Set-Based Variational Methods in Credal Networks: the SV2U Algorithm

Jaime Shinsuke Ide , Fabio Gagliardi Cozman

Escola Politécnica, University of São Paulo
Av. Prof. Mello Moraes, 2231 - São Paulo, SP, Brazil
jaime.ide@poli.usp.br, fgcozman@usp.br

Abstract. *Graphical models that represent uncertainty through sets of probability measures are often referred to as credal networks. Polynomial-time exact inference methods are available only for polytree-structured binary credal networks. In this work, we approximate potentially intractable inferences in multi-connected binary networks by tractable inferences in polytree-structures. We propose a novel set-based structural variational inference method - the SV2U algorithm. The SV2U algorithm is the first method that produces approximate inferences in large binary credal networks with theoretical solid convergence analysis and offers a promising way to handle continuous variables in credal networks.*

1. Introduction

Graphical models associated with probabilities find widespread use in artificial intelligence, to represent and think about uncertainty. In this context, Bayesian networks are the most popular tool [Pearl, 1988]. In situations where beliefs cannot be cast as sharp numeric values (that is, we have imprecise and incomplete beliefs), sets of probability are a suitable representation. Graphical models that represent uncertainty through sets of probability measures are often referred to as *credal networks* [Cozman, 2000a, Fagiuoli and Zaffalon, 1998]. A credal network could be informally defined as a *representation for a set of Bayesian networks over a fixed set of variables.*

In a credal network, a collection of sets of probability measures is associated with a directed acyclic graph. *Inference* in a credal network usually means the computation of lower and upper bounds for the conditional probability of an event. The complexity of inference in credal networks is generally high (even for tree-like networks [da Rocha and Cozman, 2002]), and approximate inference seems to be a natural solution for large networks [Cano and Moral, 2002]. Polynomial-time exact inference methods are available only for polytree-structured binary credal networks (the 2U algorithm [Fagiuoli and Zaffalon, 1998]).

The main contribution of this work is the *Structured Variational-2U* algorithm (SV2U). The SV2U algorithm mixes the *set-based variational* approximations [Ide and Cozman, 2005] with polynomial-time exact inference methods [Fagiuoli and Zaffalon, 1998], approximating potentially intractable inferences in multi-connected networks by tractable inferences in polytree-structures. We have recently in-

troduced set-based variational methods [Ide and Cozman, 2005] and in this paper we propose the novel SV2U algorithm that implements such methods. The SV2U algorithm is the first method that produces approximate inferences in large binary credal networks with theoretical solid convergence analysis and offers a promising way to handle continuous variables in credal networks. Binary credal networks are a promising model to be applied in propositional and relational Bayesian networks associated with imprecise and qualitative probabilistic assessments [Cozman et al., 2004].

In Section 2. and Section 3. we give a brief review of credal networks, variational methods and set-based approaches. The SV2U algorithm and example results are presented in Section 4.

2. Credal Networks

In this section we present a few facts on credal networks (and their basic elements, *credal sets* [Levi, 1980]); a more detailed discussion can be found elsewhere [Cozman, 2000a, Cozman, 2004, Fagiuoli and Zaffalon, 1998].

A credal set for variable X is denoted by $K(X)$; we assume that every variable is categorical and that every credal set is convex, closed, and has a finite number of vertices. Given a credal set $K(X)$ and a function $f(X)$, the *lower* and *upper* expectations of $f(X)$ are defined respectively as $\underline{E}[f(X)] = \min_{p(X)\in K(X)} E[p]\, f(X)$ and $\overline{E}[f(X)] = \max_{p(X)\in K(X)} E[p]\, f(X)$ (here $E[p]\, f(X)$ indicates standard expectation). The *lower probability* and the *upper probability* of event E are defined respectively as $\underline{P}(E) = \min_{p(X)\in K(X)} P(E)$ and $\overline{P}(E) = \max_{p(X)\in K(X)} P(E)$. A *conditional credal set* is a set of conditional distributions, obtained by applying Bayes rule to each distribution in a credal set of joint distributions [Walley, 1991]. Lower and upper conditional probabilities for a variable X given an event E are defined accordingly: $\underline{P}(X = x|E) = \min_{p(X)\in K(X)}(p(X = x, E)\,/P(E))$, $\overline{P}(X = x|E) = \max_{p(X)\in K(X)}(p(X = x, E)\,/P(E))$.

Credal networks are directed acyclic graphs associated with credal sets. An *inference* is the computation of lower and upper probabilities for an event $\{X_q = x_q\}$ given *evidence* D — here D indicates a set of observed variables. Those variables that do not belong to D are called *hidden* variables and are denoted by H. A credal network is defined by local *separately specified* credal sets $\{K(X|Y = y)\}$ when these credal sets are not related for different values of the conditioning variables Y. The *strong extension* of the network is the convex hull of the set containing all joint distributions that factorize as $\prod_i P(X_i|pa(X_i))$, where each conditional distribution $P(X_i|pa(X_i) = \pi_k)$ is selected from the local credal set $K(X_i|pa(X_i) = \pi_k)$ [Cozman, 2000b]. A strong extension defines a joint credal set where each vertex is a possible combination of conditional distributions (each vertex of this set can be represented by a Bayesian network). There is a vertex that minimizes and a vertex that maximizes $P(X = x|E)$ [Fagiuoli and Zaffalon, 1998]; thus an inference in a strong extension can be viewed as an optimization problem over the set of potential vertices.

Exact inference in general credal networks displays high complexity. Apparently the only tractable situation is represented by polytree-shaped networks with binary variables [Fagiuoli and Zaffalon, 1998]. Other than this, even inference in polytree-

shaped credal networks is a NP-complete problem [da Rocha and Cozman, 2002]. The most promising methods for exact inference are based on multilinear programming [de Campos and Cozman, 2004], but even these methods face difficulties for networks of medium size (say twenty to twenty five reasonably connected nodes). Due to the complexity of exact inference, several algorithms for approximate inference have been developed [Cano and Moral, 2002, da Rocha et al., 2003, Ide and Cozman, 2004, Tessem, 1992].

3. Set-based Variational Methods

In Section 3.1., we give a brief review of variational methods, focusing on those methods that are applied to Bayesian networks. We follow the terminology used in previous introductory material [Jaakkola, 2000, Winn, 2003]. In Section 3.2., the set-based variational method is presented.

3.1. Basic of variational methods

Modern variational methods are popular in various fields such as control theory, optimization, statistics, economics and machine learning; recently several variational approaches have been successfully used in inference and estimation of densely connected graphical probability models [Saul et al., 1996, Saul and Jordan, 1996].

Suppose that we have a directed graph associated with a joint distribution $P(X)$, where X represents the set of variables. We want to approximate $P(H|D)$ by a distribution $Q(H)$, where H is the set of hidden variables and D is the evidence — thus $X = \{H, D\}$. We choose the *Kullback-Leibler* (KL) divergence as a dissimilarity measure between $Q(H)$ and $P(H|D)$:

$$\begin{aligned} KL(Q||P) &= \sum_H Q(H) \ln \frac{Q(H)}{P(H|D)} \\ &= \sum_H Q(H) \ln Q(H) + \ln P(D) - \sum_H Q(H) \ln P(H, D) \end{aligned} \quad (1)$$

The first term of the last expression is the *negative entropy* [Jaakkola, 2000]. The second term $\ln P(D)$ is constant with respect to $Q(H)$. The last term is the expectation of $\ln P(H, D)$ with respect to $Q(H)$. From Equation (1), the divergence is null ($KL = 0$) when $Q(H) = P(H, D)$. The goal here is to find a good approximation $Q(H)$ to $P(H|D)$ by minimizing $KL(Q||P)$.

An approximate model that has been successfully used in variational methods in many areas is the fully factorized distribution; this is often called the *mean field* approximation [Parisi, 1988]. The idea is that the global behavior of distributions should be approximated by a set of independent variables [Saul et al., 1996]. Using the fully factorized distribution, we can minimize the divergence KL in a iterative and computationally efficient manner. Consider then a fully factorized distribution $Q(H) = \prod_i Q_i(H_i)$. Substituting in Equation (1) we obtain:

$$\begin{aligned} KL(Q||P) &= \sum_H \prod_i Q_i(H_i) \ln(\prod_i Q_i(H_i)) - \sum_H \prod_i Q_i(H_i) \ln P(H, D) + \ln P(D) \\ &= -\sum_i \mathbf{H}(Q_i) + \ln P(D) - \sum_H \prod_i Q_i(H_i) \ln P(H, D), \end{aligned}$$

where **H** represents the entropy. The idea is to minimize KL with respect to $Q_j(H_j)$ by assuming fixed terms $Q_i(H_i)$, $i \neq j$. Then we obtain the expression of $Q_j^*(H_j = h_j)$ that partially minimizes $KL(Q||P)$:

$$Q_j^*(H_j = h_j) \;\; = \;\; k_\lambda \exp\left(\sum_{H_{i \neq j}} \prod_i Q_i(H_i) \ln P(H, D) \right). \tag{2}$$

The normalization constant is calculated computing Equation (2) for all values of the variable H_j: $k_\lambda = \sum_{h_j \in H_j} Q_j^*(h_j)$. This is the global minimum of $KL(Q||P)$ with respect to $Q_j(H_j)$. Once we update variable H_j, we must choose another variable H_i to be updated, and so on for all hidden variables H. This process minimizes the KL divergence iteratively.

Consider the computational cost of the updating Equation (2). The key point here is that this updating expression can be simplified so that only "local" computations are needed for each hidden variable H_j ("local" in the sense that they only refer to terms that are "close" in the graph underlying the network) [Winn, 2003]. Taking into account that $P(H, D)$ can be written in terms of conditional distributions $P(X_k|pa_k)$, Equation (2) can be reformulated as:

$$\begin{aligned} Q_j^*(H_j = h_j) \;\; = \;\; & k_\lambda \exp(\sum_{H_i \in \{pa_j\}} \prod_i Q_i(H_i) \ln P(H_j|pa_j) \\ & + \sum_{k \in ch_j} \sum_{H_i \in \{X_k, pa_k\}} \prod_i Q_i(H_i) \ln P(X_k|pa_k)), \end{aligned} \tag{3}$$

where $\{pa_j\}$ is the set of parent nodes of H_j and $\{ch_j\}$ is the set of children nodes of H_j. This means that we only have to make "local" computations involving conditional distributions for variables in the Markov blanket of H_j (the *Markov* blanket of variable H_j is the set of nodes containing the parents of H_j, the children of H_j, and the parents of children of H_j).

3.2. Set-based variational methods

In this section we review the concept of "set-based" variational methods, that has been presented in [Ide and Cozman, 2005]. The goal of the *set-based* variational approach is to approximate the original joint credal set by a set of probability intervals $I_Q(H_j = h_j) = [\underline{Q}(H_j = h_j), \overline{Q}(H_j = h_j)]$ for each value of each variable H_j. These intervals are directly related to the lower and upper probabilities $\underline{P}(H_j = h_j|E)$ and $\overline{P}(H_j = h_j|E)$. The set-based approach mimics the local computations in the "standard" mean field methods, but the local computations are replaced by interval computations — the result is a process that iteratively computes probability intervals for all variables H_j. Consider the updating Equation (3), we get lower bounds $\underline{Q}(H_j = h_j)$ (upper bounds $\overline{Q}(H_j = h_j)$ are analogous):

$$\begin{aligned} \underline{Q}_j^*(H_j = h_j) \;\; \propto \;\; & \min \sum_{H_i \in \{pa_j\}} \prod_i Q_i(H_i) \ln P(H_j|pa_j) \\ & + \sum_{k \in ch_j} \sum_{H_i \in \{X_k, pa_k\}} \prod_i Q_i(H_i) \ln P(X_k|pa_k), \end{aligned} \tag{4}$$

where:

$$\begin{aligned} Q_i(H_i) &\in I_Q(H_i = h_i), \\ P(H_j|pa_j = \pi_j) &\in K_P(H_j|pa_j = \pi_j), \\ P(X_k|pa_k = \pi_k) &\in K_P(X_k/pa_k = \pi_k). \end{aligned}$$

Unfortunately, the exact computation of the intervals $I_Q(H_j = h_j)$ leads us to a global combinatorial search, because these intervals are all interrelated. A vertex that attains the minimum or maximum of $Q(H_j = h_j)$ is not necessarily the same vertex that attains the minimum or maximum of $Q(H_i = h_i)$ for $i \neq j$; enforcing such relationships leads us to combinatorial explosion.

We now consider an approximate solution that circumvents the complexity problem just discussed. In fact, we introduce a second approximation on top of the variational one. The idea is to make intervals $I_Q(H_j = h_j)$ unrelated to each other, thus reducing the computation of an approximate interval to a truly "local" combinatorial problem. To do this, we update $I_Q(H_j = h_j)$ by Equations (4) , using the outer bounds of $I_Q(H_i = h_i)$ for $i \neq j$. That is, we use approximate values $I_Q(H_i = h_i)$, computed for those vertices that $Q(H_j = h_j)$ is maximum and minimum, instead of vertices where $Q(H_i = h_i)$ is maximum or minimum. We call this approximation an *Outer Bound Step*.

The outer bound step is depicted in Figure 1. Suppose that we are approximating $I_P(H_0, H_1)$ by $I_Q(H_0, H_1)$, where H_0 and H_1 are hidden variables, using a variational method. To update the interval $I_Q(H_1 = h_1)$, we need the values of $I_Q^*(H_0 = h_0)$ for those vertices where $I_Q(H_1 = h_1)$ are attained (notice that $I_Q(H_1 = h_1)$ is an approximation of $I_P(H_1 = h_1|E)$, and so on.). For variable H_0 we just consider those vertices that lead to the extreme values of $I_Q(H_0 = h_0)$ — we take these values as an outer bound approximation for interval $I_Q^*(H_0 = h_0)$. Note that such a round of approximations happens in each iteration of the mean field scheme. The method can be summarized as follows:

- **Mean field approximation.** We approximate the original strong extension $K_P(X)$ by an approximate credal set $K_Q(X)$ that is the strong extension of local

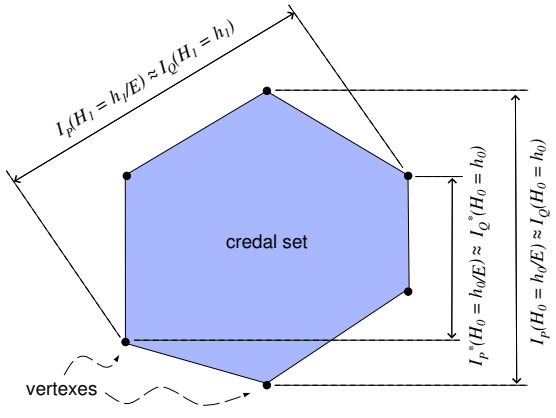

Figure 1: The outer bound step. Interval $I_Q^*(H_0 = h_0)$ is approximated by an outer bound interval $I_Q(H_0 = h_0)$.

probability intervals $I_Q(H_j = h_j) = [\underline{Q}(H_j = h_j), \overline{Q}(H_j = h_j)]$ for each value of $Q(H_j)$. Here $\underline{Q}(H_j = h_j)$ approximates the minimum value of $P(H_j = h_j|E)$ and $\overline{Q}(H_j = h_j)$ approximates the maximum value. This approximation moves us to a local computation.

- **Outer bound step.** We update the interval $I_Q(H_j = h_j)$ by Equations (4), using the outer bounds $I_Q(H_i = h_i)$, for $i \neq j$. This approximation step is crucial to keep the local computation property sought by the variational approximation.

We transform the global combinatorial problem into a local one, because: (1) The mean field approximation is based on a "local" updating mechanism (Equation 3), restricted to the Markov Blanket; (2) The outer bound step makes each updating step combinatorially unrelated to all others, thus guaranteeing locality of computation.

4. Structured Variational-2U Algorithm (SV2U)

The naive variational mean field approach described previously is computationally attractive, but it is often unable to yield sufficiently accurate results [Jaakkola, 2000, page 16]. A natural idea to improve over the naive mean field method is to combine it with exact calculations — for example, to approximate the original intractable network with tractable substructures such as trees and chains [Saul and Jordan, 1996]. We introduce a *Structured Variational-2U Algorithm* (SV2U) that uses the set-based variational approach to approximate a multi-connected binary credal network by a polytree-structure network and run the exact interval propagation 2U algorithm. Note that for binary variables the intervals $I_Q = [p_{low}, p_{high}]$ (of Equation (4)) corresponds to the local credal sets K_Q.

Consider the idea of *structured* variational method [Jaakkola, 2000, page 18], to apply the mean field approach, taking each factor Q_i over a cluster of variables $c_i = \{X_i, pa_i\}$, so as to have the approximate joint distribution: $Q(X) = \prod_i Q_i(c_i)$. We have a new updating equation for the cluster $c_j = \{X_j, pa_j\}$ [Winn, 2003, page 104]:

$$\begin{aligned} Q_j^*(X_j|pa_j) \quad \propto \quad & \sum_{k \in G_{X_j}} \sum_{l \in \{\mathbf{X}/c_j\}} \prod_l Q_l(X_l) \ln P(X_k|pa_k) \\ & - \sum_{i \in C_{X_j}} \sum_{l \in \{\mathbf{X}/c_j\}} \prod_l Q_l(X_l) \ln Q_i(X_i|pa_i), \end{aligned} \tag{5}$$

where G_{X_j} represents the set of clusters $g_k = \{X_k, pa_k\}$ that depend on X_j in the original network and C_{X_j} is the set of clusters c_i that depend on X_j in the approximate network, excluding c_j itself (in Bayesian networks, these sets of clusters are Markov blankets). Note that the expectation in Equation (5), $\mathbf{E}[\sum_{k \in G_{X_j}} \ln P(X_k|pa_k) - \sum_{i \in C_{X_j}} \ln Q_i(X_i|pa_i)]$, is computed with respect to the approximated structure $Q(X) = \prod_i Q_i(c_i)$, $i \neq j$. Also note that we need to update just those clusters c_j that don't belong to the original network cluster set G; that is, we have to update just those conditional distributions that are modified.

We use this structured variational approach described before and develop it for credal networks. Apply the set-based approach to the Equation (5), and obtain the updating expressions analogous to Equation (4):

$$\underline{Q}_j^*(X_j|pa_j) \propto \min \sum_{k\in G_{X_j}} \sum_{l\in\{\mathbf{X}/c_j\}} \prod_l Q_l(X_l|pa_l)\ln P(X_k|pa_k) - \sum_{i\in C_{X_j}} \sum_{l\in\{\mathbf{X}/c_j\}} \prod_l Q_l(X_l|pa_l)\ln Q_i(X_i|pa_i), \quad (6)$$

$$\overline{Q}_j^*(X_j|pa_j) \propto \max \sum_{k\in G_{X_j}} \sum_{l\in\{\mathbf{X}/c_j\}} \prod_l Q_l(X_l|pa_l)\ln P(X_k|pa_k) - \sum_{i\in C_{X_j}} \sum_{l\in\{\mathbf{X}/c_j\}} \prod_l Q_l(X_l|pa_l)\ln Q_i(X_i|pa_i), \quad (7)$$

where:

$$\begin{aligned} Q_l(X_l|pa_l=\pi_l) &\in K_Q(X_l=x_l|pa_l=\pi_l),\\ Q_i(X_i|pa_i=\pi_i) &\in K_Q(X_i=x_i|pa_i=\pi_i),\\ P(X_k|pa_k=\pi_k) &\in K_P(X_k/pa_k=\pi_k). \end{aligned}$$

Structured Variational 2U: SV2U

Input: Multi-connected binary credal network N.

Output: Approximations for lower/upper probabilities of some variable X_j.

01. Find a $cutset$ of the original network N,with joint distribution $P(X)=\prod_k P_k(g_k)$, to obtain an approximated polytree structure N_p.
02. Find the set of clusters $G=\{g_k\}$ of network N and $C=\{c_i\}$ of the network N_p.
03. Initialize for those clusters c_i that don't belong to G, c_j, $K_Q(X_j=x_j|pa_j=\pi_j)=[0.5,0.5]$.
03. For those clusters c_j repeat until convergence:
04. For all possible parent configurations of cluster c_j, $\{pa_j=\pi_j\}$:
05. Compute $\underline{Q}^*(X_j=x_j|pa_j=\pi_j)$ and $\overline{Q}^*(X_j=x_j|pa_j=\pi_j)$ for each value of X_j and normalize these values, from Equations (6) and (7).
06. Keep the minimum and maximum values of $Q^*(X_j=x_j|pa_j=\pi_j)$ and update $K_Q(X_j=x_j|pa_j=\pi_j)$.
07. Run the 2U algorithm in the polytree-structure associated to the joint distribution $Q(X)=\prod_k P_k(g_k)\times\prod_j Q_j(c_j)$, where k is the index of those clusters that belong to both cluster sets G and $Q_j(c_j)\in K_Q(X_j=x_j|pa_j=\pi_j)$; and keep the minimum and maximum values of probabilities of some variable X_j.

Figure 2: SV2U algorithm description.

The proposed set-based structured variational algorithm (SV2U) is described in Figure 2. Suppose that we have a multi-connected network N with joint distribution $P(X)=\prod_k P_k(g_k)$. The idea is to approximate this with a polytree-structure network N_p, finding a *loopy cutset* [Pearl, 1988], with distribution $Q(X)=\prod_i Q_i(c_i)$ (Line 01). Define the set of clusters $G=\{g_k\}$ of network N and $C=\{c_i\}$ of the network N_p. Find the approximate conditional distributions of those clusters c_i that doesn't belong to the set G. Define this set as c_j. This means that we have just to compute the distributions

$Q(c_j)$, what is done iteratively, like as described in Section 3.1., until get convergence of these distributions (Lines 03-06). During this iteration, we compute the $K_Q(X_j = x_j|pa_j = \pi_j) \in [\underline{Q}^*(X_j = x_j|pa_j = \pi_j), \overline{Q}^*(X_j = x_j|pa_j = \pi_j)]$. It means that at the end we have a polytree-structure distribution $Q(X) = \prod_k P_k(g_k) \times \prod_j Q_j(c_j)$, where k is the index of those clusters that belong to both cluster sets G and C; and $Q_j(c_j) \in K_Q(X_j = x_j|pa_j = \pi_j)$. With this polytree-structure distribution we compute the desired approximations for lower/upper probabilities, using the exact inference engine 2U (Line 07).

To illustrate and show results of SV2U algorithm, consider an example. The *Pyramid* network (Figure 3) is a multilayered graph associated with 28 binary variables (assume values "0" or "1") and local connections among layers. [Murphy et al., 1999]. We associate each variable with a binary credal set, a convex hull of the set containing all joint distributions that factorize as $\prod_i P(X_i|pa(X_i))$, $i = \{1, ..., 28\}$, where each conditional distribution $P(X_i|pa(X_i) = \pi_k)$ is selected from the local credal set $K_P(X_i|pa(X_i) = \pi_k) = [p_{low}, p_{high}]$. One possible cutset is formed by the arcs (1,6), (2,6), (2,8), (3,8), (3,10), (4,10) and (4,12). Removing these arcs we get a polytree associated to a credal set that differs from the original network for local credal sets of variables X_6, X_8, X_{10} and X_{12}. It means that local credal sets $K_P(X_6|X_1, X_2)$, $K_P(X_8|X_2, X_3)$,$K_P(X_{10}|X_3, X_4)$ and $K_P(X_{12}|X_4)$ are approximated by $K_Q(X_j)$, $j = \{6, 8, 10, 12\}$. Assuming that there are no evidence nodes, the expressions for updating Equations (6) and (7) are:

$$
\begin{aligned}
Q^*(X_6) &= k_\lambda \exp[\sum_{X_1,X_2} P(X_1)P(X_2) \ln P(X_6|X_1, X_2)], \\
Q^*(X_8) &= k_\lambda \exp[\sum_{X_2,X_3} P(X_2)P(X_3) \ln P(X_8|X_2, X_3)], \\
Q^*(X_{10}) &= k_\lambda \exp[\sum_{X_3,X_4} P(X_3)P(X_4) \ln P(X_{10}|X_3, X_4)], \\
Q^*(X_{12}) &= k_\lambda \exp[\sum_{X_4} P(X_4) \ln P(X_{12}|X_4)].
\end{aligned}
$$

From these equations we get approximated credal sets $K_Q(X_6 = 0) = [0.099, 0.346]$, $K_Q(X_8 = 0) = [0.203, 0.664]$, $K_Q(X_{10} = 0) = [0.278, 0.753]$ and $K_Q(X_{12} = 0) = [0.532, 0.810]$. Running 2U algorithm in this polytree and computing the lower/upper probabilities for all variables, we get a mean square error (MSE), between exact probabilities and those obtained by SV2U, of 2%. Results can be seen in Figure 4.

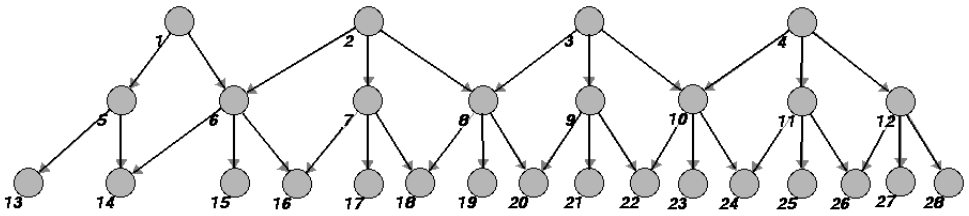

Figure 3: Pyramid network, multilayered graph associated with binary variables, used to test the SV2U algorithm.

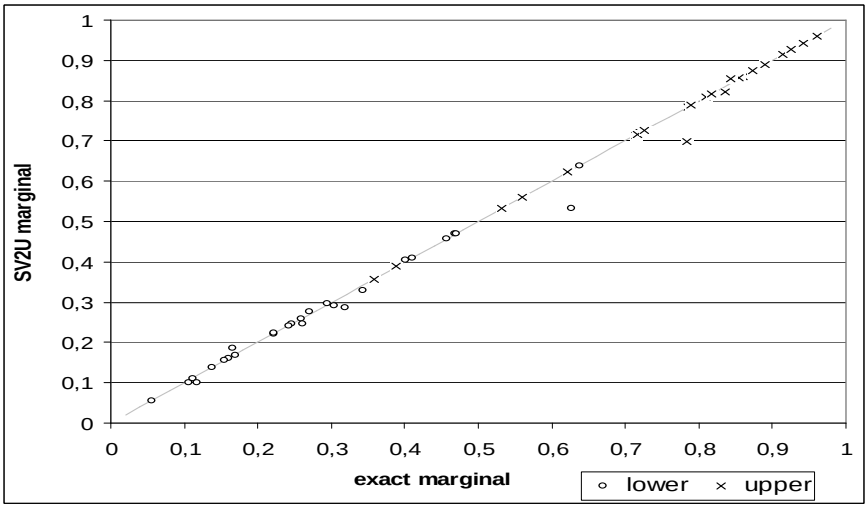

Figure 4: Results of the SV2U algorithm in the Pyramid network. Lower/upper probabilities are computed exactly and approximately, and MSE=2% is obtained.

5. Summary and Conclusions

In this paper we have introduced a structured variational approach to inference in binary credal networks. We proposed the SV2U algorithm, that make uses of a set-based structured variational method and the 2U exact interval propagation algorithm. The paper is an initial step in the construction of general variational approximation methods for credal networks. The next step is the implementation of the SV2U algorithm for general structures so as to study their empirical behavior, what requires the development of efficient methods for computing the expectation in Equations (6) and (7). From experiments in Pyramid network , the SV2U algorithm produces the result of MSE=2%, relatively worst than MSE=1.2% of the L2U algorithm [Ide and Cozman, 2004], but with solid convergence guarantees. This simple experiment shows us how the SV2U algorithm can provide good results.

Given the generality and flexibility of variational methods, the set-based methods proposed in this paper seem to be a promising approach to credal networks. Models that contain continuous variables and local credal sets defined by infinitely many constraints are not handled by most existing algorithms; they can in principle be dealt with using variational principles. Hopefully this paper will serve as a initial step in a fruitful avenue of research.

Acknowledgements

The first author has been supported by FAPESP (through grant 02/0898-2). The second author has been partially supported by CNPq (through grant 3000183/98-4). This work has received support from HP Brazil R&D.

References

Cano, A. and Moral, S. (2002). Using probability trees to compute marginals with imprecise probabilities. *International Journal of Approximate Reasoning*, 29:1–46.

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

Cozman, F. G. (2000b). Separation properties of sets of probability measures. In *Conference on Uncertainty in Artificial Intelligence*, pages 107–114.

Cozman, F. G. (2004). Graphical models for imprecise probabilities. *International Journal of Approximate Reasoning*. (to be published)

Cozman, F. G., de Campos, C. P., Ide, J. S., and da Rocha, J. C. (2004). Propositional and relational bayesian networks associated with imprecise and qualitative probabilistic assessments. In *20th Conference on Uncertainty in Artificial Intelligence*.

da Rocha, J. C. and Cozman, F. G. (2002). Inference with separately specified sets of probabilities in credal networks. In *Conference on Uncertainty in Artificial Intelligence*, pages 430–437, San Francisco, CA. Morgan Kaufmann.

da Rocha, J. C., Cozman, F. G., and de Campos, C. P. (2003). Inference in polytrees with sets of probabilities. In *Conference on Uncertainty in Artificial Intelligence*, pages 217–224, San Francisco, CA. Morgan Kaufmann.

de Campos, C. P. and Cozman, F. G. (2004). Inference in credal networks using multilinear programming. In *Proc. of the Starting Artificial Intelligence Researchers Symposium, STAIRS 2004, Valencia - Spain*, pages 50–61, Amsterdam - The Netherlands. IOS Press.

Fagiuoli, E. and Zaffalon, M. (1998). 2U: An exact interval propagation algorithms for polytrees with binary variables. *Artificial Intelligence*, 106:77–107.

Ide, J. S. and Cozman, F. G. (2004). IPE and L2U: Approximate algorithms for credal networks. In *Proc. of the Starting Artificial Intelligence Researchers Symposium, STAIRS 2004, Valencia - Spain*, pages 118–127, Amsterdam - The Netherlands. IOS Press.

Ide, J. S. and Cozman, F. G. (2005). Approximate inference in credal networks by variational mean field methods. In *4th International Symposium on Imprecise Probabilities and Their Applications*, Pittsburgh, Pennsylvania. (conditionally accepted)

Jaakkola, T. S. (2000). Tutorial on variational approximation methods. *Advanced mean field methods: theory and practice*.

Levi, I. (1980). *The Enterprise of Knowledge*. MIT Press, Cambridge, Massachusetts.

Murphy, K. P., Weiss, Y., and Jordan, M. I. (1999). Loopy belief propagation for approximate inference: An empirical study. In *In Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, pages 467–475.

Parisi, G. (1988). *Statistical Field Theory*. Addison-Wesley, MA.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems*. Morgan-Kaufman.

Saul, L. K., Jaakkola, T. S., and Jordan, M. I. (1996). Mean field theory for sigmoid belief networks. *Journal of Artificial Intelligence Research*, 4:61–76.

Saul, L. K. and Jordan, M. I. (1996). Exploiting tractable substructures in intractable networks. In Touretzky, D. S., Mozer, M. C., and Hasselmo, M. E., editors, *Advances in Neural Information Processing Systems*, volume 8, pages 486–492. MIT Press, Cambridge, MA.

Tessem, B. (1992). Interval probability propagation. *International Journal of Approximate Reasoning*, 7:95–120.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London.

Winn, J. (2003). *Variational Message Passing and its Applications*. PhD thesis, Department of Physics, University of Cambridge, Cambridge,UK.