An hybrid Bayesian network structure learning algorithm based on correlated itemset mining techniques

Zahra Kebaili, Alexandre Aussem

An hybrid Bayesian network structure learning algorithm based on correlated itemset mining techniques

Zahra Kebaili and Alexandre Aussem

Laboratoire LIESP, Université de Lyon 1
F-69622 Villeurbanne, France
{zkebaili, aaussem}@bat710.univ-lyon1.fr

RÉSUMÉ. *Cet article présente une nouvelle méthode hybride d'apprentissage de la structure des réseaux Bayésiens combinant des techniques de data mining, d'apprentissage sous contraintes et à base de score. Le but de cette méthode est d'identifier et de représenter des conjonctions de facteurs potentiellement impliqués dans des relations de dépendance, ce dont sont incapables les méthodes d'apprentissage actuelles lorsque la dépendance des variables prises deux à deux n'est pas suffisamment marquée (ex : XoR bruité). Cette méthode permet d'identifier des associations qui violent l'hypothèse de fidélité sur laquelle est basée la majorité des algorithmes sous contraintes. L'algorithme opère en deux temps : (1) découverte par niveau des motifs corrélés minimaux, (2) construction du graphe à partir de ces corrélés. La méthode est illustrée sur un benchmark qui met à défaut les algorithmes standards d'apprentissage sous contraintes et à base de score*

ABSTRACT. *In this paper, we propose a novel hybrid method for bayesian network structure learning that combines ideas from datamining, constraint-based and search-and-score learning. The aim of this method is to identify and to represent conjunctions of boolean factors implied in probabilistic dependence relationships, that may be ignored by constraint and scoring-based learning proposals when the pairwise dependencies are weak (e.g., noisy-XOR). The method is therefore able to identify some specific high order interactions that cause the violation of the faithfulness assumption on which are based most constraint-based methods. The algorithm operates in two steps: (1) extraction of supported and minimally correlated itemsets, and (2), construction of the structure by these itemsets. The method is illustrated on a simple but realistic benchmark plaguing the standard scoring and constraint-based algorithms.*

MOTS-CLÉS : *réseaux Bayésiens, apprentissage statistique, data mining.*

KEYWORDS: *Bayesian networks, statistical learning, data mining.*

1. Introduction

Bayesian Networks (BNs) are a graphical representation for probability distributions. Formally, BN are directed acyclic graphs (DAG) modelling probabilistic dependencies among variables. The DAG reflects the structure of a problem (usually a graph of causal dependencies in the modelled domain), while local interactions among neighboring variables are quantified by conditional probability distributions. Learning a Bayesian network from data requires both identifying the model structure $\mathcal{G}$ and identifying the corresponding set of model parameter values. Given a fixed structure, however, it is straightforward to estimate the parameter values. As a result, research on the problem of learning Bayesian networks from data is focused on methods for identifying "good" DAG structures from data.

Constraint-based (CB) methods search a database for conditional independence relations and constructs graphical structures called "patterns" which represent a class of statistically indistinguishable DAGs (Brown *et al.*, 2005; Cheng *et al.*, 2002; Tsamardinos *et al.*, 2006; Peña *et al.*, 2007). CB approaches are relatively quick, deterministic, and have a well defined stopping criterion. Almost all CB methods assume faithfulness : when the distribution P is faithful to a DAG $\mathcal{G}$, the d-separations in the DAG identify all and only the conditional independencies in P. However, interesting interactions may implicitly be ignored. In particular, a causal relation $X, Y \Rightarrow Z$ (e.g., $Z = \text{XOR}(X, Y)$) is ignored by CB methods if the pairwise dependencies (X, Z) and (Y, Z) are too weak to be detected by standard statistical tests. This is interpreted as $Ind_P(X, Z)$ and $Ind_P(Y, Z)$, where P stands for the underlying data distribution, so the set $\{X, Y\}$ is considered as independent on Z. Clearly, such an error may have a dramatic impact on the classifier performance if Z is the variable to be predicted.

To overcome this problem, it is necessary to handle variable sets of unspecified size and estimate their probabilistic association degree according to a statistic measure. This comes unfortunately at the expense of increased complexity hence the need for data-efficient procedures. This paper considers a novel method based on a smart level-wise algorithm that have been originally proposed to mine a collection of subsets from an item space of random variables. The chi-squared statistic is used as a measure of significance for the itemsets and the so-called χ^2-support algorithm (Silverstein *et al.*, 1998) is used to extract the minimal correlated and supported itemsets. The χ^2 leads to a measure that is upward closed in the lattice of subsets of the item space. The upward closed property reduces the mining problem to the search for a border between correlated and uncorrelated itemsets in the lattice. We use a divide-and-conquer approach : each minimally correlated itemset is passed to an exhaustive search-and-score procedure and the best scoring sub-graph is added the current DAG skeleton. The exhaustive search in the space of graphical structures is made possible by considering a selection of minimally correlated itemsets of limited size.

Many hybrid algorithms (CB and scoring) have already been developed in the past. Their aim is to improve the robustness and efficiency of the DAG search. Usually, the first stage performs a CB search and uses the resulting graph as input into a second-

stage search-and-score method. (Dash *et al.*, 1999; Dash *et al.*, 2003) for instance describe variants of hybrid algorithms using the PC algorithm as a heuristic. Our goal is different, we seek to identify random or almost-deterministic interactions of the type $X, Y \Rightarrow Z$ that render the distribution almost unfaithful. Indeed, when pairwise relationships (X, Z) and (Y, Z) are weak and the number of data is small, the dependence test may reject the dependence hypothesis even if these pair of variables are part of a more complex interaction involving other variables. In fact, the question of weak dependencies implies addressing the problem of unfaithfulness. This is known to be a difficult topic.

In Section 2, we review the necessary notions of d-separation and faithfulness. In Section 3 and 4, we describe the principles of minimal correlated itemset discovery and the corresponding χ^2 algorithm. In Section 5, we describe the new hybrid algorithm in detail. The effectiveness of the novel method is illustrated in the final Section on a benchmark plaguing standard CB and scoring methods.

2. Background

For the paper to be accessible to those outside the domain, we recall briefly the principles of Bayesian networks (BN). We denote a variable by an upper case (e.g., A,B) and its value in lower case (e.g., a, b). We denote variable sets with bold-face capitalized tokens (e.g., **A**, **B**) and corresponding sets of values by bold-face lower case tokens (e.g., **a**, **b**). A Bayesian network (BN) is a tuple $< \mathcal{G}, P >$, where $\mathcal{G} = < \mathbf{V}, \mathcal{E} >$ is a directed acyclic graph (DAG) with nodes representing the random variables **V** and P a joint probability distribution on **V**. In addition, $\mathcal{G}$ and P must satisfy the Markov condition : every variable, $X \in \mathbf{V}$ is independent of any subset of its non-descendant variables conditioned on the set of its parents. Pearl (Geiger *et al.*, 1989) provides a graphical condition called d-separation that can be used to identify any independence constraint implied by a DAG model. We use $Ind_G(X, Y | \mathbf{Z})$ to denote the assertion that DAG G imposes the constraint (via d-separation) that for all values **z** of the set **Z**, X is independent of Y given **Z**=**z**. For a probability distribution P, we use $Ind_p(X, Y | \mathbf{Z})$ to denote the assertion that for all values **z** of the set **Z**, X is independent of Y given **Z**=**z** in P. We use $Dsep_{\mathcal{G}}(X; Y | \mathbf{Z})$ to denote the assertion that DAG $\mathcal{G}$ imposes the constraint, via d-separation, that for all values **z** of the set **Z**, X is independent of Y given $\mathbf{Z} = \mathbf{z}$. We say that P is *faithful* with respect to $\mathcal{G}$ if $Ind_P(X_i; ND_X | \mathbf{Pa}_i^{\mathcal{G}})$ implies $Dsep_{\mathcal{G}}(X; Y | \mathbf{Z})$. In other words, when P is faithful to a DAG $\mathcal{G}$, the d-separations in the DAG identify all and only the conditional independencies in P.

2.1. *Association measure*

CB methods rely on a probabilistic association measure between X and Y conditionally on **Z** denoted by $Assoc(X; Y | \mathbf{Z})$. The correctness of CB algorithms is usually proved under the assumption that $Ind_P(X; Y | \mathbf{Z})$ iff $Assoc(X; Y | \mathbf{Z}) < \alpha$ where α is

the critical value used in the test. $Assoc(X;Y|\mathbf{Z})$ can implemented with a number of statistical or information theoretic measures of association (conditional mutual information, χ^2 etc.). The only requirement for $Assoc(X;Y|\mathbf{Z})$ is to return a value lower than our critical value α when $Ind_P(X;Y|\mathbf{Z})$. In this work, the χ^2 serves as a conditional independence test as well as a measure for itemset association. For example, the χ^2 value of the subset $\{A,B,C\}$ is given by :

$$\chi^2_{ABC} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} \frac{(n_{ijk}-n'_{ijk})^2}{n'_{ijk}} \qquad [1]$$

where $n'_{ijk} = n_{i..}n_{.j.}n_{..k}/n^2$ and n_{ijk} is the number of times simultaneously $A = a_i$, $B = b_j$ and $C = c_k$ in the sample, that is, the value of the cell (i,j,k) in the contingency table. The statistic is compared against a critical value to decide upon of the acceptance or rejection of the null hypothesis of conditional independence. The distribution of χ^2_{ABC} is approximatively that of chi-squared with $\nu = (a-1)(b-1)(c-1)$ degrees of freedom. In this study, all the variables are boolean, so $\nu = 1$. The association between two variables X and Y given a conditioning set $\mathbf{Z}$ is usually implemented with the χ^2 statistical measures of association. For example, the association between A and B given C is given by :

$$\chi^2_{AB|C} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} \frac{(nijk-n'^k_{ij})^2}{n'_{ijk}} \qquad [2]$$

where $n'^k_{ij} = n_{i.k}n_{.jk}/n_{..k}$. $\chi^2_{AB|C}$ follows the χ^2 law with $\nu = (a-1)(b-1)c$ freedom degrees.

3. Minimal correlated itemsets

One of the most well-studied problems in data mining is mining for association rules in market basket data. Association rules, whose significance is measured via support and confidence, are indented to identify rule of the type "A customer purchasing item A often also purchases item B". However, they have been very few applications of association rule mining algorithms to the problem of learning the BN structure. This paper is concerned with bridging the gap between level-wise mining techniques and BN learning methods. We believe the identification of correlated patterns should aid the structure discovery process and be incorporated in the graphical structure. With this view in mind, we studied an algorithm proposed by Brin, Motwani and Silvertsein (Silverstein *et al.*, 1998). Motivated by the goal of generalizing beyond market baskets and the association rules used with them, they developed the notion of mining rules that identify correlations (generalizing associations). They considered both the

absence and presence of items as a basis for generating rules and proposed measuring significance of associations via the chisquared test for correlation from classical statistics.

But first, it is necessary to clarify the concept of correlation between characters (items). Let $p(a)$ be the probability that event a occurs and $p(\overline{a}) = 1 - p(a)$ the probability that event a does not occurs. In the same way, $p(\overline{a}b)$ is the probability that b occurs but a does not. The events a and b are said to be independent if $p(ab) = p(a)p(b)$. Likewise, if $p(abc) = p(a)p(b)p(c)$, then a, b, c are independent. If any of ab, $\overline{a}b$, $a\overline{b}$, $\overline{a}\overline{b}$ are dependent, then a and b are said to be *correlated*, and so on for bigger sets. Let V a the set of variables, we call *patterns* a subset of V. $\mathcal{D}$ an experiment databases, for which each variable $A \in V$ takes the occurrence a or $\overline{a}$. An important property of correlation is that it is upward closed : if a set of items S is correlated, so every superset of S is correlated too. This leads to a measure that is upward closed in the itemset lattice. This property reduces the mining problem to the search for a border between correlated and uncorrelated itemsets in the lattice. (Silverstein *et al.*, 1998) proposed an efficient algorithm that exploit a pruning strategies. In the following section we describe this algorithm for convenience.

4. The χ^2-support algorithm

Algorithm 1 χ^2-support

Require: $\chi^2_{dep}(1)$: cutoff value at the $1 - \alpha$ significance level
    s : Threshold for the minimum support
    p : Support fraction ($p > 0.25$)
Ensure: SIG : set of minimal correlated itemsets
 1: i=1 ;
 2: $CAND_1 \leftarrow \{$set of pairs$\}$
 3: $SIG \leftarrow \emptyset$
 4: **while** $CAND_i \neq \emptyset$ **do**
 5:   $NOTSIG_i \leftarrow \emptyset$
 6:   **for all** *itemset* $\in CAND_i$ **do**
 7:     **if** more $p\%$ of the contingency cells of *itemset* have a support $> s$ **then**
 8:       **if** $\chi^2(itemset) \geq \chi^2_{dep}(1)$ **then**
 9:         $SIG \leftarrow SIG \cup itemset$
10:      **else**
11:         $NOTSIG_i \leftarrow NOTSIG_i \cup itemset$
12:      **end if**
13:     **end if**
14:   **end for**
15:   $CAND_{i+1} = $ **GenerCandidates**$(NOTSIG_i)$
     //GenerCandidats : generate a set of variables of size i+1 starting from sets of size i
16:   $i \leftarrow i + 1$
17: **end while**
18: return SIG

The support is different from that used in the support-confidence framework, because unlike in the support-confidence framework we mine for negative dependence. In other words, the support-confidence framework only look at the top-left cell in chi-squared contingency table. This definition is extended as follows : A set of items S has support s at the $p\%$ level if at least $p\%$ of the cells in the contingency table for S have value at least equal to s. The support is down-ward closed. Combining the chisquared correlation rule with pruning via support, we obtain the χ^2-support algorithm. We say that a variable set is significant if it is supported and minimally correlated. The key observation is that a variable set at level $i + 1$ can be significant only if all its subsets at level i have a support and none of its subsets at level i are correlated. Thus, for level $i + 1$, all we need is a list of the supported but uncorrelated itemsets from level i. This list is held in $NOTSIG$. The list SIG, which holds the supported and correlated variable sets, is the output set of interest. $CAND$ is a list which builds variable set candidates for level $i + 1$ from the $NOTSIG$ list at level . In the sequel, Sig_i denotes the set of supported and minimally correlated itemsets of size i.

5. A new hybrid algorithm

This paper considers a method that combines ideas from CB and search-and-score structure learning to produce a more reliable DAG in the presence of complex interactions between variables. The novel approach combines the main advantages of both methods yet avoids their difficulties. It is a two-stage hybrid algorithm, where in the first stage, the χ^2-support algorithm (Silverstein *et al.*, 1998) is run in a bottom-up fashion in the itemset lattice, to identify significantly correlated itemsets as input into a second stage hybrid algorithm. Like the TPDA algorithm (Cheng *et al.*, 2002), the DAG construction operated in three phases : drafting, thickening and thinning.

More specifically, the first phase process the pairs of correlated variables sequentially in descending order of the χ^2. Correlated patterns $\{X, Y\} \in Sig_2$ become adjacent nodes X and Y in the DAG $\mathcal{G}$ under construction unless they can be d-separated in $\mathcal{G}$. Procedure $EdgeNeeded_H(\mathcal{G}, X, Y, \mathcal{D}, \alpha)$ proposed by Cheng et al. in (Cheng *et al.*, 2002) is used. This procedure uses an heuristic to close all possible indirect pipelines between X and Y. It tries to identify a cut-set in polynomial time compared to the exponential number of CI tests 2^k that would normally be required to test each subset. The procedure tries to identify by including/excluding neighbors of X (or Y) one at a time, to find a proper cut-set $\mathbf{Z}$ such that $Ind_P(X, Y|\mathbf{Z})$.

The next stage exploits the correlated sets Sig_i of higher order ($i > 2$), in ascending order of their size i and, for equal size, in descending order of their χ^2 value. These patterns are filtered by a heuristic. The selection is carried out using the concept of *correlated equivalent* discussed below. Then an exhaustive search over the current correlated variables is carried out to find the subDAG with the highest score. The one that best explains the observed data is added to $\mathcal{G}$.

The last stage attempts to remove the false positives that may have entered in Phase 1. This is achieved by testing, using the procedure $EdgeNeeded$, whether $Ind(X, Y|\mathbf{S})$ holds for some set $\mathbf{S} \in \mathbf{V}$. If the condition holds, the link (X, Y) is removed from $\mathcal{G}$. Therefore, this stage is called the thinning phase. Procedure $EdgeNeeded_H(\mathcal{G}, X, Y, \mathcal{D}, \alpha)$ is based on a heuristic whereas $EdgeNeeded(\mathcal{G}, X, Y, \mathcal{D}, \alpha)$ is exact. It is based on the idea that two consecutive nodes in a path always close the path but require more computational time. So, just as TPDA, rather then call the full $EdgeNeeded$ procedure, the hybrid procedure uses the approximation in the first stage to speed up the algorithm and only calls $EdgeNeeded$ at the end of the third stage.

5.1. *Correlated itemset filtering*

As mentioned above, finding correlated patterns amounts to finding a border in the itemset lattice using the χ^2 support algorithm. In the worst case, when the border lies in the middle of the lattice, it is exponential in the number of variables and all the supported and minimally correlated sets on the border do not always provide useful information for the graph construction. Some of them are redundant or not interesting due to correlation transitivity. We therefore introduce a filtering technique aiming at reducing the number of correlated itemsets that have to be considered in the construction of the BN structure. Let $G(\mathbf{V}, \mathbf{E})$ be a graph. Before defining the notion of *correlated equivalent sets*, we introduce a symmetrical binary relation SUB defined on $\mathbf{V} \times \mathbf{V}$ in $\mathbf{E}$, such as $\mathbf{X}, \mathbf{Y} \subset \mathbf{V}$; SUB$(\mathbf{X}, \mathbf{Y})$ is verified if and only if : $\forall x \in \mathbf{X}$, $\exists y \in \mathbf{Y}$ such as $(u, v) \in \mathbf{E}$ and vice-versa.

Correlated equivalent sets : Let $\mathbf{Z}$ and $\mathbf{W}$ be two supported and minimally correlated sets. Define $\mathbf{U} = \mathbf{Z} \cap \mathbf{W}, \mathbf{Z}' = \mathbf{Z} \backslash \mathbf{U}$ and $\mathbf{W}' = \mathbf{W} \backslash \mathbf{U}$. Then if SUB$(\mathbf{Z}', \mathbf{W}')$, $\mathbf{W}$ and $\mathbf{Z}$ are called a correlated equivalents.

The idea behind this definition is to skip from analysis an itemset that is closely connected to another itemset that has been treated earlier in Phase II of Algorithm 2. For illustration purposes, consider an example. Let : $G = (V, E)$ a graph with vertices $\mathbf{V} = \{A, B, C, D, K, M\}$ and edges $\mathbf{E} = \{(A, B), (A, C), (B, C), (K, B), (C, M), (D, A)\}$. $\mathbf{Sig}_3 = \{ABC, BCD, BDM, KMD\}$ the correlated set of size 3 ordered in the χ^2 descending order. For the correlated BCD, ABC is a correlated equivalent, because SUB(A, D) is verified since $(A, D) \in \mathbf{E}$. For BDM, ABC is a correlated equivalent because SUB(DM, AC), since $\{(D, A), (C, M)\} \subset \mathbf{E}$. For KDM, ABC is a equivalent [SUB$(KDM, ABC) = \{(K, B), (D, A), (M, C)\} \subset \mathbf{E}$].

Algorithm 2 *Hybrid algorithm*

Require: .
    D : A database over a set of variables $\mathbf{V}$
    $Ind(X, Y|\mathbf{Z})$: a test of conditional independence
    α : the significant level of the test
    $\mathbf{SIG}$: $\{Sig_2, \cdots, Sig_n\}$ minimal correlated variable sets of increasing size
Ensure: . $\mathcal{G}$: the DAG

1: **Set** $\mathcal{G} = G(\mathbf{V}, \emptyset)$

    Phase I :*Process correlated variable pairs Sig_2*
2: **for all** $(X, Y) \in \mathbf{Sig}_2$ **do**
3:    **if** $EdgeNeeded_H(\mathcal{G}, X, Y)$ **then**
4:        Add link (X, Y) to $\mathcal{G}$
5:    **end if**
6: **end for**

    Phase II : *Process correlated itemsets of size > 2*
7: **for all** $i > 2$ **do**
8:    **for all significant** $\mathbf{X} \in \mathbf{Sig}_i$ **do**
9:        **if** There is no previously treated itemset Y such that $IsCorrEquiv(X, Y, \mathcal{G})$ **then**
10:            Find the best scoring sub-DAG $\mathcal{G}'$ over variables in Sig_i
11:            Add $\mathcal{G}'$ to $\mathcal{G}$
12:        **end if**
13:    **end for**
14: **end for**

    Phase III :*Thining*
15: **for all** $(X, Y) \in \mathcal{G}$ that was added in Phase I **do**
16:    **if** Not $EdgeNeeded(\mathcal{G}, X, Y)$ **then**
17:        Remove (X, Y) from $\mathcal{G}$
18:    **end if**
19: **end for**

20: **function** $IsCorrEquiv(\mathbf{X}, \mathbf{Y}, \mathbf{G})$ *r*eturn 1 if X and Y are correlated equivalents in $\mathcal{G}$.

---

6. Experiments

The proposed method was implemented in Matlab with the BNT and the BNT SLP Toolboxes (Murphy, 1997; Leray *et al.*, 2004). To illustrate our method, we need a specific benchmark that exhibits interactions of the type $X, Y \Rightarrow Z$ that render the distribution almost unfaithful, typically when pairwise relationships (X, Z) and (Y, Z) are loosely associated. So instead of using the benchmarks available in the BN literature (e.g., ASIA, ALARM, HAILFINDER etc.), we designed a toy problem that exhibits weak dependencies and interactions involving several variables (see Figure 1). The proposed hybrid algorithm was applied to data samples generated from this toy

BN. The boldface edges represent the weak dependencies. D and Y influence Z jointly but the marginal dependencies are weak (i.e. , $\chi^2(DZ) < \chi_{dep}$ and $\chi^2(DY) < \chi_{dep}$). Given the probability tables in figure 1, the covariance between A and C may be expressed as $cov(A, C) = cov(B, C) = a(1-a)[\alpha(1-2a) + \lambda a - \gamma(1-a)]$. In our experiments, we select the following parameters : $a = 0.1, \alpha = 1/80, \lambda = 0.8, \delta = 0.1$ in order to reduce the probabilistic dependency between A and C and between B and C. Consider the following example : if A is "alcohol", B and "smoke" and C "lung disease", we may very well suppose that each factor has marginally little influence on C (with probability 0.1) but A and B are *conjunctly* responsible for the disease (with probability 0.8).The support threshold is fixed to 10^{-3}.
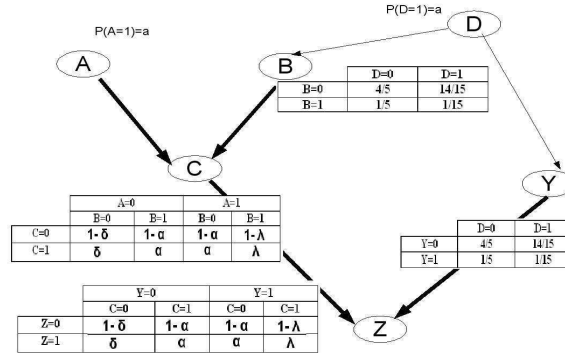

Figure 1. Our toy network with two interactions $A, B \Rightarrow C$ and $C, Y \Rightarrow Z$ shown in bold.

6.1. *Experiment 1*

For illustration, consider a single run or our algorithm on the graph depicted in Figure 1. The goal is to find the structure of the initial graph from data samples generated from this model. 7000 random examples were synthesized using a *Monte Carlo* technique called *probabilistic logic sampling* method. The $\chi^2\_support$ yields the following minimal correlated sets : $Sig_2 = \{DY(261.67), BD(203.53), BY(18.20)\}$. $Sig_3 = \{ABC(597.26), CYZ(335.13), ACD(20.14), CDZ(7.71)\}$. All edges in Sig_2 are added to the graph except BY ; The latter is not added to the graph because $\{D\}$ is shown to d-separate B and Y ($\chi^2(B, Y|D) = 2.39$). Figure 2 represents the graph obtained after of the first stage. All the CB methods based on statistical tests end up with the graph at stage 1 because all pairwise associations are too weak to be detected by statistical means.

The second stage operates on higher levels in the lattice. The sub-DAG $A \rightarrow C \leftarrow B$ is the one that maximize the BDeu. It is added to the graph. $\{CYZ\}$ is not correlated equivalent to $\{A, B, C\}$, so the best scoring sub-DAG $C \rightarrow Z \leftarrow Y$ is
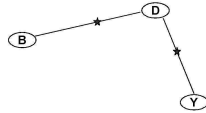
Figure 2. Skeleton obtained after Phase 1. All the CB learning methods assuming faithfulness end up here.
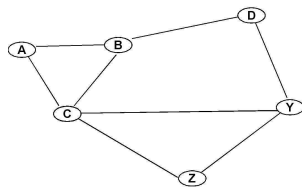

Figure 3. Skeleton obtained at Phase 2 before scoring the selected itemsets $\{A, B, C\}$ and $\{C, Y, Z\}$. These itemsets are fully connected here for sake of clarity.
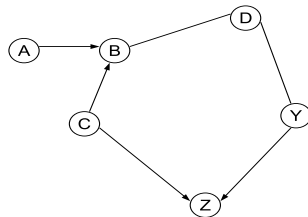

Figure 4. Final DAG.

added to $\mathcal{G}$. For the two last correlated of the level three, correlated equivalents are identified. For ACD, $(B, D) \in \mathcal{G}$, then ABC it his equivalent correlated set. At same for CDZ, $(Y, D) \in \mathcal{G}$ and ABC it his equivalent correlated set. These last sets are simply ignored from analysis.

6.2. *Experiment 2*

In the second experiments we illustrate the performance on this method against PMMS and BNPC (Brown *et al.*, 2005; Cheng *et al.*, 2002), two powerful CB algorithms and greedy scoring approach GS in Table 2 by running each algorithm 10 times for databases containing $\{5000, 10000, 15000\}$. The number of extra and missing edges is shown by symbols $(f+, f-)$. As observed, the limited number of false negatives $(f-)$ becomes at the expense of some more extra edges as shown in Tables

1 and 2 for different risk levels. Overall, the hybrid method exhibits false rate significantly lower compared to PMMS, BNPC and GS.

Size	5000							
Algo	PMMS		BNPC		GS		χ^2_sup	
Edges	f+	f-	f+	f-	f+	f-	f+	f-
Max	2	4	3	4	1	4	3	3
Min	0	2	0	2	0	3	1	0
Aver.	0.7	3.3	1.1	3.3	0.1	3.9	1.5	1.6
Size	10000							
Algo	PMMS		BNPC		GS		χ^2_sup	
Edges	f+	f-	f+	f-	f+	f-	f+	f-
Max	2	4	2	4	0	4	3	3
Min	0	2	0	2	0	4	0	0
Aver.	0.4	3.3	0.5	3.3	0	4	1.5	1.7
Size	15000							
Algo	PMMS		BNPC		GS		χ_2_sup	
Edge	f+	f-	f+	f-	f+	f-	f+	f-
Max	1	4	2	4	0	4	2	3
Min	0	1	0	2	0	4	0	0
Aver.	0.2	2.7	0.8	3	0	4	1	1.5

Tableau 1. Experimental results with risk level $\alpha = 0.05$

Size	5000							
Algo	*PMMS*		*BNPC*		*GS*		χ^2_sup	
Edges	f+	f-	f+	f-	f+	f-	f+	f-
Max	2	4	1	4	1	4	2	4
Min	0	2	0	2	0	3	0	1
Aver.	0.7	3.3	0.2	3.6	0.1	3.9	1	1.9
Size	10000							
Algo	*PMMS*		*BNPC*		*GS*		χ^2_sup	
Edges	f+	f-	f+	f-	f+	f-	f+	f-
Max	2	4	1	4	0	4	3	3
Min	0	2	0	2	0	4	0	0
Aver.	0.4	3.3	0.5	3.3	0	4	1.4	1.8
Size	15000							
Algo	*PMMS*		*BNPC*		*GS*		χ_2_sup	
Edge	f+	f-	f+	f-	f+	f-	f+	f-
Max	1	4	2	4	0	4	3	2
Min	0	1	0	2	0	4	0	0
Aver.	0.2	2.7	0.6	3	0	4	1.1	1.2

Tableau 2. Experimental results with risk level $\alpha = 0.01$

7. Conclusion and discussion

In this paper, we proposed a new method for learning structure of bayesien networks based on correlated itemset mining techniques, constraint-based and scoring methods. The key idea in this method is to identify and to represent conjunctions of factors implied in dependence relations when the pairwise dependencies are too weak. Therefore, the method is able to handle some unfaithful distributions at the price of higher temporal complexity due to a search in the lattice of all variable subsets. Despite promising results obtained on a simple toy problem, many difficulties remain in this work : the way correlated sets should be represented in the graph, the increasing number of extra edges as more correlated itemsets are selected, the unreliability of the χ^2 measure for large sets of variables.

8. Bibliographie

Brown L. E., Tsamardinos I., Aliferis C. F., « A Comparison of Novel and State-of-the-Art Polynomial Bayesian Network Learning Algorithms. », *In the Proceedings of the Twentieth National Conference on Artificial Intelligence AAAI*, p. 739-745, 2005.

Cheng J., Greiner R., Kelly J., Bell D. A., Liu W., « Learning Bayesian networks from data : An information-theory based approach. », *Artif. Intell.*, vol. 137, n° 1-2, p. 43-90, 2002.

Dash D., Druzdzel M. J., « A Hybrid Anytime Algorithm for the Construction of Causal Models From Sparse Data », *UAI*, p. 142-149, 1999.

Dash D., Druzdzel M. J., « Robust Independence Testing for Constraint-Based Learning of Causal Structure. », *UAI*, p. 167-174, 2003.

Geiger D., Verma T., Pearl J., « d-Separation : From Theorems to Algorithms. », *in* M. Henrion, R. D. Shachter, L. N. Kanal, J. F. Lemmer (eds), *UAI*, North-Holland, p. 139-148, 1989.

Leray P., Francois O., « BNT Structure Learning Package : Documentation and Experiments », 2004.

Murphy K. P., « The Bayes Net Toolbox for MATLAB », 1997.

Peña J. M., Nilsson R., Björkegren J., Tegnér J., « Towards scalable and data efficient learning of Markov boundaries », *Int. J. Approx. Reasoning*, vol. 45, n° 2, p. 211-232, 2007.

Silverstein C., Brin S., Motwani R., Ullman J. D., « Beyond Market Baskets : Generalizing Association Rules to Correlations », *VLDB*, p. 594-605, 1998.

Tsamardinos I., Brown L. E., Aliferis C. F., « The Max-Min Hill-Climbing Bayesian Network Structure Learning Algorithm. », *Machine Learning*, vol. 65, n° 1, p. 31-78, 2006.