

# Iterative Bayesian Network Implementation by Using Annotated Association Rules

Clément Fauré<sup>1,2</sup>, Sylvie Delprat<sup>1</sup>, Jean-François Boulicaut<sup>2</sup>, and Alain Mille<sup>3</sup>

<sup>1</sup> EADS CCR, Learning Systems Department, Centreda 1, F-31700 Blagnac  
{clement.faure, sylvie.delprat}@eads.net

<sup>2</sup> LIRIS UMR 5205, INSA Lyon, Bâtiment Blaise Pascal, F-69621 Villeurbanne

<sup>3</sup> LIRIS UMR 5205, Université Lyon 1, Nautibus, F-69622 Villeurbanne  
{amille, jboulica}@liris.cnrs.fr

**Abstract.** This paper concerns the iterative implementation of a knowledge model in a data mining context. Our approach relies on coupling a Bayesian network design with an association rule discovery technique. First, discovered association rule relevancy is enhanced by exploiting the expert knowledge encoded within a Bayesian network, i.e., avoiding to provide trivial rules w.r.t. known dependencies. Moreover, the Bayesian network can be updated thanks to an expert-driven annotation process on computed association rules. Our approach is experimentally validated on the Asia benchmark dataset.

## 1 Introduction

One major goal of the knowledge discovery from databases (KDD) community is to support the discovery of valuable patterns within the data. Considering 0/1 data analysis, the association rule mining technique is quite popular and we assume the reader is familiar with it [1]. It has been studied extensively both from the computational point of view and the objective interestingness perspective (i.e., using measures like frequency and confidence). Indeed, many algorithms have been designed for computing frequent and valid association rules. When the computation is tractable, it gives a huge number of rules which include many irrelevant ones: this is known to be a bottleneck for association rule based KDD processes. Application-independent redundancy has been addressed seriously by means of the closed sets and related approaches (see, e.g., [2]). Our approach to this problem is to use a well-specified subset of frequent and valid association rules called the frequent  $\delta$ -strong rules, i.e., frequent rules with at most  $\delta$  exceptions and a minimal left-hand-side property thanks to its  $\delta$ -freeness [3,4]. A second issue concerns application-dependant redundancy. Apart from rather simple template-based strategies, few authors have been considering how to remove rules which do not provide valuable information given an explicitly encoded model for available knowledge [5]. In this paper, we assume that novelty is indeed a key property to enhance subjective interestingness. In [6], the authors use a Bayesian Network (BN) to filter truly interesting frequent sets, i.e., sets

whose frequencies are somehow surprising given the expected distribution captured by the BN. In [7], we have extended this approach to support relevant association rule mining when we assume that (a) a BN captures expert knowledge about domain dependencies, and (b) we compute only our sub-collections of non redundant frequent and valid association rules (i.e., the  $\delta$ -strong rules). The intuition is that encoded dependencies can help us to filter out the patterns that reflect these dependencies. Doing so, we support the presentation of more interesting rules. In this paper, we address the obvious dynamics of knowledge discovery processes by considering that the knowledge model has to be iteratively updated. Our idea is that the initial model can be updated thanks to an expert-driven annotation of the extracted rules. Doing so, our methodology iteratively improves both the model for expert domain knowledge and the relevancy of the extracted patterns. A preliminary experimental validation on the well-known “Asia” dataset is given.

## 2 Modelling and Using Expert Knowledge

Our proposal relies on five steps and the paper concerns Steps 2 to 4:

1. Modelling an initial BN which specifies *a priori* expert domain knowledge.
2. Computing concise collections of frequent association rules with high confidence (i.e., the maximal number of exception is small w.r.t. the used frequency threshold).
3. Supporting rule post-processing (i.e., filtering) by using the knowledge model.
4. Supporting expert-driven annotation of the most interesting rules.
5. Updating the BN structure and parameters given the collected annotations.

### 2.1 Association Rule Post Processing Using a Bayesian Network

Modelling and exploiting knowledge to support the discovery of relevant association rules have been already studied. For instance, [8] considers the exploitation of expert knowledge elicited by expert rules. This has been formalized later into a belief system [9]. However, this kind of approach has a major limitation. Indeed, a rule is said to be interesting if it differs from the rules according to what is currently defined in the belief system, but not by looking at what could be inferred from these rules. Jaroszewicz et al. [6,10] have tackled this issue by modelling Bayesian networks for which “inference” is nicely integrated within the model. They describe the use of a BN to compute the added-value of frequent itemsets [1]. For each frequent itemset, the difference between its frequency value within the data and its expected frequency value inferred from the BN is computed. The more interesting patterns are the ones with the higher absolute difference value between these two measures. These itemsets can also be submitted to the expert to suggest updates of the structure and the parameters of the BN.

We have a similar approach but we exploit further the complementarities between Bayesian networks and association rules, namely dependency links between variables (directed edges of the graph, association relationship expressed

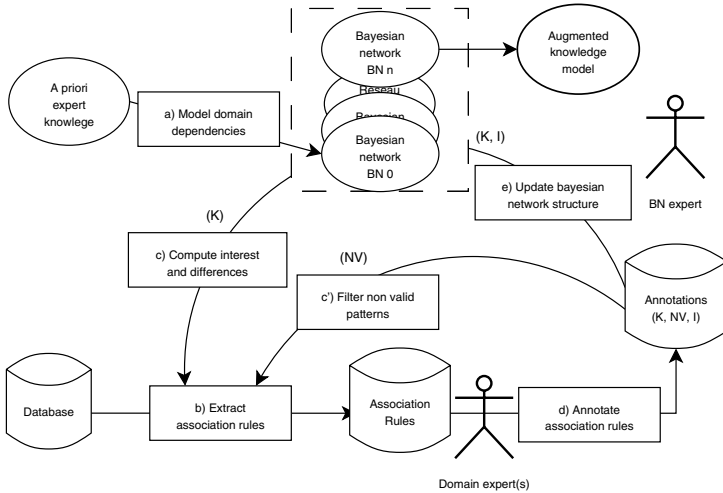


Fig. 1. A detailed view on our data mining process

by a rule) and frequencies for specific events (conditional probabilities defined in a BN, frequency of an association rule). We address “separately” these two relationships. First, we define an interestingness measure on association rule w.r.t. a BN. Then we propose an algorithm for computing relations of independence in the association rules w.r.t. the structure of the BN.

Due to the lack of space, we do not provide the needed definitions for the classical concepts related to 0/1 datasets and association rules on one hand (see, e.g., [1,4]) and Bayesian networks on the other hand (see, e.g., [11]).

**Interestingness Measure of an Association Rule Given a Bayesian Network.** Let us assume that  $DB$  is a Boolean database (i.e. a database where each record is a set of Boolean values), and  $H = \{A_1, \dots, A_n\}$  is the set of  $n$  Boolean attributes. An itemset  $I$  is just a subset from  $H$ . An association rule  $R$  is a pattern  $X \Rightarrow Y$ , where  $X$  and  $Y$  are itemsets such that  $Y \neq \emptyset$  and  $X \cap Y = \emptyset$ . The frequency of an itemset  $I$  in  $DB$  is the number of records from  $DB$  where the conjunction of all attributes from  $I$  is true. Most of the objective interestingness of association rules are based on the frequency of their set components. For instance, the confidence of  $X \Rightarrow Y$  is the frequency of  $X \cup Y$  divided by the frequency of  $X$ , i.e., the conditional probability to observe a true value for the attributes in  $Y$  when the attributes from  $X$  are true. A Bayesian network  $BN$  is a directed acyclic graph (DAG) defined by a set of nodes corresponding to the attributes in  $H$  and by  $E \subset H \times H$  its set of edges. Each node is associated with a conditional probability distribution  $P_{A_i | \Pi_{A_i}}$ , where  $\Pi_{A_i} = \{A_j | (V_{A_j}, V_{A_i}) \in E\}$  are the parents of node  $A_i$ . One of the most important properties of Bayesian networks is that they uniquely define the joint probability distribution over  $H$ . Thus, given a database  $DB$  and a Bayesian network  $BN$ , it is possible to compute the expected confidence of an association rule  $R = X \Rightarrow Y$  By extending

[6], we have defined in [7] a metric *Int* to evaluate the interest of a given association rule w.r.t. the encoded knowledge in a Bayesian network. Basically, it measures the difference between the confidence of the rule estimated on the data and the one inferred from the Bayesian network. Discussing this further is out of the scope of this paper.

**Computation of the Structural Differences Between an Association Rule and a Bayesian Network.** So far, we know how to compute collections of association rules, we have a formalism to express some expert knowledge, and finally a metric using such an encoded knowledge when ranking association rule interestingness. What is missing is a way to exploit the information of conditional independence implicitly captured by the network *BN*. Therefore, our goal is to highlight which parts of an association rule really contribute —according to *BN*— to the observation of the whole rule, and which parts are not.

Let us first introduce the D-separation property which has been formally defined by J. Pearl [11]. D-separation is a property of two sets of nodes *X* and *Y* w.r.t. another set of nodes *Z*. Informally, *X* and *Y* are said to be D-separated by *Z* if no information can flow between them when *Z* is observed.

We want to apply this notion —which is a pure graphical property— on association rules w.r.t. the *BN* structure. For any association rule  $R = X \Rightarrow Y$ , we will compute the D-separation test  $\langle X_i | X \setminus X_i | Y_j \rangle$ , where  $X_i \in X$  and  $Y_j \in Y$ . We end up with a matrix that sums up the results of all the D-separations tests. If an item of the rule ( $X_i$  or  $Y_i$ ) has a “true” value for all its D-separation tests, then it will be highlighted as being in the *D-separated part* of the rule. It means that thanks to the rule, an informative association has been found in the data which is not modelled in the current *BN* structure.

## 2.2 Post-processing and Annotation of Association Rules

Let us assume that *BN* reflects most of the domain dependencies. This network might have been defined either from scratch by an expert or through a mixed approach involving expert but also machine learning. Notice however that the initial *BN* does not have to be “complete”. For instance, it can capture only obvious dependencies, including known taxonomies over the attributes. As we go through the KDD processes, it can be updated to capture more and more domain knowledge, thus supporting the presentation of more and more valuable association rules. At each iteration, the expert might annotate the rules by labeling which parts represent what kind of information. This annotation can be used to improve the knowledge model *BN*.

Once our  $\delta$ -strong rules have been extracted (using the solver described in [4]), we compute their interest (as defined in [7]) as well as their “topological” differences w.r.t. the current instance for *BN*. These measures are used to filter uninteresting rules (interest compared with a user-defined threshold  $\epsilon$ ). It divides the rules in two classes. A first class contains the rules that do not provide further information w.r.t. *BN* (interest below  $\epsilon$ ). The expert who is inspecting the rules can decide to ignore them. The second class represents the rules that we call  $\epsilon$ -interesting. They express that some dependencies observed on the data are not

described properly by  $BN$ . The goal is to remove the rules that are  $\epsilon$ -interesting but either (a) already known by the domain expert or (b) containing non valid patterns. The idea is to refine the knowledge model by integrating step by step dependencies that were not identified at previous iterations. Understanding what information is contained by a given association rule is however a difficult task. This is why we want to highlight rule-like subpatterns of an association rule that represent a notion of D-separation between items on the left-hand side (LHS) and items on the right-hand side (RHS) given  $BN$  structure and the observation of all the LHS items. We can further divide the association among the  $\epsilon$ -interesting rules in three different types:

- K** The rule contains a pattern *already known* by the expert but that is not modelled in the current BN. It means that the structure and the parameters of the network have to be updated to integrate the causality related to this pattern. Doing so, such a pattern will not be presented as  $\epsilon$ -interesting in the next iteration.
- NV** The rule contains a pattern which appears to be *not valid* given the expert knowledge. This might be due to statistical coincidences (false positive).
- I** The rule holds a pattern that is potentially *interesting*. It has been “surprising” for the expert, and a deeper analysis has confirmed its relevancy.

In a real world data mining process, the number of association rules that fit in the categories (K) and (NV) can be huge. Moreover, a relation of association may contain mixed kinds of patterns, which might lead to tedious analysis tasks. We propose to ask the domain expert for annotations on the most interesting extracted association rules. He/she has to perform annotation following a precise method<sup>1</sup>, that will enables her/him to:

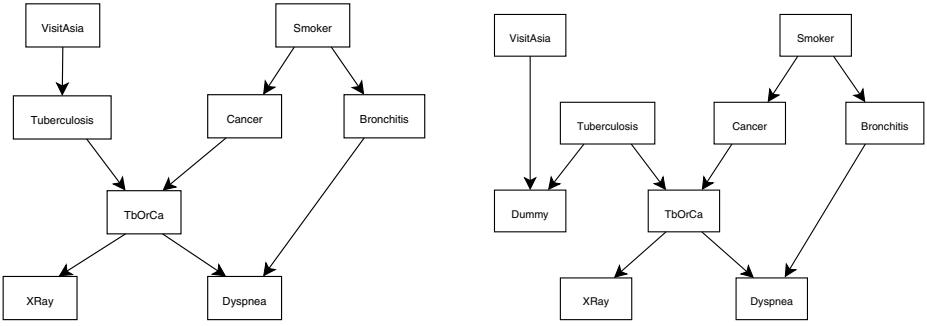
- specify whether an association rule contains one or more known patterns (K), non valid patterns (NV) or a potentially interesting one (I).
- define without ambiguity the “shape” of these patterns through the definition of a list of patterns which can only have one item in the right-hand side of the rule.
- be generic concerning the description of the detected patterns (providing only the name of the attribute or an attribute-value pair).
- define, when needed, a conjunction of attributes or items in the left-hand side of the pattern.
- associate a verbal-probability to patterns labeled as “already known” by following the idea of probability-ladder presented in [12].

These annotations can then be exploited to update the structure and the parameters of  $BN$ .

### 3 Experimental Validation

Let us consider an initial  $BN$  that already captures nicely a particular domain knowledge for the Asia dataset (i.e., a well-known benchmark within the

<sup>1</sup> This method is not detailed here due to space limitations.



**Fig. 2.** Original Asia network (left) and modified one (right)

Bayesian network community). From this network, we produced a dataset of 10,000 records. As we look for association rules, we focus on the *presence* of events. The initial *BN* structure is then modified so that the “VisitAsia” node is no longer directly connected to the “Tuberculosis” node. Our goal is to apply our methodology and see whether we can recover the right “Asia” network structure. Both networks are given in Figure 2.

Let us now follow the method described in Figure 1 step by step.

(a) The modified “Asia” BN serves as a basis for our experiments.

(b) From the generated dataset, we extract a concise collection of association rules (minimum absolute support value of 100, i.e., 0.01% of the database and maximal number of exceptions  $\delta = 10$ , i.e., a guarantee that the minimal confidence is 0.9). A total of 16 association rules are extracted immediately.

**Table 1.** Association rules extracted from the Asia dataset. Underlined items do not belong to the *D-separated* part of the rule.

Association rule	Interest	D-separated part
<u>Tuberculosis</u> $\Rightarrow$ <u>XRay</u> <u>Dyspnea</u> <u>TbOrCa</u>	0,04	
<u>VisitAsia</u> $\Rightarrow$ <u>XRay</u> <u>Dyspnea</u>	0,46	<u>VisitAsia</u> $\Rightarrow$ <u>Xray</u> <u>Dyspnea</u>
<u>Smoking</u> <u>Dyspnea</u> <u>Bronchitis</u> <u>TbOrCa</u> $\Rightarrow$ <u>XRay</u>	0,03	<u>Smoking</u> <u>Dyspnea</u> <u>Bronchitis</u> $\Rightarrow$
<u>Dyspnea</u> <u>Bronchitis</u> <u>TbOrCa</u> $\Rightarrow$ <u>XRay</u>	0,02	<u>Dyspnea</u> <u>Bronchitis</u> $\Rightarrow$
<u>Smoking</u> <u>Bronchitis</u> <u>TbOrCa</u> $\Rightarrow$ <u>XRay</u>	0,02	<u>Smoking</u> <u>Bronchitis</u> $\Rightarrow$
<u>Bronchitis</u> <u>TbOrCa</u> $\Rightarrow$ <u>XRay</u>	0,02	<u>Bronchitis</u> $\Rightarrow$
<u>Smoking</u> <u>Dyspnea</u> <u>TbOrCa</u> $\Rightarrow$ <u>XRay</u>	0,02	<u>Smoking</u> <u>Dyspnea</u> $\Rightarrow$
<u>Smoking</u> <u>TbOrCa</u> $\Rightarrow$ <u>XRay</u>	0,02	<u>Smoking</u> $\Rightarrow$
<u>Dyspnea</u> <u>TbOrCa</u> $\Rightarrow$ <u>XRay</u>	0,02	<u>Dyspnea</u> $\Rightarrow$
<u>TbOrCa</u> $\Rightarrow$ <u>XRay</u>	0,02	
<u>Smoking</u> <u>Dyspnea</u> <u>Cancer</u> $\Rightarrow$ <u>XRay</u> <u>TbOrCa</u>	0,02	
<u>Dyspnea</u> <u>Cancer</u> $\Rightarrow$ <u>XRay</u> <u>TbOrCa</u>	0,02	
<u>Dyspnea</u> <u>Bronchitis</u> <u>Cancer</u> $\Rightarrow$ <u>Smoking</u> <u>XRay</u> <u>TbOrCa</u>	0,00	
<u>Smoking</u> <u>Cancer</u> $\Rightarrow$ <u>XRay</u> <u>TbOrCa</u>	0,02	<u>Smoking</u> $\Rightarrow$
<u>Cancer</u> $\Rightarrow$ <u>XRay</u> <u>TbOrCa</u>	0,02	
<u>Bronchitis</u> <u>Cancer</u> $\Rightarrow$ <u>Smoking</u> <u>XRay</u> <u>TbOrCa</u>	0,01	

(c) Interest measure and D-separations are computed on these rules w.r.t. the modified “Asia” *BN*. (c’) Filtering out non valid patterns is optional and depends on the identification of such patterns. Obtained results are shown in Table 1. By looking at these results, let us recall that association rule mining captures only patterns over true values, i.e., presence of particular events. For example, the third rule should be read as “when we observe that a person is a smoker, a presence of dyspnea and bronchitis diagnosis and special node TbOrCa is activated, then it is often associated with abnormal x-ray result”. The underlined part of the rules denotes what we call the *core dependencies* of the rule w.r.t. *BN* structure. The last column shows a pattern of the association rule which contains missing information within the current network. Looking at these results, only one association rule has a relatively high interest value. This association rule states that “when we observe that a person has visited Asia then it is associated with abnormal x-ray results and a presence of dyspnea”. Clearly, this rule brings an information which is not modelled as a dependence in our modified *BN* structure while it was represented in the original one. It is thus possible to find rules that exhibit a difference between the available knowledge model and the data. We can however wonder whether such discovered associations are truly interesting. Furthermore, if this is the case, what are the modifications to be made to the model to reflect these observations in the data? This is of course where an expert judgment is crucially needed.

(d) An expert can now perform annotations. For our running example, assume that he/she has to put down that the rule which contains the “VisitAsia” relation belongs to the interesting category.

(e) Finally, this annotation is forwarded to the expert who is in charge of *BN* revision. By looking at the interesting pattern, it leads to a structural modification that provides the initial *BN* structure. We consider that the association rule actually found is sufficient for an expert to suggest the “right” revision. Notice that if we compute the rules on the same dataset but using the initial Asia network, we observe that the “VisitAsia” association rule no longer holds a D-separated pattern.

On the same datasets, we also applied the approach when the rules are extracted with APRIORI [1] (i.e., computing all the frequent and valid rules instead of the non redundant sub-collection of our  $\delta$ -strong rules). Doing so, 115 association rules were generated. Among all these rules, we found three different variants of the mentioned relation between “VisitAsia” and “abnormal x-ray”, including “Dyspnea” as well. Clearly, working on such redundant collections is harder: the expert will have to go through all the rules to find out patterns describing an association that involves “VisitAsia”.

## 4 Conclusion

Looking for relevant patterns in 0/1 data, we have been considering application-dependant redundancy. Our approach concerns association rule filtering when expert knowledge about attribute dependencies is encoded within a Bayesian

network and when a rule appears to be expected given this model. This paper has focused on the possible revision of such a knowledge model by using discoveries derived from the inspection and the annotation of selected association rules. The idea is that such a KDD process somehow converges towards actionable patterns: discovering new and valid statements in the data suggest refinement on the knowledge model which better captures important dependencies and thus enables to iterate on a more focused pattern discovery phase.

## References

1. Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., Verkamo, A.I.: Fast discovery of association rules. *Advances in Knowledge Discovery and Data Mining*, AAAI Press (1996) 307–328.
2. Bastide, Y., Pasquier, N., Taouil, R., Stumme, G., Lakhal, L.: Mining minimal non-redundant association rules using frequent closed itemsets. In: *Proceedings CL 2000*. Volume 1861 of LNCS., London, UK, Springer-Verlag (2000) 972–986.
3. Boulicaut, J.F., Bykowski, A., Rigotti, C.: Approximation of frequency queries by means of free-sets. In: *Proceedings PKDD 2000*. Volume 1910 of LNCS., Lyon, F, Springer-Verlag (2000) 75–85.
4. Boulicaut, J.F., Bykowski, A., Rigotti, C.: Free-sets: a condensed representation of boolean data for the approximation of frequency queries. *Data Mining and Knowledge Discovery* **7**(1) (2003) 5–22.
5. Liu, B., Hsu, W., Mun, L.F., Lee, H.: Finding interesting patterns using user expectations. *IEEE Transactions on Knowledge and Data Engineering* **11**(6) (1999) 817–832.
6. Jaroszewicz, S., Simovici, D.A.: Interestingness of frequent itemsets using bayesian networks as background knowledge. *Proceedings ACM SIGKDD 2004*, New York, USA, ACM Press (2004) 178–186.
7. Fauré, C., Delprat, S., Mille, A., Boulicaut, J.F.: Utilisation des réseaux bayésiens dans le cadre de l'extraction de règles d'association. *Proceedings EGC 2006*, Lille, F, Cepadues (2006) 569–580 In French.
8. Padmanabhan, B., Tuzhilin, A.: A belief-driven method for discovering unexpected patterns. *Proceedings KDD 1998*, New York, USA, AAAI Press (1998) 94–100.
9. Padmanabhan, B., Tuzhilin, A.: Small is beautiful: discovering the minimal set of unexpected patterns. *Proceedings ACM SIGKDD 2000*, Boston, USA, ACM Press (2000) 54–63.
10. Jaroszewicz, S., Scheffer, T.: Fast discovery of unexpected patterns in data, relative to a bayesian network. *Proceedings ACM SIGKDD 2005*, Chicago, USA, ACM Press (2005) 118–127.
11. Pearl, J.: *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann (1988).
12. Druzdzel, M.J., Diez, F.: *Criteria for combining knowledge from different sources in probabilistic networks* (2000).