A complexity analysis of MPE inference in possibilistic networks

Salem Benferhat, Amélie Levray, Karim Tabia

A complexity analysis of MPE inference in possibilistic networks

(Preprint version)

Salem BENFERHAT, Karim TABIA
*Centre de Recherche en Informatique de Lens (*CRIL*)*
Lens, France
{benferhat,tabia}@cril.fr

Amélie LEVRAY
University of Edinburgh
Edinburgh, Scotland
alevray@inf.ed.ac.uk

***Abstract*—Reasoning with uncertainty in graphical models often implies great computational cost. For example, computing the most probable explanation in Bayesian networks is known to be *NP*^{PP}-complete. Possibilistic networks represent an alternative powerful representation for uncertain information. This paper aims at showing that the computation complexity of MPE inference tasks in possibilistic networks are *NP*-complete. To that end, we provide full reduction and proof for MPE querying $\min$-based and product-based possibilistic networks. More precisely, we provide incremental proofs based on reductions to and from three well-known *NP*-complete problems: SAT, 3SAT and Weighted MaxSAT decision problems.**

***Index Terms*—Complexity, Possibilistic networks, MAP inference, MPE inference**

I. Introduction

Beliefs graphical models, such as Bayesian networks [6], credal networks [5], or possibilistic networks [3] are powerful means of compactly represent uncertainty using directed acyclic graphs and independence relationships. Typically, possibilistic networks are seen as counterparts of Bayesian networks based on possibility theory [10], where possibility degrees are more suited for handling imperfect, qualitative and partial information.

Inference in such graphical models has been extensively studied and many algorithms have emerged. On the other hand, while complexity results regarding inference in probabilistic networks are well-established [7]–[9], there is no such deep study for possibilistic networks. This paper aims at filling this gap.

Essentially, in graphical models there are three common types of queries: computing most probable (or plausible) explanation (MPE); computing a posteriori probability (or possibility) degrees (*Pr*); and computing the maximum a posteriori explanation (MAP). These tasks are known to be very hard in the probabilistic setting. Indeed, the decision problems associated to MPE, *Pr*, MAP are *NP*-complete, PP-complete and *NP*^{PP}-complete respectively (see [7], [9] for more details on complexity issues in Bayesian and credal networks). In this paper, we focus on possibility theory where we consider two interpretations of possibility theory, $\min$-based possibility theory and product-based possibility theory [11].

In [2], the authors analysed the computational complexity of MAP queries in $\min$-based and product-based possibilistic networks. They showed that the decision problem behind MAP querying is *NP*-complete for both $\min$-based and product-based possibilistic networks. Regarding MAP querying a product-based possibilistic network, only the proof of *NP*-hardness has been provided. The first part of this paper provides the proof of *NP*-completeness theorem, stated in [2], of MAP querying a product-based possibilistic network. In the second part of the paper, we address the complete analysis of the decision problems associated with the most plausible explanation (MPE) task in both $\min$-based and product-based possibilistic networks, show using a reduction to the SAT decision problem, that it is *NP*-complete. Such computational complexity outcomes favour possibility theory as an efficient alternative for reasoning with uncertainty (some results on learning possibilistic parameters over probabilistic ones can be found in [13]).

The paper is outlined as follows. In section II, we briefly recall notions on possibility theory, as well as give motivations. The third section investigates general properties on inference tasks in possibilistic networks. In particular, we provide the proof that the decision problem based on conditioning operator is the same as the one based on conjunction operator when computing MAP queries. Hence, we no longer need conditioning rule in the computation of MAP queries. The fourth section addresses the *NP*-completeness result of MAP inference in product-based possibilistic networks. In the fifth section, we establish the complexity of MPE inference. This is done by showing a reduction from 3SAT to MPE querying a binary and boolean possibilistic networks, and conversely with a reduction from MPE in a possibilistic network to a SAT problem.

II. A refresher on possibility theory and possibilistic networks

In this section, we give a short reminder of the basic notions associated to possibility theory [11] and its associated graphical models named possibilistic networks [1], [4], [12]. A possibility distribution, denoted by π, is a mapping from the set of possible worlds Ω to the unit interval $[0, 1]$. Note that we consider a finite and discrete set. For a given interpretation $\omega \in \Omega$, $\pi(\omega) = 1$ is interpreted as fully possible. $\pi(\omega) = 0$ is interpreted as impossible. π is said to be normalised if there is at least an element $\omega \in \Omega$ that is fully possible (*i.e.* such

that $\pi(\omega) = 1$).

There are two understandings of the scale $[0, 1]$ of possibility degrees, either the product-based interpretation as in probability theory or the min-based interpretation which considers degrees on an ordinal scale. These two interpretations lead to two different conditioning rules when dealing with new evidence. We call min-based conditioning $|_m$ the operation leading to $\pi(.|_m\phi)$ given by [11], [14]:

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

The product-based conditioning, denoted by $|_*$, is, as in the probabilistic setting, defined as follows:

$$\pi(\omega_i|_*\phi) = \begin{cases} \dfrac{\pi(\omega_i)}{\Pi(\phi)} & \text{if } \omega_i \in \phi; \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

A possibilistic network denoted $\mathcal{PN} =< G, \Theta >$ is specified by two components:

- A graphical component G: a directed acyclic graph (DAG) where each node represents a discrete variable and edges encode independence relations between variables.
- A numerical component Θ: a set of local normalised possibility distributions $\Theta_i = \pi_{\mathcal{PN}}(X_i|par(X_i))$ of each node X_i given its parents $par(X_i)$.

The joint possibility distribution is factorised using a chain rule, defined as:

$$\pi_{\mathcal{PN}_\otimes}(X_1, .., X_n) = \otimes_{i=1,..,n}\pi_{\mathcal{PN}_\otimes}(X_i|_\otimes\ par(X_i)). \quad (3)$$

where $\otimes = m$ in min-based possibilistic setting and $\otimes = *$ in product-based possibilistic setting.

It is well-known that inference in probabilistic models is a hard task in the general case. In particular, the decision problem associated with *MAP* in Bayesian networks is NP^{PP}-complete [7]. The next sections address the same complexity issues on product-based possibilistic networks as well as the computational complexity of *MPE* inference.

The complexity of *MAP* querying a min-based possibilistic network has already been discussed in [2] and it has been shown that *MAP* inference in this context is *NP*-complete.

But first, we recall the definition of inference tasks in possibilistic networks.

III. Inference in possibilistic networks

In this paper, we investigate two of the most common types of queries when reasoning with graphical models, that are *MAP* inference and *MPE* inference. *MAP* queries require searching for the most plausible instantiation of query variables Q given an evidence e (an instantiation of a set of variables E). While *MPE* queries search for the most plausible explanation of an evidence e. More formally,

MAP **query**: Let $\mathcal{PN}$ be a possibilistic network over the set of variables V, $Q \subset V$ be a set of query variables and $E \subset V$ be a set of evidence variables with $Q \cap E = \emptyset$. Then, given an evidence $E = e$, the aim is to compute the most plausible instantiation q of Q given the evidence e.

$$argmax_{q \in D_Q}(\Pi_{\mathcal{PN}}(q|e)). \quad (4)$$

MPE **query**: Let $\mathcal{PN}$ be a possibilistic network over the set of variables V, $E \subset V$ be a set of evidence variables. We denote X the set of remaining variables ($X = V \setminus E$). Then, given an evidence $E = e$, *MPE* query compute the most plausible instantiation x of X compatible with the evidence e. Stated otherwise by:

$$argmax_{x \in X}(\Pi_{\mathcal{PN}}(x, e)). \quad (5)$$

Note that $\Pi_{\mathcal{PN}}(x, e)$ is the possibility degree of the conjunction of x and e, especially since $X \cap E = \emptyset$. Another notation commonly used is $\Pi_{\mathcal{PN}}(x \wedge e)$.

In [2], it is stated that in the case of a *MAP* query, the problem can be reduced to finding the most plausible assignment of query variables Q compatible with the evidence e. More precisely, it can be rewritten as:

$$argmax_{q \in D_Q}(\Pi_{\mathcal{PN}}(q, e)). \quad (6)$$

Namely, given a possibilistic network $\mathcal{PN}$, Q the set of query variables and an evidence e (an instantiation of variables E), we have:

$$argmax_{q \in D_Q}(\Pi_{\mathcal{PN}}(q|e)) = argmax_{q \in D_Q}(\Pi_{\mathcal{PN}}(q, e)). \quad (7)$$

Simply put, the conditioning rule of possibility theory is not required to compute the maximum a posteriori assignment. In this section, we provide the full proof of Equation (7) of the above statement, stated in [2].

- Let us start with the min-based conditioning. Given a possibilistic network $\mathcal{PN}_m$ over V and let Q and E be two subsets of V (*s.t.* $Q \cap E = \emptyset$). Then, computing $argmax_{q \in D_Q}(\Pi(q|e))$ is equivalent to searching the instantiation q such that $\Pi(q|e) = 1$. By definition of the min-based conditioning, $\Pi(q|e) = 1$ if $\Pi(q, e) = \Pi(e)$. Assume that $argmax_{q \in D_Q}(\Pi(q, e))$ is q' then since $\Pi(e) = \max_{\omega \models e} \pi(\omega)$ or said otherwise $\Pi(e) = \max_{q \in D_Q} \Pi(q, e)$ which is given by $\Pi(q', e)$.
- Let us now consider product-based conditioning. In the same way, since the possibilistic network $\mathcal{PN}_*$ is normalised then $\forall e \in E$, $argmax_{q \in D_Q}(\Pi(q|e))$ is equivalent to searching the instantiation q such that $\Pi(q|e) = 1$. Which, by definition, is given by $\Pi(q|e) = \dfrac{\Pi(q, e)}{\Pi(e)}$, therefore, $\Pi(q|e) = 1$ if $\Pi(q, e) = \Pi(e)$. From there, assume that $argmax_{q \in D_Q}(\Pi(q, e))$ is q' then since $\Pi(e) = \max_{\omega \models e} \pi(\omega) = \Pi(q', e)$. Thus, $argmax_{q \in D_Q}(\Pi_{\mathcal{PN}}(q|e)) = argmax_{q \in D_Q}(\Pi_{\mathcal{PN}}(q, e))$.

Given this equivalence, we can focus only on the *MAP* problem redefined by Equation (6).

IV. *MAP* QUERYING PRODUCT-BASED POSSIBILISTIC NETWORKS

In [2] a computational complexity analysis of *MAP* queries in min-based and product-based possibilistic networks is provided. In particular, it is stated that the decision problem behind *MAP* querying is *NP*-complete for product-based possibilistic networks. The full proof of *NP*-hardness has been provided. This section provides the proof of *NP*-completeness theorem, stated in [2], of *MAP* querying a product-based possibilistic network. Let us first give a brief refresher on decision problems associated with *MAP* querying product-based possibilistic networks.

A. Definition of the decision problems

Let us recall the definition of the decision problem associated with a *MAP* query in product-based possibilistic networks, denoted π_***-D-MAP**.

Definition 1. *By* π_***-D-MAP**$(\mathcal{PN}_*, Q, e, t)$ *we denote the decision problem associated with MAP querying possibilistic networks that we define by:*
Input:

- $\mathcal{PN}_*$*: a product-based possibilistic network*
- e *(evidence): an instantiation of a set of variables* E
- Q *(query): a set of variables with* $Q \cap E = \emptyset$
- t*: a real number in* $(0, 1]$.

Question: Is there an instantiation q *of non observed variables* Q *such that* $\Pi_{\mathcal{PN}_\otimes}(q, e) \geq t$*?*

As said before, the decision problem we refer to in this reduction is the weighted MaxSAT problem. It is defined as follows:

Definition 2. *By* **D-WMaxSAT**(Ψ, k) *we denote the decision problem specified by:*
Input:

- Ψ*: a weighted CNF formula over boolean variables* $V = \{X_1, ..., X_n\}$ *simply represented by*
$$\Psi = \left\{ \begin{array}{c} (C_1, \alpha_1), \\ (C_2, \alpha_2), \\ ... \\ (C_m, \alpha_m). \end{array} \right\}$$
where $C_i's$ *are clauses and* $\alpha_i's$ *are positive integers.*
- k*: a positive integer*

Question: Is there an instantiation of variables V *such that the sum of weights of satisfied clauses in* Ψ *is greater or equal to* k*?*

B. From querying product-based possibilistic networks to WMaxSAT

1) Definition of a weighted CNF formula associated to a product-based possibilistic network: In what follows, we will reuse the same weighted CNF formula associated to a product-based possibilistic network and defined in [2].

Definition 3. *Let* $\mathcal{PN}_*$ *be a product-based possibilistic network over the set of boolean variables* $V = \{X_1, ..., X_n\}$. *Let* Q *be a subset of* V*,* $e = e_1, ..., e_l$ *be an instantiation of evidence variables* E *(with* $Q \cap E = \emptyset$*) and* t *be a threshold. Then* $\Psi_{\mathcal{PN}_*,Q,e,t}$ *is defined by:* $\Psi_R \cup \Psi_0 \cup \Psi_e$ *where*

$$\begin{array}{rcl} \Psi_R & = & \{(\neg x_i \vee \neg u_{ij}, \alpha_i) : \pi_{\mathcal{PN}_*}(x_i|u_{ij}) = 2^{-\alpha_i}\}, \\ \Psi_0 & = & \{(\neg x_i \vee \neg u_{ij}, M) : \pi_{\mathcal{PN}_*}(x_i|u_{ij}) = 0\}, \\ \Psi_e & = & \{(e_k, M) : k = 1, ..., l\}, \end{array} \quad (8)$$

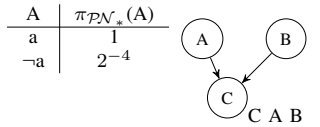
where $M > \sum\{\alpha_i : (\neg x_i \vee \neg u_{ij}, \alpha_i) \in \Psi_R\}$.

Example 1 illustrates Definition 3.

Example 1. *Let us consider the product-based possibilistic network* $\mathcal{PN}_*$ *of Figure 1. Let* $Q = \{B\}$ *be a subset of* V*, let* $e = \neg c$ *be an instantiation of evidence variables* $E = \{C\}$ *and let* $t = 2^{-2}$ *be the threshold.*

A	$\pi_{\mathcal{PN}_*}$(A)
a	1
¬a	2^{-4}

B	$\pi_{\mathcal{PN}_*}$(B)
b	2^{-8}
¬b	1



C A B	$\pi_{\mathcal{PN}_*}$(C|AB)
c a b	2^{-7}
¬c a b	1
c a ¬b	1
¬c a ¬b	2^{-2}
c ¬a b	0
¬c ¬a b	1
c ¬a ¬b	0
¬c ¬a ¬b	1

Fig. 1. Example of a product-based possibilistic network $\mathcal{PN}_*$ over A, B and C.

Let $M = 30$*. Then following Definition 3, the weighted CNF formula* $\Psi_{\mathcal{PN}_*,\{B\},\neg c,2^{-2}}$ *is*

$$\Psi_{\mathcal{PN}_*,\{B\},\neg c,2^{-2}} = \left\{ \begin{array}{c} (a, 4), \\ (\neg b, 8), \\ (\neg c \vee \neg a \vee \neg b, 7), \\ (c \vee \neg a \vee b, 2), \\ (\neg c \vee a \vee b, 30), \\ (\neg c \vee a \vee \neg b, 30), \\ (\neg c, 30) \end{array} \right\} \begin{array}{l} \Psi_R \\ \\ \\ \\ \Psi_0 \\ \\ \Psi_e \end{array}$$

2) Reduction from a product-based possibilistic network to a weighted CNF formula: Theorem 1, given in [2], provides the result that the decision problem π_***-D-MAP**$(\mathcal{PN}_*, Q, e, t)$ can be reduced into **D-WMaxSAT** $(\Psi_{\mathcal{PN}_*,Q,e,t}, k)$.

Theorem 1. *[2] Let* $\mathcal{PN}_*$ *be a product-based possibilistic network. Let* Q *be a subset of* V*,* e *be an instantiation of variables* E *and* t *be a threshold. Let* $\Psi_{\mathcal{PN}_*,Q,e,t}$ *be the CNF formula given by Definition 3. Then,* π_***-D-MAP**$(\mathcal{PN}_*, Q, e, t)$ *answers "yes" if and only if* **D-WMaxSAT**$(\Psi_{\mathcal{PN}_*,Q,e,t}, X + log_2\ t + M * (Z + |E|))$ *answers "yes" where* π_***-D-MAP** *is given by Definition 4 and* **D-WMaxSAT** *is given by Definition 2.*

In this section, we provide the full proof of Theorem 1 and illustrate it with examples.

Proof. Let us first define the parameters of the WMaxSAT decision problem, **D-WMaxSAT**$(\Psi_{\mathcal{PN}_*,Q,e,t}, k)$. Namely,

- $\Psi_{\mathcal{PN}_*,Q,e,t}$ is the weighted CNF formula given by Definition 3.
- k is the threshold for the problem and it is given by:

$$k = X + log_2 t + M * ((\sum \Pi_{\mathcal{PN}_*}(x_i|u_i) = 0) + 1) \quad (9)$$

where M is defined in Definition 3. And X is given by the sum of weights in Ψ_R: $X = \sum\{\alpha_i : (\neg x_i \vee \neg u_{ij}, \alpha_i) \in \Psi_R\}$.

The second part of the proof consists in showing that the two decision problems as defined are equivalent. Let the query associated to **D-WMaxSAT** be: Does **D-WMaxSAT**($\Psi_{\mathcal{PN}_*,Q,e,t}, X + log_2 t + M * (Z + 1)$) answer "yes"? More precisely, is there an instantiation of all variables that satisfies a subset of clauses in $\Psi_{\mathcal{PN}_*,Q,e,t}$ having the sum of the degrees of the satisfied clauses greater or equal to k?

Recall that π_***-D-MAP** decision problem is: Given an instantiation e of evidence variables, is there an instantiation q of query variables Q such that $\Pi(q, e) \geq t$?

For the sake of clarity, in this proof we simply write Ψ instead of $\Psi_{\mathcal{PN}_*,Q,e,t}$.

$\star$ Assume that **D-WMaxSAT**(Ψ, k) answers "yes". This means that there exists a subset $A \subseteq \Psi$ such that:

- $\{(\phi_i, \alpha_i) \in A\}$ is consistent and
- $\sum_{(\phi_i,\alpha_i)\in A} \alpha_i \geq k$

Note that we can state that $\{(e_k, M) : k = 1, ..., l\}$ is included in A. Indeed, if some (ϕ_i, M) of Ψ is not in A then $\sum_{(\phi_i,\alpha_i)\in A}$ cannot be greater than $M * (Z + |E|)$. Let us denote by $A^* = A \setminus \{(\phi_i, M) : (\phi_i, M) \in A\}$ then we can also state that:

- $\{(\phi_i, \alpha_i) \in A^*\}$ is consistent,
- $\sum_{(\phi_i,\alpha_i)\in A^*} \alpha_i \geq X + log_2\ t$

Let ω be a model of $\{\phi_i : (\phi_i, \alpha_i) \in A\}$ and $\{\phi_i : (\phi_i, \alpha_i) \in A^*\}$. Since $X = \sum\{\alpha_i : (\phi_i, \alpha_i) \in \Psi$ and $\alpha_i \neq M\}$. Then the latter equation implies that:

$$\sum\{\alpha_i : (\phi_i, \alpha_i) \notin A^*\} \leq -log_2\ t$$

This can be rewritten as:

$$\sum\{\alpha_i : (\phi_i, \alpha_i) \in \Psi \setminus A,\ \omega \nvDash \phi_i\} \leq -log_2\ t$$

It is enough now to consider the following immediate simplified inequalities to get the desirable result.

$$\begin{array}{rl}
\sum\{\alpha_i : (\phi_i, \alpha_i) \in \Psi \setminus A,\ \omega \nvDash \phi_i\} & \leq -log_2\ t \\
-\sum\{log_2 2^{-\alpha_i} : (\phi_i, \alpha_i) \in \Psi \setminus A,\ \omega \nvDash \phi_i\} & \leq -log_2\ t \\
-log_2(*\{2^{-\alpha_i} : (\phi_i, \alpha_i) \in \Psi \setminus A,\ \omega \nvDash \phi_i\}) & \leq -log_2\ t \\
-log_2(*\{2^{-\alpha_i} : \omega \nvDash \neg x_i \vee \neg u_{ij}\}) & \leq -log_2\ t \\
-log_2(*\{2^{-\alpha_i} : \omega \vDash x_i \wedge u_{ij}\}) & \leq -log_2\ t \\
-log_2\ \pi_{\mathcal{PN}_*}(\omega) & \leq -log_2\ t \\
\pi_{\mathcal{PN}_*}(\omega) & \geq t
\end{array}$$

with $\omega \vDash e$. Hence the answer to π_***-D-MAP**($\mathcal{PN}_*, Q, e, t$) is also "yes" by taking q such that $\omega \models q$.

$\star$ Assume that **D-WMaxSAT**($\Psi_{\mathcal{PN}_*,Q,e,t}, k$) answers "no". Then, for all consistent subset of clauses A that include Ψ_0 and Ψ_e we have

$$\sum\{\alpha_i : (\phi_i, \alpha_i) \in A\} < k.$$

Let us consider such a subset A_*. Let ω be a model of A^*, then following the same previous steps we have:

$$\begin{array}{rl}
\sum\{\alpha_i : (\phi_i, \alpha_i) \in \Psi \setminus A_*\ s.t\ \omega \nvDash \phi_i\} & > -log_2 t \\
-log_2(*\{2^{-\alpha_i} : \omega \nvDash \neg x_i \vee \neg u_{ij}\}) & > -log_2\ t \\
-log_2(*\{2^{-\alpha_i} : \omega \vDash x_i \wedge u_{ij}\}) & > -log_2\ t \\
-log_2\ \pi_{\mathcal{PN}_*}(\omega) & > -log_2\ t \\
\pi_{\mathcal{PN}_*}(\omega) & < t
\end{array}$$

with $\omega \vDash e$. Hence the answer to π_***-D-MAP**($\mathcal{PN}_*, Q, e, t$) is also "no". □

Example 2. *Let $Q = \{B\}$ and $E = \{C\}$ be the set of query variables and evidence variables respectively. Let us consider the evidence $e = \neg c$. Let $\Psi_{\mathcal{PN}_*,Q,e,t}$ be the weighted CNF formula associated to $\mathcal{PN}_*$ given by Definition 3. The MAP query over $\mathcal{PN}_*$ is: Is there an instantiation q of the variables Q such that $\Pi_{\mathcal{PN}_*}(q, e) \geq 2^{-2}$. Hence, the corresponding problem* **D-WMaxSAT**$(\Psi_{\mathcal{PN}_*,Q,e,t}, k)$ *is given by: Is there an instantiation of the variables such that the sum of the degrees of the satisfied clauses is greater or equal to k?*

Let us set the values of the variables X, M and Z: $X = 21$, $M = 30$, and $Z = 2$. Then, $k = X + log_2 t + 30 * (Z + 1) = 109$. Given this configuration, **D-WMaxSAT**$(\Psi_{\mathcal{PN}_*,\{B\},\neg c,2^{-2}}, 109)$ *answers "yes". Indeed, it is enough to consider A such that*

$$A = \left\{ \begin{array}{c} (a, 4), \\ (\neg b, 8), \\ (\neg c \vee \neg a \vee \neg b, 7), \\ (\neg c \vee a \vee b, 30), \\ (\neg c \vee a \vee \neg b, 30), \\ (\neg c, 30) \end{array} \right\}$$

The sum of the weights in A is equal to 109. A model of formulas in A can be $a\neg b\neg c$ for which using the product-based chain rule has a possibility degree of $\Pi_{\mathcal{PN}_*}(a\neg bc) = 2^{-2}$. Hence, π_*-**D-MAP**$(\mathcal{PN}_*, \{B\}, \neg c, 2^{-2})$ *answers "yes" as well.*

C. Complexity of MPE inference

In this subsection, we analyse the complexity of *MPE* inference in product-based possibilistic networks. As we have mentioned in the definition of the query, we search for the assignment of **all** variables compatible with the evidence. Which means that the only difference with a *MAP* query as redefined in Equation (6) is that instead of a subset of variables we use all of them.

Based on what has just been proven for *MAP* inference, we argue that by choosing a set of query variables corresponding to the remaining variables $V \setminus E$, the complexity results follows from Theorem 1. This is formally stated in the following proposition.

Proposition 1. π_***-D-MPE** *is NP-complete.*

We provide a more rigorous proof for *MPE* inference in min-based possibilistic networks in the next section.

V. Complexity analysis of *MPE* inference in min-based possibilistic networks

This section focuses on *MPE* query in min-based possibilistic networks. Contrary to *MAP* inference, where the complexity analysis have shown that *MAP* inference in possibilistic networks costs less than in Bayesian networks; Here we show that *MPE* querying a possibilistic network is *NP*-complete as in Bayesian networks.

Basically, to prove that π_m**-D-MPE** is *NP*-complete, we follow the same steps as in the analysis of *MAP* in min-based possibilistic networks given in [2]. More precisely,

- we first show the *NP*-hardness of π_m**-D-MPE** by providing a reduction from the **D-3SAT** decision problem to π_m**-D-MPE** decision problem. In particular, in this reduction we build a special possibilistic network that only takes into account boolean variables and binary values. This network is called a binary and boolean possibilistic network (see [2] for more details).
- we provide a reduction of the π_m**-D-MPE** decision problem, defined for min-based possibilistic networks, to the **D-SAT** decision problem.

A. Definition of the decision problems

We first formally define the decision problem associated with a *MPE* query in min-based possibilistic networks, denoted π_m**-D-MPE** as well as the decision problem associated with querying a B&B possibilistic network.

Definition 4. *We denote π_m-**D-MPE**$(\mathcal{PN}_m, e, t)$ the decision problem associated with MPE querying a* min*-based possibilistic network. It is defined by:*
Input:

- *$\mathcal{PN}_m$: a* min*-based possibilistic network*
- *e (evidence): an instantiation of a set of variables E*
- *t: a real number in $(0, 1]$.*

Question: Is there an instantiation x of the variables X such that $\Pi_{\mathcal{PN}_m}(x, e) \geq t$?

Definition 5 considers a particular case of Definition 4 where degrees are either 0 or 1 (hence no need to explicitly specify t).

Definition 5. *By **B&B**$_m$-**D-MPE**$(\mathcal{PN}_{B\&B_m}, e)$ we denote the decision problem associated with MPE querying a* min*-based Boolean and Binary possibilistic network that we define by:*
Input:

- *$\mathcal{PN}_{B\&B_m}$: a* min*-based binary and boolean possibilistic network over $V = \{X_1, ..., X_n\}$*
- *e (evidence): an instantiation of a set of observation variables E*

Question: Is there an instantiation x of variables X such that $\Pi_{\mathcal{PN}_{B\&B_m}}(x, e) = 1$?

In [2], the authors gave the reasons why we can afford to only consider the case $\Pi_{\mathcal{PN}_{B\&B_m}}(x, e) = 1$.

Let us now recall the boolean satisfiability decision problem denoted **D-SAT**.

Definition 6. *By **D-SAT**(Ψ) we denote the decision problem associated to determining if there exists an assignment ω that satisfies Ψ. It is defined by:*
Input: Ψ a formula in a conjunctive normal form
Question: Is Ψ satisfiable?

A restricted version of the SAT problem involves a 3CNF and is called **D-3SAT** decision problem. A 3CNF is a formula in a conjunctive normal form for which each clause is a disjunction of at most 3 literals. Thus,

Definition 7. *By **D-3SAT**(Ψ) we denote the decision problem defined by:*
Input: Ψ a 3CNF formula
Question: Is Ψ satisfiable?

B. From 3SAT to MPE querying over B&B possibilistic networks

As in [2], we build a B&B possibilistic network from a 3CNF. The definition of this reduction is recalled by Definition 8. For more details, the reader can refer to [2].

Definition 8. *Let $\Psi = C_1 \wedge C_2 \wedge ... \wedge C_m$ be a 3CNF formula. Let $V = \{X_1, ..., X_n\}$ be the set of propositional variables appearing in Ψ. The B&B possibilistic network associated with Ψ, denoted by $\mathcal{PN}_\Psi$ is defined as follows:*

1) *For each propositional symbol X_i appearing in Ψ, we add a boolean node variable in the graph. Each variable X_i is associated with a possibility distribution given by: $\pi_{\mathcal{PN}_\Psi}(x_i) = 1$ and $\pi_{\mathcal{PN}_\Psi}(\neg x_i) = 1$.*
2) *For each clause C_j of Ψ, we add a conditional node variable, C_j. Parents of C_j are the rooted variables X_i that are involved in C_j. Each C_j is associated with a conditional possibility distribution given by: $\forall u_{jk}$ an instance of parents of C_j that models the satisfiability of the clause.*
3) *Lastly, we add a single boolean node, E_Ψ, which represents the satisfiability of the overall formula Ψ. It has all nodes $C'_j s$ as parents. Intuitively, E_Ψ is set to true if all clauses are satisfied.*

Theorem 2 provides the reduction from the decision problem **D-3SAT**(Ψ) into **B&B**$_m$**-D-MPE**$(\mathcal{PN}_\Psi, e)$ where the input e is let to e_Ψ. More formally:

Theorem 2. *Let Ψ be a 3CNF formula. Let $\mathcal{PN}_\Psi$ be the B&B possibilistic network given by Definition 8. Let $V_{\mathcal{PN}_\Psi}$ be the set of variables in $\mathcal{PN}_\Psi$, namely $\{X_1, ..., X_n\} \cup \{C_1, ..., C_m\} \cup \{E_\Psi\}$. Then,* **D-3SAT**$(\Psi)$ *answer is "yes" if and only if the* **B&B**$_m$**-D-MPE**$(\mathcal{PN}_\Psi, e_\Psi)$ *answers "yes" where* **D-3SAT** *is given in Definition 7 and* **B&B**$_m$**-D-MPE** *is given by Definition 5.*

C. From querying min-based possibilistic networks to SAT

In this section, we no longer restrict ourselves to binary possibility distributions. Namely, (conditional) possibility degrees can take any value in the unit interval $[0, 1]$. However, for the sake of simplicity, we still only consider boolean variables. This is not a restriction and the proof can be adapted by encoding a non-boolean variable by a set of boolean variables. We propose to reduce the decision problem π_m**-D-MPE** to the decision problem **D-SAT**.

The following gives the definition of the CNF formula associated with the network a MPE query, denoted by $\Psi_{\mathcal{PN}_m,e,t}$.

Definition 9. *Let $\mathcal{PN}_m$ be a* min*-based possibilistic network over the set of boolean variables $V = \{X_1, ..., X_n\}$. Let $e = e_1, ..., e_l$ be an instantiation of evidence variables E and let t be a threshold. Then $\Psi_{\mathcal{PN}_m,e,t}$ over the set of variables $V = \{X_1, ..., X_n\}$, is given by:*

$$\begin{array}{rcl} \Psi_{\mathcal{PN}_m,e,t} & = & \{(\neg x_i \vee \neg u_{ij}) : \pi_{\mathcal{PN}_m}(x_i|u_{ij}) < \mathbf{t}\} \\ & \cup & \{\mathbf{e_k} : \mathbf{k} = \mathbf{1}, ..., \mathbf{l}\} \end{array}$$

This reduction is done in polynomial time (and space) with respect to the size of $\mathcal{PN}_m$.

The following theorem states that π_m**-D-MPE** can be reduced to **D-SAT**.

Theorem 3. *Let $\mathcal{PN}_m$ be a* min*-based possibilistic network, e be an instantiation of evidence variables E and t be a real number in $(0, 1]$. Let $\Psi_{PN_m,e,t}$ be the CNF formula given by Definition 9. Then, π_m-**D-MPE**$(\mathcal{PN}_m, e, t)$ says "yes" if and only if* **D-SAT**$(\Psi_{\mathcal{PN}_m,e,t})$ *says "yes" where π_m-**D-MPE** is given by Definition 4 and* **D-SAT** *is given by Definition 6.*

Proof. The proof is similar to the one provided in [2] for analyzing the complexity of map-querying possibilistic networks. We need to prove that when $\Psi_{\mathcal{PN}_m,e,t}$ is satisfiable then $\Pi_{\mathcal{PN}_m}(x, e) \geq t$ and that when $\Psi_{\mathcal{PN}_m,e,t}$ is unsatisfiable then $\Pi_{\mathcal{PN}_m}(x, e) < t$ for all assignments of all variables compatible with e.

- Assume that $\Psi_{\mathcal{PN}_m,e,t}$ is satisfiable. This means that there exists an instantiation of all variables, denoted by ω^*, that satisfies all clauses of $\Psi_{\mathcal{PN}_m,e,t}$ including $e = e_1, ..., e_l$. Then we have $\pi_{\mathcal{PN}_m}(x_i|u_{ij}) < t$ by construction of $\Psi_{\mathcal{PN}_m,e,t}$. So if ω^* satisfies all clauses in $\Psi_{\mathcal{PN}_m,e,t}$ then ω^* falsifies each of the formulas in $\{(x_i \wedge u_{ij}) : (\neg x_i \vee \neg u_{ij}) \in \Psi_{\mathcal{PN}_m,e,t}\}$. Thus, all conditionals $\pi_{\mathcal{PN}_m}(x_i|u_{ij})$ applied in chain rule to compute $\pi_{\mathcal{PN}_m}(\omega^*)$ have a possibility degree greater or equal to t. Therefore, $\pi_{\mathcal{PN}_m}(\omega^*) \geq t$. Hence the answer to π_m-**D-MPE**$(\mathcal{PN}_m, e, t)$ is also "yes".
- Assume that $\Psi_{\mathcal{PN}_m,e,t}$ is unsatisfiable. Then for all instantiation of variables ω such that $\omega \models e(= e_1 \wedge .. \wedge e_l)$, there exists at least a clause $C_i = \neg x_i \vee \neg u_{ij}$ that is falsified by ω (and hence $\omega \models x_i \wedge u_{ij}$). And by construction of $\Psi_{\mathcal{PN}_m,e,t}$, we have $\pi_{\mathcal{PN}_m}(x_i|u_{ij}) < t$, so using the min-based chain rule we have $\forall \omega \models e$, $\pi_{\mathcal{PN}_m}(\omega) < t$. Hence π_m-**D-MPE**$(\mathcal{PN}_m, e, t)$ is also "no".

□

To summarise Proposition 1 together with Theorems 2 and 3 show that the decision problem associated with MPE inference is *NP*-complete for both min-based and product-based possibilistic networks.

VI. Conclusions

This paper dealt with the computational complexity of inference in possibilistic networks. The main results shown in this work proved that possibilistic networks offer interesting advantages for reasoning with uncertain information. These results have an impact on the complexity of reasoning in the interval-based possibilistic setting. A future work concerns the computation of the a posteriori possibility degree of an event in both min-based and product-based possibilistic networks.

VII. Acknowledgements

This work is supported by the european project H2020 Marie Sklodowska-Curie Actions (MSCA) research and Innovation Staff Exchange (RISE): AniAge (High Dimensional Heterogeneous Data based Animation Techniques for Southeast Asian Intangible Cultural Heritage).

References

[1] Nahla Ben Amor, Didier Dubois, Héla Gouider, and Henri Prade. Preference modeling with possibilistic networks and symbolic weights: A theoretical study. In *ECAI 2016 - , The Hague, The Netherlands*, pages 1203–1211, 2016.

[2] Salem Benferhat, Amélie Levray, and Karim Tabia. Possibilistic networks: MAP query and computational analysis. In *IEEE 30th ICTAI, 5-7 November 2018, Volos, Greece.*, pages 916–923. IEEE, 2018.

[3] Christian Borgelt, Jörg Gebhardt, and Rudolf Kruse. *Possibilistic graphical models*. Springer, 2000.

[4] Christian Borgelt and Rudolf Kruse. *Graphical models - methods for data analysis and mining*. Wiley, 2002.

[5] F. G. Cozman. Credal networks. *Artificial Intelligence*, 120(2):199 – 233, 2000.

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

[7] Adnan Darwiche and James D. Park. Complexity results and approximation strategies for MAP explanations. *CoRR*, abs/1107.0024, 2011.

[8] Cassio P. De Campos. New complexity results for map in bayesian networks. In Toby Walsh, editor, *IJCAI*, pages 2100–2106. IJCAI/AAAI, 2011.

[9] Cassio P. De Campos and Fábio G. Cozman. The inferential complexity of bayesian and credal networks. In *Proceedings of the 19th IJCAI, Edinburgh, Scotland, UK, July 30 - August 5*, pages 1313–1318. Professional Book Center, 2005.

[10] Didier Dubois. Possibility theory and statistical reasoning. *Computational statistics & data analysis*, 51(1):47–69, 2006.

[11] Didier Dubois and Henri Prade. *Possibility Theory: Qualitative and Quantitative Aspects*, pages 169–226. Springer Netherlands, Dordrecht, 1998.

[12] Laurent Garcia and Régis Sabbadin. Complexity results and algorithms for possibilistic influence diagrams. *Artif. Intell.*, 172(8-9):1018–1044, 2008.

[13] Maroua Haddad, Philippe Leray, Amélie Levray, and Karim Tabia. Learning the parameters of possibilistic networks from data: Empirical comparison. In *Proceedings of the 30th International FLAIRS Conference 2017, Marco Island, Florida.*, pages 736–741. AAAI Press, 2017.

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