Approximating MAP inference in credal networks using probability-possibility transformations

Salem Benferhat, Amélie Levray, Karim Tabia
Univ Lille Nord de France, F-59000 Lille, France
UArtois, CRIL UMR CNRS 8188, F-62300 Lens, France
Email: {benferhat, levray, tabia}@cril.fr

***Abstract*—This paper focuses on belief graphical models and provides an efficient approximation of MAP inference in credal networks using probability-possibility transformations. We first present two transformations from credal networks to possibilistic ones that are suitable for MAP inference in credal networks. Then we present four criteria to evaluate our approximate MAP inference. The last part of the paper provides experimental studies that compare our approach with both standard exact and approximate MAP inference in credal networks. The paper also provides a brief analysis of MAP inference complexity using possibilistic networks and the results definitely open new perspectives for MAP inference in credal networks.**

I. Introduction

MAP (**M**aximum **A** **P**osteriori) inference in probabilistic graphical models is a problem of great interest and has been investigated for years [13], [14], [19], [21], [22]. Thus, there exists a variety of methods and algorithms to compute the configuration of query variables with the highest probability given some observed variables. However, Bayesian networks, which are the most widely used probabilistic graphical models, might seem unfit for representing some kinds of information such as the knowledge of a group of experts, or incomplete knowledge. This is why more general frameworks are needed for allowing more flexibility especially regarding the model parameters. Credal networks [4] have been designed to generalize Bayesian networks and offer more expressiveness as they represent uncertain information by means of credal sets instead of single probability values. The problem when reasoning with such general and expressive models is that they entail higher computational complexity. Methods and algorithms to compute MAP inference in credal networks exist and give good results in terms of accuracy [17]. However, these methods are not very efficient in terms of computational complexity especially when dealing with problems having many variables.

The aim of this paper is to provide a new and efficient method for MAP inference in credal networks based on probability-possibility transformations. Different probability-possibility transformations have been proposed in the literature [2], [3], [6], [8], [12], [23], [24]. In this paper, we focus on two transformations from credal network to possibilistic networks that are suitable for MAP inference. This paper also provides a brief complexity analysis of MAP inference in possibilistic network and performs an extensive experimental study on MAP inference using probability-possibility transformations.

The first part of this paper gives the general context of this study by recalling the basic notions of graphical models used in this paper, the definition of MAP inference in credal networks, and the definition of the criteria used to compute the results of MAP requests. The second part introduces our approach, the transformations and the used evaluation criteria. Lastly, the paper presents the experimental study and provides an analysis and a discussion of the results.

II. A brief refresher on credal and possibilistic networks

Let us briefly present in this section the main belief graphical models we are dealing with, namely standard Bayesian networks, credal networks and possibilistic networks.

A. Bayesian networks

Bayesian networks ($\mathcal{BN}$) are well-known probabilistic graphical models [5], specified by two components:

- a *graphical component*: a directed acyclic graph (DAG) with nodes representing variables A_i and edges encoding (in)dependence relationships between variables.
- a *quantitative component*: where each variable A_i is associated with a local probability distribution $p(A_i|par(A_i))$ for each variable A_i in the context of its parents, denoted $par(A_i)$.

This representation, by means of graphical models, allows to compactly represent a probability distribution over a set of variables. The joint probability distribution encoded by a $\mathcal{BN}$ is computed using the chain rule:

$$p(A_1, .., A_n) = \prod_{i=1}^{n} p(A_i|par(A_i)). \qquad (1)$$

B. Credal networks

Credal networks [4], [18] are also probabilistic graphical models, based on imprecise probability theory [15], [26]. A key notion in this theory is the one of credal set which is often interpreted as a set of imprecise beliefs, in the sense that the true probability measure (if it exists) is in this set but there is no way to determine it exactly due to lack of knowledge. In order to characterize a credal set, one can use a set of extreme points [20], probability intervals or linear constraints. In this paper, we use interval-based probability distributions (IPD for short) which are a very natural and common way to specify imprecise and ill-known information. In an imprecise probability distribution IP, every state of the world $\omega_i \in \Omega$ is associated with a probability interval $IP(\omega_i) = [l_i, u_i]$ where l_i (resp. u_i) denotes the lower (resp. upper) bound of the probability of ω_i. Note that while a standard probability distribution p induces a complete order over the set of possible worlds Ω, an imprecise probability distribution IP may induce a partial order since some interpretations may be incomparable in case of overlapping intervals.

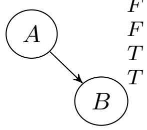
Definition 1 (Credal network)**.** *A credal network $\mathcal{CN} = \langle G, K \rangle$ is a probabilistic graphical model where*

- *$G = \langle V, E \rangle$ is a directed acyclic graph as for $\mathcal{BN}$ with $V = \{A_1, .., A_n\}$ is the set of variables and E is the set of edges.*
- *$K = \{IP_1, IP_2, .., IP_n\}$ is a collection of local IPDs, each IP_i is associated with the variable A_i in the context of its parents $par(A_i)$.*

Note that in practice, in local tables, one can also specify a set of extreme points instead of an imprecise probability distribution as in JavaBayes[1] software. Regarding the semantics of credal networks, the most common one is to view a credal network $\mathcal{CN}$ as an encoding of a set of joint probability distributions, called extensions where each distribution p is encoded by a compatible Bayesian network $\mathcal{BN}$ [4].

Example 1. *This is a small example of a credal network $\mathcal{CN}$. In the third column of the local distributions, there is an example of a compatible distribution of the imprecise distribution IP.*

A	$IP(A)$	$p(A)$
F	[.5, .9]	.7
T	[.1, .5]	.3



A	B	$IP(B\|A)$	$p(B\|A)$
F	F	[.36, .4]	.4
F	T	[.35, .35]	.35
T	F	[.25, .25]	.25
T	T	[0, .04]	0

Fig. 1. Example of a credal network $\mathcal{CN}$ and a compatible Bayesian network.

C. Possibilistic networks

Possibilistic networks [25] are belief graphical model models based on possibility theory, alternative uncetainty theory particularly suited for handling incomplete and qualitative knowledge. A possibilistic network $\mathcal{PN} = \langle G, \Theta \rangle$ involves:

- *A graphical component* G which is a DAG exactly as in Bayesian networks.
- *A numerical component* Θ, a set of local possibility distributions $\pi(A_i|par(A_i))$ associated with each variable A_i in the context of its parents.

The semantics associated to a possibilistic network is a possibility distribution π, matching every world $\omega_i \in \Omega$ to a real number in $[0, 1]$. Contrary to probability theory, possibility degrees can be seen as upper probabilities, consonant plausibility functions, degrees of potential surprise, etc. [10]. Possibility theory in this sense is an alternative theory to represent and handle uncertainty. Nonetheless, those two theories can be related as a possibility measure can encode a family a probability measures and which can be seen as a credal set.

In the quantitative[2] possibilistic setting, the joint possibility distribution is factorized using the following possibilistic counterpart of the chain rule:

$$\pi(a_1, a_2, .., a_n) = *_{i=1}^{n}(\pi(a_i|par(a_i))) \qquad (2)$$

where $*$ denotes here the product operator (for more details, see [9]).

D. MAP inference in $\mathcal{CN}$

Inference in probabilistic graphical models generally consists in computing the probability of an event. In credal networks, this equivalently comes down to computing lower or upper probabilities of an event. Let $V = \{A_1..A_n\}$ be the set of variables of the model. Let $\mathcal{O} \subseteq V$ be the set of observed variables and let $o \in \mathcal{O}$ be an instantiation (or configuration). Let also $\mathcal{Q} \subseteq V$ be the set of query variables and $q \in \mathcal{Q}$. For **M**aximum **A** **P**osteriori (*MAP*), given a assignment o of observed variables $\mathcal{O}$, the objective is to compute the most probable instantiation q of the query variables $\mathcal{Q}$. In general, $\mathcal{Q} \cap \mathcal{O} = \emptyset$. Note that when $\mathcal{Q}$ and $\mathcal{O}$ span over all variables, the problem is known as the most probable explanation (*MPE*).

More formally, the *MAP* inference problem comes down to compute:

$$argmax_{q \in D_{\mathcal{Q}}}(IP(q \mid o)) \qquad (3)$$

where $argmax$ denotes a decision criterion allowing to choose the set of "most probable configurations" of query variables. In the following, we will give some of the most used decision criteria that can be used for answering *MAP* requests in credal networks.

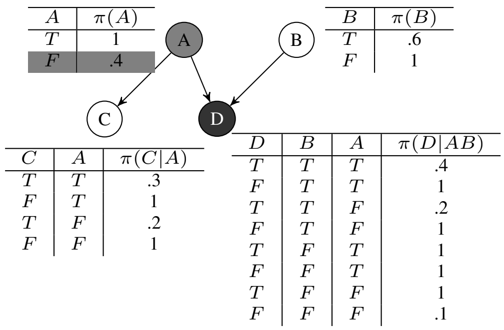
Example 2. *Let us see an example of MAP inference. In this case, over a possibilistic network to make it simpler. Consider the following possibilistic network (Figure 2) over the set of variables $V = \{A, B, C, D\}$. In this example, we want to compute the MAP request on D given that $A = F$. Many algorithms exist to answer this query, like variable elimination or junction tree algorithm. Here, if we compute* $\Pi(D{=}T|A{=}F){=}1$ *and* $\Pi(D{=}F|A{=}F){=}.6$*, then the result of MAP query over the variable D given A is $D{=}T$.*

We need decision criteria to answer *MAP* queries in credal networks due to the representation by means of intervals. A natural criterion is the one of *Interval-dominance* (used for instance in [1] for classification, decision tasks, etc.) which refers to non-dominated instantiations of query variables.

Definition 2 (*Interval-dominance*)**.** *An instantiation q_i of query variables $\mathcal{Q}$ dominates another instantiation q_j iff*

[1] http://www.cs.cmu.edu/~javabayes/Home/

[2] In this paper, we interpret possibility degrees as upper probabilities, hence the use of the product operator.



A	$\pi(A)$
T	1
F	.4

B	$\pi(B)$
T	.6
F	1

C	A	$\pi(C\|A)$
T	T	.3
F	T	1
T	F	.2
F	F	1

D	B	A	$\pi(D\|AB)$
T	T	T	.4
F	T	T	1
T	T	F	.2
F	T	F	1
T	F	T	1
F	F	T	1
T	F	F	1
F	F	F	.1

Fig. 2. Example of a possibilistic network

$\underline{IP}(q_i|o)>\overline{IP}(q_j|o)$ *where* o *is an instantiation of observation variables* $\mathcal{O}$.

This criterion is not enough informative and often of little use in practice. Indeed, this criterion often results in large amounts of outcomes also called answer set (too many query variable instances are not dominated), making it difficult to make decisions for instance in classification where only one class (outcome) should be returned. In the current experiments, we propose to reduce the number of outcomes returned by *MAP* queries using other criteria.
We use the well-known criteria *Maximax*, *Maximin*, and *Hurwicz*. These three criteria are commonly used in decision making under uncertainty since the early 1950's. The *Maximax* criterion can be viewed as an optimistic criterion. It examines the maximum payoffs of alternatives and chooses the alternative whose outcome is the best. Definition 3 gives a formal definition of *Maximax* criterion for the imprecise probability setting.

Definition 3 (*Maximax* criterion)**.** *An instantiation* q_i *of query variables* $\mathcal{Q}$ *is a result of MAP inference iff* $\overline{IP}(q_i|o)\geq\max\{1-\sum_{q_j\neq q_k}\underline{IP}(q_j|o),\forall q_k\}$*, where* o *is an instantiation of observed variables* $\mathcal{O}$.

The *Maximin* criterion also known as the Walds *Maximin* criterion is a pessimistic criterion. It suggests that the decision maker examines only the minimum payoffs of alternatives and chooses the alternative whose outcome is the least worst.

Definition 4 (*Maximin* criterion)**.** *An instantiation* q_i *of query variables* $\mathcal{Q}$ *is a result of MAP inference iff* $\underline{IP}(q_i|o)\geq\max\{1-\sum_{q_j\neq q_k}\overline{IP}(q_j|o),\forall q_k\}$*, where* o *is an instantiation of observed variables* $\mathcal{O}$.

The last criterion we review is the well-known *Hurwicz* criterion which attempts to find a trade-off between the extremes, posed by the optimistic and pessimistic criteria, by assigning a certain weight, a to optimism and the balance $1-a$ to pessimism. This index reflects the decision maker attitude towards risk taking. A cautious decision maker will set $a=1$ which reduces the *Hurwicz* criterion to the *Maximin* criterion. An adventurous decision maker will set $a=0$ which reduces the *Hurwicz* criterion to the *Maximax* criterion.

Definition 5 formally defines the *Hurwicz's* criterion with imprecise probabilities using the coefficient $a=0.5$.

Definition 5 (*Hurwicz's* criterion)**.** *Let* q_i *be an instantiation of query variables* $\mathcal{Q}$ *with* o *an instantiation of observed variables of* $\mathcal{O}$, $a=\{0.5*(1-\sum_{q_i\neq q_k}\overline{IP}(q_i|o))+0.5*(1-\sum_{q_i\neq q_k}\underline{IP}(q_i|o)),\forall q_k\}$*. Then* q_i *is a result of MAP inference iff* $a=\max_{\forall q_j}\{0.5*(1-\sum_{q_j\neq q_k}\overline{IP}(q_j|o))+0.5*(1-\sum_{q_j\neq q_k}\underline{IP}(q_j|o)),\forall q_k\}$.

Example 3. *Let us show an example of these criteria over the following imprecise distribution (Table I).*

ω	$IP(\omega)$
ω_1	$[.25;.3]$
ω_2	$[.27;.32]$
ω_3	$[.26;.33]$
ω_4	$[.07;.12]$

TABLE I
EXAMPLE OF AN IMPRECISE DISTRIBUTION

On this example, clearly the only world that can be excluded from the results with the interval-dominance criterion is ω_4*. The outcomes that can be obtained using the different criteria are listed in the following items.*

- *Interval-dominance: Answer set =* $\{\omega_1,\ \omega_2,\ \omega_3\}$,
- *Maximax: Answer set =* $\{\omega_3\}$,
- *Maximin: Answer set =* $\{\omega_2\}$,
- *Hurwicz: Answer set =* $\{\omega_2,\ \omega_3\}$.

As shown in this example, one of the main problems of *MAP* inference in credal networks is that the number of outcomes may be very large especially when the interval-dominance criterion is used. The second big problem is the one of computational complexity of *MAP* inference in credal networks.

E. Complexity of MAP inference in credal networks

The computational complexity of *MAP* inference in credal networks have been studied in [17]. To sum up, *MAP* inference in credal network has been established to be NP^{PP}-Hard for multiply-connected network and Σ_2^P-Complete for polytrees. This extra computational cost in comparison with Bayesian networks (where *MPE* problem is *NP*-complete and *MAP* is NP^{PP}-complete in multiply-connected networks) is not surprizing since in credal networks, there is need to deal with both upper and lower bounds.

As mentioned in [25], there is no study on the complexity of *MAP* inference in possibilistic networks. We can safely assume that inference in quantitative possibilistic networks is not worse than in Bayesian ones. In fact, answering queries comes down to applying the chain rule and marginalization in both Bayesian networks and possibilistic ones. Moreover, some probabilistic network inference algorithms like variable elimination and the junction tree algorithm have been adapted from the probabilistic setting and seem to show the same complexity.

In practice, the size of credal networks is often large and given the high complexity of *MAP* inference, it is then fundamental to have approximate *MAP* inference ensuring a good compromise between accuracy and computational complexity. This paper proposes a new approximate inference method for *MAP* in $\mathcal{CN}$s by transforming a credal network into a possibilistic one $\mathcal{PN}$. This transformation will keep as much as possible the information encoded by the credal network but then there is no need to deal with upper and lower bounds since a possibilistic network encodes a unique possibility distribution. Of course, one could select (using some criteria) a Bayesian network $\mathcal{BN}$ that is compatible with the credal network $\mathcal{CN}$ and use the $\mathcal{BN}$ to answer queries, but we want to empirically assess our approach based on probability-possibility transformations which is well principled and could ensure better tradeoff between accuracy and computational efficiency. The next section deals with the transformation of a credal network into a possibilistic one.

III. A PROBABILITY-POSSIBILITY TRANSFORMATION BASED APPROACH

Several probability-possibility transformations have been proposed as we recall in the following. Methods generalizing to imprecise probabilities have been proposed by Masson and Denoeux [16] and others in [6]. In this section, we present two of such transformations that are appropriate for approximating MAP inference in credal networks.

A. Probability-possibility transformations

Probability and possibility theories have both been deeply studied and some bridges have been proposed to link these two settings [12], [27]. We now have some transformations passing from probability theory to possibility theory and vice versa. Dubois and Prade [8] have, for instance, proposed the Optimal Transformation (**OT**) which is defined as:

$$\pi(\omega_i) = \sum_{j/p(\omega_j)\leq p(\omega_i)} p(\omega_j) \tag{4}$$

Transformations are required to satisfy basic principles to preserve as much as possible the information and **OT** is proven to be the optimal one satisfying such principles. More works on transformations can be found in [2], [3], [6], [8], [12], [23], [24]. Turning a probability measure into a possibilistic one is useful when dealing with weak sources of information, or even when computing with possibilities is simpler than with probabilities as claimed in [11].

B. From interval-based probability distributions to possibilistic ones

The first transformation we study is the one of Masson and Denoeux [16], where the authors learn possibility distributions from empirical data by transforming confidence intervals into possibility distributions. The first point is to consider an imprecise probability distribution as a means of encoding a partial order $\mathcal{M}$ over Ω. Let $\mathcal{M}$ be the partial order encoded by an imprecise probability distribution IP and let $\mathcal{C}$ be the set of linear extensions (complete orders) that are compatible with the partial order $\mathcal{M}$. MD transformation proceeds as follows. For each linear extension $\mathcal{C}_l \in \mathcal{C}$ and for each interpretation $\omega_i \in \Omega$, we find the compatible probability distribution which will give the most specific possibility distribution when transforming with **OT**.

$$\pi^{C_l}(\omega_i) = \max_{p_1..p_n}\big(\sum_{p_j\leq p_i} p_i\big) \tag{5}$$

Indeed, MD transformation can be reduced to **OT** when we consider single values instead of intervals.

Then, to compute the possibility distribution taking into account each possibility distribution built for each linear extension, for each interpretation we use the maximum value of this interpretation in the set of possibility distributions.

$$\pi(\omega_i) = \max_{\mathcal{C}_l\in\mathcal{C}}(\pi^{C_l}(\omega_i)) \tag{6}$$

This transformation tries on one hand to preserve the order of interpretations induced by IP and the dominance principle requiring that $\forall\phi\subseteq\Omega$, $P(\phi)\leq\Pi(\phi)$ on the other hand.

The second transformation, called *CD* stands for *Cumulative Distribution*, is related to upper and lower cumulative distributions. In the current work, we transform an imprecise probability distribution into two possibility distributions. In [7], the authors discussed the connection that one can make between generalized p-box and possibility distributions and gave a representation of a p-box by two possibility distributions. Given a set of probability intervals and an ordering relation $\leq_{\mathcal{C}_I}$ on a linear extension $\mathcal{C}_l$ between elements ω_i, we can easily build a generalized p-box [7], $[\underline{F}, \overline{F}]$ defined by two cumulative distributions $\underline{F}$ and $\overline{F}$. Given the consecutive sets $A_i = \{\omega_i, \forall\omega_i \in \Omega \text{ and s.t. } \omega_i \leq_{\mathcal{C}_I} \omega_j \text{ iff } i < j\}$, lower and upper generalized cumulative distributions corresponding to Ω are, respectively:

$$\underline{F}(\omega_i) = \underline{P}(A_i) = \max\Big(\sum_{\omega_i\in A_i} l_j, 1 - \sum_{\omega_i\notin A_i} u_j\Big)$$

$$\overline{F}(\omega_i) = \overline{P}(A_i) = \min\Big(\sum_{\omega_i\in A_i} u_j, 1 - \sum_{\omega_j\notin A_i} l_j\Big)$$

From this two cumulative distributions, we can compute two possibility distributions $\pi_{\overline{F}}$ and $\pi_{\underline{F}}$ where:

$$\pi_{\underline{F}}(\omega_i) = 1 - \max\{\underline{F}(\omega_j) < \underline{F}(\omega_i) : j = 0..n\} \tag{7}$$

$$\pi_{\overline{F}}(\omega_i) = \overline{F}(\omega_i) \tag{8}$$

These two equations are written as they have been defined in [7]. But as for the use of $\pi_{\underline{F}}$, we will simply consider as a possibility distribution, the lower generalized cumulative distribution $\pi_{\underline{F}}(\omega_i) = \underline{F}(\omega_i)$ and we normalize it. Now let us now see how to apply these transformations on credal networks.

C. From credal networks to possibilistic networks

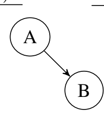
A direct method to transform a credal network into a possibilistic one is to transform only local probability tables into local possibilistic ones. This has the advantage of preserving the independence relationships.

Definition 6 (Credal-possibilistic network transformation)**.** *Let $\mathcal{CN}$ be a credal network, $\mathcal{PN}_{\mathcal{CN}}$ is a possibilistic network obtained from $\mathcal{CN}$ and defined by:*

- *A graphical component G which is the same graph as the credal network hence $\mathcal{PN}_{\mathcal{CN}}$ encodes the same independence relations as $\mathcal{CN}$.*
- *A collection of local possibility tables π_i obtained by transforming local credal sets IP_i with TR, a transformation from interval-based probability distributions into possibilistic ones.*

Example 4. *Let $\mathcal{CN}$ be the credal network of Figure 1 over two binary variables A and B. Using the MD transformation of Equation 6, the credal network $\mathcal{CN}$ of Figure 1 will be transformed to the possibilistic network $\mathcal{PN}$ of Figure 3.*

A	$\pi_{MD}(A)$
F	1
T	1

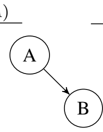


A	B	$\pi_{MD}(B\|A)$
F	F	1
F	T	.64
T	F	.29
T	T	.04

Fig. 3. The possibilistic network $\mathcal{PN}_{CN}$ obtained from the credal network $\mathcal{CN}$ of Fig. 1.

Using CD transformation, we obtain two possibilistic networks (Figure 4), the upper one π_u on this example matches the one obtained with MD transformation. The lower one, π_l corresponds to the one obtained using Equation 7, needing normalizing the obtained local possibility tables in order to draw inferences.

A	$\pi_l(A)$	$\pi_u(A)$
F	.5	1
T	.1	.5



A	B	$\pi_l(B\|A)$	$\pi_u(B\|A)$
F	F	.96	1
F	T	.6	.64
T	F	.25	.29
T	T	0	.04

Fig. 4. The possibilistic network $\mathcal{PN}_{CN}$ obtained from the credal network $\mathcal{CN}$ of Fig. 1.

We have studied some principles of credal-to-possibilistic network transformations in [3] where two main issues were answered: i) Does the distribution $\pi_{\mathcal{PN}}$ dominate $IP_{\mathcal{CN}}$ (the joint interval-based distribution encoded by $\mathcal{CN}$)? and ii) Is the partial order of interpretations induced by $IP_{\mathcal{CN}}$ preserved by the transformation TR?

Regarding the first issue, for elementary worlds $\omega_i \in \Omega$, we ensure that the computed possibility distribution dominates the corresponding probability degrees in case where the credal network $\mathcal{CN}$ is a Bayesian network (namely, all the intervals in $\mathcal{CN}$ are singletons). Regarding arbitrary events $\phi \subseteq \Omega$, the issue is still open. If we use the optimal transformation OT, the obtained joint possibility distribution does not guarantee to dominate the joint probability distribution. On the second issue, there is no guarantee that the interpretations' order encoded by the joint distribution is the same after the transformation. For more details, see [3].

Nevertheless, this approach can still be considered as an approximate method. The following section will highlight empirically how accurate is this approximate approach for *MAP* inference in credal networks.

IV. Experimental Studies

In this section, we give the results of our experimental studies where we have used new criteria to assess *MAP* requests accuracy in credal networks.

A. Experimentation setup

Before giving a detailed record of what we have implemented for the experimental study, let us recall that there exists no platform or implemented algorithm that can compute *MAP* inference in possibilistic networks. Furthermore, there is also no platform that computes *MAP* inference in credal networks. Yet, there exist packages to perform some inference tasks. Precisely, those packages return the probability degree or interval of a variable given an evidence. We implemented:

- the transformation of a credal network into a possibilistic network,
- an inference algorithm in possibilistic networks,
- the procedure to compute *MAP* outcomes from the results of the inference algorithm in credal networks and possibilistic networks.

B. Evaluation criteria

The benchmarks used in the current work are presented in Table II, such benchmarks are publicly available at http://ipg.idsia.ch/software/CNsBench.zip.

Networks	Topology	#Nodes	max |domain|
Alarm	Multiply-connected	37	4
Insurance	Multiply-connected	27	5
Poly	Polytree	10	4
Multi	Multiply-connected	6	4

TABLE II
Credal networks used in the experimentations.

In order to compare the results of *MAP* inference in credal networks and their possibilistic counterparts, each query Q is submitted to a credal network $\mathcal{CN}$ (using JavaBayes) then to the corresponding possibilistic network $\mathcal{PN}$ obtained from $\mathcal{CN}$. And in the same way, Q is submitted to a credal network through JavaBayes and to the same credal network using GL2U, a package for approximate inference in credal networks. The results are compared through the accuracy measure defined as follows:

$$accuracy(Q_1, Q_2...Q_n) = \frac{1}{n}\sum_{i:1..n}\frac{|CN_{MAP}(Q_i) \cap PN_{MAP}(Q_i)|}{|CN_{MAP}(Q_i) \cup PN_{MAP}(Q_i)|}, \quad (9)$$

where $CN_{MAP}(Q_i)$ (resp. $PN_{MAP}(Q_i)$) denotes the results of the query Q_i submitted to the network $\mathcal{CN}$ (resp. $\mathcal{PN}$). This measure evaluates the agreement between the results of $\mathcal{CN}$ to the *MAP* queries and the ones of $\mathcal{PN}$. Thus, the experiment provides:

- Accuracy rates compared to the exact algorithm implemented in JavaBayes software:
  - The accuracy of the approximate inference algorithm used in *GL2U* software.
  - The accuracy of *MAP* requests using possibilistic networks obtained by transforming the credal network using *MD* and *CD* transformations.
- Inclusion rates: The inclusion rate is a measure showing how much of the outcomes returned by one network are included in the outcomes returned by another network. In our case, we compute the proportion of outcomes returned by the approximate approach that are included in the outcomes of the exact approach.
- Size of outcomes set: we compare the number of outcomes to the number of possible outcomes.

C. Results

This subsection can be divided into two types of results, quantitative ones and qualitative ones. We carried out experiments with different numbers of query variables (more precisely, we vary the number of query variables between 1 to 5 and for each case, we tested around 200 networks).

1) Quantitative results: One of the main objectives of this experiment was to show that our approach could considerably reduces the computation time of *MAP* inference and this is what we present in Table III. Indeed, this table shows the number of files handled successfully by the different tested approaches. We can notice that *GL*, in terms of number of query variables, cannot handle queries with more that 3 variables. On the contrary, our approach based on the transformations is always better in terms of the number of networks answered even when we vary the number of query variables from 1 to 5. Note that when we say that an algorithm was not able to answer a query on a given network, we mean that it reached a timeout. Clearly, our approach handles bigger networks and queries without reaching the timeout.

2) Qualitative results: We have shown that our approach outperforms the other approximate approach (*GL*) in terms of computation time. So a natural question is about the actual quality of the results. To answer this question, we provide in Table IV some results regarding the number of outputs returned over the number of possible outcomes and the percentage of configurations returned that are included in the answer sets returned by the exact approach given by JavaBayes.

In Table IV, there are three main results that show the efficiency of our method:

i) When using *Interval-dominance* criterion, the number of configurations returned by JavaBayes as the result of *MAP* inference is around 80% of possible outcomes. These results clearly show a lot of confusion and make it hard to make decisions with such number of outcomes.

Criterion	*MD*	*CD*	*GL*	
.794	.685	.36	.88	$\%\ answers/all$
Inter-dom	.967	1	.891	% Inclusion
.36	.685	.36	.88	$\%\ answers/all$
Maximax	.546	.74	.629	% Inclusion

TABLE IV
PROPORTION OF RETURNED ANSWERS OVER ALL POSSIBLE OUTCOMES VS PROPORTION OF INCLUDED ANSWER SETS

On the other hand, the *Maximax* criterion ensures a narrower proportion of outcomes (around 36%). The method using the *CD* transformation gives similar results.

ii) Regarding the transformation *MD* and information preservation, the proportion of returned outcomes combined to the proportion of included outcomes show that *MD* is the transformation that preserves the information the better. These results hold when considering *Interval-dominance* criterion.

iii) Table IV finally shows that the approximate approach *GL* generally gives sets of outcomes larger than the exact approach. And even more, as the number of requested variables increases, *GL* tends to return all possible outcomes.

In the following, we show graphically the accuracy of each method *MD*, *CD* and *GL*. The axis x is to be read as A# for Alarm file and P# for Poly file with # is the number of requested variables. We present the results of two types of networks, polytrees and multiply-connected networks, and with three different criteria, *Interval-dominance*, *Maximax* and *Hurwicz*. Indeed, we omit *Maximin* criterion due to the similarity in terms of accuracy with *Maximax* and *Hurwicz* criteria.

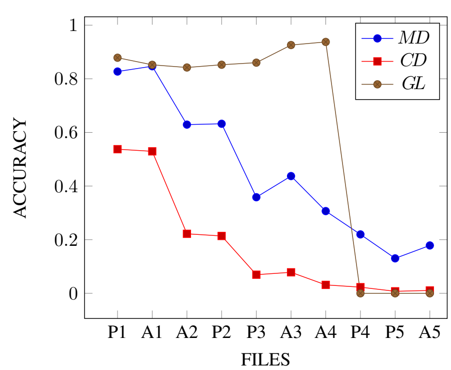


Fig. 5. Comparison between *MD*, *CD* and *GL* using *Interval-dominance* criterion

On Figure 5, the approximate method *GL* gives better results for both types of networks, except queries with more than 4 request variables where it can no longer answer. This problem can be explained by the fact that the variables are chosen randomly and it can affect the difficulty of the *MAP* inference algorithm implemented. The results of *GL* are in agreement with the previous results presented in Table IV. Indeed, by the fact that this method returns around 88% of the outcomes, it

# query variables	1			2			3			4			5		
Algorithm	*MD*	*CD*	*GL*	*MD*	*CD*	*GL*	*MD*	*CD*	*GL*	*MD*	*CD*	*GL*	*MD*	*CD*	*GL*
Alarm	187	187	**143**	149	149	149	77	77	68	63	63	**0**	43	43	**0**
Insurance	180	180	164	152	152	152	116	116	**52**	55	55	**0**	_	_	_
Poly	200	200	**140**	200	200	190	200	200	180	200	200	**0**	180	180	**0**
Multi	200	200	**110**	200	200	200	200	200	**120**	200	200	**0**	_	_	_

TABLE III
NUMBER OF FILES ANSWERED BY THE DIFFERENT ALGORITHMS.

is more likely to be in the 79% of the results returned by the exact method.

As for the possibilistic approach using *MD* transformation, if we correlate the accuracy results observed in the graphics, with the proportion given in Table IV, than *MD* is slightly better than *GL*. Indeed, by returning less outcomes than the exact approach and having a better proportion of included outcomes, it balances the accuracy rate which is still better than *CD*. As well, this approach is not sensitive to the size of the network nor by the size of the request variables.

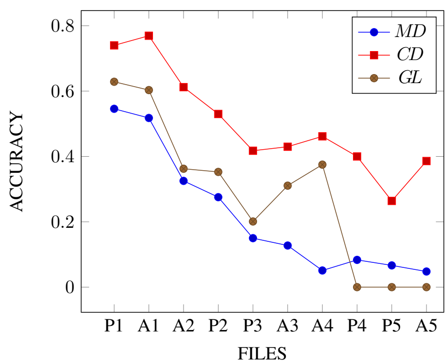


Fig. 6. Comparison between *MD*, *CD* and *GL* using *Maximax* criterion

Now, considering *Maximax* criterion, we observe on Figure 6 that *CD* gives the best results in terms of accuracy but also in terms of inclusion (cf. Table IV). Still, it decreases when the number of requested variables increases.

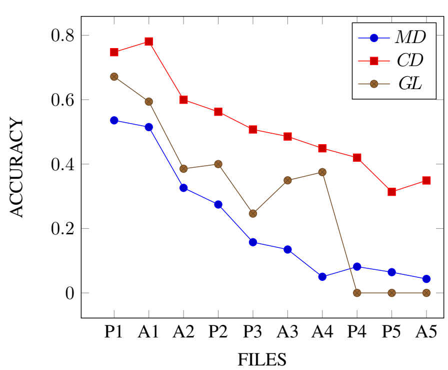


Fig. 7. Comparison between *MD*, *CD* and *GL* using *Hurwicz* criterion

Finally, we also conducted our experiment using *Hurwicz* criterion with the 0.5 degree associated to each evaluation. In terms of results (Figure 7), they are more or less the same as *Maximax* criterion. This is why, in the last graphic (Figure 8), we compare these 3 criteria with *CD* method.

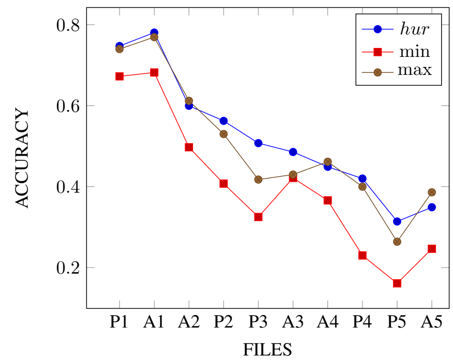


Fig. 8. Comparison between *Maximax*, *Maximin*, *Hurwicz* criteria for *CD* method

What we can see from Figure 8 is that the three criteria mostly behave the same way. We can conclude that from those three criteria, one should choose *Hurwicz* criterion, and if we would like to favor an optimistic (resp. pessimistic) evaluation, we could increase the degree of *Hurwicz* criterion (resp. decrease). Overall, the three approximate algorithms show the same behavior towards the number of requested variables, the accuracy rates all decrease as the number of variables increases.

This section shows empirically that possibilistic networks ensure an interesting trade-off in terms of accuracy and computational time. This led us to start investigating the issue of computational complexity in possibilistic networks. The following section provides some preliminary findings.

D. A note on the complexity of inference in possibilistic networks

As said earlier in this paper, there is no systematic study of complexity issues for inference in possibilistic networks and most of the works assume that the same complexity results in Bayesian networks still hold in the possibilistic setting. This section briefly shows that inference in possibilistic networks is less costly than in Bayesian networks. Let us start with the *MPE* problem.

Definition 7. *Let $\mathcal{PN}$ be a possibilistic network and e be an evidence. Let* **D-***MPE be the decision problem: Is there a complete instantiation q of* **all non observed variables** *$\mathcal{Q}$ such that $\Pi(q, e) > t$? with $t \in [0, 1]$.*

Recall that in *MPE* queries, $\mathcal{Q}{=}X\backslash E$. Intuitively, the decision problem for *MPE* comes down to answering whether the possibility degree $\Pi(q,e)$ is greater than a rational number t.

Theorem 1. **D-**MPE *is* NP*-complete.*

The membership of **D-**MPE to NP and it is hardness can be shown very easily and similarly to the way they are shown in Bayesian networks (for lack of space, the proof of the theorem is not provided in this paper but it can be found following this link: https://www.dropbox.com/s/oo8q3aim4rm2nm5/PN-complexity.pdf?dl=0).

Now, regarding the complexity of *MAP* queries, we recall that in the possibilistic setting, we are given an evidence e and the problem is to compute the most plausible configuration of some variables $\mathcal{Q}$. Namely, the answer is $argmax_{q_i \in \mathcal{Q}} \Pi(q_i|e)$. Recall that in the possibilistic setting, $argmax_{q_i \in \mathcal{Q}} \Pi(q_i|e){=}\max_{x \in \Omega \cap q \cap e} \Pi(x)$. Hence, the decision problem of *MAP* inference in possibilistic networks is exactly the one of Definition 7, namely, the decision problem for *MAP* (noted *D-MAP*) here comes down to answer whether the statement: Is there a complete instantiation $(x_1, .., x_n)$ of the network variables $(X_1, .., X_n)$ that is compatible with q and e and such that $\Pi(x_1, .., x_n){>}t$. Clearly, the complexity of *D-MAP* in possibilistic networks is also NP-complete.

V. Concluding discussions

We provide a new and efficient approach to perform *MAP* inference in credal networks by transforming them into possibilistic ones. We carried out experiments to compare our approach to both exact and approximate approaches for *MAP* inference in credal networks (*GL*). The benefits of our approach are i) reducing the computational time of *MAP* inference while ii) ensuring narrower answer sets. Experimental results showed that, first, using the approximate algorithm (*GL*) on credal networks was not computationally interesting due to the limits it has shown when the number of request variables increases. Then, when using criteria like *Hurwicz*, *CD* algorithm performed quite efficiently on numerous networks and numerous request variables. One thing that we have not been mentioning so far, is the complexity of our transformation *MD* and *CD*, this is to be taken into account when choosing an approach. And in this matter, *CD* is quite a direct translation and does not imply a high complexity, contrary to *MD* transformation. This supports even more the choice of *CD* that gives a good alternative to approximate *MAP* inference in credal networks. As future works, we plan to investigate new algorithm for *MAP* inference in possibilistic networks. As shown in the last section, the complexity of inference in possibilistic networks is less costly than in Bayesian and credal networks. This will definitely open new perspectives for *MAP* inference especially for credal networks.

References

[1] A Antonucci, G Corani, and M Zaffalon. Bayesian networks with imprecise probabilities: Theory and application to classification. Technical report, IDSIA - Dalle Molle Institute for Artificial Intelligence, 2010.

[2] S Benferhat, A Levray, and K Tabia. On the analysis of probability-possibility transformations: Changing operations and graphical models. In *ECSQARU*, volume 9161 of *Lecture Notes in Computer Science*, pages 279–289. Springer, 2015.

[3] S Benferhat, A Levray, and K Tabia. Probability-possibility transformations: Application to credal networks. In *SUM*, volume 9310 of *Lecture Notes in Computer Science*, pages 203–219. Springer, 2015.

[4] F G Cozman. Credal networks. *Artif. Intell.*, 120(2):199–233, 2000.

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

[6] S. Destercke, D. Dubois, and E. Chojnacki. Transforming probability intervals into other uncertainty models. In *EUSFLAT 2007 proc.*, volume 2, pages 367–373, Ostrava, Czech Republic, 2007.

[7] S Destercke, D Dubois, and E Chojnacki. Unifying practical uncertainty representations - I: generalized p-boxes. *Int. J. Approx. Reasoning*, 49(3):649–663, 2008.

[8] D Dubois, L Foulloy, G Mauris, and H Prade. Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliable Computing*, 10(4):273–297, 2004.

[9] D. Dubois and H. Prade. *Possibility Theory: An Approach to Computerized Processing of Uncertainty*. Plenum Press, New York, 1988.

[10] D Dubois and H Prade. When upper probabilities are possibility measures. *Fuzzy Sets and Systems*, 49(1):65 – 74, 1992.

[11] K. D Jamison and W A. Lodwick. The construction of consistent possibility and necessity measures. *Fuzzy Sets and Systems*, 132(1):1–10, 2002.

[12] G. J. Klir and J. F. Geer. Information-preserving probability- possibility transformations: Recent developments. In *Fuzzy Logic*, pages 417–428. 1993.

[13] J Kwisthout. Treewidth and the computational complexity of MAP approximations. In *Probabilistic Graphical Models*, volume 8754 of *Lecture Notes in Computer Science*, pages 271–285. Springer, 2014.

[14] J Kwisthout, H L. Bodlaender, and L C. van der Gaag. The necessity of bounded treewidth for efficient inference in bayesian networks. In *ECAI*, volume 215 of *Frontiers in Artificial Intelligence and Applications*, pages 237–242. IOS Press, 2010.

[15] I. Levi. *The enterprise of knowledge : an essay on knowledge, credal probability, and chance*. MIT Press Cambridge, Mass, 1980.

[16] M.-H. Masson and T. Denoeux. Inferring a possibility distribution from empirical data. *Fuzzy Sets and Systems*, pages 319–340, 2006.

[17] D D Mauá, C P de Campos, A Benavoli, and A Antonucci. On the complexity of strong and epistemic credal networks. In *UAI*. AUAI Press, 2013.

[18] D D Mauá, C P de Campos, A Benavoli, and A Antonucci. Probabilistic inference in credal networks: New complexity results. *J. Artif. Intell. Res. (JAIR)*, 50:603–637, 2014.

[19] D D Mauá, C P de Campos, and F G Cozman. The complexity of MAP inference in bayesian networks specified through logical languages. In *IJCAI*, pages 889–895. AAAI Press, 2015.

[20] E Miranda and S Destercke. Extreme points of the credal sets generated by elementary comparative probabilities. In *ECSQARU*, volume 7958 of *Lecture Notes in Computer Science*, pages 424–435. Springer, 2013.

[21] J Pearl. *Probabilistic reasoning in intelligent systems - networks of plausible inference*. Morgan Kaufmann series in representation and reasoning. Morgan Kaufmann, 1989.

[22] S Eyal Shimony and C Domshlak. Complexity of probabilistic reasoning in directed-path singly-connected bayes networks. *Artif. Intell.*, 151(1-2):213–225, 2003.

[23] Y Ben Slimen, R Ayachi, and N Ben Amor. Probability-possibility transformation: - application to bayesian and possibilistic networks. In *WILF*, volume 8256 of *Lecture Notes in Computer Science*, pages 122–130. Springer, 2013.

[24] T. Sudkamp. On probability-possibility transformations. *Fuzzy Sets and Systems*, pages 73–81, 1992.

[25] K Tabia. Possibilistic graphical models for uncertainty modeling. In *SUM*, volume 9858 of *Lecture Notes in Computer Science*, pages 33–48. Springer, 2016.

[26] P Walley. Towards a unified theory of imprecise probability. *Int. J. Approx. Reasoning*, 24(2-3):125–148, 2000.

[27] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 100:9–34, 1999.