Uncertain Evidence in Bayesian Networks : Presentation and Comparison on a Simple Example

Ali Ben Mrad[1], Veronique Delcroix[2], Mohamed Amine Maalej[1], Sylvain Piechowiak[2], and Mohamed Abid[1]

[1] University of Sfax, Tunisia
ENIS, CES, Sfax, Tunisia
[2] Université Lille Nord de France, F-59000 Lille, France
UVHC, LAMIH UMR 8201, F-59313 Valenciennes, France

Abstract. We consider the problem of reasoning with *uncertain evidence* in Bayesian networks (BN). There are two main cases: the first one, known as *virtual evidence*, is evidence *with* uncertainty, the second, called *soft evidence*, is evidence *of* uncertainty. The initial inference algorithms in BNs are designed to deal with one or several hard evidence or virtual evidence. Several recent propositions about BN deal with soft evidence, but also with ambiguity and vagueness of the evidence. One of the proposals so far advanced is based on the fuzzy theory and called *fuzzy evidence*. The original contribution of this paper is to describe the different types of uncertain evidence with the help of a simple example, to explain the difference between them and to clarify the appropriate context of use.

Keywords: bayesian networks, uncertain evidence, virtual evidence, likelihood evidence, soft evidence, fuzzy evidence

1 Introduction

Bayesian networks (BN) [Pe88,Je96] are powerful tools for knowledge representation and inference under uncertainty. They combine multiple sources of information to provide a formal framework within which complex systems can be represented and processed. The different sources of information are not always perfect, therefore, the observation can be uncertain and imprecise. For the purpose of this paper, we present five types of evidence: *hard evidence*, *virtual evidence* (VE), also called likelihood evidence, that is evidence *with* uncertainty [Pe88], *soft evidence* (SE) that is evidence *of* uncertainty [VK02], and two approaches of *fuzzy evidence* [MM11,TL07]. These methods are applied and presented on a simple example. A result of this work is to clarify the distinction between these different types of evidence. The presence of several soft evidences has to be treated using specific algorithms. We detail the case of a single evidence and briefly explain the case of several evidences.

2 Different Types of Evidence in Bayesian Networks

2.1 Definitions and Vocabulary

Evidence in BN may be regular or uncertain. Regular evidence, called also hard evidence specifies which value a variable is in. This is the usual way to enter an observation to be propagated in a BN [Pe88,Je96]. Uncertain evidence specifies the probability distribution of a variable. We focus on two types of uncertain evidences. According to [VK02,PZ10], we use the terms *virtual evidence* and *soft evidence* as follows: *virtual evidence* [Pe88] can be interpreted as evidence *with* uncertainty, and can be represented as a likelihood ratio. This kind od evidence is also called *likelihood evidence*. *Soft evidence* [VK02], can be interpreted as evidence *of* uncertainty, and is represented as a probability distribution of one or more variables.

Many BN engines accept a probability distribution as input for the update. Most existing implementations of uncertain evidence are virtual evidence, but the literature is not consistent about naming uncertain evidence. The term *soft evidence* is used in many cases incorrectly as indicated in Table 1. In the cases listed in Table 1, the evidence is considered to be a virtual evidence and is propagated by adding a virtual node.

Table 1. Different Names of the Virtual Evidence (VE) in the BN Engines

BN engines	**Names of the VE**	**Web Site**
BNT	Soft evidence	http://www.cs.ubc.ca/~murphyk/Software/BNT/bnt.html
Bayesialab	Likelihood distribution	http://www.bayesia.com
NETICA	Likelihood	http://www.norsys.com
HUGIN	Likelihood findings	http://www.hugin.com
GeNIe	Soft evidence	http://genie.sis.pitt.edu

2.2 Algorithms Dealing with Uncertain Evidence

The issue of how to deal with uncertain evidence in BN appears in [Pe88] and has recently been the subject of many algorithms developments as indicated in Table 2. To clarify the distinction between the different types of evidence in BN we present in the following sections an illustrative example and the modeling of the different types of evidence.

Table 2. Algorithms dealing with uncertain evidence (VE: virtual evidence, SE: soft evidence)

Algorithms	**Type of evidence**	**References**
VE method	VE	[Pe90]
Jeffrey's rule	VE and single SE	[Je83]
IPFP (Iterative Proportional Fitting Procedure)	SE	First appeared in [Kr37] Studied in [Cs75,Fi70,PD05] Extended in [Bo89,Cr00]
The big Clique algorithm and extension	SE	[VK02]
Derived Algorithm Combining VE method, Jeffrey's rule and IPFP	VE and SE	[PZ10]

3 Comparison of Different Types of Evidence with a Simple Example

3.1 Presentation of the "Snow Example"

Our example models the influence of the amount of snow on the congestion (Fig. 1). The variable S represents the "amount of snow in mm" with values in $[0, 120]$ and the variable C represents the "congestion" with values in $\{yes, no\}$.

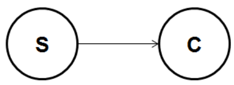

Fig. 1. Bayesian network graph of the snow example

The conditional probability of C given S is defined by the equation:

$$P(C = yes \mid S) = e^{-\frac{1}{2}\times(\frac{S-60}{40})^2} \quad (1)$$

This probability function can be understood as follows: under the threshold of 60 mm of snow, the congestion is all the more probable that the amount of snow is important. Beyond this threshold, people leave their homes less and less and subsequently the probability of congestion decreases. Whereas some BN engines deal with continuous variables, or even mixted variables, the most common way is to use a discretization of the variable S (see Table 3). The BN engine Netica [Ne] offers the possibility to obtain the Conditional Probability Table (CPT) of the node S discretized from the equation (1). In the following cases, the CPT $P(C|S)$ is conformed to the equation (1), to ensure a proper comparison of results of the different methods. Our starting model is described by the graph presented in Fig. 1 and by the probabilities given in Tables 3 and 4 for the probabilities.

Table 3. $P(S)$

S_1	0	0.6
S_2	]0,40]	0.22
S_3	]40,80]	0.14
S_4	]80,120]	0.04

Table 4. $P(C \mid S)$

	S_1	S_2	S_3	S_4
yes	0.3247	0.6058	0.96	0.6058
no	0.6753	0.3942	0.04	0.3942

We propagate the same observation for the different types of evidence to ensure a good comparison. Assume that the amount of snow is effectively 80 mm. We are going to compute $P(C = yes|e)$, $P(C = yes|ve)$, $P(C = yes|se)$ where e represents an hard evidence, ve denotes a virtual evidence and se denotes a soft evidence. The last two parts concern the case of ambiguity.

3.2 Junction Tree Algorithm

We apply the junction tree inference algorithm [La88,Je90] to different types of evidence. We will trace the changes in calculation and graph during the successive stages of this algorithm. It can be summarized as follows:

- Construction process (or transformation of the graph): moralizing the graph, triangulating the graph, forming the junction tree.
- Initialization process: initializing the potential of cliques and separators.
- Propagation process: ordered series of local manipulations, called message passing, on the join-tree potentials. The result is a consistent join tree.
- Marginalization process: from the consistent join tree, compute the posterior probability $P(V)$ for each variable of interest V.

After construction of the junction tree, the network of our example is reduced to a single clique $\{SC\}$. The construction process is valid for all types of BN. However, the other three phases are different depending on the presence or absence of observations. In this paper, we study the case of presence of observation.

3.3 Hard Evidence

Hard evidence is an observation of a random variable having with certainty a particular value $V = v$. To encode observations on a variable V, we consider the notion of likelihood Λ_V as indicated in [HD96], which is encoded as follows: If V is observed, then $\Lambda_V(v) = 1$ when v is the observed value of V and $\Lambda_V(v) = 0$ otherwise. If V is not observed, then $\Lambda_V(v) = 1$ for each value v. The likelihood encoding of evidence is used in the initialization process to enter the observation in the junction tree. A hard evidence is represented by a likelihood where $\Lambda_V(v) = 1$ for exactly one value v. Assume that the amount of snow observed is 80 mm. This observation is interpreted as $S = S_3$ and illustrated in Table 5.

Table 5. Likelihood encoding an hard evidence

	S				C	
v	S_1	S_2	S_3	S_4	yes	no
$\Lambda_V(v)$	0	0	1	0	1	1

The probability $P(C = yes \mid e)$ where e denotes the hard evidence $S = S_3$ is a part of the definition of the BN: $P(C = yes \mid e) = 0.96$ (see Table 4).

The drawback of discretization is that all values in the same interval are treated in the same way no matter their position in the interval. In our example, the observation of 41 mm or 79 mm provides the same results. Since the chosen discretization is coarse with only 4 states, the result may be not satisfying. The finer the discretization, the more relevant are the results, and the larger is the CPT. A discretization of S with 13 states provides $P(C = yes \mid e) = 0.93$ and with 60 states we obtain $P(C = yes \mid e) = 0.893$, where e represents the hard evidence $S = S_i$ and S_i is the interval containing 80 in the chosen discretization.

3.4 Virtual Evidence

Virtual Evidence (VE), proposed by Pearl in [Pe88], provides a convenient way of incorporating evidence *with* uncertainty. A VE on a variable V is represented by a likelihood Λ_V where each $\Lambda_V(v)$ is a real number in $[0, 1]$. Pearl's method extends the given BN by adding a binary virtual node which is a child of V. In our example, we add a node S_{obs} and a directed edge from S toward S_{obs} (see Fig. 2).

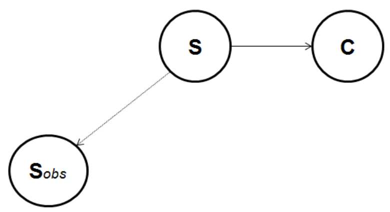

Fig. 2. Bayesian Network graph encoding a virtual evidence on S.

In our example, we consider now the virtual evidence ve on S represented by the likelihood given in Table 7. This translates that the observed amount of snow is about 80 mm, and since the measure is rough, both intervals S_3 and S_4 must be considered. In the BN, the main value of S_{obs}, say *yes*, corresponds to the virtual evidence ve that is represented by the CPT of the virtual node S_{obs} (Table 6).

Table 6. CPT of $P(S_{obs} \mid S)$

	S_1	S_2	S_3	S_4
S_{obs} = yes	0	0	0.5	0.5
S_{obs} = no	1	1	0.5	0.5

Table 7. Likelihood encoding *ve*

	S				C	
v	S_1	S_2	S_3	S_4	yes	no
$\Lambda_V(v)$	0	0	0.5	0.5	1	1

The virtual evidence ve can be easily propagated by propagating the hard evidence $S_{obs} = yes$ in the augmented BN; we obtain $P(C = yes \mid ve) = P(C = yes \mid S_{obs} = yes) = 0.881$. This result can be obtained as follows:

$$P(C = yes \mid ve) = \frac{\sum_i P(C = yes \mid S_i) \times Q(S = S_i)}{\sum_{c,i} P(C = c \mid S_i) \times Q(S = S_i)} \quad (2)$$

where $Q(S = S_i) = P(S = S_i \mid S_{obs} = yes)$ represents the quantified values as indicated in [TA04]. In the VE method, the propagated value is the marginal on S, represented by the quantified values Q.

3.5 Soft Evidence

Soft evidence (SE), named by Valtorta in [VK02], can be interpreted as evidence of uncertainty. SE is given as a probability distribution of one or several variables $R(Y)$, $Y \subseteq X$, where X is the set of variables. Therefore, there is uncertainty about the specific state Y is in but we are sure of the probability distribution $R(Y)$. Since $R(Y)$ is a certain observation, this distribution should be preserved by updating belief. This is the main difference with virtual evidence for which this is not required.

In case of a single soft evidence, Chan and Darwiche [CD05] showed that a soft evidence can be converted into a virtual evidence and updating can be carried out by virtual evidence method as detailed in [PZ10]. This method is not directly applicable to the situation in which multiple soft evidences are presented since it does not guarantee that the soft evidence $R(Y)$ will be preserved after updating. Updating several SE requires specific algorithms to preserve the initial distribution (see Table 2). An interesting use of soft evidence regarding discretization is proposed in [DB08].

In our example, we assume that we have a soft evidence se on S, given by the distribution $R(S) = (0, 0, 0.5, 0.5)$. This means that we are sure that the amount of snow is in the interval $[40, 120]$. The likelihood ratio is $L(S) = \frac{R(S)}{P(S)}$ where $P(S)$ is the marginal probability of S given in Table 3. Thus, in our example, $L(S) = 0 : 0 : \frac{0.5}{0.14} : \frac{0.5}{0.04}$. After normalization, we obtain $(0 : 0 : 0.222 : 0.778)$. Eventually, these values of the likelihood ratio are considered as a virtual evidence; we obtain $P(C = yes \mid se) = 0.783$.

The comparison of the values of $P(C = yes \mid ve) = 0,881$ and $P(C = yes \mid se) = 0,783$ obtained in our example can be explained as follows. The soft evidence $(0, 0, 0.5, 0.5)$ has been converted into the virtual evidence $(0, 0, 0.222, 0.778)$ by using the likelihood ratio. Thus, the probability 0.222 associated to S_3 is less than the initial probability 0.5, in order to compensate the influence of the prior distribution over S, in which S_3 is more probable than S_4 (see table 1). Since S_3 leads to congestion with a higher probability than S_4, updating the VE $(0, 0, 0.5, 0.5)$ leads to a higher probability of congestion than updating the ve $(0, 0, 0.222, 0.778)$.

4 Fuzzy Evidence

In this section, we consider the proposition presented in [MM11] to model fuzzy evidence by using the fuzzy logic theory. In that aim, we consider the BN of the Fig. 1 in which the node S is replaced by the node S^f whose possible states are representing the amount of snow in natural language: S_1^f =*'not at all'*, S_2^f =*'a little'*, S_3^f =*'some'* and S_4^f =*'a lot'*. We substitute vagueness by a membership degree and we model the relationship between the amount of snow and the linguistics states of snowing by the fuzzy sets shown in Fig. 3. The CPT

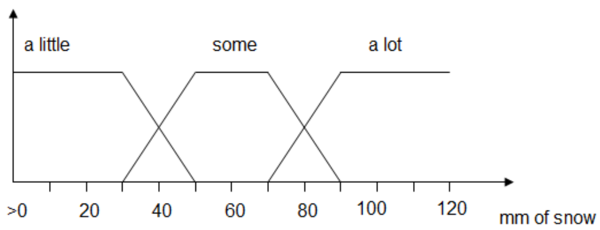

Fig. 3. Fuzzy sets of S^f.

of $P(C \mid S^f)$ is given in Table 8. In order to be consistent with the previous example, it has been computed according to equation 1 and Fig. 3. The algorithm used to propagate ambiguous observations is based on the junction tree algorithm. Assume that the amount of snow is 80 mm. The fuzzy evidence fe can be expressed thanks to the following membership degrees:

Table 8. $P(C \mid S^f)$

	S_1^f	S_2^f	S_3^f	S_4^f
$P(C = yes \mid S_i^f)$	0.3247	0.67	0.9137	0.6707

$$\mu_1(80) = 0,\ \mu_2(80) = 0,\ \mu_3(80) = 0.5,\ \mu_4(80) = 0.5.$$

Table 9 shows the likelihood encoding of the fuzzy evidence fe. It is used in the junction tree algorithm as in the previous section.

Table 9. Likelihood encoding fe

	S				C	
v	S_1^f	S_2^f	S_3^f	S_4^f	yes	no
$\Lambda_V(v)$	0	0	0.5	0.5	1	1

The result is obtained by:

$$P(C = yes \mid fe) = \frac{\sum_i P(C = yes \mid S_i^f) \times \mu_i(80)}{\sum_{c,i} P(C = c \mid S_i^f) \times \mu_i(80)} \tag{3}$$

This method applied to the snow example provides $P(C = yes \mid fe) = 0.792$.

The important distinction between uncertain evidence and fuzzy evidence is that with a fuzzy evidence, there is no uncertainty about the value *'not at all'*, *'a little'*, *'some'* and *'a lot'* of the snowfall but rather an ambiguity about the degree to which a value matches the category *'not at all'*, *'a little'*, *'some'* and *'a lot'*. This ambiguity is treated by fuzzy sets (Fig. 3).

5 Fuzzy Reasoning in Bayesian Networks

In this section, we introduce fuzzy Bayesian equations, as down in [TL07]. According to Zadeh's definition [Za68], the probability of a fuzzy event $\tilde{A}$ in X is given by

$$P(\tilde{A}) = \sum_{x \in X} \mu_{\tilde{A}}(x) \times P(x) \tag{4}$$

In our context, X denotes a variable from the Bayesian network and x is one of its value. $\mu_{\tilde{A}}$ is the membership function of $\tilde{A}$ and $\mu_{\tilde{A}}(x)$ is the grade of membership of x into $\tilde{A}$.

We present fuzzy Bayesian equations through our snow example. The Bayesian network considered is described by Fig. 1 and Tables 3 and 4. We are interested in fuzzy events consisted of S_i; for example the fuzzy event $\widetilde{S}$ = *'some'* (see Fig. 3) is described by the membership function: $\mu_{\tilde{S}}(S_1) = 0$, $\mu_{\tilde{S}}(S_2) = 0.0625$, $\mu_{\tilde{S}}(S_3) = 0.9375$, $\mu_{\tilde{S}}(S_4) = 0.0625$. In this case, following eq. 11 in [TL07], Bayesian equation is

$$P(C = yes \mid \tilde{S}) = \sum_{i \in I} \mu_{\tilde{S}}(S_i) \times P(S = S_i \mid C = yes) \times P(C = yes)/P(\tilde{S}) \tag{5}$$

The marginal fuzzy probability is $P(\tilde{S}) = \sum_{i \in I} \mu_{\tilde{S}}(S_i) \times P(S = S_i)$ (see eq. 12 in [TL07]) Thus we can figure out $P(C = yes \mid \tilde{S}) = 0.92$.

We may also be interested in fuzzy events of C_j, for example $\widetilde{C} =$ *'little congestion'*. Then, following eq. 10 in [TL07], Bayesian equation is

$$P(\tilde{C} \mid S = S_i) = \sum_j \mu_{\tilde{C}}(C_j) \times P(S = S_i \mid C = C_j) \times P(C = C_j)/P(S = S_i) \quad (6)$$

If $\tilde{S}$ and $\tilde{C}$ are both fuzzy events, then we have (see eq. 13 in [TL07])

$$P(C = \tilde{C} \mid S = \tilde{S}) = \sum_{j \in J} \sum_{i \in I} \mu_{\tilde{C}}(C_j) \times \mu_{\tilde{S}}(S_i) \times P(S = S_i \mid C = C_j) \times P(C = C_j)/P(\tilde{S}) \quad (7)$$

The advantage of this method is threefold. First, we can insert fuzzy observation as shown in (5). Second, we can calculate the probability of a fuzzy event as shown in (6). Finally, we can calculate the probability of a fuzzy event conditional to another fuzzy event as shown in (7).

6 Conclusion

This paper considers the problem of reasoning with *uncertain evidence* in Bayesian Networks. The use of *uncertain evidence* significantly extends the power of Bayesian networks since it is needed in lots of real applications. A key contribution of our work is to describe and to explain the different ways of modeling and updating uncertain evidences. We also propose so-called *fuzzy evidence* which are pertinent for ambiguous observations. Our comparison between *hard evidence*, *virtual evidence*, *soft evidence* and *fuzzy evidence*, showed that each kind of evidence is adequate for a specific context.

Even if the term "soft evidence" has been used in a confusing way in the literature, it is clear that virtual evidence reflects an observation *with* uncertainty whereas soft evidence expresses an observation *of* uncertainty.

This throw light on the posterior probability of the observed node which not change in the case of soft evidence because we are certain about this probability distribution, but it changes in virtual evidence because we are uncertain of the probability distribution which is thus modified by the marginalization process. Concerning fuzzy evidence, the observation can belong in the same time to more than one class (80 mm is considered in the same time as 'some' and 'a lot' snowfall with membership degrees).

In the last section, we presented fuzzy reasoning in Bayesian network. This method allows to insert fuzzy evidence, to calculate the probability of fuzzy event, and to calculate the probability of fuzzy event conditional to a fuzzy observation.

References

[MM11] Ben Mrad, A., Maalej, M. A., Delcroix, V., Piechowiak,S., Abid, M.: Fuzzy Evidence in Bayesian Networks. Proc. of Soft Computing and Pattern Recognition, Dalian, China, (2011).

[Bo89] Bock,H. H.: A conditional iterative proportional fitting (CIPF) algorithm with applications in the statistical analysis of discrete spatial data. Bull. ISI, Contributed papers of 47th Session in Paris. 1, 141–142 (1989).

[CD05] Chan, H. Darwiche, A.: On the revision of probabilistic beliefs using uncertain evidence. Artificial Intelligence. 163, 67–90 (2005).

[Cr00] Cramer, E.: Probability measures with given marginals and conditionals: I-projections and conditional iterative proportional fitting. Statistics and Decisions. 18, 311–329 (2000).

[Cs75] Csiszar, I.: I-divergence geometry of probability distributions and minimization problems. The Annuals of Probability. 3(1), 146–158 (1975).

[DB08] Di Tomaso, E., Baldwin, J. F.: An approach to hybrid probabilistic models. International Journal of Approximate Reasoning. 47(2), 202–218, (2008).

[Fi70] Fienberg, S. E.: An iterative procedure for estimation in contingency tables. Ann. Math. Statist. 41(3), 907-917 (1970).

[HD96] Huang, C. , Darwiche, A.: Inference in belief networks: A procedural guide. International Journal of Approximate Reasoning 15(3), 225–263 (1996).

[Je83] Jeffrey, R.: The Logic of Decisions. 2nd Edition, University of Chicago Press, Chicago, (1983).

[Je96] Jensen, F.: An Introduction to Bayesian Networks, Springer, New York, (1996).

[Je90] Jensen, F. , Lauritzen, S., Olesen, K.: Bayesian updating in recursive graphical models by local computations. Computational Statistical Quaterly, 4, 269–282, (1990).

[Kr37] R. Kruithof, R.: Telefoonverkeer Rekening. De Ingenieur, 52, 15–25 (1937).

[La88] Lauritzen, S.: Local computation with probabilities on graphical structures and their application to expert systems. Journal of the Royal Statistical Society. 50:157, (1988).

[Ne] http://www.norsys.com

[Pe88] Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufman, San Mateo, CA. (1988).

[Pe90] Pearl, J.: Jeffrey's rule, passage of experience, and neo-Bayesianism. Knowledge Representation and Defeasible Reasoning. eds. H. E. Kyburg, Jr., R. P. Loui and G. N. Carlson, Kluwer Academic Publishers, Boston, 245–265 (1990).

[PD05] Peng, Y., Ding, Z.: Modifying Bayesian networks by probability constraints. in Proc. 21st Conf. Uncertainty in Artificial Intelligence, Edinburgh, (2005).

[PZ10] Peng,Y., Zhang, S., Pan, R.: Bayesian Network Reasoning with Uncertain Evidences. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. 18(5), 539–564 (2010).

[TL07] Tang, H. and Liu, S.: Basic Theory of Fuzzy Bayesian Networks and Its Application in Machinery Fault Diagnosis. In Proc. of FSKD. 4, 132-137 (2007).

[TA04] Tu, H., Allanach, J., Singh,S., Pattipati K. R., Willett,P.: The adaptive safety analysis and monitoring system. Proc. SPIE 5403, 153 (2004).

[VK02] Valtorta, M., Kim, Y., Vomlel, J.: Soft Evidential Update for Probabilistic Multiagent Systems. International Journal of Approximate Reasoning. 29(1), 71–106, (2002).

[Za68] Zadeh, L.A.: Probability measures of fuzzy events, Journal of Mathematical Analysis and Applications 23, 421–427 (1968).