Decision Making by Credal Nets

Alessandro Antonucci
IDSIA
Lugano (Switzerland)
alessandro@idsia.ch

Cassio Polpo de Campos
IDSIA
Lugano (Switzerland)
cassio@idsia.ch

Abstract—**Credal nets are probabilistic graphical models which extend Bayesian nets to cope with sets of distributions. This feature makes the model particularly suited for the implementation of classifiers and knowledge-based systems. When working with sets of (instead of single) probability distributions, the identification of the optimal option can be based on different criteria, some of them eventually leading to multiple choices. Yet, most of the inference algorithms for credal nets are designed to compute only the bounds of the posterior probabilities. This prevents some of the existing criteria from being used. To overcome this limitation, we present two simple transformations for credal nets which make it possible to compute decisions based on the *maximality* and *E-admissibility* criteria without any modification in the inference algorithms. We also prove that these decision problems have the same complexity of standard inference, being NP^{PP}-hard for general credal nets and NP-hard for polytrees.**

I. INTRODUCTION

A Bayesian net is a probabilistic graphical model which allows the specification of a joint probability mass function by means of a collection of conditional mass functions, each collection associated only with a single variable. Such a decomposition is made possible by the conditional independence relations among the model variables $\mathbf{X} := (X_1, \ldots, X_n)$, which are graphically depicted by feeding a directed acyclic graph $\mathcal{G}$ (nodes are in one-to-one correspondence with the variables) with the Markov condition.[1] Under these assumptions, a set of conditional mass functions (one for each joint value of the parent variables, for each variable) is sufficient to compute the joint $P(\mathbf{x}) := \prod_{i=1}^{n} P(x_i | \pi_i)$, where the state π_i of the parents of X_i is consistent with the state $\mathbf{x}$ of $\mathbf{X}$ [1]. To gain robustness with this modeling, a *credal net* can be specified instead [2]. It is obtained by replacing each conditional mass function $P(X_i | \pi_i)$ with a closed convex set of distributions (or *credal set*) $K(X_i | \pi_i)$. Similar to a Bayesian net, which defines one *joint* mass function, a credal net defines one *joint* credal set, as follows:

$$K(\mathbf{X}) = \text{CH} \left\{ P(\mathbf{X}) \left| \begin{array}{l} P(\mathbf{x}) := \prod_{i=1}^{n} P(x_i | \pi_i) \\ P(X_i | \pi_i) \in \text{ext}[K(X_i | \pi_i)] \\ \forall i = 1, \ldots, n \quad \forall \pi_i \end{array} \right. \right\}, \quad (1)$$

where the notation $\text{ext}[\cdot]$ is used for the extreme points of the credal sets and CH for the convex hull operator. A typical inferential task on credal nets consists in the *updating* of the

probabilities for a variable of interest $X \in \mathbf{X}$, given some evidence x_E about other variables $X_E \subseteq \mathbf{X} \setminus \{X\}$. For credal nets, this is intended as the computation of the posterior credal set $K(X|x_E)$, which is obtained from the joint credal set (1) by elementwise marginalisation and conditioning.

Classifiers and knowledge-based expert systems are typical applications of credal nets [3], [4]. In both cases, one is typically interested in deciding which one is the most likely value of X among the elements of its possibility space $\mathcal{X}$, on the basis of the observed evidence x_E. This decision should be based on $K(X|x_E)$. For Bayesian nets, where $K(X|x_E)$ collapses into a single conditional mass function $P(X|x_E)$, this simply corresponds to the identification of the most probable option $x^* := \arg\max_{x \in \mathcal{X}} P(x|x_E)$.[2] Unlike the Bayesian case, such a decision task can be addressed in many ways in the credal case, where different criteria, some of them eventually returning more than a single option, have been proposed (Sect. II). Yet, most of the inference algorithms for credal nets are not designed to compute the posterior credal set, but only its bounds, i.e., $\underline{P}(x|x_E) := \min_{P(X|x_E) \in K(X|x_E)} P(x|x_E)$ and (similarly) $\overline{P}(x|x_E)$. This prevents some of those criteria from being adopted. It basically means that, when coping with credal nets, the higher modeling freedom provided by using credal sets is partially wasted when making decisions, because of the lack of suitable inference algorithms. The goal of this paper is to fill this gap. This is achieved by specific transformations augmenting the credal net with auxiliary variables (Sect. III) that make possible to determine, with standard algorithms, the set of optimal options according to two of the major decision criteria, namely *maximality* (Sect. IV) and *E-admissibility* (Sect. V).[3] Apart from some specific applications to maximality-based classification [4], decision making based on these two criteria for general credal nets is an almost unexplored field of investigation. The computational complexity of these decision problems should be therefore considered. Notably, as it happens for the standard decision problem consisting in the *evaluation* of the bounds of the posterior, the problems discussed in this paper are proved to be NP^{PP}-hard for general and NP-hard for polytree-shaped (i.e., singly-connected graphs) credal nets (Sect. VI).

---

[1]According to the Markov condition for directed graphs any variable is conditionally independent of its non-descendants non-parents given its parents.

[2]This corresponds to the maximisation of the expected utility with a zero-one loss. Utility functions are not explicitly reported for notational ease.

[3]For maximality, this issue has been already discussed in [5]. Yet, the solution proposed in that paper is approximate and referred only to queried variables associated to nodes without parents or without children.

II. Decisions based on Credal Sets

First let us review the main decision criteria proposed within the framework of credal sets.[4] We consider decisions based on the posterior credal set $K(X|x_E)$ for a generic credal net, but the definitions below can be referred to any credal set. Generally speaking, different distributions in a credal set can assign the highest probability to different options. This basically means that there is not a single straightforward way to extend to the credal case the decision criterion adopted in the Bayesian framework. The so-called Γ-maximin criterion represents a pessimistic approach which takes the option maximising the lower probability, i.e.,

$$x^* := \arg\max_{x \in \mathcal{X}} \underline{P}(x|x_E), \qquad (2)$$

while the optimistic version, with the upper instead of the lower probability, is called Γ-maximax. Another approach consists in evaluating the overlapping among the posterior probability intervals and reject the states whose upper probabilities are smaller than the lower probability of some other state. This is called *interval dominance*, and generally leads to a set $\mathcal{X}^* \subseteq \mathcal{X}$ of optimal options, where:

$$\mathcal{X}^* := \left\{ x' \in \mathcal{X} \,|\, \nexists x'' \in \mathcal{X} : \underline{P}(x''|x_E) > \overline{P}(x'|x_E) \right\}. \quad (3)$$

The above mentioned criteria require only the evaluation of the posterior lower and upper probabilities for each $x \in \mathcal{X}$. A number of algorithms have been proposed to compute these probabilities. Other, more sophisticated, criteria for decision have been proposed, but, at least in their formulation, they need the posterior credal set available. In particular, according to the *maximality* criterion, a state is rejected if, for each element (or extreme point) of the credal set, there is another state with higher probability, i.e., the maximal states are:

$$\mathcal{X}^* := \left\{ x' \in \mathcal{X} \,\middle|\, \nexists x'' \in \mathcal{X} : \begin{array}{l} P(x''|x_E) > P(x'|x_E), \\ \forall P(X|x_E) \in K(X|x_E) \end{array} \right\}. \quad (4)$$

A more cautious approach is *E-admissibility*, where only options which are optimal for at least a distribution in the posterior credal set are adopted, i.e.,

$$\mathcal{X}^* := \left\{ x' \in \mathcal{X} \,\middle|\, \begin{array}{l} \exists P(X|x_E) \in K(X|x_E) : \\ P(x'|x_E) > \max_{\substack{x'' \in \mathcal{X} \\ x'' \neq x'}} P(x''|x_E) \end{array} \right\}. \quad (5)$$

Despite a substantial lack of unanimity about the best way to take decisions based on credal sets, these two latter criteria can be regarded as a reasonable compromise between the necessary caution when taking decisions and the need of avoiding a too high number of possible options [7].[5] Yet, the available algorithms for credal nets inference cannot be directly employed to identify the sets in (4) or (5), and this substantially reduces the advantages of a credal modeling. In the next sections we provide a solution to this problem.

III. Computing Expectations

Let us first consider the very simple case where the credal net we work with is made of a single (queried) variable X, while the set of observed variables is empty. We consider a real function $g : \mathcal{X} \to \mathbb{R}$, which is also called *gamble* [8]. The lower bound for the expectation of this gamble with respect to the credal set $K(X)$ is:

$$\underline{E}(g) := \min_{P(X) \in K(X)} \sum_{x \in \mathcal{X}} P(x) \cdot g(x). \qquad (6)$$

Now, let us augment the net with a binary variable Y, with states $\{0, 1\}$, which is assumed to be a child of X. For the conditional credal sets of Y given X we adopt a "Bayesian" specification: each conditional credal set is made of a single element, which can be completely specified by the values of $P(Y = 0|x)$, for each $x \in \mathcal{X}$. After this quantification, the lower probability for the first state of the auxiliary variable with respect to the augmented credal net is:

$$\underline{P}(Y = 0) = \min_{P(X) \in K(X)} \sum_{x \in \mathcal{X}} P(x) \cdot P(Y = 0|x). \quad (7)$$

Thus, setting $P(Y = 0|x) := g(x)$ for each $x \in \mathcal{X}$, we have that (7) coincides with (6), i.e., $\underline{E}(g) = \underline{P}(Y = 0)$. In other words, by means of an appropriate quantification of the conditional probabilities for the auxiliary node, the lower (and similarly the upper) expectation of a gamble corresponds to a lower (upper) probability of the auxiliary node.

Note that in the above derivation we are implicitly assuming that the gamble g has values only in the $[0, 1]$ interval. If this is not the case, we can easily bypass the problem by considering the "normalised" gamble:

$$\tilde{g}(x) := \frac{g(x) - \min_{x \in \mathcal{X}} g(x)}{\max_{x \in \mathcal{X}} g(x) - \min_{x \in \mathcal{X}} g(x)}, \qquad (8)$$

whose values can be clearly interpreted as probabilities. This means that we can always compute the expectation of gamble $\tilde{g}$ as in (7) and then obtain the lower expectation for the original gamble as follows:[6]

$$\underline{E}(g) = \min_{x \in \mathcal{X}} g(x) + [\max_{x \in \mathcal{X}} g(x) - \min_{x \in \mathcal{X}} g(x)] \cdot \underline{E}(\tilde{g}). \quad (9)$$

Finally, it is easy to note that the augmentation of Y to X is a local transformation which can be equivalently performed if the net is defined over more than a single variable, by simply replacing $K(X)$ with $K(\mathbf{X})$ as in (1). Similarly, the above derivation can be performed in the same way even if the set of observed variables X_E is not empty. Summarising, in a generic credal net over $\mathbf{X}$, we can express the conditional lower expectation for a (normalised) gamble over a queried variable X given evidence x_E as the posterior lower probability, on the augmented net, for the first state of the auxiliary child Y, i.e., $\underline{E}(\tilde{g}|x_E) = \underline{P}(Y = 0|x_E)$, where the latter quantity can be computed by any standard updating algorithm for credal nets.

---

[4]See [6] or [7] for a more detailed overview and discussion on this topic.
[5]Note that Γ-maximin/max return a single option, while the optimal options for interval dominance include those for E-admissibility and maximality.

[6]This is simply based on the fact that, given a gamble f, if α and β are real constants, with α positive, $\underline{E}[\alpha f + \beta] = \alpha \underline{E}[f] + \beta$. This relation can be easily proved (or found in [8]).

IV. SOLVING MAXIMALITY

In the previous section we showed how to compute the posterior lower and upper expectations for a generic gamble associated to a variable in a credal net. As a simple corollary of this result, we can address the maximality criterion as in (4). In fact, in order to detect these optimal options, we should check, for each pair $x', x'' \in \mathcal{X}$, whether or not $P(x''|x_E) > P(x'|x_E)$ for each $P(X|x_E) \in K(X|x_E)$. Notably, this corresponds to decide whether or not:

$$\underline{E}(I_{\{x''\}} - I_{\{x'\}}|x_E) > 0, \qquad (10)$$

where $I_{\{x\}}$ is an indicator function (i.e., the gamble which is non-zero only for $X = x$). Thus, in other words, maximality can be checked by evaluating, for each $x', x'' \in \mathcal{X}$, the lower expectation of the gamble:

$$g_{x',x''}(x) = \begin{cases} -1 & \text{if} \quad x = x' \\ +1 & \text{if} \quad x = x'' \\ 0 & \text{otherwise} \end{cases} . \qquad (11)$$

This task can be solved by exploiting the transformation outlined in the previous section. Algorithm 1 can be therefore used to address maximality with credal nets. Note that the condition in the fourth row is equivalent to that in (10), and can be tested by evaluating $\underline{P}(Y = 0|x_E)$ with some updating algorithm for credal nets.

Algorithm 1
1. $\mathcal{X}^* := \mathcal{X}$
2.   **for** $x' \in \mathcal{X}$
3.   **for** $x'' \in \mathcal{X}$
4.   **if** $\underline{E}(\tilde{g}_{x',x''}) > \frac{1}{2}$
5. $\mathcal{X}^* := \mathcal{X}^* \setminus \{x'\}$

V. SOLVING E-ADMISSIBILITY

Now let us consider how to address the E-admissibility. In order to evaluate whether or not x' belongs to $\mathcal{X}^*$ defined as in (5), we have to check, for each $x' \in \mathcal{X}$, whether or not:[7]

$$\max_{P(X|x_E) \in K(X|x_E)} \left[P(x'|x_E) - \max_{x'' \in \mathcal{X}} P(x''|x_E) \right] > 0. \quad (12)$$

Note that, unlike the case of maximality as in (10), the left-hand side of the E-admissibility test in (12) cannot be reformulated as the (upper) expectation of a gamble. Yet, the same graphical transformation presented in Sect. III, with a credal quantification of the conditional probabilities for Y can be used to obtain a similar result and check E-admissibility with standard algorithms for credal nets.

Let us first show how to do that in a simple situation where the net has a single variable X and is Bayesian, i.e., we have a probability mass function $P(X)$ associated to X. In this case, (12) rewrites as:

$$\max_{x'' \in \mathcal{X}} [P(x') - P(x'')] > 0. \qquad (13)$$

---

[7]Following (5), we should add condition $x'' \neq x'$ in the maximum. Yet, as this case makes the difference zero, it has no effect on the overall test.

Now let us augment this Bayesian net with an auxiliary binary child Y as in Sect. III. Yet, this time we consider a credal quantification of the conditional probability table $P(Y|X)$. In particular, the quantification is *extensive*, this meaning that the table takes values in a collection $\{P_{x''}(Y|X)\}_{x'' \in \mathcal{X}}$. For the corresponding credal net, we have:

$$\overline{P}(Y = 0) := \max_{x'' \in \mathcal{X}} \sum_{x \in \mathcal{X}} P_{x''}(Y = 0|x) \cdot P(x). \qquad (14)$$

In particular, for the quantification of the tables we assume $P_{x''}(Y = 0|x) = \tilde{g}_{x'',x'}(x)$, for each $x, x'' \in \mathcal{X}$. In this case, by a derivation similar to that considered in Sect. III, it is easy to see that (14) becomes equivalent to the left-hand side of (13). Again, by considerations similar to those reported at the end of the previous section, we can extend the result to the case where the net is credal, with multiple variables and non-empty evidence. Thus, we can also use E-admissibility for decision making in credal nets. Algorithm 2 shows the procedure, with the test in the third row to be checked by credal nets inference algorithms.

Algorithm 2
1. $\mathcal{X}^* := \emptyset$
2.   **for** $x' \in \mathcal{X}$
3.   **if** $\overline{P}(Y = 0|x_E) > \frac{1}{2}$
4. $\mathcal{X}^* := \mathcal{X}^* \cup \{x'\}$

VI. COMPLEXITY ISSUES

The computational complexity characterising inference on credal nets has been deeply investigated in [9]. Regarding updating, the considered problem is to decide whether or not $\overline{P}(x') > k$ (in the original proof $X_E := \emptyset$; in fact the proof with non-empty evidence follows analogously), which means the existence of at least a $P(X) \in K(X)$ such that $P(x') > k$, for given $x' \in \mathcal{X}$ and (non-negative) k. We call this task *evaluation*. We perform some (polynomial-time) reductions of evaluation to its counterparts for interval dominance, maximality and E-admissibility. If we can solve any of these problems, then we can also solve evaluation, which implies the hardness of decision making based on these criteria with credal nets.

First, we rewrite the decision problems we are treating in a convenient way. Again, we consider a credal net over $\mathbf{X}$, a state x' of the queried variable X, and no evidence, i.e., $X_E := \emptyset$. Regarding interval dominance as in (3), we say that x' is *interval-dominance optimal* if and only if

$$\forall x'' \neq x' : \exists P_1, P_2 : P_1(x') > P_2(x''), \qquad (15)$$

that is, for each x'', we can freely choose P_1 to maximize the probability of x' and P_2 to minimize the probability of x'' such that we try to satisfy (15).

For maximality as in (4), we say that x' is *maximal* if and only if there is no $x'' \in \mathcal{X}$ such that (10) is satisfied. This means that, for every $x'' \in \mathcal{X}$, $\underline{E}(I_{\{x''\}} - I_{\{x'\}}) \leq 0$, and hence:

$$\forall x'' \neq x' : \exists P : P(x'') \leq P(x'), \qquad (16)$$

that is, there is no x'' that dominates x'. Differently from interval dominance as in (15), here the same P has to be used to compute $P(x'')$ and $P(x')$.

Finally, we say that x' is *E-admissible* if and only if

$$\exists P: \ \forall x'' \neq x': \ P(x') > P(x''). \qquad (17)$$

Note now that the same P has to be used over all the alternatives x'' to x', while distinct P (for each alternative x'') could be employed for the maximality criterion.

Now we will show that any of the three decision problems: (i) *is x' interval-dominance optimal?*, (ii) *is x' maximal?*, and (iii) *is x' E-admissible?*, can be used to solve evaluation, i.e., decide whether or not $\overline{P}(x') > k$, which is the hard problem from where the reduction is made. For this purpose, as in the previous sections, we augment the net with an auxiliary variable Y, which has the queried variable X as sole parent. Unlike the previous cases, Y is assumed to be ternary, with possibility space $\mathcal{Y} := \{0, 1, 2\}$, and the model conditional probability mass functions such that:

$$P(Y=0|x) := \frac{I_{\{x'\}} + 1}{k+3}, \quad P(Y=1|x) := \frac{k+1}{k+3}, \quad (18)$$

for each $x \in \mathcal{X}$, while the probabilities for $Y = 2$ are obtained by requiring the mass function to sum one. By simple manipulations, we have, for each $P(X) \in K(X)$:

$$P(Y=0) = \sum_{x \in \mathcal{X}} P(Y=0|x) \cdot P(x) = \frac{P(x')+1}{k+3}, \quad (19)$$

$$P(Y=1) = \sum_{x \in \mathcal{X}} P(Y=1|x) \cdot P(x) = \frac{k+1}{k+3}, \quad (20)$$

$$P(Y=2) = \sum_{x \in \mathcal{X}} P(Y=2|x) \cdot P(x) = \frac{\sum_{x \neq x'} P(x)}{k+3}. \quad (21)$$

An important fact is that, for each $P(X) \in K(X)$, we have:

$$P(Y=1) = \frac{k+1}{k+3} \geq \frac{\sum_{x \neq x'} P(x)}{k+3} = P(Y=2). \quad (22)$$

Let us consider the problem of deciding whether or not the state $Y = 0$ is interval-dominance optimal. The only required test by (15) is to check whether

$$\exists P_1, P_2 : P_1(Y=0) > P_2(Y=1), \qquad (23)$$

because we already know that $P(Y = 1) \geq P(Y = 2)$ by (22). This is equivalent to the evaluation problem as:

$$\exists P_1 : \ \frac{P_1(x')+1}{k+3} > \frac{k+1}{k+3} \iff \exists P_1 : \ P_1(x') > k. \quad (24)$$

Finally, let us note that when augmenting the credal net with the auxiliary child Y, the topology and the treewidth of the graph remain the same. In particular, if the original net was a *polytree* (i.e., associated to a singly connected graph), the net remains a polytree also after the transformation. Because of the above inclusion, on the basis of the complexity results in [9], we conclude that interval-dominance optimality is NP-hard for polytrees and NP$^{\text{PP}}$-hard for general credal nets.

Now let us discuss the maximality decision for $Y = 0$. In this case, (16) reduces to decide whether or not exists $P(X) \in K(X)$ such that $P(Y = 1) \leq P(Y = 0)$. Again this is the only required test, because of (22), as if this is true, then we also have that exists $P(X) \in K(X)$ such that $P(Y = 2) \leq P(Y = 0)$. But testing whether or not $P(Y = 0) \geq P(Y = 1)$ is equivalent to check:

$$\frac{P(x')+1}{k+3} \geq \frac{k+1}{k+3} \iff P(x') \geq k, \qquad (25)$$

from which the analogous complexity result follows.[8]

Finally, the same argumentation holds for E-admissibility, as (17) reduces to $\exists P: \ P(Y = 0) > P(Y = 1)$ (using (22)), which is again equal to deciding whether $\exists P: \ P(x') > k$.

VII. CONCLUSIONS AND OUTLOOKS

Two simple graphical transformations allowing for the identification of the set of optimal options according to the maximality and E-admissibility criteria for the updated beliefs about a variable in a credal network are presented. Inferences based on these decision criteria (and similarly for interval dominance) are proved to have the same complexity of the evaluation of the bounds of the posterior probability, the latter problem being NP$^{\text{PP}}$-hard for general credal nets and NP-hard for polytrees. These findings should be adopted when performing classification or decisions making based on credal nets. As a future work, we intend to extend these results to situations where more than a single variable is queried.

REFERENCES

[1] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kauf., 1988.
[2] F. G. Cozman, "Credal networks," *Artificial Intelligence*, vol. 120, pp. 199–233, 2000.
[3] A. Piatti, A. Antonucci, and M. Zaffalon, "Building knowledge-based systems by credal networks: a tutorial," in *Advances in Mathematics Research*, A. R. Baswell, Ed. Nova Science Publishers, 2010, vol. 11.
[4] M. Zaffalon and E. Fagiuoli, "Tree-based credal networks for classification," *Reliable computing*, vol. 9, no. 6, pp. 487–509, 2003.
[5] A. Antonucci and F. Cuzzolin, "Credal sets approximation by lower probabilities: application to credal networks," in *Proceedings of IPMU 2010*, 2010, pp. 716–725.
[6] R. Hable, "Data-based decisions under complex uncertainty," Ph.D. dissertation, University of Bayreuth, 2009.
[7] M. Troffaes, "Decision making under uncertainty using imprecise probabilities," *International Journal of Approximate Reasoning*, vol. 45, no. 1, pp. 17–29, 2007.
[8] P. Walley, *Statistical Reasoning with Imprecise Probabilities*. London: Chapman and Hall, 1991.
[9] C. P. de Campos and F. G. Cozman, "The inferential complexity of Bayesian and credal networks," in *Proceedings of IJCAI '05*, Edinburgh, 2005, pp. 1313–1318.

---

[8]Strictly speaking, the test in (25) is not a strict inequality as in the evaluation problem. Yet, the same complexity results of [9] hold in this case.