On the use of min-based revision under uncertain evidence for possibilistic classifiers

Salem Benferhat[1] Karim Tabia[1]

1.CRIL CNRS UMR8188
Artois University, France
Email: benferhat@cril.univ-artois.fr, tabia@cril.univ-artois.fr

Abstract— *Possibilitic networks, which are compact representations of possibility distributions, are powerful tools for representing and reasoning with uncertain and incomplete knowledge. According to the operator conditioning is based on, there are two possibilistic settings: quantitative and qualitative. This paper deals with qualitative possibilistic network classifiers under uncertain inputs. More precisely, we first present and analyze Jeffrey's rule for revising possibility distributions by uncertain observations in the qualitative setting. Then, we propose an efficient algorithm for revising possibility distributions encoded by naive possibilistic networks for classification purposes. This algorithm consists in a series of efficient and equivalent transformations of initial naive possibilistic classifiers.*

Keywords— Min-based possibilistic networks, classification under uncertain inputs

1 Introduction

Graphical models such as Bayesian networks [1][2], influence diagrams and possibilistic networks [3][4] are well-known formalisms widely used for representing and reasoning with uncertain and incomplete knowledge. Possibilistic networks are like Bayesian ones but lie on possibility theory [5][6] to handle imprecise and incomplete knowledge. They allow to factor a global joint possibility distribution into a set of local possibility distributions. There are two types of possibilistic networks: quantitative and qualitative. In the former, conditioning is based on the $product$ operator while it is based on the min-operator in the latter.

Classification is an important task in many real world applications consisting in predicting the class instance corresponding to an observation. This task is a special kind of inference: given an observed instance of each observable variable A_i, it is required to determine the class instance c_k for the instance to classify among a predefined set of class labels. It is important to note that possibilistic classifiers have not been sufficiently studied in spite of the fact that they are very appropriate for problems where knowledge is imprecise or missing as in real-time classification problems or problems where some inputs are missing or uncertain. These problems require classifiers under uncertain inputs. However, only few works used possibilistic classifiers [7][8][9][10] and to the best of our knowledge, there is only one preliminary work [11] addressing possibilistic network classifiers under uncertain inputs in the quantitative setting. More precisely, this work proposes an algorithm suitable of classification under uncertain inputs using product-based possibilistic networks.

This paper also addresses possibilistic-based classification with uncertain observations but in the qualitative setting. While quantitative possibilistic networks are quite similar to the Bayesian ones, the qualitative ones show significant differences. For instance, normalizing conditional possibility distributions in product-based possibilistic networks is the same as in the probabilistic setting while it is significantly different in min-based networks. We first recall the qualitative possibilistic counterpart of Jeffrey's rule [12] for revising possibility distributions by uncertain observations and show that this rule cannot be directly applied for revising possibilistic knowledge encoded by a possibilistic network. Indeed, classification by directly using the possibilistic counterpart of Jeffrey's rule is exponential in the number of attributes and attribute domains. Then, we proposed an efficient method for revising naive min-based possibilistic networks suitable for classification with uncertain inputs. This algorithm is based on a series of equivalent and polynomial transformations of initial possibilistic networks taking into account uncertain inputs.

The rest of this paper is organized as follows: Section 2 briefly presents basic background about possibility theory and possibilistic networks. In section 3, we address possibilistic belief revision based on the possibilistic counterpart of Jeffrey's rule. Section 4 proposes a new efficient algorithm for naive possibilistic network classification with uncertain inputs. Finally, section 5 concludes this paper.

2 Basic background on possibility theory and possibilistic networks

Let us first fix our notations: V=$\{A_1, A_2, .., A_n\}$ denotes the set of variables. D_A=$\{a_1, a_2, .., a_m\}$ denotes the finite domain of variable A. a_i denotes an instance (value) of variable A_i. A, X,.. denote subsets of variables from V. D_X=$\times_{A_i \in X} D_{A_i}$ represents the cartesian product relative to variables A_i involved in subset X. Ω=$\times_{A_i \in V} D_{A_i}$ denotes the universe of discourse and consists in the cartesian product of all variable domains involved in V. A tuple w=$(a_1, a_2, .., a_n)$ or w=$a_1 a_2..a_n$ which is an instance of Ω represents a possible state of the world. ϕ, φ denote subsets of Ω called events while $\overline{\phi}$ denotes the complementary of ϕ in Ω ($\overline{\phi}$=Ω-ϕ).

2.1 Possibility theory

Possibility theory was introduced by Zadeh [5] and developed by Dubois and Prade [6]. It is an uncertainty theory based on a pair of dual measures in order to evaluate knowledge/ignorance relative to event in hand. The concept of possibility distribution π is one of the important building blocks of possibility theory: It is a mapping from the universe of discourse Ω to the unit scale $[0, 1]$ which can be either quantitative or qualitative. In both these settings, a possibility degree $\pi(w_i)$ expresses to what extent it is consistent that w_i can be

ISBN: 978-989-95079-6-8

the actual state of the world. In particular, $\pi(w_i)$=1 means that w_i is totally possible and $\pi(w_i)$=0 denotes an impossible event. The relation $\pi(w_i)>\pi(w_j)$ means that w_i is more possible than w_j. A possibility distribution π is said to be normalized if $max_{w_i \in \Omega}(\pi(w_i))$=1. It is said to be sub-normalized otherwise.

The second important concept in possibility theory is the one of possibility measure denoted $\Pi(\phi)$ and computing the possibility degree relative to an event $\phi \subseteq \Omega$. It evaluates to what extent ϕ is consistent with the current knowledge encoded by possibility distribution π on Ω. It is defined as follows:

$$\Pi(\phi) = max_{w_i \in \phi}(\pi(w_i)). \quad (1)$$

The term $\Pi(\phi)$ denotes the possibility degree relative to having one of the events involved in ϕ as the actual state of the world.

The necessity measure is the dual of possibility measure and evaluates the certainty implied by the current knowledge of the world. Namely, $N(\phi)$=1-$\Pi(\overline{\phi})$ where $\overline{\phi}$ denotes the complementary of ϕ.

Given a possibility distribution π on Ω, marginal distributions π_X relative to subset of variables X ($X \subseteq V$) are computed using the max operator as follows:

$$\pi_X(x) = max_{w_i \in \Omega}(\pi(w_i) : w_i[X] = x), \quad (2)$$

where term $w_i[X] = x$ denotes the fact that x is the instantiation of X in w_i.

According to the interpretation underlying the possibilistic scale [0,1], there are two variants of possibility theory:

- **Qualitative possibility theory**: In this case, the possibility distribution is a mapping from the universe of discourse Ω to an "ordinal" scale where only the "ordering" of values is important.
- **Quantitative possibility theory**: In this case, the possibilistic scale [0,1] is numerical and possibility degrees are like numeric values that can be manipulated by arithmetic operators.

In this paper, we only focus on qualitative setting. Conditioning is a fundamental notion concerned with updating the current knowledge (encoded by a possibility distribution π) when an evidence (a sure event) is observed.

In the qualitative setting, conditional possibility degree of w_i given an event ϕ is computed as follows (we assume that $\Pi(\phi) \neq 0$) [13]:

$$\pi_m(w_i|\phi) \begin{cases} 1 & \text{if } \pi(w_i)\text{=}\Pi(\phi) \text{ and } w_i \in \phi; \\ \pi(w_i) & \text{if } \pi(w_i)< \Pi(\phi) \text{ and } w_i \in \phi; \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

2.2 *Brief description of possibilistic networks*

Like Bayesian networks, possibilistic ones [4] involves two components:

1. A graphical component consisting in a DAG[1] which encodes direct influence relationships existing between domain variables.

2. A numerical component which is a "numerical" component composed of a set of local a priori and conditional possibility distributions. The latter measure the influence endured by each domain variable A_i in the context of its parents U_{A_i}. Local possibility distributions should satisfy the normalization condition denoted as follows:

$$max_{a_{i_j} \in D_{A_i}}(\pi(a_{i_j}|U_{A_i})) = 1 \quad (4)$$

The joint possibility distribution encoded by the network is computed using the min-based chain rule. Namely,

$$\Pi(A_1, A_2, .., A_n) = min_{i=1..n}(\pi(A_i|U_{A_i})) \quad (5)$$

In classification problems, there exists one node associated with the class variable C which is not observable (it is the target variable) while the remaining nodes represent the attributes A_1, A_2,..,A_n that may be observable. Classification is ensured by computing the most plausible class instance given the instance to classify. Namely, given an observation denoted A=(a_1,a_2,..,a_n) of $\{A1, A_2, .., A_n\}$, the predicted class c is determined as follows:

$$c = argmax_{c_k \in D_C}(\Pi(c_k|A)) \quad (6)$$

Note that term $\Pi(c_k|A)$ denotes the possibility degree of having c_k the actual class instance given the observation A=(a_1,a_2,..,a_n). It is important to note that a class instance c_k is candidate for a given observation $a_1..a_n$ implies that $\Pi(c_k|a_1..a_n)$=1. Note also that several class instances may be totally possible for the observation to classify.

2.3 *Min-based naive possibilistic network classifiers*

A naive network classifier is the simplest form of possibilistic network classifiers. It lies on the strong independence assumption of attributes in the context of the parent node: attributes are assumed independent in the context of the class node. As it is shown in Figure 1, the only dependencies allowed in naive classifiers are from the class node C to each attribute A_i. As for the quantitative component of a naive possibilistic network, it involves the prior possibility distribution relative to the class node and the conditional possibility distributions relative to attributes given the class node. Figure 2 gives the joint possibility distribution encoded by network of Figure 1.

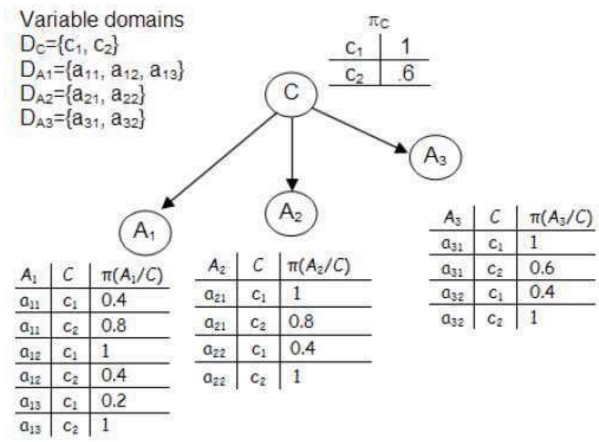

Figure 1: Naive possibilistic network structure

Classification is ensured using min-based conditioning (see Equation 3) by computing the a posteriori possibility degree for each class instance c_k given the observation $A_1..A_n$ (namely $\Pi(c_k|A_1..A_n)$). Namely,

$$\Pi(c_k|A_1..A_n) = min(\pi(c_k, A_1..A_n), \Pi(A_1..A_n)) \quad (7)$$

[1] Direct Acyclic Graph

ISBN: 978-989-95079-6-8

A_1	A_2	A_3	C	$\pi(CA_1A_2A_3)$
a_{11}	a_{21}	a_{31}	c_1	0.4
a_{11}	a_{21}	a_{31}	c_2	0.6
a_{11}	a_{21}	a_{32}	c_1	0.4
a_{11}	a_{21}	a_{32}	c_2	0.6
a_{11}	a_{22}	a_{31}	c_1	0.4
a_{11}	a_{22}	a_{31}	c_2	0.6
a_{11}	a_{22}	a_{32}	c_1	0.4
a_{11}	a_{22}	a_{32}	c_2	0.6
a_{12}	a_{21}	a_{31}	c_1	1
a_{12}	a_{21}	a_{31}	c_2	0.4
a_{12}	a_{21}	a_{32}	c_1	0.4
a_{12}	a_{21}	a_{32}	c_2	0.4

A_1	A_2	A_3	C	$\pi(CA_1A_2A_3)$
a_{12}	a_{22}	a_{31}	c_1	0.4
a_{12}	a_{22}	a_{31}	c_2	0.4
a_{12}	a_{22}	a_{32}	c_1	0.4
a_{12}	a_{22}	a_{32}	c_2	0.4
a_{13}	a_{21}	a_{31}	c_1	0.2
a_{13}	a_{21}	a_{31}	c_2	0.6
a_{13}	a_{21}	a_{32}	c_1	0.2
a_{13}	a_{21}	a_{32}	c_2	0.6
a_{13}	a_{22}	a_{31}	c_1	0.2
a_{13}	a_{22}	a_{31}	c_2	0.6
a_{13}	a_{22}	a_{32}	c_1	0.2
a_{13}	a_{22}	a_{32}	c_2	0.6

Figure 2: Joint possibility distribution encoded by naive classifier of Figure 1

Note that according to min-based conditioning, Equation 7 is only valid when term $\pi(c_k, A_1..A_n) < \Pi(A_1..A_n)$ (See min-based conditioning of Equation 3). To the best of our knowledge, there is no work that addresses the problem of classification under uncertain inputs using min-based possibilistic networks.

3 Min-based revision for classification under uncertain inputs

In our context, uncertainty relative to uncertain/missing attribute A_i is represented by a possibility distribution π'_{A_i} given for example by the expert. Moreover, uncertainty can bear on any attribute subset or on the whole attribute set.
Jeffrey proposed in [12] a method based on probability kinematics for revising a probability distribution p into p' given uncertainty bearing on a set of mutually exclusive and exhaustive events λ_i. In this method, uncertainty is of the form (λ_i, α_i) with α_i=$p'(\lambda_i)$. Jeffrey's rule states that although there is uncertainty about events λ_i, conditional probability of any event $\phi \subseteq \Omega$ given any uncertain event λ_i. The possibilistic counterpart of this rule has been investigated in [14]. In the possibilistic framework, revised possibility distribution π' must comply with the principle stating that uncertainty about events λ_i must not alter the conditional possibility degree of any event $\phi \subseteq \Omega$ given any event λ_i. Namely,

$$\forall \lambda_i \in \Omega, \;\; \forall \phi \subseteq \Omega, \Pi'(\phi|\lambda_i) = \Pi(\phi|\lambda_i) \tag{8}$$

Note that contrary to the probabilistic and quantitative possibilistic settings, Equation 8 may have several solutions. In this case, the least specific distribution is selected according to the principle stating that if an event is not explicitly discarded, then it must remain possible. Min-based conditioning allows to obtain from Equation 8 the following one:

$$\forall \phi, \Pi'(\phi) = max_{\lambda_i}(min(\pi'(\lambda_i, \phi), \pi(\lambda_i, \phi), \Pi(\phi)) \tag{9}$$

In classification with uncertain inputs problems, the possibilistic counterpart of Jeffrey's rule cannot directly be applied to revise the possibility distribution encoded by the classifier because of two major problems:
1- The first problem concerns the fact that Jeffrey's rule can be applied only if uncertainty bears on a set of exhaustive and mutually exclusive events while in classification with uncertain inputs problems, uncertainty is bearing on a set of attributes A_i which are not mutually exclusive. Indeed, if inputs A_1,..,A_n are uncertain, then this uncertainty is encoded by $\pi'(A_1)$,..,$\pi'(A_n)$ respectively. In order to apply Jeffrey's rule, we must compute a joint possibility distribution $\pi'(A_1..A_n)$ relative to $A_1..A_n$ where uncertain events $a_1..a_n$ are exhaustive and mutually exclusive. Obviously, the term $\pi'(A_1..A_n)$ is function of $\pi'(A_1)$,..,$\pi'(A_n)$ and will be henceforth denoted by $\pi'(A_1..A_n)$=$f(\pi'(A_1)$,..,$\pi'(A_n))$. Uncertain inputs $\pi'(A_1)$,..,$\pi'(A_n)$ can be combined using a combination operator according to the problem constraints and objectives. It is important to note that the algorithm we propose in next section works regardless the combination function f provided that this latter satisfies the following natural propriety (called henceforth unanimity):
i) If $\forall$i=1..n, $\pi'(a_i)$=1 then $\pi'(a_1..a_n)$=1.
ii) If $\exists a_i \in D_{A_i}$ such that $\pi'(a_i)$=α<1, then $\pi'(a_1..a_i..a_n)$<1 whatever are the possibility degrees of the other variable values.
One can easily check that the $product$ and min (and more generally $t\text{-}norm$ operators), which are the most used combination operators in the possibilistic framework satisfy this natural propriety. However, max and more generally, $t\text{-}conorm$ operators do not satisfy it. In this paper, we will use the min operator in order to induce the joint possibility distribution $\pi'(A_1..A_n)$ from uncertain inputs $\pi'(A_1)$,..,$\pi'(A_n)$. Namely,

$$\pi'(a_1..a_n) = min_{k=1..n}(\pi'(a_k)) \tag{10}$$

2- The second problem concerns the computational complexity of performing classification under uncertain inputs based on revising the possibility distribution encoded by the classifier by uncertain inputs using the possibilistic counterpart of Jeffrey's rule. Namely, given the initial possibility degrees of class instances c_k, we want to revise this possibility distribution given uncertain attributes A_1,..,A_n using Jeffrey's rule. More precisely, we need to compute $\pi'(c_k)$=$\pi(c_k|\pi'(A_1, .., A_n))$ defined as follows:

$$\pi'(c_k) = max_{A_1..A_n}(min(\Pi(c_k|A_1..A_n), \pi'(A_1..A_n))) \tag{11}$$

It is clear that Equation 11 cannot be used to compute $\pi'(c_k)$ because this computation is exponential in the number of attributes and attribute domains.

Example:
Let us illustrate possibilistic classification with uncertain inputs using Jeffrey's rule on the naive possibilistic classifier of Figure 1. The uncertain inputs to classify are provided by Figure 3. Classification is ensured by determining the most

π'_{A1}	
a_{11}	1
a_{12}	.4
a_{13}	1

π'_{A2}	
a_{21}	1
a_{22}	1

π'_{A3}	
a_{31}	1
a_{32}	.4

Figure 3: Uncertain inputs to classify

plausible class instance(s) given these uncertain inputs. Figure 4 gives the results of revising the initial joint possibility distribution (see Figure 2) using the possibilistic counterpart of Jeffrey's rule (see Equation 11). One can easily deduce from joint possibility distribution π' of Figure 4 that only class instance c_2 is totally plausible ($\pi'(c_2)$=1) given the uncertain observations of Figure 3. However, in order to obtain such a result, we computed the revised joint possibility distribution. We would like to ensure the same classification without computing the joint possibility distribution. Next section proposes

ISBN: 978-989-95079-6-8

A_1	A_2	A_3	C	$\pi'(CA_1A_2A_3)$	A_1	A_2	A_3	C	$\pi'(CA_1A_2A_3)$
a_{11}	a_{21}	a_{31}	c_1	0.4	a_{12}	a_{22}	a_{31}	c_1	0.4
a_{11}	a_{21}	a_{31}	c_2	1	a_{12}	a_{22}	a_{31}	c_2	0.4
a_{11}	a_{21}	a_{32}	c_1	0.4	a_{12}	a_{22}	a_{32}	c_1	0.4
a_{11}	a_{21}	a_{32}	c_2	0.4	a_{12}	a_{22}	a_{32}	c_2	0.4
a_{11}	a_{22}	a_{31}	c_1	0.4	a_{13}	a_{21}	a_{31}	c_1	0.2
a_{11}	a_{22}	a_{31}	c_2	1	a_{13}	a_{21}	a_{31}	c_2	1
a_{11}	a_{22}	a_{32}	c_1	0.4	a_{13}	a_{21}	a_{32}	c_1	0.2
a_{11}	a_{22}	a_{32}	c_2	0.4	a_{13}	a_{21}	a_{32}	c_2	0.4
a_{12}	a_{21}	a_{31}	c_1	0.4	a_{13}	a_{22}	a_{31}	c_1	0.2
a_{12}	a_{21}	a_{31}	c_2	0.4	a_{13}	a_{22}	a_{31}	c_2	1
a_{12}	a_{21}	a_{32}	c_1	0.4	a_{13}	a_{22}	a_{32}	c_1	0.2
a_{12}	a_{21}	a_{32}	c_2	0.4	a_{13}	a_{22}	a_{32}	c_2	0.4

Figure 4: Revised joint possibility distribution π'

an efficient algorithm suitable for naive min-based possibilistic network classification with uncertain inputs.

4 A polynomial algorithm for naive possibilistic classifier under uncertain inputs

Classification based on possibilistic networks is ensured by determining if each class instance $c_k \in D_C$ is totally possible given the inputs ($\Pi(c_k|A_1..A_n)$=1). This is the basic idea of our algorithm: only search for class instances having a posteriori possibility degrees equal to 1 on the basis of uncertain inputs. Our algorithm ensures this search through a series of equivalent transformations on the initial possibilistic network taking into account the inputs to classify. The five steps of our algorithm are detailed in the following.

4.1 Step 1: Eliminating not totally possible observations

In our context, classification is ensured by revising the possibility distribution encoded by the classifier according to the possibilistic counterpart of Jeffrey's rule (using Equation 11). Since we are only interested in determining if a given class label c_k has an a posteriori possibility degree equal to 1, it is possible to discard all instances of $A_1A_2..A_n$ where $\pi'(A_1A_2..A_n)<1$ because such instances force the value of $\pi'(c_k)$ to be less than 1. According to the *unanimity* propriety, each input $a_i \in D_{A_i}$ which is not totally possible ($\pi'(a_i)<1$) will prevent every input configuration $a_1..a_i..a_n$ from being totally possible. It is clear that such configurations are useless for the classification task and can be discarded. This leads to eliminating from each attribute domain $D^{\Pi G}_{A_i}$ (relative to attribute A_i in initial network ΠG) values whose possibility degrees in π'_{A_i} are less than 1. In this step, attribute domain $D^{\Pi G}_{A_i}$ is changed to $D^{\Pi G_{S_1}}_{A_i}$ which denotes A_i's domain in network ΠG^{S_1} obtained from ΠG after Step 1. Namely, $D^{\Pi G_{S_1}}_{A_i} = D^{\Pi G}_{A_i} - \{a_i, if \ \ \pi'_{A_i}(a_i)<1\}$. All the remaining attribute instances are totally possible after this step. Then we have the following proposition:

Proposition 1 Let ΠG be the naive possibilistic classifier encoding the initial knowledge and π'_{A_1}, π'_{A_2},..,π'_{A_n} be the possibility distributions encoding uncertainty relative to attributes A_1, A_2,.., A_n respectively. Let ΠG^{S_1} be the naive possibilistic network obtained by eliminating not totally possible instances. Then,
$\pi'^{\Pi G}(c_k) = max_{A_1..A_n}(\Pi^{\Pi G}(c_k|a_1..a_n) * \pi'(a_1..a_n))$=1
if and only if
$\pi^{\Pi G_{S_1}}(c_k) = max_{A_1..A_n}(\Pi^{\Pi G_{S_1}}(c_k|a_1..a_n)) = 1$

This proposition states that if there is a class instance c_k which is totally possible in the initial network ΠG given the uncertain inputs then it is also totally possible in network ΠG^{S_1} obtained after eliminating not totaly possible inputs.

Example (continued): Network of Figure 5 is obtained by respectively substituting $D^{\Pi G}_{A_1}$=$\{a_{11}, a_{12}, a_{13}\}$, $D^{\Pi G}_{A_2}$=$\{a_{21}, a_{22}\}$ and $D^{\Pi G}_{A_3}$=$\{a_{31}, a_{32}\}$ by $D^{\Pi G^{S_1}}_{A_1}$=$\{a_{11}, a_{13}\}$, $D^{\Pi G^{S_1}}_{A_2}$ = $\{a_{21}, a_{22}\}$ and $D^{\Pi G^{S_1}}_{A_3}$=$\{a_{31}\}$.

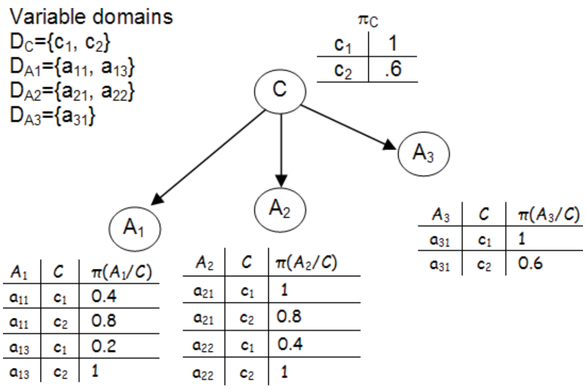

Figure 5: ΠG^{S_1}: Possibilistic network after Step 1

It is important to note that the possibility distribution relative to node C in ΠG^{S_1} is exactly the same as in ΠG.

4.2 Step 2: Eliminating unary variables

After the instance elimination step, some attributes may have their domains only containing one instance. In the example of Figure 5, attribute A_3's domain $D^{\Pi G_{S_1}}_{A_3}$ exactly contains one element (which is a_{31}). Such unary variables can be eliminated provided that class node distribution is adapted to keep the joint possibility distribution unchanged. The possibility degree of any instance $c_k a_1 a_2..a_n$ in network ΠG^{S_1} is computed using the min-based chain rule as follows:

$$\Pi^{\Pi G_{S_1}}(c_k a_1..a_n) = min(\pi^{\Pi G_{S_1}}(c_k), \pi^{\Pi G_{S_1}}(a_1|c_k), .., \pi^{\Pi G_{S_1}}(a_n|c_k) \tag{12}$$

Unary attributes can be eliminated giving a new network ΠG^{S_2} where class node distribution has to be adjusted in order to guarantee that $\pi^{\Pi G_{S_1}}(c_k a_1..a_n)$ and $\pi^{\Pi G_{S_2}}(c_k a_1..a_n)$ remain equal. Hence, for an unary attribute A_i whose domain $D^{\Pi G_{S_1}}_{A_i}$ contains only one value a_i, it is possible to achieve this transformation by substituting each $\pi^{\Pi G_{S_1}}(c_k)$ with $\pi^{\Pi G_{S_2}}(c_k)$=min($\pi^{\Pi G_{S_1}}(c_k), \pi^{\Pi G_{S_1}}(a_i|c_k)$). This finding is formalized by the following proposition:

Proposition 2 Let ΠG^{S_1} be the naive possibilistic classifier whose nodes involve class node C and attribute nodes A_1, A_2,..,A_n. Assume that A_1 is a unary attribute whose domain only contains the instance a_1 and let ΠG^{S_2} be the naive possibilistic network involving C, A_2,..,A_n such that: $\pi^{\Pi G_{S_2}}(a_i)$= $\pi^{\Pi G_{S_1}}(a_i)$ for i=2,..,n and $\pi^{\Pi G_{S_2}}(c_k)$=min($\pi^{\Pi G_{S_1}}(a_1|c_k), \pi^{\Pi G_{S_1}}(c_k)$). Then, $\forall c_k \in D_C$, $\forall a_i \in D^{\Pi G_{S_1}}_{A_i}$ for i=1..n,
$\pi^{\Pi G_{S_1}}(c_k a_1 a_2..a_n)$=$\pi^{\Pi G_{S_2}}(c_k a_2..a_n)$.

Example (continued): Let us continue our example. In network of Figure 5, attribute A_3 is unary and will be removed. After Step 2, we obtain network of Figure 6.

ISBN: 978-989-95079-6-8

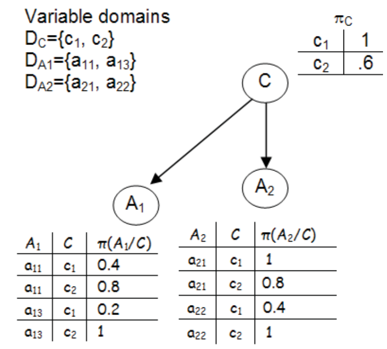

Figure 6: ΠG^{S_2}: Possibilistic network after Step 2

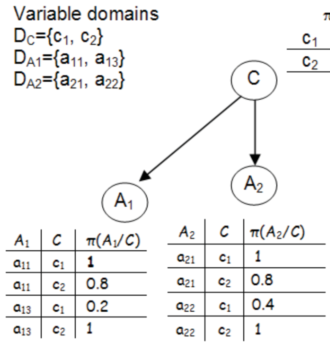

Figure 7: ΠG^{S_3}: Possibilistic network example after Step 3

4.3 *Step 3: Re-normalizing local possibility distributions*

As a consequence of the instance elimination step (Steps 1), local possibility distributions of some attributes might not be normalized. Indeed, it may exist a variable A_i and a class label c_k such that $max_{A_i}(\pi^{\Pi G_{S_2}}(a_i|c_k))$=$\alpha$ (α<1). Step 3 deals with this problem considering two cases:

- If α=0, then $\forall a_i \in D_{A_i}^{\Pi G_{S_2}}$, $\pi^{\Pi G_{S_2}}(c_k a_1..a_i..a_n)$=0 meaning that whatever is the value of A_i, this forces $\pi^{\Pi G_{S_2}}(c_k a_1 a_2..a_n)$=0. Hence, the class c_k cannot be among plausible ones, and can be removed from $D_C^{\Pi G_{S_2}}$.
- If 0<α<1, re-normalizing the conditional possibility distribution relative to attribute A_i can be done by building a new network ΠG^{S_3} by substituting $\pi^{\Pi G_{S_2}}(a_i|c_k)$ with
$$\begin{cases} 1 & \text{if } \pi(a_i|c_k)=max_{a_j \in D_{A_i}}(\pi^{\Pi G_{S_2}}(a_j|c_k)); \\ \pi(a_i|c_k) & \text{otherwise.} \end{cases}$$
and substituting $\pi^{\Pi G_{S_2}}(c_k)$ with $min(\pi^{\Pi G_{S_2}}(c_k), max_{a_j \in D_{A_i}}(\pi^{\Pi G_{S_2}}(a_j|c_k))$. After this transformation, $\pi^{\Pi G_{S_3}}(c_k a_1..a_n)$ is conserved while sub-normalized local possibility distribution becomes normalized.

Proposition 3 Let ΠG^{S_2} be the naive possibilistic network involving nodes C,A_1,A_2,..,A_n obtained from Step 2. Assume that conditional possibility distribution relative to node A_1 is not normalized ($\exists c_k$ such that $max_{A_1}(\pi^{\Pi G_{S_2}}(a_1|c_k))$=$\alpha$ and 0<α<1). Let ΠG^{S_3} be the naive possibilistic network having same structure as ΠG^{S_2} where $\pi^{\Pi G_{S_3}}(a_1|c_k)$= 1 if $\pi(wa_i|c_k)$=$max_{a_j \in D_{A_i}}(\pi^{\Pi G_{S_2}}(a_j|c_k))$, $\pi(a_i|c_k)$ otherwise and $\pi^{\Pi G_{S_3}}(a_i|c_k)$=$\pi^{\Pi G_{S_2}}(a_i|c_k)$ for i=2,..,n and $\pi^{\Pi G_{S_3}}(c_k)$=min($\pi^{\Pi G_{S_2}}(c_k), max_{a_j \in D_{A_i}}(\pi^{\Pi G_{S_2}}(a_j|c_k))$. Then, $\forall c_k \in D_C^{\Pi G_{S_3}}$, $\forall a_i \in D_{A_i}^{\Pi G_{S_3}}$, $\pi^{\Pi G_{S_3}}(c_k a_1 a_2..a_n)$=$\pi^{\Pi G_{S_2}}(c_k a_1 a_2..a_n)$

Note that contrary to product-based possibilistic networks, normalizing conditional possibility distribution of min-based networks is significantly different from normalizing local probability distributions of Bayesian networks.

Example (continued): In network ΠG^{S_2} of Figure 6, local possibility distribution of attribute A_1 is not normalized ($max_{a_i \in D_{A_1}}(\pi^{\Pi G_{S_2}}(a_i|c_1))$=.4). After re-normalizing ΠG^{S_2}, we obtain network of Figure 7. Possibility distribution relative to A_1 has been re-normalized and possibility distribution relative to class node C has been adjusted accordingly. One can easily check that joint possibility distributions encoded by networks of Figure 6 and Figure 7 are equal.

4.4 *Step 4: Prior totally possible class lookup*

The aim of Steps 1, 2 and 3 is simplifying and re-normalizing the initial network. Step 4 allows to search for class instances which are totally possible in network NP^{S_3}. Once the network is simplified and re-normalized, it is immediate that each class instance c_k having the utmost prior possibility degree in network NP^{S_3} (c_k=$argmax_{c_j \in D_C^{NP_{S_3}}}(\pi^{NP_{S_3}}(c_j))$, is totally possible given the uncertain inputs to classify. In particular, if $\pi^{NP_{S_3}}(c_k)$=1, then we can assert that there exists an attribute configuration $a_1..a_n$ allowing c_k to be totally plausible. This result is formalized in the following proposition:

Proposition 4 Let NP^{S_3} be the naive possibilistic network obtained after Steps 1, 2 and 3. Then, $\forall c_k \in D_C^{\Pi G_{S_3}}$ such that $\pi^{\Pi G_{S_3}}(c_k)$=1 or c_k=$argmax_{c_j \in D_C^{NP_{S_3}}}(\pi^{NP_{S_3}}(c_j))$, then $max_{A_1..A_n}(\pi^{\Pi G_{S_3}}(c_k|a_1..a_n)) = 1$

The attribute configuration guaranteeing that c_k is totally possible is the one where $\pi^{\Pi G_{S_3}}(a_i|c_k)$=1 for i=1..n.

Example (continued): Proposition 4 allows to assert that in network ΠG^{S_3} of Figure 7, class instance c_2 is totally possible since c_2=$argmax_{c_j \in D_C^{NP_{S_3}}}(\pi^{NP_{S_3}}(c_j))$. The attribute configuration allowing this result is $a_{13}a_{22}$. One can easily check that $\Pi^{\Pi G_{S_3}}(c_2|a_{13}a_{22})$=1.

4.5 *Step 5: Conditionally totally possible class lookup*

We assume in this step that the class variable C is a binary variable (namely, $D_C^{\Pi G_{S_3}}$=$\{c_1, c_2\}$) and class label c_2 is totally possible ($\pi^{\Pi G_{S_3}}(c_2)$>$\pi^{\Pi G_{S_3}}(c_1)$).

In Step 4, only a subset of plausible class instances are found. Indeed, other class instances c_k having $\pi^{\Pi G_{S_3}}(c_k)$=α (0<α<1) and $c_k \neq argmax_{c_j \in D_C^{NP_{S_3}}}(\pi^{NP_{S_3}}(c_j))$ can be totally possible given the uncertain inputs to classify. Namely, it may exist an attribute configuration $a_1 a_2..a_n$ such that $\Pi^{\Pi G_{S_3}}(c_k|a_1 a_2..a_n)$=1 even if $\pi^{\Pi G_{S_3}}(c_k)$<1 and $c_k \neq argmax_{c_j \in D_C^{NP_{S_3}}}(\pi^{NP_{S_3}}(c_j))$. Hence, there is a need to check if every class instance c_k which is not found totaly possible in Step 4 can be conditionally totally possible without exploring all configurations of $\Pi^{\Pi G_{S_3}}(c_k|a_1 a_2..a_n)$. If class instance c_1 is totally possible, then this implies

ISBN: 978-989-95079-6-8

that there exists an attribute configuration $a_1..a_n$ where $\Pi^{\Pi G_{S_3}}(c_1|a_1a_2..a_n)$=1. In order to find such a configuration, we can use the following decomposition:

$\Pi^{\Pi G_{S_3}}(c_1|a_1..a_n)$=1 if and only if
$\pi^{\Pi G_{S_3}}(c_1a_1..a_n)$=$\Pi^{\Pi G_{S_3}}(a_1..a_n)$. Recall that
$\Pi^{\Pi G_{S_3}}(a_1..a_n)$=$max(\pi^{\Pi G_{S_3}}(c_1a_1..a_n), \pi^{\Pi G_{S_3}}(c_2a_1..a_n))$
Hence, $\pi^{\Pi G_{S_3}}(c_1a_1..a_n) \geq \pi^{\Pi G_{S_3}}(c_2a_1..a_n)$. Then
$\frac{min(\pi(c_1),\pi(a_1|c_1),..,\pi(a_n|c_1))}{min(\pi(c_2),\pi(a_1|c_2),..,\pi(a_n|c_2))} \geq 1$.
Let us define $\pi_{c_1}(a_i/c_1)$=1 if
$\pi(a_i|c_1) \geq \pi(a_i|c_2)$ and $\pi_{c_1}(a_i/c_1)$=$\pi(a_i|c_2)$ otherwise. Then
$\frac{min(\pi(c_1),\pi(a_1|c_1),..,\pi(a_n|c_1))}{min(\pi(c_2),\pi(a_1|c_2),..,\pi(a_n|c_2))} \geq 1$ implies that

$$\pi(c_1) \leq min(\pi_{c_1}(a_1|c_1), .., \pi_{c_1}(a_n|c_1)). \quad (13)$$

Decomposition of Equation 13 can be used to build a new network ΠG^{S_5} by transforming conditional possibility distributions relative to each attribute A_i in ΠG^{S_3} by substituting every term $\pi^{\Pi G_{S_3}}(a_i|c_1)$ by 1 if $\pi^{\pi G_{S_5}}(a_i|c_1) \geq \pi^{\Pi G_{S_3}}(a_i|c_2))$ and by $\pi^{\Pi G_{S_5}}(a_i|c_1)$ otherwise and discarding c_2 since it is known to be totally possible. After this transformation, some local conditional possibility distributions in network ΠG may not be normalized. Repeating Step 2 (Re-normalization) on sub-normalized local distributions in network ΠG^{S_5}) allows to re-normalize them. Once re-normalization accomplished, the new possibility distribution $\pi_C^{\Pi G_{S_5}}$ relative to class node shows whether class instance c_1 is totally possible. Then we have the following proposition:

Proposition 5 Let ΠG^{S_3} be the naive possibilistic network obtained after Steps 1, 2 and 3, and let $D_C^{\Pi G_{S_3}}$=$\{c_1, c_2\}$, $\pi^{\Pi G_{S_3}}(c_1) < \pi^{\Pi G_{S_3}}(c_2)$. Let ΠG be the possibilistic network having same structure as ΠG^{S_3} where $D_C^{\Pi G^{S_5}}$=$\{c_1\}$. Let for i=1..n $\pi^{\Pi G^{S_5}}(a_i|c_2)$=1 if $\pi^{\Pi G_{S_3}}(a_i|c_1) \geq \pi^{\Pi G_{S_3}}(a_i|c_2))$ and $\pi^{\Pi G^{S_5}}(a_i|c_2)$= $\pi^{\Pi G^{S_3}}(a_i|c_2)$ otherwise. Then, $\pi^{\Pi G^{S_5}}(c_1)$=1 if and only if there exists an attribute configuration $a_1..a_n$ such that $\Pi^{\Pi G_{S_3}}(c_1|a_1..a_n)$=1.

It is important to note that in comparison with the lookup for conditionally totally possible class instances in product-based networks (see [11]), this step shows significant differences in min-based networks.

Example (continued): Transformation of Step 5 on network ΠG^{S_3} and its re-normalization gives network ΠG^{S_5} of Figure 8 where network of the left side gives the network obtained af-

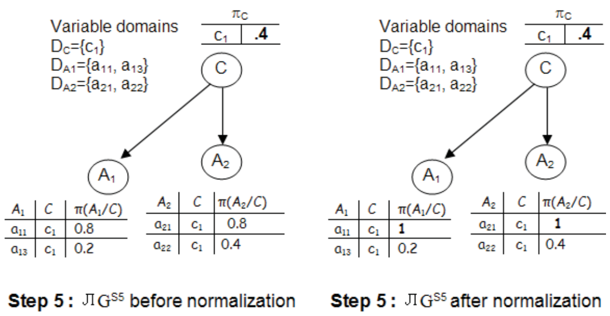

Figure 8: ΠG^{S_5}: Possibilistic network after Step 5

ter Step 5 before re-normalization while network of right side represents ΠG^{S_5} after re-normalization according to Step 3. After this transformation, we can assert that class instance c_1 is not totally possible ($\pi^{\Pi G^{S_5}}(c_1) < 1$). This result confirms the one obtained by directly applying the possibilistic counterpart of Jeffrey's rule (see Figure 4).

5 Conclusion

This paper dealt with min-based possibilistic network classifiers under uncertain inputs. It first addressed the min-based possibilistic counterpart of Jeffrey's rule for revising possibilistic knowledge encoded by a qualitative naive possibilistic network. After showing that classification based on the min-based revision of a possibility distribution encoded by a qualitative possibilistic classifier is exponential in the number of uncertain inputs, we proposed a polynomial algorithm for revising a naive min-based possibilistic network given uncertain inputs. This algorithm applies a series of equivalent and polynomial transformations on the initial network taking into account the uncertain evidence to classify. In future works, we will address general possibilistic classifiers under uncertain inputs.

References

[1] F. V. Jensen. *An Introduction to Bayesian Networks*. UCL press, London, 1996.

[2] A. Mittal. *Bayesian Network Technologies: Applications and Graphical Models*. IGI Publishing, USA, 2007.

[3] S. Benferhat, D. Dubois, L Garcia, and H. Prade. Possibilistic logic bases and possibilistic graphs. In *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence UAI 99*, pages 57–64, Stockholm, 1999.

[4] C. Borgelt and R. Kruse. *Graphical Models: Methods for Data Analysis and Mining*. John Wiley & Sons, Inc., USA, 2002.

[5] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1(3), 1978.

[6] D. Dubois and H. Prade. *Possibility theory*. Plenum Press, New-York, 1988.

[7] B.a Haouari, Z. Elouedi N. Ben-Amor, and K. Mellouli. Naive possibilistic network classifier. In *Proceedings of The Thirteenth Congress of International Association for Fuzzy-Set Management and Economy, SIGEF 2006*, 2006.

[8] T. Denoeux and L. M. Zouhal. Handling possibilistic labels in pattern classification using evidential reasoning. *Fuzzy Sets and Systems*, 122(3):409 – 424, 2001.

[9] H. Eyke. Possibilistic induction in decision-tree learning. In *ECML '02: Proceedings of the 13th European Conference on Machine Learning*, pages 173–184, London, UK, 2002.

[10] I. Jenhani, N. Benamor, and Z. Elouedi. Decision trees as possibilistic classifiers. *Int. J. Approx. Reasoning*, 48(3):784–807, 2008.

[11] S. Benferhat and K. Tabia. An efficient algorithm for naive possibilistic classifiers with uncertain inputs. In *The Second International Conference on Scalable Uncertainty Management(SUM'08)*, Naples, Italy, oct 2008. Springer LNAI.

[12] R. C. Jeffrey. *The Logic of Decision*. McGraw Hill, NY, 1965.

[13] E. Hisdal. Conditional possibilities independence and non interaction. *Fuzzy Sets and Systems*, pages 283–297, 1978.

[14] D. Dubois and H. Prade. A synthetic view of belief revision with uncertain inputs in the framework of possibility theory. *Int. J. of Approximate Reasoning*, 17(2–3):295–324, aot 1997.

ISBN: 978-989-95079-6-8