tend to investigate different horizons:

- Extend score based methods to deal with evidence data.
- Develop detailed experimental results and compare the efficiency of structure learning approaches in the evidence framework.

References

1. Bach Tobji, M.A., Ben Yaghlane, B., Mellouli, K.: A new algorithm for mining frequent itemsets from evidential databases. In: Proceedings of Information Processing and Management of Uncertainty (IPMU 2008), pp. 1535–1542, Malaga (2008)
2. Ben Hariz, N., Ben Yaghlane, B.: Learning parameters in directed evidential networks with conditional belief functions. In: Cuzzolin, F. (ed.) BELIEF 2014. LNCS, vol. 8764, pp. 294–303. Springer, Heidelberg (2014)
3. Ben Yaghlane, B.: Uncertainty Representation and Reasoning in Directed Evidential Networks. Ph.D. thesis, Institut Supérieur de Gestion de Tunis (2002)
4. Ben Yaghlane, B., Mellouli, K.: Inference in directed evidential networks based on the transferable belief model. *IJAR* **48**(2), 399–418 (2008)
5. Chickering, D.: Optimal structure identification with greedy search. *J. Mach. Learn. Res.* **3**, 507–554 (2002)
6. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc. Ser. B* **39**, 1–38 (1977)
7. Dencœur, T.: Maximum likelihood estimation from uncertain data in the belief function framework. *Knowl. Data Eng.* **25**, 113–119 (2013)
8. Hewawasam, K., Premaratne, K., Subasingha, S., Shyu, M.-L.: Rule mining and classification in imperfect databases. In: Proceedings of the 8th International Conference on Information Fusion, vol. 1, pp. 661–668, Philadelphia (2005)
9. Jordan, M.: Learning in Graphical Models. Kluwer Academic Publisher, New York (1998)
10. Krause, P.J.: Learning probabilistic networks. *Knowl. Eng. Rev.* **13**(4), 321–351 (1998)
11. Lauritzen, S.L., Spiegelhalter, D.J.: Local computation with probabilities and graphical structures and their application to expert systems. *J. Roy. Stat. Soc. B* **50**, 157–224 (1988)
12. Naim, P., Wuillemin, P.H., Leray, P., Pourret, O., Becker, A.: Réseaux Bayésiens. Eyrolles, Paris (2004)
13. Neapolitan, R.: Learning Bayesian Networks. Prentice Hall, New York (2003)
14. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, New York (1988)
15. Pearl, J., Verma, T.: A theory of inferred causation. In: Allen, J., Fikes, R., Sandewall, E. (eds.) Proceedings of the Second International Conference on Knowledge Representation and Reasoning, pp. 441–452. Morgan Kaufmann, New York (1991)
16. Petitrenaud, S.: Independence tests for uncertain data with a frequentist method. In: Proceedings of the 5th International Conference on Soft Methods in Probability and Statistic (SMPS 2010), pp. 519–526 (2010)
17. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, Princeton (1976)
18. Smets, P.: Jeffrey’s rule of conditioning generalized to belief functions. In: Proceedings of the Ninth international conference on Uncertainty in artificial intelligence (UAI 1993), pp. 500–505, Washington (1993)
19. Smets, P., Kennes, R.: The transferable belief model. *Artif. Intell.* **66**, 191–234 (1994)
20. Spirtes, P., Glymour, C., Scheines, R.: Causation, Prediction, and Search. Springer-Verlag, New York (2000)

21. Tang, Y., Zheng, J.: Dempster conditioning and conditional independence in evidence theory. In: Australian Conference on Artificial Intelligence, vol. 3809, pp. 822–825, Sydney (2005)
22. Verma, T., Pearl, J.: Equivalence and synthesis of causal models. In: Proceedings of Sixth Conference on Uncertainty in Artificial Intelligence, pp. 220–227 (1990)
23. Xu, H., Smets, Ph.: Evidential reasoning with conditional belief functions. In: Heckerman, D., et al. (eds.) Proceedings of Uncertainty in Artificial Intelligence (UAI 1994), pp. 598–606, Seattle (1994)