

Constraining and Summarizing Association Rules in Medical Data

Carlos Ordonez
Teradata, NCR
San Diego, CA 92127, USA

Norberto Ezquerra
Georgia Institute of Technology
Atlanta, GA 30332, USA

Cesar A. Santana
Emory University
Atlanta, GA 30322, USA *

Abstract

Association rules are a data mining technique used to discover frequent patterns in a data set. In this work association rules are used in the medical domain, where data sets are generally high dimensional and small. The chief disadvantage about mining association rules in a high dimensional data set is the huge number of patterns that are discovered, most of which are irrelevant or redundant. Several constraints are proposed for filtering purposes, since our aim is to discover only significant association rules and accelerate the search process. A greedy algorithm is introduced to compute rule covers in order to summarize rules having the same consequent. The significance of association rules is evaluated using three metrics: support, confidence and lift. Experiments focus on discovering association rules on a real data set to predict absence or existence of heart disease. Constraints are shown to significantly reduce the number of discovered rules and improve running time. Rule covers summarize a large number of rules by producing a succinct set of rules with high quality metrics.

1 Introduction

Association rules [1, 2] are a data mining technique used to discover frequent patterns in a large data set [25]. In particular, association rules are suitable for discovering predictive rules relating subsets of attributes from a data set with numeric and categorical attributes [38], since interesting patterns may exist in subsets of the attributes, but not on all attributes taken together. Our goal is to find significant association rules with high predictive power. We have identified three main issues. First, in the case of a high dimensional data set, a significant fraction of association rules is not useful because they are irrelevant or redundant. Second, most relevant rules with high quality metrics appear only at low frequency levels. Third, the number of discovered rules becomes extremely large at low frequencies. Therefore, we focus on defining constraints to reduce the number of rules and summarizing rules to solve these issues.

We introduce an application of association rules in the medical domain. This work comes from a long-term project to apply data mining, image processing and artificial intelligence techniques, to enhance automated heart disease diagnosis [14, 13, 12, 6, 39, 38]. Association rules are proposed and preliminary results are reported in [12]. Neural networks are used to predict reversibility (heart response) images, based on stress and myocardial (heart muscle) thickening images [6]. In [39] we explore the idea of constraining association rules in binary data for the first time and report preliminary findings from a data mining perspective. The knowledge base rules from an expert system used for heart disease diagnosis, are validated using association rules [13]. Some theoretical results about constrained association rules are presented in [38].

1.1 Contributions and Article Outline

Association rules are constrained to reduce the number of discovered patterns and they are summarized to get a concise set of rules. The significance of association rules is evaluated using support, confidence and

*© Springer-Verlag, 2006. This is the author's version of the work. The official version of this article was published in Knowledge and Information Systems (KAIS Journal). 9(3):259-283, 2006. DOI: 10.1007/s10115-005-0226-5

lift. The proposed constraints include maximum association size, an attribute grouping constraint and an antecedent/consequent rule filtering constraint. Association rules are summarized using rule covers with a greedy algorithm. We show that the problem of mining association rules in a high dimensional data set with numeric and categorical attributes is challenging, due to the large number of patterns rather than data set size. On the other hand a large number of patterns impacts the running time of the data mining algorithm, making result interpretation difficult.

We have organized this article in the following manner. Section 2 introduces the definition of association rules, as well as the metrics, support, confidence and lift. Section 3 describes the main problems encountered using association rules, introduces filtering and search constraints and explains how to summarize association rules with covers. Section 4 presents three sets of experiments with a real data set. The first set of experiments studies the impact of each constraint on the number of rules and performance, varying program parameters. The second set of experiments focuses on finding predictive rules with high confidence and high lift. The third set of experiments presents covers summarizing a large number of discovered rules. Section 5 discusses related research work. Section 6 presents conclusions and directions for future work.

2 Definitions

2.1 Association Rules

Let $D = \{T_1, T_2, \dots, T_n\}$ be a data set of n transactions and let \mathcal{I} be a set of items, $\mathcal{I} = \{i_1, i_2, \dots, i_m\}$, where $T_i \subseteq \mathcal{I}$. In general, a subset of \mathcal{I} is called an itemset. Let X and Y be two itemsets s.t. $X \subset \mathcal{I}$, $Y \subset \mathcal{I}$ and $X \cap Y = \emptyset$. An association rule is an implication of the form $X \Rightarrow Y$, where X is called the antecedent and Y is called the consequent of the rule. In the context of this work, a data record consists of numeric and categorical attribute values. A data record is transformed into a binary vector and then such vector is transformed into a transaction (itemset), where the presence of item i_j indicates the j th coordinate is 1; when item i_j is not present then the j th binary coordinate is zero. The itemset representation is used for efficiency purposes because transactions represent vectors with most coordinates equal to zero.

2.2 Support, Confidence and Lift

Let $P(X)$ be the probability of appearance of X in D and let $P(Y|X)$ be the conditional probability of appearance of Y , given X appears. For an itemset X , $support(X)$ is defined as the fraction of transactions $T_i \in D$ such that $X \subseteq T_i$. Thus, $P(X)$ can be estimated as $P(X) = support(X)$. The support of a rule $X \Rightarrow Y$ is defined as $support(X \Rightarrow Y) = P(X \cup Y)$. An association rule $X \Rightarrow Y$ has a measure of reliability called confidence, defined as $confidence(X \Rightarrow Y) = support(X \cup Y) / support(X)$. Confidence can be used to estimate $P(Y|X)$: $P(Y|X) = P(X \cup Y) / P(X) = confidence(X \Rightarrow Y)$. Viewing support and confidence as probabilistic metrics is explained in more detail in [24, 28]. The standard problem of mining association rules [1] is to find all rules whose metrics are equal to or greater than some specified minimum support threshold ψ and minimum confidence threshold α . A k -itemset with support above the minimum threshold is called frequent. We use a third significance metric for association rules called *lift* [4] (also called interest [7]), defined as $lift(X \Rightarrow Y) = P(Y|X) / P(Y) = confidence(X \Rightarrow Y) / support(Y)$. Lift measures how much the presence of Y depends on the presence or absence of X , or viceversa. Lift values greater than 1 indicate that the presence of Y depends on the presence of X .

3 Constraining and Summarizing Association Rules

We briefly analyze the problem of mapping information from categorical and numerical attribute values to binary format. We introduce constraints to find predictive association rules with preferred attribute combi-

nations and to accelerate the rule discovery process. We explain how to summarize association rules having the same consequent using rule covers.

3.1 Mapping Attributes

A data set with numeric and categorical attributes must be transformed into transactions in order to apply association rules. Numeric attributes are binned into intervals and each interval is mapped to an item. Categorical attributes are transformed into items, mapping each categorical value to one item. If an attribute has negation there are additional items corresponding to each negated categorical value or corresponding to each negated numeric range. In short, each transaction is a set of items and each item corresponds to the presence or absence of one categorical value or one numeric interval.

We explain how data records are transformed into transactions containing items. Let A_1, A_2, \dots, A_p be all the attributes to be analyzed. Let $R = \{r_1, r_2, \dots, r_n\}$ be the input data set with n records whose values are taken from $dom(A_1) \times dom(A_2) \times \dots \times dom(A_p)$, where $dom(A_i)$ is either a categorical or numerical domain. The data set size is n and its dimensionality is p . We transform R into a data set $D = \{T_1, T_2, \dots, T_n\}$ of n transactions containing subsets of m items, by means of the transformation process to be explained below. Items are identified by consecutive integers starting with one, i.e. $1, 2, \dots, m$.

The following transformation algorithm transforms numerical and categorical values into binary data, where each binary dimension corresponds to one item, as defined in Section 2. Each binary dimension corresponds to one numerical range or one categorical value. The transformation process is divided into two phases. In the first phase a transformation table M is constructed based on user's requirements. In the second phase, attribute values in each tuple r_j are mapped to items based on M . Each tuple r_j becomes a transaction T_j , suitable for association rule mining. A transaction only includes the subscripts (items) of binary dimensions equal to 1 (for instance, in $T_1 = \{1, 5, 10\}$, it is assumed items 2 and 8 are not present in T_1). For a categorical attribute A_i each categorical value is mapped to one item. If negation is desired for a categorical attribute A_i with more than two distinct values, then each negated value is mapped to an item. So if a categorical attribute A_i has h_i distinct values ($h_i \geq 3$) the transformation process produces h_i items without negation or $2h_i$ items with negation. The domain expert specifies N_i cutoff points corresponding to numerical attribute A_i producing $N_i + 1$ intervals. Then each interval is mapped to one item producing $N_i + 1$ items. If negation is desired the number of items is doubled, producing in total $2(N_i + 1)$ items. Negation on numerical attributes significantly increases the potential number of associations. Therefore, it must be used on an individual attribute basis after careful consideration. Once table M has been constructed the second phase is straightforward. There are two important points about M . First, constraints are specified on attributes, but constraints are actually used on items. Second, it is not possible to apply the standard association rule [2] directly on a data set with alphanumeric records without a transformation into a binary data set. In general, R can be a high dimensional data set and the transformation process produces an even higher dimensional data set D .

3.2 Constraining and Summarizing Association Rules

Several problems come up when trying to discover association rules in a high dimensional data set. For each problem we propose a solution that is in the form of a constraint, using a metric or summarizing patterns. Our primary goal is to discover significant rules for predictive purposes, but also making the data mining process faster.

Rule complexity. Association rules that involve many items are hard to generalize, are hard to interpret and can potentially generate a large number of redundant rules. Therefore, there should be a default threshold for association size. Most approaches are exhaustive in the sense that they find *all* rules above the user-specified thresholds, but in general that produces a huge amount of rules as it will be shown in the experimental section. The largest size of discovered associations is a practical bottleneck for algorithm performance. Another reason to limit size is that if there are two rules $X_1 \Rightarrow Y$ and $X_2 \Rightarrow Y$ s.t. $X_1 \subset X_2$, the

first rule is probably more interesting because it is simpler and it is likely to have higher support. However, the second rule may be more interesting if it summarizes several rules similar to the first rule. This aspect will be explained below in more detail.

Irrelevant rules for predictive purposes. By the definition of association rules, an item can appear in the antecedent or in the consequent of a rule. But from a practical standpoint we would like items corresponding to predicted attributes to appear in the consequent and conditions, causes, independent attributes to appear in the antecedent. Given the rule $X \Rightarrow Y$, the association $X \cup Y$ must be a frequent itemset because $support(X \cup Y)$ defines rule support. But there are many rules derived from $X \cup Y$ that are irrelevant. In other words, minimum support helps pruning the search space for significant associations, but minimum confidence is not enough to filter out irrelevant rules, because there may be rules having high confidence containing forbidden items in the antecedent or in the consequent. Therefore, items, derived from attributes, need to be constrained to appear either on the antecedent or the consequent of each rule.

Association rules having irrelevant or redundant groups of items. This is the case where certain itemsets have such high support or represent well-known attribute-value combinations, that they do not really say anything new. Consider items i_j and $i_{j'}$. If the association $X_1 = \{i_j, i_{j'}\}$ is not interesting because it is already known that i_j and $i_{j'}$ commonly appear together, then any other association X_2 s.t. $X_1 \subset X_2$ will not be interesting. The association X_2 may be less interesting than X_1 because in general it will have lower support and it will be more complex since it has more items. That is, simpler association rules with higher support are preferred. Therefore, many of the items (if not all) can be carefully grouped by the domain expert to discard uninteresting associations. If no grouping is done then item i_j is always relevant, no matter which other items $i_{j'}$ appear together with it. Small groups of uninteresting item combinations can be identified either automatically by running a straight association rules algorithm or by domain knowledge. This constraint has the purpose of reducing the number of redundant association rules which grow in a combinatorial manner as the minimum support threshold is lowered.

Decreasing minimum support. It has been shown that minimum support is the performance bottleneck for association rules [37]. It is desirable to run the algorithm once with a very low support, thus avoiding repeated runs with decreasing supports. However, rules with very low support may be interesting for exploratory purposes, but they may not be reliable. In particular, rules involving only one transaction should be filtered out.

Weak implication. For some rules with high confidence, the appearance of the consequent may not be necessarily linked to the appearance of the antecedent. Recall the definition of $lift(X \Rightarrow Y)$ from Section 2. If $lift(X \Rightarrow Y) \approx 1$ then $P(X \cup Y) \approx P(X)P(Y)$. Then this is evidence that X and Y are independent. On the other hand, if $lift(X \Rightarrow Y) \geq \lambda > 1$, then $P(X \cup Y) \geq \lambda P(X)P(Y)$. This is saying that the actual probability of X and Y appearing together is much higher than the product of their respective probabilities if they were independent. That is, X and Y are dependent on each other. Notice that $lift(X \Rightarrow Y)$ is symmetric (i.e. $lift(X \Rightarrow Y) = lift(Y \Rightarrow X)$), but $confidence(X \Rightarrow Y)$ is not. Lift can be used to identify rules where the dependence between X and Y is weak or strong and to further reduce the number of rules. Recall from Section 2 we are using a minimum confidence threshold α to filter rules. So if a rule $X \Rightarrow Y$ has $confidence(X \Rightarrow Y) \geq \alpha$ then $lift(X \Rightarrow Y) \geq \alpha/support(Y) \geq \alpha$ because $support(Y) \leq 1$. Therefore, if $\alpha \approx 1$ then $lift(X \Rightarrow Y) \approx 1$. For instance, if we are using a minimum confidence threshold $\alpha = 0.8$ then all rules will have lift metrics starting at 0.8, which is fairly close to 1. Such rules will not be considered interesting since they indicate X and Y are independent. By a similar reasoning, lift numbers close to zero or below α correspond to low confidence rules. These facts support the idea of introducing a minimum lift threshold above 1 to filter out rules with independent itemsets as well as low confidence rules. Lift is a well-known measure used for predictive purposes, but it is not sufficient by itself to evaluate the implication strength of a rule, because it is symmetric and it does not provide a probability. In this work, lift is used in combination with confidence to evaluate the significance of association rules. One of the goals for discovered association rules is to enrich the knowledge base of an expert system [14]. Confidence is a measure that is directly used as the confidence factor (CF) of a

prediction rule used by the expert system.

Large number of discovered rules. If the number of discovered rules is large, even incorporating filtering constraints, interpretation becomes difficult. It is desirable to get a few representative rules so that many rules can be derived from them. We propose to generate a cover of association rules with the same consequent (predicted disease) in order to generate a rule summary that can be interpreted by the domain expert in less time. Given two rules with the same consequent, $r_1 : A_1 \Rightarrow B$ and $r_2 : A_2 \Rightarrow B$, such that $A_1 \subseteq A_2$, we say that r_2 covers r_1 . Then we can include r_2 and omit r_1 . When there are several rules covered by r_2 this will produce a concise summary. We will apply this approach to each set of rules with the same consequent, but different antecedent itemsets.

Search and filtering constraints

Based on the difficulties outlined above we introduce the following improvements to the standard association rule search algorithm [2], which has two phases. Phase 1 finds all associations having minimum support, proceeding bottom-up, generating frequent 1-itemsets, 2-itemsets and so on, until there are no frequent itemsets. Phase 2 produces all rules having minimum support, minimum confidence and minimum lift. Two of our constraints work on Phase 1 and one works on Phase 2.

The first constraint is κ , the user-specified maximum association size. Associations are generated up to size κ , eliminating the search for associations of size $\kappa + 1$, $\kappa + 2$ and so on. This constraint is simple, but essential to get simple rules and reduce output size. Notice that each discovered rule will also have at most κ items.

In the remainder of this section we explain how attributes are constrained. Input attributes are extended with two constraints. Let $\mathcal{I} = \{i_1, i_2, \dots, i_m\}$ be the set of items to be mined, obtained by the transformation process from the attributes $\mathcal{A} = \{A_1, \dots, A_p\}$. Let *attribute* : $\mathcal{I} \rightarrow \mathcal{A}$ be a function that returns the attribute corresponding to one item.

Let $\mathcal{C} = \{c_1, c_2, \dots, c_p\}$ be a set of antecedent/consequent constraints for each attribute A_j as motivated above. Observe that constraints are specified on attributes and not on items. Each constraint c_j can have one out of three values: 1 if item A_j can only appear in the antecedent of a rule, 2 if it can only appear in the consequent and 0 if it can appear in either place. We define the function antecedent/consequent $ac : \mathcal{A} \rightarrow \mathcal{C}$ as $ac(A_j) = c_j$ to make reference to one such constraint. Let X be a k -itemset; X is said to be antecedent-interesting if for all $i_j \in X$ then $ac(attribute(i_j)) \neq 2$; X is said to be consequent-interesting if for all $i_j \in X$ then $ac(attribute(i_j)) \neq 1$.

Let $\mathcal{G} = \{g_1, g_2, \dots, g_p\}$ be a set of p group constraints, one for each attribute A_j ; g_j is a positive integer if A_j is constrained to belong to some group or 0 if A_j is not group-constrained at all. We define the function $g : \mathcal{A} \rightarrow \mathcal{G}$ as $g(A_j) = g_j$. Since each attribute belongs to one group then the group numbers induce a partition on the attributes. Note that if $g_j > 0$ then there should be two or more attributes with the same group value of g_j . Otherwise that would be equivalent to having $g_j = 0$. The itemset X is said to be group-interesting if for each item pair $\{i_j, i_{j'}\}$ it is true $g(attribute(i_j)) \neq g(attribute(i_{j'}))$.

3.3 Algorithm to Compute Rule Covers

The algorithm to compute a cover for a set of rules with the same consequent is as follows. Rule generation in Phase 2 is modified to generate sets of rules having the same consequent. We then apply the following greedy algorithm to quickly find a cover. For a set of rules $\mathcal{R} = \{r_1 : A_1 \Rightarrow B, r_2 : A_2 \Rightarrow B, \dots, r_d : A_d \Rightarrow B\}$ we compute how many rules are covered by r_i . If $A_i \supseteq A_j$ we say r_i covers r_j . Let $cover(r_i)$ be the number of rules covered by r_i . Let the rule cover \mathcal{R}' be initially empty. Then if rule r_i has maximum $cover(r_i)$ it is added to the rule cover and all its covered rules are marked and discarded. Ties are broken in favor of the rule with more items. The process is recursively repeated for the rest of uncovered rules. For completeness a rule can cover itself if it has not been covered yet. Then by the above algorithm, the rule cover will have at most the original number of rules and it is guaranteed that every rule will be

covered. This algorithm is fast, but possibly not optimal with respect to the number of rules selected, but for our purposes this is acceptable. The cover for a set of rules with the same consequent B represents a subset of rules, $\mathcal{R}' \subseteq \mathcal{R}$, $\mathcal{R}' = \{r'_1 : A'_1 \Rightarrow B, r'_2 : A'_2 \Rightarrow B, \dots, r'_c : A'_c \Rightarrow B\}$, s.t. $cover(r'_i) \geq cover(r'_j)$ for $i < j$ and for each $r_i \in \mathcal{R} \exists r'_j \in \mathcal{R}'$ s.t. $A_i \subseteq A'_j$.

The algorithm above is used to classify and summarize rules. This algorithm works in a post-processing fashion and not in Phase 1 or Phase 2 for several reasons. Each cover rule has minimum quality metrics, but we are interested in particular rules that have high confidence and a cover rule cannot provide such information. As explained above, the antecedent/consequent constraint cannot be pushed into Phase 1. Cover rules tend to be too specific, with very low support levels.

Our proposed rule covers are related to maximal frequent itemsets and closed itemsets. A maximal frequent itemset is an itemset in which adding an item will make it fall below the minimum support threshold [34, 23]. By the downward closure property, all proper subsets of a maximal frequent set are frequent and not maximal. Cover rules may correspond to maximal frequent itemsets when there is no threshold on rule size and their consequent is not a subset of the consequent of other cover rules. Cover rules do not correspond to maximal frequent itemsets when their consequent is a subset of the consequent of longer rules. Covered rules do not correspond to maximal frequent itemsets because the itemset from the covering rule is a superset. In general, itemsets from cover rules are a subset of maximal frequent itemsets. Therefore, mining all maximal frequent itemsets would produce a larger set of rules than required. Even further, computing support for their subsets would be required anyway to get all metrics and a filtering phase would be required to eliminate rules not complying with the group constraints and minimum interestingness metrics. An itemset is closed [41] if there is no superset with the same support; closed itemsets represent a more restricted pattern than maximal frequent itemsets. But closed itemsets are a more concise representation than all frequent itemsets, but less concise and more precise than maximal frequent itemsets. In general cover rules correspond to closed itemsets if there is no threshold on rule size; the exception happens when the cover rule itemset is a proper subset of the itemset from another cover rule and has the same support. Many covered rules may correspond to closed itemsets. Mining all closed itemsets instead of frequent itemsets would lead to a similar algorithm, but all rules derivable from subsets of closed itemsets need to be generated anyway.

3.4 Algorithm to Search and Summarize Constrained Association Rules

Based on the search constraints and rule filters introduced in Section 3.2, and rule covers introduced in Section 3.3, we assemble all pieces together into one algorithm that goes from converting input records to computing rule covers. The transformation process using the given cutoffs for numeric attributes and desired negated attributes, produces the input data set for Phase 1. Each data record becomes a transaction T_i ; refer to definitions from Section 2. After the data set is transformed, certain items are further filtered out depending on the prediction goal. Notice items can only be filtered after attributes are transformed because they depend on the numeric cutoffs and negation. This will be explained in more detail in Section 4. In Phase 1 we use the $g()$ constraint to avoid searching for irrelevant itemsets. Phase 1 generates all associations up to size κ . Phase 2 generates all rules filtering using the $ac()$ constraint. Phase 3 summarizes association rules using covers. The main input parameters are κ , as well as the support threshold ψ , the confidence threshold α and minimum lift λ . The algorithm major steps are:

- Preprocessing:
Transform n records with numeric and categorical attributes into a transaction data set to get $D = \{T_1, \dots, T_n\}$, based on specified cutoffs and negation for numeric attributes.
Filter items depending on predictive goal to generate 1-itemsets.
- Phase 1:
Search for frequent itemsets from size 1 up to size κ using ψ and the $g()$ constraint.
- Phase 2:
Generate rules in sets of rules having the same consequent, using the $ac()$ constraint having confidence at least α and lift above λ . Each set of rules is of the form: $\mathcal{R} = \{A_1 \Rightarrow B, A_2 \Rightarrow B, \dots, A_d \Rightarrow B\}$.

- Phase 3:
Summarize results computing a set of cover rules \mathcal{R}' for each set of rules \mathcal{R} having the same consequent. The output is a set of rules: $\mathcal{R}' = \{r'_1 : A'_1 \Rightarrow B, r'_2 : A'_2 \Rightarrow B, \dots, r'_c : A'_c \Rightarrow B\}$ such that $cover(r'_i) \geq cover(r'_j)$ for $i < j$ and $A'_i \in \{A_1, \dots, A_d\}$, where $cover(r'_i)$ is the number of covered rules by rule r'_i .

Preprocessing is semi-manual, requiring the user to specify cutoffs, negation, constraints and items to be discarded. Phases 1 and 2 are automatic requiring the user to specify the minimum support, confidence and lift thresholds. Phase 3 is automatic and has no parameters.

Some theoretical results describing properties of the proposed constraints can be found in [38]. We provide a summary of them. Group constraints produce an overall speedup in the frequent itemset generation because they reduce both number of associations in Phase 1 and the number of rules in Phase 2. As it has been noted in the literature, in general Phase 1 is the most time-demanding. The algorithm cannot take advantage of $ac(i)$ constraints in Phase 1, only in Phase 2. That is, an association involving items that appear either in the antecedent or the consequent must be generated anyway. This causes Phase 1 to be slow even though the number of generated rules is reduced. Maximum association size κ , significantly improves performance, given the large number of subsets in a long itemset. Assume given the minimum support there is a frequent k -itemset X s.t. $\kappa < k$. Then there are $2^k - \binom{k}{\kappa}2^\kappa$ pruned associations by the proposed algorithm. That is, the number of pruned associations is large when the X has many items. The size constraint eliminates the need of searching many subsets of X .

4 Experiments

Our experiments were run on a Sun computer running at 800 MHz with 256 MB of main memory and 100GB of disk space. The algorithm was implemented in the C++ language. Performance experiments were repeated five times and the average is reported.

4.1 Real Data Set Description

Our experiments were based on a real data set obtained from a hospital, shown in Table 1. The data set contained the profiles of several patients being treated for heart disease. There were 655 patients having 113 attributes, combining numeric, categorical and image data. Medical records were collected during the course of several years. Records were created when the patient was admitted to the hospital for treatment and were subsequently updated as the patient received medical treatment. We selected 25 medical attributes corresponding to all numeric and categorical attributes listed in Table 1. The remaining attributes that were not used in our experiments, involved image data for the patient during stress (exercise), rest (no exercise) and reversibility for the heart (referred to a condensed map of 9 regions from the heart of a normal person), and attributes with highly skewed distributions. For instance, race was not used because 82% of patients had race=W and 10% had missing race information. Only global cholesterol levels were available, with most patients having no measurements on LDL (bad) or HDL (good) cholesterol levels. Image attributes were already summarized into the 9 perfusion measurements that were used in the experiments, which in the context of this work were the basic and most important criteria to evaluate the patient's heart health. Therefore, in our experiments, $p = 25$ and $n = 655$. These attributes provide a complete picture about the patient based on perfusion measurements (obtained with a medical procedure to digitize the pattern of a colored substance given to the patient) for specific regions of the heart, known risk factors for heart disease (if available) and the degree of disease in four arteries (artery narrowing in percentage terms). For each attribute we give its usual abbreviation in the medical domain, its data type, what type of medical information (MI) it contains and a complete description. Attributes are classified into three types according to the information they contain; 'P' attributes correspond to perfusion measurements on specific regions of the heart; 'R' attributes correspond to risk factors and 'D' attributes correspond to heart disease (artery narrowing) measurements.

No	Name	Data Type	Medical Info	Description	Constraints		
					Neg	g	ac
1	<i>AGE</i>	N	R	Age of patient	N	0	1
2	<i>LM</i>	N	D	Left Main	Y	0	2
3	<i>LAD</i>	N	D	Left Anterior Desc.	Y	0	2
4	<i>LCX</i>	N	D	Left Circumflex	Y	0	2
5	<i>RCA</i>	N	D	Right Coronary	Y	0	2
6	<i>AL</i>	N	P	Antero-Lateral	N	1	1
7	<i>AS</i>	N	P	Antero-Septal	N	1	1
8	<i>SA</i>	N	P	Septo-Anterior	N	1	1
9	<i>SI</i>	N	P	Septo-Inferior	N	1	1
10	<i>IS</i>	N	P	Infero-Septal	N	1	1
11	<i>IL</i>	N	P	Infero-Lateral	N	1	1
12	<i>LI</i>	N	P	Latero-Inferior	N	1	1
13	<i>LA</i>	N	P	Latero-Anterior	N	1	1
14	<i>AP</i>	N	P	Apical	N	1	1
15	<i>SEX</i>	C	R	Gender	N	0	1
16	<i>HTA</i>	C	R	Hyper-tension Y/N	N	0	1
17	<i>DIAB</i>	C	R	Diabetes Y/N	N	0	1
18	<i>HYPLPD</i>	C	R	Hyperlipoidemia Y/N	N	0	1
19	<i>FHCAD</i>	C	R	Family hist. of disease	N	0	1
20	<i>SMOKE</i>	C	R	Patient smokes Y/N	N	0	1
21	<i>CLAUDI</i>	C	R	Claudication Y/N	N	0	1
22	<i>PANGIO</i>	C	R	Previous angina Y/N	N	0	1
23	<i>PSTROKE</i>	C	R	Prior stroke Y/N	N	0	1
24	<i>PCARSUR</i>	C	R	Prior carotid surgery Y/N	N	0	1
25	<i>CHOL</i>	N	R	Cholesterol level	N	0	1

Table 1: Attributes of medical data set. Constraints: g (group), ac (antecedent/consequent).

The goal is to relate perfusion measurements and risk factors to disease measurements. The image data represents the local degree of blood distribution (perfusion) in the heart muscle (myocardium). Heart disease measurements include 4 numeric attributes that store the percentage of heart disease caused by a specific artery of the heart (*LM*, *LAD*, *LCX*, and *RCA*) and 9 numeric attributes (*AL*, *IS*, *SA*, *AP*, *AS*, *SI*, *LI*, *IL*, *LA*) that store a perfusion measurement, which is a value in the range $[-1, 1]$. Each of the four artery measurements has a value between 0% and 100% that represents the percentage of vessel narrowing. A perfusion measurement closer to 1 indicates a more severe defect, whereas a value closer to -1 indicates absence of a perfusion defect. An abbreviation that is used in several column names is CAD, which stands for Coronary Artery Disease. Risk factors are self-explanatory. The right part of Table 1 contains the default values of constraints for association rule mining. Constraints are explained in Section 4.2.

4.2 Default Parameter Settings

In this section we explain the default settings for program parameters, that were based on the domain expert opinion. To understand the meaning of attribute names given below please refer to Table 1.

Transformation parameters

In order to set the transformation parameters we must discuss attributes corresponding to heart vessels. The *LAD*, *RCA*, *LCX* and *LM* numbers represent the percentage of vessel narrowing (or blockage) compared to a healthy artery. Attributes *LAD*, *LCX* and *RCA* were partitioned by cutoff points at 50% and 70%. In the cardiology field a 70% value or higher indicates significant coronary disease and a 50% value indicates borderline disease. A value lower than 50% means the patient is healthy. The most common cutoff value used by the cardiology community to distinguish healthy from sick patients is 50%. The *LM* artery is treated different because it poses higher risk than the other three arteries. Attribute *LM* was partitioned at 30% and 50%. The reason behind these numbers is both the *LAD* and the *LCX* arteries branch from the *LM* artery

and then a defect in *LM* is more likely to cause a larger diseased heart region. That is, narrowing (blockage) in the *LM* artery is likely to produce more disease than blockages on the other arteries. That is why its cutoff values are set 20% lower than the other vessels. The 9 heart regions (*AL, IL, IS, AS, SI, SA, LI, LA, AP*) were partitioned into 2 ranges at a cutoff point of 0.2, meaning a perfusion measurement greater or equal than 0.2 indicated a severe defect. *CHOL* was partitioned with cutoff points 200 (warning) and 250 (high). These values correspond to known medical settings. Missing values were assigned one item during the transformation process, but they were filtered out before rule generation. This is important since we did not use any medical field that had missing information. Also, we did not substitute any missing value for some imputed value. In our data set missing information was indicated with a question mark ("?"). The question mark for each attribute was mapped to an item. Items in the mapped data set corresponding to missing information, were ignored during the association search process. That is, no associations containing items corresponding to missing information were generated. All discovered associations referred to records without missing information in the corresponding attributes. Finally, only the four artery measurements had negation to find rules referring to healthy patients and sick patients. The rest of attributes did not require negation. Since most risk factors were binary and perfusion measurements were divided into two ranges negation on them was eliminated. Negation was not considered useful for age and cholesterol level.

Search and filtering constraints

The maximum association size κ varied from 1 to 5 to study the impact of constraints. In the second set of experiments, to get simple rules, $\kappa = 4$. For the last set of experiments, computing rules covers, we increased κ to 10. A lower κ produces fewer and simpler rules. A higher κ produces many more rules that are also more complex, but which can be used to derive simpler rules using covers.

Association rule mining had the following thresholds for metrics. The support threshold was fixed at $\psi = 1\% \approx 7$. That is, rules involving 6 or fewer patients were discarded. This eliminated rules that were statistically unreliable or were probably particular to our data set. From a medical point of view, rules with high confidence are desirable, but unfortunately, they are infrequent. Based on previous experiments [39, 38] and the domain expert opinion, minimum confidence was $\alpha = 70\%$. From the medical point of view, rules with 90% or higher confidence are preferable, but they are infrequent and generally do not involve all risk factors. Rules in the 80%-90% range may be acceptable depending on which measurements, arteries and risk factors are involved. Rules in the 70%-80% range have borderline reliability and they indicate potential patterns to be discovered in subsets of patients. Rules with confidence lower than 70% are not medically reliable; in particular, rules with confidence 50% or lower are never considered interesting. For the first set of experiments we use a minimum lift threshold equal to zero to effectively use only support and confidence for filtering purposes. Then for the second set, we use a lift threshold slightly higher than 1 to filter out rules where X and Y are very likely to be independent. All rules with $lift(X \Rightarrow Y) \leq 1$ turned out to have $lift(X \Rightarrow Y) \geq 0.9$. This is explained by $\alpha = 0.7$ as explained in Section 3.2. Finally, we use a high lift threshold equal to 2 to get rules where there is a strong dependence between X and Y , to explore with higher κ and to reduce the number of rules. In short, for the set of experiments measuring number of patterns and time minimum lift was 0, to get simple predictive rules minimum lift was 1.2, and to get rule covers minimum lift was 2.0.

The group constraints and antecedent/consequent constraints were set as follows. This set of constraints is by no means definitive or optimal, but it represents what our experience has shown to be most useful. Refer to Table 1 to understand the attribute meanings. The 4 main coronary arteries *LM, LAD, LCX*, and *RCA* were constrained to appear in the consequent of the rule; that is, $ac(i) = 2$. All the other attributes were constrained to appear in the antecedent, i.e. $ac(i) = 1$. In other words, R (risk factors) and P (perfusion measurements) appear in the antecedent, whereas D (disease) medical attributes, appear in the consequent of a rule. It is evident that from a medical point view, we are not interested in determining the likelihood of presenting a risk factor based on artery disease. The nine regions of the heart

Settings	Number of associations	Number of rules	Time in minutes
Negation=N size=4			
Group=Y Ac=Y	15,165	468	3
Group=N Ac=Y	26,263	665	10
Group=Y Ac=N	15,165	14,525	59
Group=N Ac=N	26,263	168,950	210
Negation=Y size=4			
Group=Y Ac=Y	36,982	2,987	13
Group=N Ac=Y	60,042	4,610	41
Group=Y Ac=N	36,982	164,056	463
Group=N Ac=N	60,042	N/A	N/A
Negation=Y Group=Y Ac=Y			
size=2	692	17	0
size=3	6,635	371	1
size=4	36,982	2,987	13
size=5	124,506	60,688	472

Table 2: Impact of constraints on number of associations, number of rules and execution time.

(*AL, IS, SA, AP, AS, SI, LI, IL, LA*) were constrained to be in the same group (group 1). The group settings for the rest of attributes varied depending on the type of rules being mined. Recall that combinations of items in the same group are not considered interesting and are eliminated from further analysis. The nine heart regions were constrained to be on the same group since we were interested in finding their interaction with risk factors and not among them. Group constraints were used for certain risk factors to decrease the number of rules predicting absence of heart disease. Group constraints on risk factors were not necessary for rules predicting existence of heart disease. The default constraints are summarized in Table 1.

4.3 Impact of Constraints on the Number of Patterns

We carried out experiments to evaluate the impact of constraints on the number of patterns found and on running time, at the lowest possible support level. Minimum support was set at $\psi = 0.2\%$ in order to capture all potential associations involving at least two patients, and the maximum rule size was $\kappa = 4$. Minimum confidence was $\alpha = 0.7$ and $\lambda = 0$ (for lift). This support level will show the output size of all potentially interesting association rules. However, later in the article we will focus on prediction rules involving at least 1% of patients. The reason behind such higher setting is that those rules will be more reliable, from a medical point of view. The experiments show the change in the number of rules and elapsed time when negation, antecedent/consequent constraints and group constraints are turned on and off.

Table 2 summarizes the impact of constraints on number of patterns and running time. Notice that our data set is small, but the problem of finding significant association rules is difficult given the large number of patterns. The first part of Table 2, where there is no negation, shows that removing the $ac()$ constraint produces an order of magnitude increase in the number of rules when $g()$ constraints are used. But when group constraints are not used, removing the $ac()$ constraint increases the number of rules by two orders of magnitude. Times are affected accordingly. The last run took more than three hours to finish. The second part of Table 2 introduces negation for the arteries. The rest of numeric attributes are not negated. The number of rules increases an order of magnitude on average. The last run, where $ac()$ and $g()$ constraints are not used, could not be completed in less than one day. Only Phase 1 (association generation) was finished. The number of rules was probably in the order of one million. We emphasize again that rules pruned by constraints are not medically interesting. The third part of Table 2 shows the combinatorial explosion on the number of rules as association size increases even with all constraints turned on. Some association rules

involving 5 attributes may provide valuable information, but it would require significant effort to analyze 60k rules. In short, these experiments prove that constraining is essential to get a manageable number of rules and to have acceptable running times.

4.4 Mining Prediction Rules

The goal of the experiments was to relate perfusion measurements and risk factors to vessel disease (also known as stenosis), to validate and improve actual diagnosis rules used by an expert system [19]. Some rules were expected, confirming valid medical knowledge, and some rules were surprising, having the potential to enrich the expert system knowledge base. Discovered rules were classified into two groups; first, those that express that if there is a low perfusion measurement or no risk factor, then there is no heart disease; second, those that express that if there exists a risk factor or a high perfusion measurement, then there is heart disease. The maximum association size κ was 4. That is, we mined rules involving at most 4 attributes; that way we get simpler rules with higher support.

There are two main measurements to quantify the quality of medical findings: sensitivity and specificity. Sensitivity refers to the probability of correctly identifying sick patients. Specificity is the probability of correctly identifying healthy individuals. These measures rely on a standard that can be trusted by a medical doctor. That is, a measurement that tells with very high accuracy if the person is sick or not. Getting such ideal measurement may involve doing invasive (dangerous) medical procedures on the patient. In the context of this paper the medical standard was catheterization. Unfortunately, in a few cases a clinician reading was taken, but in general it was not available. The general guideline we used was to increase sensitivity and decrease specificity. That is, we preferred to increase false positives at the risk of including people with borderline heart disease. We used lift, in addition to confidence, as a measure to balance sensitivity and specificity. So our experiments show a bias towards finding rules predicting disease with higher lift and rules predicting no disease with lower lift.

Support, confidence and lift were used as the main filtering parameters. Minimum lift in this case was $\lambda = 1.2$. Support was used to discard low probability patterns. Confidence was used to look for reliable prediction rules. Lift was used to compare similar rules with the same consequent and to select rules with higher predictive value. Confidence, combined with lift, was used to evaluate the significance of each rule. Rules with confidence $\geq 90\%$ were considered significant. Rules with lift ≥ 2 were considered significant. Rules with two or more items in the consequent were considered significant. Rules with high support, only risk factors, low lift or borderline confidence were considered interesting, but not significant. Rules with borderline cases were discarded, but are potentially interesting. Rules with artery figures in wide intervals (more than 70% of the attribute range) were not considered interesting like rules having a measurement in the 30-100 range for *LM*.

Rules predicting absence of heart disease

The default program parameter settings are described in Section 4.2. Refer to Table 1 to understand the meaning of abbreviations for attribute names. Perfusion measurements for the 9 regions were in the same group (group 1). Rules relating no risk factors (equal to "n") with healthy arteries were considered interesting. Risk factors *HTA*, *DIAB*, *HYPLPD*, *FHCAD*, *CLAUDI* were in the same group (group 2). Risk factors describing previous conditions for disease (*PANGIO*, *PSTROKE*, *PCARSUR*) were in the same group (group 3). The rest of the risk factor attributes did not have any group constraints. Since we were after rules relating negative risk factors and low perfusion measurements to healthy arteries, several items were filtered out to reduce the number of patterns. The discarded items involved arteries with values in the higher (not healthy) ranges (e.g. [30, 100], [50, 100], [70, 100]), perfusion measurements in [0.2, 1] (no perfusion defect), and risk factors equal to "y" for the patient (person presenting risk factor). Minimum support was $\psi = 1\%$ and minimum confidence was $\alpha = 70\%$.

```

Confidence = 1:
IF 0 <= AGE < 40.0 - 1.0 <= AL < 0.2 PCARSUR = n THEN 0 <= LAD < 50, s=0.01 c=1.00 l=2.1
IF 0 <= AGE < 40.0 - 1.0 <= AS < 0.2 PCARSUR = n THEN 0 <= LAD < 50, s=0.01 c=1.00 l=2.1
IF 40.0 <= AGE < 60.0 SEX = F 0 <= CHOL < 200 THEN 0 <= LCX < 50, s=0.02 c=1.00 l=1.6
IF SEX = F HTA = n 0 <= CHOL < 200 THEN 0 <= RCA < 50, s=0.02 c=1.00 l=1.8
Two items in the consequent:
IF 0 <= AGE < 40.0 - 1.0 <= AL < 0.2 THEN 0 <= LM < 30 0 <= LAD < 50, s=0.02 c=0.89 l=1.9
IF SEX = F 0 <= CHOL < 200 THEN 0 <= LAD < 50 0 <= RCA < 50, s=0.02 c=0.73 l=2.1
IF SEX = F 0 <= CHOL < 200 THEN 0 <= LCX < 50 0 <= RCA < 50, s=0.02 c=0.73 l=1.8
Confidence >= 0.9:
IF 40.0 <= AGE < 60.0 - 1.0 <= LI < 0.2 0 <= CHOL < 200 THEN 0 <= LCX < 50, s=0.03 c=0.90 l=1.5
IF 40.0 <= AGE < 60.0 - 1.0 <= IL < 0.2 0 <= CHOL < 200 THEN 0 <= LCX < 50, s=0.03 c=0.92 l=1.5
IF 40.0 <= AGE < 60.0 - 1.0 <= IL < 0.2 SMOKE = n THEN 0 <= LCX < 50, s=0.01 c=0.90 l=1.5
IF 40.0 <= AGE < 60.0 SEX = F DIAB = n THEN 0 <= LCX < 50], s=0.08 c=0.92 l=1.5
IF HTA = n SMOKE = n 0 <= CHOL < 200 THEN 0 <= LCX < 50, s=0.02 c=0.92 l=1.5
Only risk factors:
IF 0 <= AGE < 40.0 THEN 0 <= LAD < 50, s=0.03 c=0.82 l=1.7
IF 0 <= AGE < 40.0 DIAB = n THEN 0 <= LAD < 50, s=0.03 c=0.82 l=1.7
IF 40.0 <= AGE < 60.0 SEX = F DIAB = n THEN 0 <= LAD < 50, s=0.07 c=0.72 l=1.5
IF 40.0 <= AGE < 60.0 SMOKE = n THEN 0 <= LCX < 50, s=0.11 c=0.75 l=1.2
IF 40.0 <= AGE < 60.0 SMOKE = n THEN 0 <= RCA < 50, s=0.11 c=0.76 l=1.3
Support >= 0.2:
IF - 1.0 <= IL < 0.2 DIAB = n THEN 0 <= LCX < 50, s=0.41 c=0.72 l=1.2
IF - 1.0 <= LA < 0.2 THEN 0 <= LCX < 50, s=0.39 c=0.72 l=1.2
IF SEX = F THEN 0 <= LCX < 50, s=0.23 c=0.73 l=1.2
IF 40.0 <= AGE < 60.0 - 1.0 <= IL < 0.2 THEN 0 <= RCA < 50, s=0.21 c=0.73 l=1.3

```

Figure 1: Association rules predicting disease=N.

The program produced a total of 9,595 associations and 771 rules in slightly over one minute. Although most of these rules provided valuable knowledge, we only describe some of the most surprising ones. Figure 1 shows rules predicting no heart disease in groups. These rules have the potential to improve the expert system. The group with confidence=1 shows some of the few rules that had 100% confidence. It was surprising that some rules referred to young patients, but not older patients. The rules involving *LAD* had high lift with localized perfusion defects. The rules with *LM* had low lift, confirming that other risk factors may imply a healthy artery. The group with two items shows the only rules predicting absence of disease in two arteries. They include combinations of all the arteries and have high lift. These rules highlight low cholesterol level, female gender and young patients. It turned out all of them refer to the same patients. The 90% confidence group shows fairly reliable rules. Unfortunately, their lift is not high. The group with only risk factors shows rules that do not involve any perfusion measurements. These rules highlight the importance of smoking habits, diabetes, low cholesterol, gender and age in having no heart disease. The last group describes rules with high support. Most of them involve the *LCX* artery, the *IL* region and some risk factors. These rules had low lift stressing the importance of many other factors to have healthy arteries. Summarizing, these experiments show *LCX* is more likely to be healthy given absence of risk factors and low perfusion measurements. Lower perfusion measurements appeared in heart regions *IL* and *LI*. Some risk factors have less importance because they appear less frequently in the rules. But age, sex, diabetes and cholesterol level appear frequently, a clear indication of their importance.

Rules predicting existence of heart disease

The default program parameter settings are described in Section 4.2. Refer to Table 1 to understand the meaning of abbreviations for attribute names. The four arteries (*LAD*, *LCX*, *RCA*, *LM*) had negation. Rules relating presence of risk factors (equal to "y") with diseased arteries were considered interesting. There were no group constraints for any of the attributes, except for the 9 regions of the heart (group 1). This allowed finding rules combining any risk factors with any perfusion defects. Since we were after rules relating risk factors and high perfusion measurements indicating heart defect to diseased arteries, several unneeded items were filtered out to reduce the number of mined patterns. The filtered items involved arteries with values in the lower (healthy) ranges (e.g. [0, 30), [0, 50), [0, 70)), perfusion measurements in [-1, 0.2) (no perfusion defect), and risk factors equal to "n" for the patient (person not presenting risk factor). In a similar manner to predicting absence of disease, $\psi = 1\%$ and $\alpha = 70\%$.

The program produced a total of 10,218 associations and 552 rules in less than one minute. Most of these rules were considered important and about one third were medically significant. Most rules refer to patients with localized perfusion defects in specific heart regions and particular risk factors with the *LAD* and *RCA*

```

confidence = 1:
IF 0.2 <= SA < 1.0 HYPLPD = y PANGIO = y THEN 70 <= LAD < 100, s=0.01 c=1.0 l=3.2
IF 60 <= AGE < 100 0.2 <= SA < 1.0 FHCAD = y THEN not(0 <= LAD < 50), s=0.02 c=1.0 l=1.9
IF 0.2 <= IS < 1.0 CLAUDI = y PSTROKE = y THEN not(0 <= RCA < 50), s=0.02 c=1.0 l=2.3
IF 60 <= AGE < 100 0.2 <= IS < 1.0 250 <= CHOL < 500 THEN 70 <= RCA < 100, s=0.02 c=1.0 l=3.2
IF 0.2 <= IS < 1.0 SEX = F 250 <= CHOL < 500 THEN 70 <= RCA < 100, s=0.01 c=1.0 l=3.2
IF 0.2 <= IS < 1.0 HTA = y 250 <= CHOL < 500 THEN 70 <= RCA < 100, s=0.011 c=1.0 l=3.2
Two items in the consequent:
IF 0.2 <= AL < 1.1 PCARSUR = y THEN 70 <= LAD < 100 not(0 <= RCA < 50), s=0.01 c=0.70 l=3.9
IF 0.2 <= AS < 1.1 PCARSUR = y THEN 70 <= LAD < 100 not(0 <= RCA < 50), s=0.01 c=0.78 l=4.4
IF 0.2 <= AP < 1.1 PCARSUR = y THEN 70 <= LAD < 100 not(0 <= RCA < 50), s=0.01 c=0.8 l=4.5
IF 0.2 <= AP < 1.1 PCARSUR = y THEN not(0 <= LAD < 50) not(0 <= RCA < 50), s=0.01 c=0.8 l=2.8
confidence >= 0.9:
IF 0.2 <= SA < 1.1 PANGIO = y THEN 70 <= LAD < 100, s=0.023 c=0.938 l=3.0
IF 0.2 <= SA < 1.0 SEX = M PANGIO = y THEN 70 <= LAD < 100, s=0.02 c=0.92 l=2.9
IF 60 <= AGE < 100 0.2 <= IL < 1.1 250 <= CHOL < 500 THEN 70 <= RCA < 100, s=0.02 c=0.92 l=2.9
IF 0.2 <= IS < 1.0 SMOKE = y 250 <= CHOL < 500 THEN 70 <= RCA < 100, s=0.02 c=0.91 l=2.9
Only risk factors:
IF SEX = M PSTROKE = y 250 <= CHOL < 500 THEN not(0 <= LAD < 50), s=0.01 c=0.73 l=1.4
IF 40.0 <= AGE < 60.0 SEX = M 250 <= CHOL < 500 THEN not(0 <= LCX < 50), s=0.02 c=0.83 l=2.2
IF SMOKE = y PANGIO = y 250 <= CHOL < 500 THEN not(0 <= RCA < 50), s=0.01 c=0.80 l=1.9
Support >= 0.15:
IF 0.2 <= IL < 1.1 THEN not(0 <= LAD < 50), s=0.25 c=0.71 l=1.4
IF 0.2 <= AP < 1.1 THEN not(0 <= LAD < 50), s=0.24 c=0.78 l=1.5
IF 0.2 <= IL < 1.1 SEX = M THEN not(0 <= LAD < 50), s=0.19 c=0.72 l=1.4
IF 0.2 <= AP < 1.1 SEX = M THEN not(0 <= LAD < 50), s=0.18 c=0.75 l=1.5
IF 60 <= AGE < 100 0.2 <= AP < 1.1 THEN not(0 <= LAD < 50), s=0.18 c=0.87 l=1.7

```

Figure 2: Association rules predicting disease=Y.

arteries. It was surprising there were no rules involving *LM* and only 9 with *LCX*. Tomography or coronary catheterization are the most common ways to detect heart disease. Tomography corresponds to myocardial perfusion studies. Catheterization involves inserting a tube into the coronary artery and injecting a substance to measure which regions are not well irrigated. These rules characterize the patient with coronary disease. There are three basic elements for analysis: perfusion defect, risk factors and coronary stenosis.

Figure 2 shows groups of rules predicting heart disease. Hypertension, diabetes, previous cardiac surgery and male sex constitute higher risk factors. The 100% confidence group shows some of the only 22 rules with 100% confidence. They show a clear relationship of perfusion defects in the *IS*, *SA* regions, certain risk factors and both the *RCA* and *LAD* arteries. The rules having *RCA* have very high lift pointing to specific relationships between this artery and cholesterol level and the *IS* region. It was interesting the rule with *LAD* ≥ 70 also had very high lift, but referred to different risk factors and region *SA*. The group of rules with two items in the consequent, shows the only rules involving two arteries. They show a clear link between *LAD* and *RCA*. It is interesting these rules only involve a previous surgery as a risk factor. These four rules are surprising and extremely valuable. This is confirmed by the fact that two of these rules had the highest lift among all discovered rules (above 4). The 90% confidence group shows some outstanding rules out of the 35 rules that had confidence 90-99%. All of these rules have very high lift with a narrow range for *LAD* and *RCA*. These rules show that older patients of male gender, high cholesterol levels and localized perfusion measurements, are likely to have disease on the *LAD* and *RCA* arteries. The group involving only risk factors in the antecedent shows several risk factors and disease on three arteries. Unfortunately their support is relatively low, but they are valuable as they confirm medical knowledge. The rule with lift=2.2 confirms that gender and high cholesterol levels may lead to disease in the *LCX* artery. The group with support above 0.15 shows the rules with highest support. All of them involved the *LAD* artery and a variety of risk factors. Their lift was low-medium, confirming more factors are needed to have a more accurate prediction. There were no high-support rules involving *LCX*, *RCA* or *LM* arteries, confirming they have a lower probability of being diseased.

4.5 Summarizing Association Rules using Rule Covers

The following experiments focus on summarizing results with rule covers and have two goals. The first goal is to discover a few representative rules from which simpler ones, like the ones discussed in Section 4.4, can be derived. The second goal is to get a summarized representation of a large set of rules, that has a significantly lower number of rules, but which comprises all discovered rules. We applied the greedy algorithm introduced in Section 3, along with parameter settings that will generate a significantly higher number of

Prediction goal	disease=N	disease=Y
Time in minutes	39	6
Number of associations	88,520	37,759
Number of rules	4,280	721
Number of rules in cover	1,617	344
Number of distinct consequents	10	14

Table 3: Summary of rule covers.

```

IF -1.0 <= IS < 0.2 SEX = F SMOKE = n CLAUDI = n PSTROKE = n 0 <= CHOL < 200
THEN 0 <= LAD < 50, rulesCovered:4 s=0.01 c=1.00 l=2.1
IF -1.0 <= AP < 0.2 SEX = F SMOKE = n CLAUDI = n PSTROKE = n 0 <= CHOL < 200
THEN 0 <= LAD < 50, rulesCovered:6 s=0.01 c=1.00 l=2.9
IF -1.0 <= AP < 0.2 SEX = F SMOKE = n CLAUDI = n PSTROKE = n 0 <= CHOL < 200
THEN 0 <= LM < 30.0 0 <= LAD < 50, rulesCovered:6 s=0.01 c=1.00 l=2.2
IF 40 <= AGE < 60 -1.0 <= AP < 0.2 SEX = F HYPLPD = n PSTROKE = n 0 <= CHOL < 200
THEN 0 <= LAD < 50 0 <= LCX < 50, rulesCovered:20 s=0.01 c=1.00 l=2.8
IF 40 <= AGE < 60 -1.0 <= LL < 0.2 SEX = F HTA = n PANGIO = n 0 <= CHOL < 200
THEN 0 <= LCX < 50 0 <= RCA < 50, rulesCovered:10 s=0.01 c=1.00 l=2.4
IF 40 <= AGE < 60 -1.0 <= AP < 0.2 SEX = F HYPLPD = n PSTROKE = n 0 <= CHOL < 200
THEN 0 <= LAD < 50 0 <= RCA < 50, rulesCovered:24 s=0.01 c=1.00 l=2.9
IF 40 <= AGE < 60 -1.0 <= LA < 0.2 SEX = M CLAUDI = n PSTROKE = n 0 <= CHOL < 200
THEN 0 <= LM < 30 0 <= LAD < 50 0 <= LCX < 50, rulesCovered:8 s=0.01 c=0.82 l=2.4
IF 40 <= AGE < 60 -1.0 <= AP < 0.2 SEX = F HYPLPD = n PSTROKE = n 0 <= CHOL < 200
THEN 0 <= LM < 30 0 <= LAD < 50 0 <= RCA < 50, rulesCovered:24 s=0.01 c=1.00 l=3.0
IF 40 <= AGE < 60 -1.0 <= AP < 0.2 SEX = F DIAB = n PANGIO = n 0 <= CHOL < 200
THEN 0 <= LM < 30 0 <= LAD < 50 0 <= LCX < 50 0 <= RCA < 50, rulesCovered:14 s=0.01 c=1.0 l=3.8

```

Figure 3: Cover rules predicting disease=N.

patterns than previous experiments. Minimum support was 0.01 and minimum confidence was 0.7, like in previous experiments. Minimum lift was 2.0 in order to include only rules with high predictive value and to produce a manageable number of rules. This setting allowed us to increase the maximum association size κ because many rules with low lift were automatically filtered out. So the maximum association size κ was 10 in order to generate a large number of subrules to be covered. All attributes used the antecedent/consequent constraint (Section 4.2) and the group constraints (Section 4.4).

There were two separate runs to compute rule covers with the same parameters, but some items had to change in the pre-processing step from the algorithm from Section 3.4, depending on the prediction goal. The first run involved rules predicting absence of disease (disease=N). The second run was for rules predicting existence of disease (disease=Y). There is a rule cover computation for each different consequent. Results are summarized in Table 3. It is clear that increasing κ produces a big impact on running time compared to $\kappa = 4$, used in previous experiments. There are more distinct consequents for rules predicting existence of disease, but fewer rules, compared to their counterpart. The data set contains many more patterns predicting absence of disease. The reduction in output size is significant for the covers of rules predicting absence of disease, where the number of rules decreases by 62%. The decrease for the covers of rules predicting disease is 52%. It is outstanding that rule covers produce a reduction in size because of the constraints we are using. Reduction would be higher if constraints were not used since patterns would be more redundant. Thus, rules predicting existence of heart disease are more complex and based on their covers, have less redundancy. The table indicates that the number of rules is the main factor impacting the total running time; the number of associations plays a smaller role.

Figure 3 and Figure 4 show some selected cover rules, predicting absence of disease and predicting existence of disease, respectively. These rules have high lift, high confidence and a high number of rules covered. In general, we picked one of the first top rules in each cover. Notice that each cover rule also has minimum metrics above the thresholds specified in Section 4.2, since they are themselves rules produced by our algorithm.

We observed that rules predicting absence of disease had higher confidence on average than those predicting disease. In fact, there were many cover rules predicting absence of disease with confidence equal to 1. Most of them had lift between 2 and 3. We found rules predicting no disease for all arteries and their

```

IF 60 <= AGE < 100 0.2 <= AP < 1.1 SEX = F HTA = y SMOKE = y PSTROKE = y
THEN 70 <= LAD < 100, rulesCovered:17 s=0.01 c=1.00 l=3.2
IF 40 <= AGE < 60 0.2 <= LI < 1.1 SEX = M DTAB = y HYPLPD = y PSTROKE = y
THEN 70 <= LCX < 100, rulesCovered:3 s=0.01 c=0.88 l=3.3
IF 60 <= AGE < 100 0.2 <= IS < 1.0 SEX = M HYPLPD = y SMOKE = y PSTROKE = y
THEN 70 <= RCA < 100 rulesCovered:4 s=0.02 c=0.71 l=2.3
IF 60 <= AGE < 100 0.2 <= IL < 1.1 SEX = M HTA = y HYPLPD = y CLAUDI = y
THEN not(0 <= LAD < 50) 70 <= RCA < 100 rulesCovered:9 s=0.01 c=0.89 l=4.0
IF 60 <= AGE < 100 0.2 <= IL < 1.1 SEX = M HTA = y HYPLPD = y PSTROKE = y
THEN not(0 <= LAD < 50) not(0 <= LCX < 50), rulesCovered:2 s=0.02 c=0.71 l=2.8
IF 60 <= AGE < 100 0.2 <= IL < 1.1 SEX = M HTA = y HYPLPD = y SMOKE = y PSTROKE = y
THEN not(0 <= LAD < 50) not(0 <= RCA < 50) rulesCovered:8 s=0.01 c=0.75 l=2.6
IF 60 <= AGE < 100 0.2 <= LI < 1.1 SEX = M HTA = y CLAUDI = y
THEN not(0 <= LAD < 50) not(0 <= LCX < 50) 70 <= RCA < 100, rulesCovered:1 s=0.01 c=0.70 l=5.2
IF 60 <= AGE < 100 0.2 <= LA < 1.1 SEX = M CLAUDI = y PSTROKE = y
THEN not(0 <= LAD < 50) not(0 <= LCX < 50) not(0 <= RCA < 50), rulesCovered:1 s=0.01 c=0.7 l=4.3

```

Figure 4: Cover rules predicting disease=Y.

combinations with most of them having middle age, female gender and low cholesterol as factors for no disease. It was interesting there were 2,508 rules with 3 items in the consequent. The highest coverage (i.e. $c(r_i)$) rules were found in rules predicting no disease. These numbers confirm such rules have higher redundancy.

Rules predicting existence of disease had lower confidence as noted above. Most of them had lift between 2 and 3. Most rules with high coverage had confidence less than 1. The sets of cover rules with more items involved the *LAD* and the *RCA* artery. There were fewer cover rules with *LCX* in the consequent. There were no rules involving the *LM* artery. Most cover rules involved older age, high cholesterol, hypertension, diabetes, hyperlipidemia and being male. In comparison to rules predicting no disease, there were very few cover rules having 3 items in the consequent: only three such rules were discovered. Given their high lift metric they were considered valuable, even though they only covered themselves and had borderline confidence.

Our experiments show that rule covers tend to be more useful for rules predicting absence of disease because cover rules have high confidence and high lift, and also there is a higher reduction in output size. On the other hand, cover rules predicting existence of disease tend to have lower confidence figures. Therefore, some of the simpler (covered) rules predicting existence of disease, presented in Section 4.4, are preferred because they have both high confidence and high lift. In short, rule covers were more useful to provide a summary for association rules predicting absence of disease.

5 Related Work

The following articles are related to using data mining with medical data. Applying data mining to medical data creates unique challenges [47]. Some particular issues include fragmented data collection, stricter privacy concerns, rich attribute types (image, numeric, categorical, missing information), complex hierarchies behind attributes and an already rich and complex knowledge base. A well-known program to help heart disease diagnosis (called HDP) is described in [20, 36, 35]. This research work shows a computer program can improve differential diagnosis made by medical doctors for cardiovascular disease. The program is based on Bayesian networks combined with artificial intelligence inference mechanisms [35]. Association rules have been used to help infection detection and monitoring [8, 9], to understand what drugs are co-prescribed with antacids [11], to discover frequent patterns in gene data [5, 15], to understand interaction between proteins [40] and to detect common risk factors in pediatric diseases [18]. Association rules have also been extended with fuzzy sets and alternative metrics to improve their application [17]. The problem of significantly reducing the number of association rules by using constraints was not studied before in the medical domain. To our knowledge, there is no previous work on using association rules to find prediction rules for artery disease, relating risk factors and perfusion measurements.

Association rules were proposed in the seminal paper [1]. There is work on mining Quantitative Association Rules (QAR) [49], but they have not been shown to be superior to decision trees. The I/O complexity

of early association rule algorithms as well as some statistical support metrics are studied in [26]. Clustering association rules, rather than transactions, once they are mined, is analyzed in [33]. The output is a summary of association rules. Both [49] and [33] use different approaches to automatically bin numeric attributes. In our project it was preferred to use well-known medical cutoffs for binning numeric attributes, to improve result interpretation and validation. Mining association rules combining numeric and categorical attributes is studied in [46]; those rules are allowed to contain disjunctions, which are not useful for our purposes because they do not provide a conjunction of risk factors and measurements of sick patients; the optimized confidence approach is interesting, but it may not produce medically meaningful ranges for arteries. Incorporating constraints into the proposed optimized confidence problem is an issue for future research.

More closely related work to ours includes the following. Using rule templates to understand association rules in a post-processing phase is proposed in [29]. Two types of templates are introduced: inclusive and restrictive templates. Inclusive templates are similar to the *ac* constraint, but there is an important difference: the *ac* constraint is specified on original attributes and then it is converted into constraints on items (mapped numeric and categorical attributes); this mapping of constraints is dynamic given different numeric cutoffs and different selected attributes. Restrictive templates are similar to the item filtering performed on the preprocessing phase. Our *group* constraint has no equivalent with inclusive or restrictive templates. A further difference is that we incorporate the *group* constraint into Phase 1 to accelerate association generation and we show that the *ac* constraint can only be used as a filtering process during rule generation. An optimized algorithm that incorporates constraints into the mining process is proposed in [27], but it does not consider group constraints. Our association rule constraints are related to [37, 32, 50]. Constraining association rules is studied in depth in [37], where constraints are defined as item boolean expressions involving two variables. This approach was later improved to incorporate queries as predicates on itemsets [32]. In [50] there is a proposal of algorithms that can incorporate constraints to include or exclude certain items in the association generation phase; they focus only in two types of constraints: items constrained by a taxonomy [48] or associations which include certain items. We do not use hierarchies, and even though we perform filtering on the preprocessing phase, excluding/including items is not enough to filter medically relevant rules. However, taxonomies could be useful classifying attributes into types of risk factors and types of perfusion measurements. This approach is based on succinctness (simple item selection predicates) and 2-variable constraints (boolean expressions); it is different from ours because it does not deal with predictive rules, attribute transformation, rule size and only uses support as the fundamental interestingness metric. It is well-known that simple constraints on support can be used for pruning the search space in Phase 1 [52]. Several association rule constraints are classified and studied in [42]. We discuss how these constraints relate to ours. Item constraints are related to our *group* constraint, but item constraints do not produce a partition of attributes as *group* constraints do. The length constraint is similar to the rule size constraint, with the difference that we only consider rules with a maximum size. Model-based constraints are somewhat similar to the *ac* constraint, but they are specified on itemsets and not on raw attributes. Aggregate constraints are not applicable since we are dealing with a non-uniform set of items, produced by the pre-processing phase. Our item filtering can be considered a succinct constraint. Further work on constrained association rules is discussed in [24]. Association rules and prediction rules from a decision tree are contrasted in [21]. The lift measure was introduced in [7] and was called interest. This measure is now commonly called lift and can be used to evaluate association rules for predictive purposes [4].

Rule covers [30, 31, 16, 45], maximal frequent itemsets [23, 34], closed itemsets [41, 43, 22] and basis [51, 3, 44, 10] represent alternatives to get condensed representations of association rules. Maximal frequent itemsets and closed sets represent itemsets, where adding an item makes them infrequent or makes them decrease their support. As explained in Section 3.3, our rule covers are related to them. The general idea is to find frequent closed itemsets or sets of representative associations from which the rest can be easily generated. An inference rule to generate a small set of non-redundant rules is introduced in [16]. The set of rules in a cover is minimized by allowing small errors in support [45]; in this manner more rules

are covered by fewer closed itemsets. The representation discussed in [30] involves a set of rules derived from rules having a form with a disjunction in the consequent. Such rules are not good for predictive purposes because they provide a complex profile of the attributes appearing in the antecedent. Our approach is different in several aspects from previous work. We use covers as a way to summarize and classify sets of association rules with the same consequent, whereas several approaches generate covers of rules having different consequents. Covers help understanding many rules, but they are not a substitute for each covered rule because we are more interested in rules having higher confidence and as shown in the experiments, some cover rules have lower confidence than individual rules. Our cover rules can be used to derive all mined rules, but the cover may derive some rules that may not be interesting based on the group constraint.

6 Conclusions

We studied the problem of constraining and summarizing association rules. We focused on using association rules for predicting combinations of several target attributes. The main reason behind using association rules for predictive purposes, instead of other data mining techniques, is that they are adequate to discover combinatorial patterns that exist in subsets of the data set attributes. Association rules are filtered using support, confidence and lift. Lift helps selecting rules with high predictive power and is used in conjunction with confidence to evaluate the significance of discovered rules. Several computational problems related to association rule mining were presented to motivate the introduction of search and filtering constraints. Such constraints had three main purposes: finding only predictive rules, reducing the number of discovered rules down to a manageable size and improving running time. Thus, attributes are constrained to belong to groups, in order to avoid uninteresting combinations among them and to reduce the combinatorial explosion of patterns. They are also constrained to appear either in the antecedent or in the consequent to discover predictive rules. Associations are constrained to have a maximum number of attributes to produce fewer and simpler rules. In our case, constraints are not enough to produce a small set of rules. Therefore, we use covers to produce a summary. We introduce a greedy algorithm that computes a cover for each set of rules having the same consequent. Covers include rules whose antecedent is a superset of the antecedent of simpler rules. This results in a smaller set of representative rules. Experiments with a real data set, studied the proposed constraints and the summarization with rule covers. Constraints were shown to significantly reduce both the number of discovered rules and running time. We were able to show interesting association rules predicting absence or existence of disease in four specific heart arteries, with high confidence and high lift. We used rule covers to summarize rules with a higher number of attributes and higher lift. We showed some representative rules selected from the rule covers, predicting absence or existence of disease as well, that have high quality metrics. Our proposed framework can be applied in other domains where prediction based on subsets of attributes is required on combinations of several target attributes.

There are several aspects for future research. The interaction between constraints and summarization with covers needs to be understood in more depth. The number of rules grows as support, confidence or lift are lowered. The data set may be post-processed with a clustering algorithm to explain why certain rules have low or high confidence. We intend to validate prediction rules with a different medical data set, but collecting such information is a difficult task given privacy regulations. Association rules and decision trees can be compared in the context of heart disease prediction with respect to their reliability, usefulness and simplicity. Finally, we need to study the impact on performance of each constraint from a theoretical point of view.

Acknowledgements

We thank the Emory University Radiology Department for providing the medical data set used in this article. We also thank Ernie Garcia, David Cooke and Elizabeth Krawczynska, from the Emory University Hospital, and Edward Omiecinski and Levien de Braal, from the Georgia Institute of Technology, for their

participation in the medical data mining project. The first author would like to thank Dr. Hiroshi Oyama, from the University of Tokyo, for his invitation to talk about this research work at the medical data mining seminar and for the excellent feedback provided. The first author would also like to thank the anonymous reviewers for their constructive comments.

References

- [1] R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. In *ACM SIGMOD Conference*, pages 207–216, 1993.
- [2] R. Agrawal and R. Srikant. Fast algorithms for mining association rules in large databases. In *VLDB Conference*, pages 487–499, 1994.
- [3] Y. Bastide, N. Pasquier, R. Taouil, and G. L. Lakhal. Mining minimal non-redundant association rules using frequent closed itemsets. In *Computational Logic*, pages 972–986, 2000.
- [4] R. Bayardo and R. Agrawal. Mining the most interesting rules. In *ACM KDD Conference*, pages 145–154, 1999.
- [5] C. Becquet, S. Blachon, B. Jeudy, J.F. Boulicaut, and O. Gandrillon. Strong association-rule mining for large-scale gene-expression data analysis: a case study on human SAGE data. *Genom Biol.*, 3(12), 2002.
- [6] L. Braal, N. Ezquerria, E. Schwartz, and Ernest V. Garcia. Analyzing and predicting images through a neural network approach. In *Proc. of Visualization in Biomedical Computing*, pages 253–258, 1996.
- [7] S. Brin, R. Motwani, J.D. Ullman, and S. Tsur. Dynamic itemset counting and implication rules for market basket data. In *Proc. ACM SIGMOD Conference*, pages 255–264, 1997.
- [8] S.E. Brossette, A.P. Sprague, J.M. Hardin, K.B. Waites, W.T. Jones, and S.A. Moser. Association rules and data mining in hospital infection control and public health surveillance. *J Am Med Inform Assoc. (JAMIA)*, 5(4):373–381, 1998.
- [9] S.E. Brossette, A.P. Sprague, W.T. Jones, and S.A. Moser. A data mining system for infection control surveillance. *Methods Inf Med.*, 39(4):303–310, 2000.
- [10] A. Bykowski and C. Rigotti. DBC: a condensed representation of frequent patterns for efficient mining. *Information Systems*, 28(8):949–977, 2003.
- [11] T.J. Chen, L.F. Chou, and S.J. Hwang. Application of a data mining technique to analyze coprescription patterns for antacids in Taiwan. *Clin Ther*, 25(9):2453–2463, 2003.
- [12] D. Cooke, C. Ordonez, E.V. Garcia, E. Omiecinski, E. Krawczynska, R. Folks, C. Santana, L. de Braal, and N. Ezquerria. Data mining of large myocardial perfusion SPECT (MPS) databases to improve diagnostic decision making. *Journal of Nuclear Medicine*, 40(5), 1999.
- [13] D. Cooke, C. Santana, T. Morris, L. de Braal, C. Ordonez, E. Omiecinski, N. Ezquerria, and E.V. Garcia. Data mining of large myocardial perfusion SPECT (MPS) databases: Validation of expert system rule confidences. *Journal of Nuclear Medicine*, 41(5):187, 2000.
- [14] D. Cooke, Cesar Santana, Tahia Morris, Leven de Braal, C. Ordonez, E. Omiecinski, N. Ezquerria, and Ernest V. Garcia. Validating expert system rule confidences using data mining of myocardial perfusion SPECT databases. In *Computers in Cardiology Conference*, pages 116–119, 2000.

- [15] C. Creighton and S. Hanash. Mining gene expression databases for association rules. *Bioinformatics*, 19(1):79–86, 2003.
- [16] L. Cristofor and D.A. Simovici. Generating an informative cover for association rules. In *ICDM*, pages 597–600, 2002.
- [17] M. Delgado, D. Sanchez, M.J. Martin-Bautista, and M.A. Vila. Mining association rules with improved semantics in medical databases. *Artificial Intelligence in Medicine*, 21(1-3):241–5, 2001.
- [18] S.M. Down and M.Y. Wallace. Mining association rules from a pediatric primary care decision support system. In *Proc of AMIA Symp.*, pages 200–204, 2000.
- [19] N. Ezquerra and R. Mullick. Perfex: An expert system for interpreting myocardial perfusion. *Expert Systems with Applications*, 6:455–468, 1993.
- [20] H.S. Fraser, W.J. Long, and S. Naimi. Evaluation of a cardiac diagnostic program in a typical clinical setting. *J Am Med Inform Assoc. (JAMIA)*, 10(4):373–381, 2003.
- [21] A. Freitas. Understanding the crucial differences between classification and association rules - a position paper. *SIGKDD Explorations*, 2(1):65–69, 2000.
- [22] K. Gade, J. Wan, and G. Karypis. Efficient closed pattern mining in the presence of tough block constraints. In *ACM KDD Conference*, pages 138–147, 2004.
- [23] K. Gouda and M.J. Zaki. Efficiently mining maximal frequent itemsets. In *ICDM Conference*, pages 163–170, 2001.
- [24] J. Han and M. Kamber. *Data Mining: Concepts and Techniques*. Morgan Kaufmann, San Francisco, 1st edition, 2001.
- [25] J. Han, J. Pei, and Y. Yin. Mining frequent patterns without candidate generation. In *Proc. ACM SIGMOD Conference*, pages 1–12, 2000.
- [26] J.L. Han. Background for association rules and cost estimate of selected mining algorithms. In *ACM CIKM*, pages 73–80, 1996.
- [27] J.L. Han. Pushing constraints in templates for mining association rules. In *Florida AI Research Symp*, pages 375–379, 1996.
- [28] T. Hastie, R. Tibshirani, and J.H. Friedman. *The Elements of Statistical Learning*. Springer, New York, 1st edition, 2001.
- [29] M. Klemettinen, H. Mannila, P. Ronkainen, H. Toivonen, and A.I. Verkamo. Finding interesting rules from large sets of discovered association rules. In *ACM CIKM*, pages 401–407, 1994.
- [30] M. Kryszkiewicz. Concise representation of frequent patterns based on disjunction-free generators. In *Proc. IEEE ICDM Conference*, pages 305–312, 2001.
- [31] M. Kryszkiewicz. Reducing borders of k-disjunction free representations of frequent patterns. In *ACM SAC Conference*, pages 559–563, 2004.
- [32] Laks V.S. Lakshmanan, R. Ng, J. Han, and Alex Pang. Optimization of constrained frequent set queries with 2-variable constraints. In *ACM SIGMOD Conference*, pages 157–168, 1999.
- [33] B. Lent, A. Swami, and J. Widom. Clustering association rules. In *Proc. IEEE ICDE Conference*, pages 220–231, 1997.

- [34] D.I. Lin and Z.M. Kedem. Pincer-search: a new algorithm for discovering the maximum frequent itemset. In *EDBT Conference*, pages 105–119, 1998.
- [35] W.J. Long. Medical reasoning using a probabilistic network. *Applied Artificial Intelligence*, 3:367–383, 1989.
- [36] W.J. Long, H.S. Fraser, and S. Naimi. Reasoning requirements for diagnosis of heart disease. *Artificial Intelligence in Medicine*, 10(1):5–24, 1997.
- [37] R. Ng, Laks Lakshmanan, and J. Han. Exploratory mining and pruning optimizations of constrained association rules. In *Proc. ACM SIGMOD Conference*, pages 13–24, 1998.
- [38] C. Ordonez, E. Omiecinski, Leven de Braal, Cesar Santana, and N. Ezquerra. Mining constrained association rules to predict heart disease. In *IEEE ICDM Conference*, pages 433–440, 2001.
- [39] C. Ordonez, C.A. Santana, and L. Braal. Discovering interesting association rules in medical data. In *Proc. ACM SIGMOD Data Mining and Knowledge Discovery Workshop*, pages 78–85, 2000.
- [40] T. Oyama, K. Kitano, T. Satou, and T. Ito. Extraction of knowledge on protein-protein interaction by association rule discovery. *Bioinformatics*, 18(5):705–714, 2002.
- [41] N. Pasquier, Y. Bastide, R. Taouil, G., and L. Lakhal. Discovering frequent closed itemsets for association rules. In *ICDT Conference*, pages 398–416, 1999.
- [42] J. Pei and J. Han. Constraints in data mining: Constrained frequent pattern mining: a pattern-growth view. *SIGKDD Explorations*, 4(1):31–39, 2002.
- [43] J. Pei, J. Han, and R. Mao. CLOSET: An efficient algorithm for mining frequent closed itemsets. In *ACM DMKD Workshop*, pages 21–30, 2000.
- [44] V. Phan-Luong. The representative basis for association rules. In *IEEE ICDM Conference*, pages 639–640, 2001.
- [45] V. Pudi and J.R. Haritsa. Reducing rule covers with deterministic error bounds. In *PAKDD Conference*, pages 313–324, 2003.
- [46] K. Shim R. Rastogi R. Mining optimized association rules with categorical and numeric attributes. In *Proc. IEEE ICDE Conference*, pages 503–512, 1998.
- [47] J.F. Roddick, P. Fule, and W.J. Graco. Exploratory medical knowledge discovery: Experiences and issues. *SIGKDD Explorations*, 5(1):94–99, 2003.
- [48] R. Srikant and R. Agrawal. Mining generalized association rules. In *VLDB Conference*, pages 407–419, 1995.
- [49] R. Srikant and R. Agrawal. Mining quantitative association rules in large relational tables. In *Proc. ACM SIGMOD Conference*, pages 1–12, 1996.
- [50] R. Srikant, Q. Vu, and R. Agrawal. Mining association rules with item constraints. In *Proc. ACM KDD Conference*, pages 67–73, 1997.
- [51] R. Taouil, N. Pasquier, Y. Bastide, and L. Lakhal. Mining bases for association rules using closed sets. In *IEEE ICDE Conference*, page 307, 2000.
- [52] K. Wang, Y. He, and J. Han. Pushing support constraints into association rules mining. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 15(3):642–658, 2003.

Author Biographies

Carlos Ordonez received a degree in Applied Mathematics (Actuarial Sciences) and an M.S. degree in Computer Science, both from the UNAM University, Mexico, in 1992 and 1996 respectively. He got a Ph.D. degree in Computer Science from the Georgia Institute of Technology, USA, in 2000. Dr. Ordonez currently works for Teradata (NCR) conducting research on database and data mining technology. He has published more than 20 research articles and holds three patents.

Norberto Ezquerra obtained his undergraduate degree in Mathematics and Physics from the University of South Florida, and his doctoral degree from Florida State University, USA. He is an Associate Professor in the College of Computing at the Georgia Institute of Technology and an adjunct faculty member in the Emory University School of Medicine. His research interests include computer graphics, computer vision in medicine, AI in medicine, modeling of physically-based systems, medical informatics and telemedicine. He is Associate Editor of the IEEE Transactions on Medical Imaging Journal, and a member of the American Medical Informatics Association and the IEEE Engineering in Medicine Biology Society.

Cesar A. Santana received his M.D. degree in 1984 from the Institute of Medical Science, in Havana, Cuba. In 1988 he finished his residency training in Internal Medicine, and in 1991, completed a fellowship in Nuclear Medicine in Havana, Cuba. Dr. Santana received a Ph.D. in Nuclear Cardiology in 1996 from the Department of Cardiology of the Vall d' Hebron University Hospital in Barcelona, Spain. Dr. Santana is an Assistant Professor in the Emory University School of Medicine and conducts research in the Radiology Department at the Emory University Hospital.