Understanding soft evidence as probabilistic evidence : illustration with several use cases

Ali Ben Mrad
University of Sfax, Tunisia
ENIS, CES, Sfax, Tunisia
LAMIH UMR CNRS 8201
Université de Valenciennes et du Hainaut Cambrésis
59313 Valenciennes FRANCE
Email: BenMradAli2@gmail.com

Véronique Delcroix
LAMIH UMR CNRS 8201
Université de Valenciennes et du Hainaut Cambrésis
59313 Valenciennes FRANCE
Email: Veronique.Delcroix@univ-valenciennes.fr

Sylvain Piechowiak
LAMIH UMR CNRS 8201
Université de Valenciennes et du Hainaut Cambrésis
59313 Valenciennes FRANCE
Email: Sylvain.Piechowiak@univ-valenciennes.fr

Mohamed Amine Maalej
University of Sfax, Tunisia
ENIS, CES, Sfax, Tunisia
Email: m.maalej@gmail.com

Mohamed Abid
University of Sfax, Tunisia
ENIS, CES, Sfax, Tunisia
Email: mohamed.abid_ces@yahoo.fr

***Abstract*—This paper aims to get a better understanding of the notions of evidence, probabilistic evidence and likelihood evidence in Bayesian Networks. Evidence comes from an observation of one or several variables. Soft evidence is probabilistic evidence, since the observation consists in a local probability distribution on a subset of variables that has to replace any former belief on these variables. It has to be clearly distinguished from likelihood evidence, also called virtual evidence, for which the evidence is specified as a likelihood ratio. Since the notion of soft evidence is not yet widely understood, most of the Bayesian Networks engines do not propose related propagation functions and the terms used to describe such evidence are not stabilised. First, we present the different types of evidence on a simple example with an illustrative context. Then, we discuss the understanding of both notions in terms of knowledge and observation. Next, we propose to use soft evidence to represent certain evidence on a continuous variable, after fuzzy discretization.**

***keywords*: Bayesian networks, uncertain evidence, virtual evidence, likelihood evidence, soft evidence, probabilistic evidence.**

I. Introduction

Probabilistic inference in Bayesian network (BN) is the computation of posterior probability distribution in order to update the belief about some target variables after receiving evidence. In a BN, *evidence* is a collection of findings, and a *finding* (or observation) on a variable X of the BN is represented by a vector of 0s and 1s of length k, with $\{x_1, x_2, \dots, x_k\}$ being the set of states of X [1]. For example the vector $(1, 0, 0, \dots, 0, 1)$ represents the statement that X can be in state x_1 or in state x_k, and cannot be in other states. A finding is the statement that some states of the observed variable are impossible. The terms *hard finding* and *negative finding* are also used in literature. More commonly, a finding on the variable X (or *positive finding* or *specific finding*) is often understood as an instantiation $X = x_i$, represented by a vector with a single 1, at the i^{th} position.

978-1-4673-5814-9/13/$31.00 ©2013 IEEE

For 25 years, numbers of papers have dealt with belief updating after receiving uncertain evidence, represented by vectors of real values [2]–[11]. They concern both the specification of uncertain evidence and their propagation.

Likelihood evidence (also called *virtual evidence*) refers to Pearl's (1988) idea of interpreting uncertain evidence on a set of events as a hard evidence on some virtual events that only depend on this set of events. The uncertainty of evidence is specified in the conditional probability table of the added virtual node.

The term *soft evidence* is employed in two different meanings depending on the authors, leading to a confusion. A clear distinction between likelihood evidence and soft evidence is presented in [4], however, the term *soft evidence* is frequently used as a synonym of likelihood evidence. For example, most BN engines support Pearl's virtual evidence method, but some of them use the term soft evidence.

Soft evidence can be seen in many places. For example, one may not be able to observe the precise state of a variable but may know its distribution in a particular situation; meaning that this is not the marginal probability distribution embedded in the BN. Also sometimes it is more important to know the distribution of a variable than its precise state. When two BNs interact with each other, the information exchanged between them is often in the form of probability distributions of shared variables (see [12], [13]). In this paper we focus on the problem of belief update from non-deterministic information on one (or more) variables of a given model. This kind of information cannot be represented by the instantiation of one or more variables in the model. Our aim is to make clear how to specifiy and propagate likelihood evidence and soft evidence, but not to revise the model, i.e. changing the structure or parameters of BN.

Despite numerous publications on soft evidence [4]–[8], [10], [13]–[16], the majority of BN engines does not provide algorithms to propagate soft evidence.

The rest of the paper is organized as follows. In Section 2, we explain the distinction between virtual evidence and soft evidence. Then, we present a fairly complete review of the literature on the topic. Section 3 is devoted to present a simple and fully worked out example. In section 4, we propose to use soft evidence to represent certain evidence on a continuous variable, after fuzzy discretization (see [16]). Section 5 contains a discussion of the difference between model revision and soft evidential update.

II. Uncertain evidence: definition and vocabulary

Bayesian networks [1], [2], [14], [17] are widely used in decision-making, mainly for problems with uncertainty.

Now, we present in details the definitions of different types of uncertain evidence, both in terms of specification and of propagation. In the following, variables are denoted by upper case letters; a state of a variable is denoted by a lower case letter; a set of variables is denoted by an upper case boldface letter, and a configuration of states over a set $\mathbf{X}$ is denoted by $\mathbf{x}$, a lower case boldface letter.

The virtual evidence method, proposed by Pearl in [2], provides a convenient way of incorporating evidence *with* uncertainty. This kind of uncertain evidence is also called *likelihood evidence* as it is specified by a likelihood ratio $L(X)$ representing the strength of confidence toward the observed event. Likelihood ratio $L(X)$ is defined as

$$L(X) = (P(Obs \mid x_1) : \ldots : P(Obs \mid x_n))$$

$P(Obs \mid x_i)$ is interpreted as the probability of the observed state given X is in the state x_i.

Pearl's method to propagate such evidence extends the given BN by adding a binary virtual node which is a child of X. The uncertain evidence on X is replaced by a certain evidence on the added node. The hard evidence on the added node is propagated using a classical inference algorithm in BN. The uncertainty of evidence is specified in the conditional probability table of the added virtual node.

Soft evidence (SE), following Valtorta in [4], can be interpreted as evidence *of* uncertainty. SE is given as a probability distribution $R(\mathbf{Y})$ of one or several variables $\mathbf{Y} \subseteq \mathbf{X}$, where $\mathbf{X}$ is the set of variables of the model. Therefore, there is uncertainty about the specific state $\mathbf{Y}$ is in but we are sure of the probability distribution $R(\mathbf{Y})$. As $R(\mathbf{Y})$ is a certain observation, this distribution should be preserved when updating belief. This is the main difference with likelihood evidence for which this is not required. Another important difference between likelihood ratio and SE is that the first one is given by L(X) that is not a probability distribution whereas the second is given by a probability distribution. Because of these characteristics, soft evidence is a *strong probabilistic evidence*.

Propagating SE requires updating a probability distribution $P(\mathbf{X})$ by another lower dimensional probability distribution $R(\mathbf{Y})$ where $\mathbf{Y} \subseteq \mathbf{X}$. As mentionned by Peng [8], the difficulty stems from the fact that the Bayes' rule cannot be directly applied here because $R(\mathbf{Y})$, although acting as a condition for the update, is not itself an event.

The approach proposed by Jeffrey in [3] for this problem is known as "probability kinematics"; it is based on the requirements that:

1) the posterior distribution $Q(\mathbf{Y})$ is unchanged: $Q(\mathbf{Y}) = R(\mathbf{Y})$,
2) the conditional probability of other variables given $\mathbf{Y}$ remain invariant under the observation: $Q(\mathbf{X}\setminus\mathbf{Y} \mid \mathbf{Y}) = P(\mathbf{X} \setminus \mathbf{Y} \mid \mathbf{Y})$.

In other words, even if P and Q disagree on $\mathbf{Y}$, they agree on the consequences of $\mathbf{Y}$ on other variables.

However, Jeffrey's rule cannot be directly applied to BNs, because their operations are defined on full joint probability distributions. This can be overcome by converting a soft evidence to a likelihood evidence as suggested in [6]. $R(\mathbf{Y})$ can be converted to a likelihood ratio

$$L(\mathbf{Y}) = \frac{R(\mathbf{y_1})}{P(\mathbf{y_1})} : \ldots : \frac{R(\mathbf{y_n})}{P(\mathbf{y_n})}.$$

The authors show that applying this likelihood evidence to $P(\mathbf{X})$ provides the same results as Jeffrey's rule. Jeffrey's rule specifies evidence using posterior probabilities, while Pearl's method specifies evidence using likelihood ratios. Likelihood evidence is a way to specify uncertain evidence without taking into account the belief on the variable before receiving this evidence. It is *uncertain evidence without a priori*.

Another difficulty happens when considering several pieces of uncertain evidence. Chan and Darwiche raised the question [6] "Should, and do, iterated belief revisions commute?". Iterated revisions commute with Pearl's method, but not with Jeffrey's rule. In case of multiple pieces of soft evidence, converting each of them into likelihood ratios and using Pearl's method lead to different results regarding the order of findings. The initial probability distributions are not all preserved by updating.

Some authors claim that when several pieces of evidence carrying the "All things considered" interpretation must not be commutative [6], [18]. Others argue that soft evidences are true observations of distributions of some events, and as such, they all should be preserved in the updated posterior distribution [8]. If the belief on $\mathbf{Y}$'s states, represented by $R(\mathbf{Y})$, is susceptible to be improved by new evidence, then we can speak of *weak probabilistic evidence* and use Jeffrey's rule for their propagation. Else, soft evidence is a *strong probabilistic evidence*.

The unified approach proposed in [4] handles the several types of soft evidence by introducing observation variables: soft evidence can be given by marginal probability distributions defined for values of a single variable, or on the

Cartesian product of the values of several variables, conditional probability distributions, or probability assignments to a logical or probabilistic function. Several algorithms have been proposed to propagate Soft evidence in a BN: Big clique [4], [12], BN-IPFP1 and BN-IPFP2 [7], lazy big clique [19] and SMOOTH [20], that accept inconsistent soft evidence. Another series of algorithms have been proposed in order to revise the parameters of a BN after soft evidence: E-IPFP, E-IPFP-SMOOTH and D-IPFP [21]. In this problem, soft evidence is considered as new knowledge that has to be (definitely) included in the model.

III. Illustration on the Snow-Lateness example

This example takes place in a university campus located in a mountain city, on which, there is not enough way to have lunch. Hence, a large proportion of students leave the campus at noon to eat, either walking or driving. When it snows, walking and driving are more difficult: traffic is slowed down and the risk of falling when walking is higher as the streets of the campus are steep. Figure 1 shows the BN with the four variables Snow, High Risk of Falling, Slow Traffic and Lateness.

Snow (S)**:**
'$s1 =$ not at all' means there is no snow on campus,
'$s2 =$ a little' means that $0 <$ amount of snow ≤ 40,
'$s3 =$ some' means that $40 <$ amount of snow ≤ 80,
'$s4 =$ a lot' means that $80 <$ amount of snow

High Risk of Falling (F)**:** the risk of falling while travelling on foot is high (f) or normal ($\bar{f}$).

Normal Traffic (T)**:** $P(t)$ is the probability that a vehicle circulating on the campus has a normal speed.

Lateness (L)**:** $\bar{l}$ means that students are not late (what is normally found). The event l occurs when the number of late students exceeds a threshold.

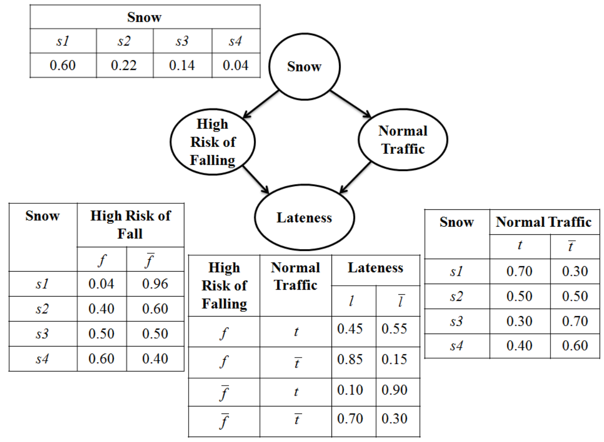
The conditional probability $P(\bar{t} \mid S = s_i)$ can be understood as follows: under the threshold of $60cm$ of snow, the Slow Traffic is all the more probable that the amount of snow is important. Beyond this threshold, people leave their homes less and less and subsequently the probability of Slow Traffic decreases.

The marginal probabilities are given below, because we refer to them later:

$P(S) = (0.6, 0.22, 0.14, 0.04)$
$P(F) = (0.206, 0.794)$
$P(T) = (0.588, 0.412)$
$P(L) = (0.396, 0.604)$

We denote respectively he_X, le_X and se_X the values of hard evidence, likelihood evidence and soft evidence on a variable X. Let's consider the following cases.

a) One piece of likelihood evidence: we consider, in this case, that the observation comes from a person who has only a partial observation of the traffic state on the campus that day, and who has no precise measure of traffic conditions. It is therefore an evidence with uncertainty and we do not require that it remains unchanged by the inference. The likelihood evidence le_T on a slow traffic is specified by the likelihood



Snow			
s1	*s2*	*s3*	*s4*
0.60	0.22	0.14	0.04

Snow	High Risk of Fall	
	f	$\bar{f}$
s1	0.04	0.96
s2	0.40	0.60
s3	0.50	0.50
s4	0.60	0.40

High Risk of Falling	Normal Traffic	Lateness	
		l	$\bar{l}$
f	t	0.45	0.55
f	$\bar{t}$	0.85	0.15
$\bar{f}$	t	0.10	0.90
$\bar{f}$	$\bar{t}$	0.70	0.30

Snow	Normal Traffic	
	t	$\bar{t}$
s1	0.70	0.30
s2	0.50	0.50
s3	0.30	0.70
s4	0.40	0.60

Figure 1. BN of the snow-lateness example.

ratio L(T) = (0.3 : 0.7), representing $P(obs \mid t) : P(obs \mid \bar{t})$. Thus, by applying Pear'l method of virtual evidence, we obtain:

$P(S \mid le_T) = (0.542, 0.237, 0.175, 0.046)$
$P(F \mid le_T) = (0.232, 0.768)$
$P(T \mid le_T) = (0.38, 0.62)$
$P(L \mid le_T) = (0.519, 0.481)$

The propagation of this uncertain observation illustrates that the likelihood ratio given on the traffic variable is modified after propagating this likelihood evidence : $L(T) = le_T$ is not equal to $P(T \mid le_T)$. on the slow traffic variable is modified after inference.

b) One piece of soft evidence: let us now try to better understand the meaning of soft evidence with a simple example. Recall that the event $\bar{t}$ means that a vehicle, circulating on campus, has an average speed less than normal. Imagine that ten speed sensors are installed throughout the campus. We assume that it is a temporary installation (the sensors are installed temporarily). The problem consists in propagating information given by the sensors into the existing model without changing it (neither the structure nor the parameters of the BN). For each, the average speed recorded on the last hour is considered. We can reasonably assume that if all sensors that traffic is slowed down on their measuring point, then $P(\bar{t}) = 1$. If the traffic is uniformly distributed on campus, we can assume that the number n_{normal} of sensors that indicate a normal traffic is a good measure of the probability that the traffic is normal on campus, i.e. $P(t) = n_{normal}/10$.

In this context, assume that 7 speed sensors indicate that traffic is slowed down, and 3 indicate a normal speed. This observation is certain. We denote this soft evidence se_T represented by the probability distribution $R(T) = (0.3, 0.7)$. As we have an absolute confidence in the result given by the speed sensors, no further evidence can improve this observation. The observed probability distribution on T should not be modified neither by inference of this observation, nor by any observation on other variables. For this reason, the distribution $R(T)$ has to replace the marginal distribution

$P(T)$ of the observed variable. The need to preserve the observed distribution in the case of soft evidence indicates that soft evidence should not be propagated as likelihood evidence. However, as explained above with jeffrey's rule, the virtual evidence method can be used to propagate a single piece of soft evidence, by making a simple transformation from soft evidence to likelihood evidence.

This is illustrated below. First, we compute the likelihood of T:

$L(T) = R(T)/P(T) = (0.3/0.588 : 0.7/0.412)$.

After normalization, $L(T) = (0.231 : 0.769)$; this is propagated by the virtual evidence method and we obtain:

$P(S \mid se_T) = (0.52, 0.243, 0.188, 0.049)$
$P(F \mid se_T) = (0.241, 0.759)$
$P(T \mid se_T) = (0.3, 0.7)$
$P(L \mid se_T) = (0.565, 0.435)$

The notation $P(X \mid se_T)$, while not rigourous since a soft evidence is not an event, is a notation to design the updated probability distribution after the observation of the soft evidence se_T.

By comparison, the notation $P(X \mid le_Y)$ is relevant since the likelihood evidence is propagated as a hard evidence (thus an event) on a virtual node).

Let's remark that the updated probability on T equals the observed probability distribution $R(T)$, as specified by soft evidence definition.

When there are several pieces of evidences including at least one piece of soft evidence, this technique does not preserve the probability distribution that specifies the soft evidence. This is illustrated in the following example.

We treat now the case of multiple pieces of evidences. First, we start with two pieces of likelihood evidence then we treat the case of two pieces of soft evidence, treated either simultaneously or successively.

c) Several pieces of likelihood evidence: successive or simultaneous propagation of several pieces of likelihood evidence gives the same results. Take the example of two uncertain findings: le_T is specified by the likelihood ratio $(0.3 : 0.7)$ and le_F is specified by the likelihood ratio $(0.7 : 0.3)$; we obtain:

$P(S \mid le_T, le_F) = (0.436, 0.277, 0.222, 0.064)$
$P(F \mid le_T, le_F) = (0.413, 0.587)$
$P(T \mid le_T, le_F) = (0.35, 0.65)$
$P(L \mid le_T, le_F) = (0.573, 0.427)$

These results illustrate that belief on slow traffic T is amended by the evidence (even uncertain) about the high risk of falling (F), despite the fact that evidence with uncertainty has been entered on the variable T. This is explained as follows: the observer who locally observed the traffic state revises his judgement after a second (even uncertain) observation that the risk of falling seems higher than usual.

d) Several pieces of soft evidence: take the example of two pieces of soft evidence, se_T and se_F, specified by the probability distribution $R(T) = (0.3, 0.7)$ and $R(F) = (0.7, 0.3)$. To satisfy both se_T and se_F, the updated distribution Q has to preserve $Q(T) = R(T)$ and $Q(F) = R(F)$.

Table I
BELIEF UPDATE ON BN OF THE SNOW-LATENESS EXAMPLE WITH TWO PIECES OF SOFT EVIDENCE, (USING THE CONVERSION INTO LIKELIHOOD EVIDENCE).

Evidence	posterior Belief on f	posterior Belief on t
se_T and se_F	0.741	0.23
se_T then se_F	0.741	0.23
se_T then se'_F	0.7	0.236
se_F then se'_T	0.73	0.3

Table I shows the posterior probability of the two variables (T) and (F) obtained by propagating simultaneously or successively the two pieces of soft evidence se_T and se_F by using the transformation of a soft evidence into a likelihood evidence and then using Pearl's virtual evidence method. Rows 3 and 4 of Table I show the result after recalculating a new likelihood ratio for the second piece of soft evidence using the distribution updated by the first one, as proposed in [8].

Thus, in all cases, the propagation alters at least one value of the two pieces of soft evidence. This illustrates the need of specific algorithms to update a probability distribution from several pieces of soft evidence. In that purpose, several algorithms have been proposed [4], [8], [15].

IV. DISCUSSION ABOUT SOFT EVIDENCE: KNOWLEDGE OR OBSERVATION

Since a soft evidence $R(X)$ has to be preserved by the propagation, it is not combined with the prior probability $P(X)$ during the process of inference. The soft evidence $R(X)$ is considered to replace $P(X)$ and thus we need to define a new probability distribution Q over $\mathcal{X}$. This probability distribution is such that $Q(X) = R(X)$ and the inference algorithm try to minimize the distance between P and Q. Therefore, considering a soft evidence se_X about the variable X and another variable $Y \in \mathcal{X}$, the notation $P(Y|se_X)$ is incorrect and should be replaced by the notation $Q(Y)$.

Thus we can maintain the following statements:

- a likelihood evidence on a variable does not cast doubt on the joint probability distribution. It does not constitute a knowledge contribution.
- a soft evidence casts doubt on the joint probability distribution since it replaces the marginal probability distribution over the concerned variable.

As a consequence, should we consider that a soft evidence is a knowledge contribution ? In that case, why not changing the model rather than propagating the soft evidence ? We now discuss that question.

The problem of model revising (graph and / or probabilities) and the problem of belief update are two different problems. The addressed problem in this paper concerns the belief update in Bayesian network, and not the model revision, which lead to change the graph structure or parameters of the model. In the case of soft evidence, information is probabilistic in nature and leads to replace (temporarily) the a priori existing in the model by the provided information. Revising the model leads to change the BN when new (probabilistic) knowledge on one (or more) variable has to be definitely integrated in the model.

Coin example: when the coin falls in a dark place, with difficult access, I cannot see it correctly and my observation is uncertain. It is a likelihood evidence that do not modify the prior distribution on the coin, for example $(0.5, 0.5)$ when the coin is normal. On the other hand, if I know that the coin has been exchanged with a rigged coin with $P(side = head) = 0.7$, this is a soft evidence since I'm sure of that distribution. In such a situation, why not modifying the prior probability of the coin in the model, or adding a parent node to the node coin, in order to represent the type of coin (normal, rigged with 0.7 for head, etc.).

No model is complete. In this paper, we address the problem of propagating information in BN about a variable that is not in the BN. A model is always a compromise: on the one hand, over model parameter (for example the type of piece fake or not) and therefore obtain a heavier model thus a less efficient model; on the other hand, choosing a smaller model, less detailed, but more efficient.

Snow and lateness example: we consider the model presented above about the traffic on a campus. The soft evidence over the traffic state given by the ten speed sensors could be taken into account by adding ten nodes representing the sensor measures in different places of the campus. The variable representing the state of Traffic T should be a common child of these nodes. Furthermore, the influence of the snow (S) on the state of the traffic (T) has to be reconsidered. Another way to model that situation is to consider that we have two distinct models: the first one concerns the way the snow influences lateness on the campus; the second one represents a kind of expert of the traffic on the campus, represented by the speed sensors. The expert is providing the soft evidence about the state of the traffic (T). The modification of the initial Bayesian network in order to integrate the soft evidence requires to integrate the model of the *expert* in the first model which is generally not an easy task. All the more so that the expert's model can change.

Considering a distinct model for any soft evidence is the approach proposed in AEBN [12], [13]. This way to consider a soft evidence clearly dissuades from modifying the initial BN.

These examples aims to illustrate that managing soft evidence cannot be reduced to modifying the model.

V. Soft evidence for continuous variables

A common way of making a variable which has a continuous value more comprehensible is to discretise it. For example, the amount of snow may be discretised into not at all, a little, some and a lot. This is known as a hard discretization. One problem with hard discretization is that all the points of an interval are considered as if they were the same discrete value. Thus, they are treated in the same way whether they are in the centre or on the edges of the interval. Consider the model presented above about the traffic on a campus. On the one hand, all amount of snow less than $40cm$ and not null may be discretised as a little. This discretization does not distinguish between the amount of snow $2cm$ and $40cm$, thus,

$P(l|S = 2cm) = P(l|S = 40cm) = P(l|S = s2) = 0.5$.

On the other hand, the amount of snow $40cm$ and $41cm$ are treated differently, thus,

$P(l|S = 41cm) = P(l|S = s3) = 0.625$.

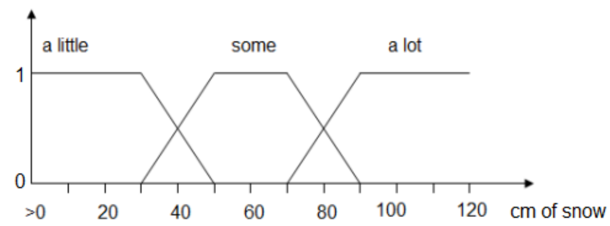


Figure 2. Membership function of S.

For this reason slightly different ways of choosing the partition, can give very different results. The set of thresholds may be supplied by an expert. Techniques have been applied to find the optimal set of thresholds while inducing the model from data but they can become computationally very expensive when applied to large databases [22]. Soft evidence is a type of uncertain evidence that occurs when uncertainty is represented through probability. When a continuous variable is discretized in a BN, the choice of the number of intervals is sometimes a difficult compromise: many intervals can generate a large CPT, especially if the variable has several parent nodes, or if the child nodes have others parents. Conversely, a coarse discretization leads to a loss of precision. One solution consists in choosing a small number of intervals, and using soft evidence to represent certain evidence on a continuous variable, after fuzzy discretization [16].

The probability distribution of soft evidence can be extracted from the relationship between the amount of snow and the states of the variable Snow (see Figure 2).

Thus,
$P(l|S = 40cm) = P(l|S = se_S) = 0.562$
such as $R(S) = (0, 0.5, 0.5, 0)$.

The appropriateness of the use of soft evidence rather than likelihood evidence is that likelihood evidence represents a subjective statement that can be improved by something observed later, while soft evidence represents an observation that cannot be improved by anything observed later.

VI. Discussion and conclusion

Likelihood evidence does not cast doubt on the probability distribution. It does not constitute a knowledge contribution. Soft evidence casts doubt on the probability distribution as it replaces the marginal probability distribution over the concerned variables.

As a consequence, should we consider that soft evidence is a knowledge contribution ? In that case, why not changing the model rather than propagating the soft evidence ? Let's examine a situation where the modification of the model to manage soft evidence is not an appropriate solution. We consider the example of "Snow and lateness" presented above. The soft evidence given by the ten speed sensors could be taken into account by adding ten nodes representing the sensor measures. The variable representing the state of Traffic T should be a common parent of these nodes. In cases the information provided by the sensors is available only temporary, modifying the structure of the graph is not an appropriate solution.

Another way to model that situation is to consider that we have two distinct models: the first one concerns the way the snow influences lateness on the campus; the second one represents a kind of expert of the traffic on the campus, represented by the speed sensors. The expert is providing soft evidence about the state of the traffic (T). Considering a distinct model for any piece of soft evidence is the approach proposed in AEBN [12], [13]. The modification of the initial Bayesian network in order to integrate the soft evidence requires to integrate the model of the *expert* in the initial model.

Another example of use of soft evidence concerns fuzzy discretization. Thus, soft evidence can be used to represent certain evidence on a continuous variable.

These examples aim to illustrate that managing soft evidence cannot be reduced to modifying the model.

References

[1] F. V. Jensen and T. D. Nielsen, *Bayesian Networks and Decision Graphs*, 2nd ed. Springer Publishing Company, Incorporated, 2007.

[2] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1988.

[3] R. C. Jeffrey, *The Logic of Decision*. (2nd edition) University of Chicago Press, 1990, 246 pages.

[4] M. Valtorta, Y.-G. Kim, and J. Vomlel, “Soft evidential update for probabilistic multiagent systems,” *International Journal of Approximate Reasoning*, vol. 29, no. 1, pp. 71–106, 2002.

[5] J. Vomlel, “Probabilistic reasoning with uncertain evidence,” *Neural Network World, International Journal on Neural and Mass-Parallel Computing and Information Systems*, vol. 14, no. 5, pp. 453–465, 2004.

[6] H. Chan and A. Darwiche, “On the revision of probabilistic beliefs using uncertain evidence,” *Artificial Intelligence*, vol. 163, no. 1, pp. 67–90, 2005.

[7] R. Pan, Y. Peng, and Z. Ding, “Belief update in bayesian networks using uncertain evidence,” in *ICTAI*, 2006, pp. 441–444.

[8] Y. Peng, S. Zhang, and R. Pan, “Bayesian network reasoning with uncertain evidences,” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 18, no. 5, pp. 539–564, 2010.

[9] A. Ben Mrad, M. A. Maalej, V. Delcroix, S. Piechowiak, and M. Abid, “Fuzzy evidence in bayesian networks,” in *Proc. of Soft Computing and Pattern Recognition, Dalian, China,*, 2011.

[10] A. Ben Mrad, V. Delcroix, M. A. Maalej, S. Piechowiak, and M. Abid, “Uncertain evidence in bayesian networks : Presentation and comparison on a simple example,” in *Procceedings of the 14th Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU,2012*, July 2012, pp. 39–48.

[11] S. Benferhat and K. Tabia, “Inference in possibilistic network classifiers under uncertain observations,” *Annals of Mathematics and Artificial Intelligence*, pp. 1–41, 2012.

[12] M. Bloemeke, “Agent-encapsulated bayesian networks,” Ph.D. dissertation, Columbia, SC, USA, 1998, aAI9918909.

[13] S. Langevin, M. Valtorta, and M. Bloemeke, “Agent-encapsulated bayesian networks and the rumor problem,” in *AAMAS ’10 Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems*, vol. 1, 2010, pp. 1553–1554.

[14] A. Darwiche, *Modeling and Reasoning with Bayesian Networks*. Cambridge University Press, 2009.

[15] Y. Peng and Z. Ding, “Modifying bayesian networks by probability constraints,” in *UAI*, 2005, pp. 459–466.

[16] E. D. Tomaso and J. F. Baldwin, “An approach to hybrid probabilistic models,” *International Journal of Approximate Reasoning*, vol. 47, no. 2, pp. 202–218, 2008.

[17] K. Korb and A. Nicholson, *Bayesian Artificial Intelligence*, 2nd ed. Chapman and Hall, 2010.

[18] C. G. Wagner, “Probability kinematics and commutativity,” *Philosophy of Science*, vol. 69, no. 2, pp. 266–278, 2002.

[19] S. Langevin and M. Valtorta, “Performance evaluation of algorithms for soft evidential update in bayesian networks: First results,” in *SUM*, 2008, pp. 284–297.

[20] S. Zhang, Y. Peng, and X. Wang, “An efficient method for probabilistic knowledge integration,” in *Tools with Artificial Intelligence, 2008. ICTAI’08. 20th IEEE International Conference on*, vol. 2. IEEE, 2008, pp. 179–182.

[21] Y. Peng, Z. Ding, S. Zhang, and P. Rong, “Bayesian network revision with probabilistic constraints,” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 20, no. 03, pp. 317–337, 2012.

[22] N. Friedman and M. Goldszmidt, “Discretizing continuous attributes while learning bayesian networks,” in *In Proc. ICML*. Morgan Kaufmann, 1996, pp. 157–165.