Learning Parameters in Directed Evidential Networks with Conditional Belief Functions

Narjes Ben Hariz and Boutheina Ben Yaghlane

LARODEC Laboratory - Institut Supérieur de Gestion de Tunis, Tunisia
narjes.benhariz@gmail.com,
boutheina.yaghlane@ihec.rnu.tn

Abstract. Directed evidential networks with conditional belief functions are one of the most commonly used graphical models for analyzing complex systems and handling different types of uncertainty. A crucial step to benefit from the reasoning process in these models is to quantify them. So, we address, in this paper, the issue of estimating parameters in evidential networks from evidential databases, by applying the maximum likelihood estimation generalized to the evidence theory framework.

Keywords: Belief Functions, Parameters Estimation, Evidential models, Evidential DataBases.

1 Introduction

Evidential graphical models have gained, in recent years, an expanding interest as a powerful tool for modeling and analyzing complex systems and reasoning under different types of uncertainty based on the belief functions theory.

One of the most commonly applied models in the evidential framework are the Directed EVidential Networks with conditional belief functions (DEVNs) [3]. On one hand, these models generalize the evidential networks with conditional belief functions [18] by handling n-ary relations between variables, on the other hand, unlike probabilistic models such as Bayesian networks [10], they are able to handle different levels of uncertainty in data.

A DEVN is based on two parts: the graphical part that consists on a directed acyclic graph with a set of nodes and a set of edges and the numeric parameters represented by conditional belief functions. Another point of interest of these networks is their flexibility in representing beliefs. In fact, conditional beliefs in these models can be expressed according to two different manners: for each node in the context of its parents (per child node) or for each dependency relation between a parent node an a child node (per edge).

The majority of works concerning DEVNs address inference algorithms and reasoning in these networks [3,2,8]. Nevertheless, an essential step before being able to reason with evidential networks is to quantify them. The data needed in the quantification process are generally derived from expert opinions or from data stored in databases.

F. Cuzzolin (Ed.): BELIEF 2014, LNAI 8764, pp. 294–303, 2014.

Thus, we address in this paper the problem of learning parameters in DEVNs from uncertain data stored in Evidential DataBases (EDB) [1], and this, by applying the maximum likelihood principle, one of the most statistical methods generally used in learning BNs [6,7,12].

The paper is organized as follow: In Section 2, we remaind briefly the most important background notions regarding the belief functions theory. The evidential data bases are recalled in Section 3. We recall the basic concepts related to the directed evidential networks with conditional belief functions in section 4. Section 5 concerns the maximum likelihood and its use in learning BNs. In Section 6, we present the main purpose of the paper which is the algorithm of learning parameters in DEVNs with its two variants, per child node and per edge. In the last Section, we explain the proposed approach through an illustrative example.

2 Belief Functions Theory: Basic Concepts

The belief functions theory, also known as evidence theory or Dempster-Shafer theory is a general and flexible framework for handling and modeling different types of uncertainty [13]. In the following, we remind some basic concepts of this theory, more details can be found in [13,15].

Let Ω be a finite set of exclusive and exhaustive elements called the frame of discernment and 2^{Ω} its power set.

The portion of belief supporting exactly a proposition A is called the basic belief assignment (*bba*), which is a function from 2^{Ω} to $[0,1]$ such that:

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

Any subset $A \in \Omega$ with $m^{\Omega}(A) > 0$ is called a focal element, and the set of all these elements is denoted by $\mathcal{F}(m^{\Omega})$.

With each mass function m^{Ω} is associated a belief (bel^{Ω}) and plausibility (pl^{Ω}) functions from 2^{Ω} to $[0,1]$, which give the minimum and maximum amount of support attributed to A, respectively. These functions are defined as follows:

$$bel^{\Omega}(A) = \sum_{\emptyset \neq B \subseteq A} m^{\Omega}(B) \tag{2}$$

$$pl^{\Omega}(A) = \sum_{\emptyset \neq B \cap A} m^{\Omega}(B) \tag{3}$$

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

$$m^{\Omega}[B](A) = \sum_{C \subseteq \overline{B}} m^{\Omega}(A \cap C), \tag{4}$$

where $\bar{B}$ is the complement of the proposition B. More details about the rules of conditioning in the belief functions theory can be found in [14,16].

3 Evidential DataBases

An Evidential DataBase (EDB) or a Dempster-Shafer (DS) database is a database storing certain or/and uncertain data modeled using the belief functions framework [1].

In an EDB with L lines and C columns (attributes), each attribute $c \in [1, C]$ has a frame of discernment Ω_c including its possible values.

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

$$m_{lc}(\emptyset) = 0 \; and \sum_{A \subseteq \Omega_c} m_{lc}(A) = 1 \tag{5}$$

Data in an EDB, can take different levels of imperfection:

- Certain data: when the focal element is a singleton with a mass equal to one.
- Probabilistic data: when all focal elements are singletons.
- Possibilistic data: when focal elements are nested.
- Missing data: when the total amount of evidence is affected to one focal element which is the frame of discernment.
- Evidential data: including any other type of information.

4 Directed Evidential Networks with Conditional Belief Functions

Directed EVidential Networks with conditional belief functions (DEVNs) are proposed in [3] to generalize the evidential networks with conditional belief functions (ENCs) [18] that generalize Bayesian Networks (BNs) [10] for handling different types of uncertainty using evidence theory framework.

As it is derived from ENCs and BNs, a DEVN is based on two principal parts:

The qualitative level which is modeled by a Directed Acyclic Graph (DAG) $G = (N, E)$, where $N = \{N_1, ..., N_x\}$ is the set of nodes (variables), and $E = \{E_1, ..., E_y\}$ is the set of edges coding the different conditional dependencies between variables.

The quantitative level which is represented by a set of parameters θ modeled by conditional belief functions. Each node in the DEVN is associated with an a priori mass function. If it is a root node, adding to this function, the node is associated with a conditional mass function defined per edge or per child node.

Each node N_i in a DEVN is a representation of a random variable taking its values on a frame of discernment Ω_{N_i}. Let $PA(N_i)$ and $CH(N_i)$ denote the set of its parent nodes and the set of its child nodes, respectively. Like in BNs, each root node in a DEVN is associated with an a priori *bbm*, but unlike in BNs, child nodes in DEVNs are associated with both an a priori mass function and a conditional one.

Conditional belief functions in DEVNs can be defined in two manners:

- **Per child node**, as in BNs: for each child node N_c is associated a conditional belief function given all its parent nodes $PA(N_c)$. This conditional mass is denoted by $m^{\Omega_{N_c}}[PA(N_c)](N_c)$.
- **Per edge**, as in ENCs: the conditional relation between a child node N_c and a parent node $N_p \in PA(N_c)$, represented by an edge, is weighted with a conditional mass function $m^{\Omega_{N_c}}[N_p](N_c)$.

These to ways of modeling conditional beliefs makes DEVNs more flexible than BNs and ENCs and make the quantification of the network easier to an expert.

5 Maximum Likelihood and Learning in BNs

The issue of parameter estimation from data sets remains an important subject in statistics and knowledge management problems. One of the well known statistic methods for estimating parameters of a statistical model is the Maximum Likelihood (ML) principle [11]. This method is the center of the majority of approaches of learning parameters in probabilistic models from databases containing both complete and missing data.

When all variables are observed perfectly, the simplest and most used method for estimating probabilities in BNs is the ML which measure the probability of an event by its frequency of occurrence in the database. The estimated probability to a random variable[1] X_i conditionally to its parent nodes $PA(X_i)$ is calculated as follows:

$$P(X_i = x_k | PA(X_i) = x_j) = \frac{N_{i,j,k}}{\sum_k N_{i,j,k}}, \tag{6}$$

where $N_{i,j,k}$ is the number of events for which X_i takes the value x_k and its parents takes the configuration of values x_j. More details concerning the statistical learning in BNs can be found in [6,7,12].

Many other learning approaches are developed to estimate parameters from databases containing missing data, one of the most popular is the Expectation Maximization (EM) algorithm [4] which is based mainly on the ML estimation.

The likelihood principle and the EM algorithm were generalized, under the belief functions framework, to the Credal EM [17] and the Evidential EM [5] in order to handle the imprecision and the uncertainty in data.

The main idea of the extension of the likelihood notion to the evidence framework is to take the classical likelihood, defined originally in the probability framework, weighted by the mass function associated to each variable [5].

Thus, we apply, in the rest of the paper, the maximum likelihood principle and its generalization in the evidence theory, to develop a new algorithm for estimating the a priori mass function and the conditional beliefs in a directed evidential network from data bases storing different types of data: complete, missing, certain and/or uncertain.

---

[1] Each random variable corresponds to a node in the Bayesian network.

6 Learning Parameters in DEVNs

We present in this part, the main purpose of the paper which is learning parameters in directed evidential networks with conditional belief functions from an evidential database by applying the maximum likelihood estimation principle.

As mentioned previously, in a DEVN, each node is associated with an a priori *bba* and each child node is quantified by a conditional belief function modeled according to two different approaches: per child node, when the mass function of each child node is calculated given all its parent nodes, or per edge, when the relation between a child node and a parent node is evaluated by a conditional belief.

Let us consider a DEVN with a set of nodes N and a set of edges E and an EDB with L lines and C columns such that each column corresponds to a random variable (node) in the DEVN.

Building on the generalization idea of the likelihood principle in the evidence theory and analogically to the ML in the probability framework expressed by equation (6), the a priori mass function of a node $N_i \in N$ can be calculated as follows:

$$m^{\Omega_{N_i}}(N_i = A_k) = \frac{\sum_{l=1}^{|L|} m_{lc}^{\Omega_{N_i}}(N_i = A_k)}{\sum_{l=1}^{|L|} m_{lc}^{\Omega_{N_i}}(N_i)}, \tag{7}$$

where A_k is a proposition from $2^{\Omega_{N_i}}$, c denotes the column corresponding to the node N_i and $m_{lc}^{\Omega_{N_i}}$ is the mass function defining the cell in the l^{th} line and c^{th} column.

Similarly, we define in equation (8) the conditional mass function of a node N_i given its parent nodes $PA(N_i) = \{pa_1(N_i), ..., pa_z(N_i)\}$:

$$m^{\Omega_{N_i}}[PA(N_i) = x](N_i = A_k) = \frac{\sum_{l=1}^{|L|} m_{lc}^{\Omega_{N_i}}(N_i = A_k) * \prod_j m_{lcj}^{\Omega_{pa_j}}(pa_j(N_i) = x_j)}{\sum_{l=1}^{|L|} \prod_j m_{lcj}^{\Omega_{pa_j}}(pa_j(N_i) = x_j)}, \tag{8}$$

where x is a configuration of values in which each parent node takes a possible proposition from its frame of discernment.

These equations are the core of the learning parameters algorithms in directed evidential networks.

6.1 Learning Algorithm Per Child Node

The process of estimating parameters per child node in a DEVN is based on two main steps: the estimation of an a priori mass function for each node and the estimation of the conditional mass function of each child node given all its parent nodes. This process is detailed formally by Algorithm 1.

Note that this algorithm can be used for learning parameters in Bayesian networks from probabilistic data.

Algorithm 1. Learning parameters per child node

Require: $DAG = (N, E)$, $Data$
Ensure: $DEVN = (N, E, \theta_p, \theta_c)$

```
for each node N_i ∈ N do
  1. Calculate the a priori mass function m^{Ω_{N_i}}(N_i)
  c ← SelectColumn(Ni, C)

  for each proposition A_k in 2^{Ω_{N_i}} do
    m^{Ω_{N_i}}(A_k) ←Result_of_equation(7)
  end for
  N_i.θ_p ← m^{Ω_{N_i}}(A)
  2. Calculate the conditional mass function m^{Ω_{N_i}}[PA](N_i)

  if N_i is a child node then
    cPA = SelectColumns(PA, C)

    for each proposition A_k in 2^{Ω_{N_i}} do

      for each possible configuration conf_j do
        m^{Ω_{N_i}}[PA = conf_j](A_k) ← Result_of_equation(8)
      end for
    end for
    N_i.θ_c ← m^{Ω_{N_i}}[PA](A)
  else
    N_i.θ_c ← ∅
  end if
end for
```

6.2 Learning Algorithm Per Edge

The approach of learning parameters per edge, described in Algorithm 2, aims to quantify each dependency relation between a parent node and a child node by a conditional mass function. The step of estimation the a priori mass function for each node is similar to the first step in Algorithm 1.

Note that this algorithm can be also used for learning parameters in evidential networks with conditional belief functions from any type of data.

7 Illustrative Example

In order to explain the learning algorithm detailed previously, we present in the following an illustrative example focusing on a part from "ASIA" network[2] and a part from a corresponding evidential database modeled in figure 1.

[2] The Bayesian network of the classical problem Asia Chest Clinic first described in [9]

Algorithm 2. Learning parameters per edge

Require: $DAG = (N, E)$, $Data$
Ensure: $DEVN = (N, E, \theta_p, \theta_c)$

for each node $N_i \in N$ **do**
  1. Calculate the a priori mass function $m^{\Omega_{N_i}}(N_i)$
  2. Calculate the conditional mass function $m^{\Omega_{N_i}}[pa(Ni)](N_i)$

  if N_i is not a root node **then**
    $PA \leftarrow Parents(N_i)$

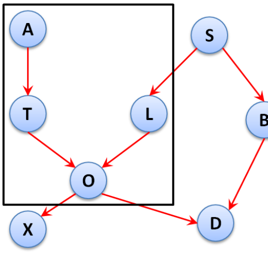
    for each proposition A_j in $2^{\Omega_{N_i}}$ **do**

      for each parent node $pa \in PA$ **do**
        $c \leftarrow SelectColumn(pa, C)$
        $m^{\Omega_{N_i}}[pa = x_q](A_j) \leftarrow Result\_of\_equation(8)$
        $N_i.pa.\theta_c \leftarrow m^{\Omega_{N_i}}[pa](A)$
      end for
    end for
  else
    $N_i.\theta_c \leftarrow \emptyset$
  end if
end for

All variables in "ASIA" network are binary, we consider in this example, as shown in figure 1, four variables $\{A, T, O, L\}$ having the power sets, respectively: $\{a, \bar{a}, a \cup \bar{a}\}$; $\{t, \bar{t}, t \cup \bar{t}\}$; $\{o, \bar{o}, o \cup \bar{o}\}$ and $\{l, \bar{l}, l \cup \bar{l}\}$.

Data used in this example are composed from 20 instances and contain different levels of imperfection: uncertain attributes, certain attributes and imprecise attributes.

The different results of applying the learning process to the selected part of "ASIA" network are shown in figure 2. For each root node (A and L) is associated an a priori mass function. The node T is quantified by an a priori mass function $m^{\Omega_T}(T)$ and a conditional belief knowing its parent node $m^{\Omega_T}[A](T)$. Note that in the case of a node having one parent (such as T), the result of learning parameters per child node or per edge is the same. For the node O is associated an a priori mass function $m^{\Omega_O}(O)$, a conditional mass function given all its parent nodes $m^{\Omega_O}[T, L](O)$ and a conditional mass function given one parent node $m^{\Omega_O}[T](O)$ and $m^{\Omega_O}[L](O)$.




```
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
1(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
```

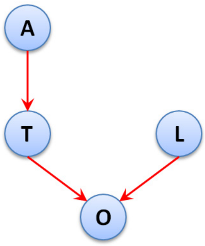
Fig. 1. The graphical structure and the EDB of the network for the Asia Chest Clinic problem

In the following we present some examples of calculation details in order to further clarify equations (7) and (8):

- $m^{\Omega_A}(A=a) = \frac{\sum_1^{20} m_{lc}^{\Omega_A}(A=a)}{\sum_1^{20}(m_{lc}^{\Omega_A}(A=a)+m_{lc}^{\Omega_A}(A=\bar{a})+m_{lc}^{\Omega_A}(A=a\cup\bar{a}))} = \frac{0+0+0.5+\ldots+1}{1+1+0.5+0.5+\ldots+1} = 0.2185$

$m^{\Omega_A}(A)$

a	0.2185
$\bar{a}$	0.558
$a \cup \bar{a}$	0.2235



	$m^{\Omega_T}(T)$	$m^{\Omega_T}[A](T)$: a	$\bar{a}$	$a \cup \bar{a}$
t	0.45	0.736	0.376	0.353
$\bar{t}$	0.55	0.264	0.624	0.647
$t \cup \bar{t}$	0	0	0	0

$m^{\Omega_L}(L)$

l	0.3
$\bar{l}$	0.6
$l \cup \bar{l}$	0.1

Per child node

	$m^{\Omega_O}(O)$	$m^{\Omega_O}[T,L](O)$: l			$\bar{l}$			$l \cup \bar{l}$		
		t	$\bar{t}$	$t \cup \bar{t}$	t	$\bar{t}$	$t \cup \bar{t}$	t	$\bar{t}$	$t \cup \bar{t}$
o	0.5145	1	0.52	0	0.57	0.425	0	0	0	0
$\bar{o}$	0.3555	0	0.48	0	0.2	0.325	0	1	1	0
$o \cup \bar{o}$	0.13	0	0	1	0.23	0.25	1	0	0	1

Per edge

	$m^{\Omega_O}(O)$	$m^{\Omega_O}[T](O)$: t	$\bar{t}$	$t \cup \bar{t}$	$m^{\Omega_O}[L](O)$: l	$\bar{l}$	$l \cup \bar{l}$
o	0.45	0.554	0.48	0	0.6	0.558	0
$\bar{o}$	0.55	0.267	0.42	0	0.4	0.22	1
$t \cup \bar{o}$	0	0.18	0.09	1	0	0.216	0

Fig. 2. The result of applying the learning parameters algorithms on a part of ASIA network

- $m^{\Omega_T}[A=a](T=\bar{t}) = \frac{\sum_1^{20} m_{lc}^{\Omega_T}(T=\bar{t})*m_{lc}^{\Omega_A}(A=a)}{\sum_1^{20} m_{lc}^{\Omega_A}(A=a)} = \frac{1*0.5+1*0.2+1*0.45}{0.5+1+0.22+0.2+0.45+1+1} = 0.264$
- $m^{\Omega_O}[T=\bar{t}, L=l](O=o) = \frac{\sum_1^{20} m_{lc}^{\Omega_O}(O=o)*m_{lc}^{\Omega_T}(T=\bar{t})*m_{lc}^{\Omega_L}(L=l)}{\sum_1^{20} m_{lc}^{\Omega_T}(T=\bar{t})*m_{lc}^{\Omega_L}(L=l)} = \frac{1*1*1+1*1*1+1*1*0.6}{1*1+1*1+1*1+1*1+1*1} = 0.52$

Note that if we use probabilistic data (as in node T), then these equations give the same result as equation (6) which explains the fact that applying our first algorithm (for the case per child node) to learn the parameters of the DEVNs from a complete database gives the same results as the algorithm based in the maximum likelihood for learning parameters in Bayesian networks.

It is important to mention that if a configuration value of a parent node does not exist in the database, then the total amount of belief will be assigned to the total ignorance. For instance the proposition $\{t \cup \bar{t}\}$ does not appear in the database (the value $\{0, 1\}$ in T column), this makes $m^{\Omega_O}[T = t \cup \bar{t}, L = l](O = o \cup \bar{o})$ equal to $m^{\Omega_O}[T = t \cup \bar{t}, L = \bar{l}](O = o \cup \bar{o})$ equal to $m^{\Omega_O}[T = t \cup \bar{t}, L = l \cup \bar{l}](O = o \cup \bar{o})$ equal to 1. This can be a simple solution for the problem of zero counts in the data.

8 Conclusion

We have proposed, in this paper, new algorithms for learning parameters in directed evidential networks with conditional belief functions by applying the generalization of the maximum likelihood estimation in the evidence theory for handling uncertainty in data, stored in evidential databases.
As a future work, we intend to study the complexity of the proposed algorithm by applying it to complex systems and big databases. Another center of interest will be to improve the equation used in the proposed algorithms to deal with some problems such as zero counts and overfitting.

References

1. Bach Tobji, M.A., Ben Yaghlane, B., Mellouli, K.: A new algorithm for mining frequent itemsets from evidential databases. In: Proceeding of Information Processing and Management of Uncertainty (IPMU 2008), Malaga, Spain, pp. 1535–1542 (2008)
2. Ben Yaghlane, B., Mellouli, K.: Updating directed belief networks. In: Hunter, A., Parsons, S. (eds.) ECSQARU 1999. LNCS (LNAI), vol. 1638, pp. 43–54. Springer, Heidelberg (1999)
3. Ben Yaghlane, B., Mellouli, K.: Inference in Directed Evidential Networks Based on the Transferable Belief Model. IJAR 48(2), 399–418 (2008)
4. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum Likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, SERIES B 39, 1–38 (1977)
5. Denœux, T.: Maximum Likelihood Estimation from Uncertain Data in the Belief Function Framework. Knowledge and Data Engineering 25, 119–113 (2013)

6. Jordan, M.: Learning in Graphical Models. Kluwer Academic Publisher (1998)
7. Krause, P.J.: Learning probabilistic networks. The Knowledge Engineering Review 13(4), 321–351 (1998)
8. Laâmari, W., Ben Yaghlane, B.: Reasoning in Singly-Connected Directed Evidential Networks with Conditional Beliefs. In: Likas, A., Blekas, K., Kalles, D. (eds.) SETN 2014. LNCS, vol. 8445, pp. 221–236. Springer, Heidelberg (2014)
9. Lauritzen, S.L., Spiegelhalter, D.J.: Local computation with probabilities and graphical structures and their application to expert systems. J. Royal Statistical Society B 50, 157–224 (1988)
10. Murphy, K.: Probabilistic Graphical Models. Michael Jordan (2002)
11. Myung, I.J.: Tutorial on maximum likelihood estimation. Journal of Mathematical Psycology 47, 90–100 (2003)
12. Naim, P., Wuillemin, P.H., Leray, P., Pourret, O., Becker, A.: Réseaux Bayésiens. Eyrolles (2004)
13. Shafer, G.: A Mathematical Theory of Evidence. Princeton Univ. Press, Princeton (1976)
14. 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), Washington, DC, USA, pp. 500–505 (1993)
15. Smets, P., Kennes, R.: The transferable belief model. Artificial. Intelligence 66, 191–234 (1994)
16. Tang, Y., Zheng, J.: Dempster Conditioning and Conditional Independence in Evidence Theory. In: Zhang, S., Jarvis, R.A. (eds.) AI 2005. LNCS (LNAI), vol. 3809, pp. 822–825. Springer, Heidelberg (2005)
17. Vannoorenberghe, P., Smets, P.: Partially supervised learning by a **cr**edal **EM** approach. In: Godo, L. (ed.) ECSQARU 2005. LNCS (LNAI), vol. 3571, pp. 956–967. Springer, Heidelberg (2005)
18. Xu, H., Smets, P.: Evidential Reasoning with Conditional Belief Functions. In: Heckerman, D., et al. (eds.) Proceedings of Uncertainty in Artificial Intelligence (UAI 1994), Seattle, Washington, USA, pp. 598–606 (1994)