Evaluating Association Rules and Decision Trees to Predict Multiple Target Attributes

Carlos Ordonez, Kai Zhao
University of Houston
Houston, TX 77204, USA *

Abstract

Association rules and decision trees represent two well-known data mining techniques to find predictive rules. In this work, we present a detailed comparison between constrained association rules and decision trees to predict multiple target attributes. We identify important differences between both techniques for such goal. We conduct an extensive experimental evaluation on a real medical data set to mine rules predicting disease on multiple heart arteries. The antecedent of association rules contains medical measurements and patient risk factors, whereas the consequent refers to the degree of disease on one artery or multiple arteries. Predictive rules found by constrained association rule mining are more abundant and have higher reliability than predictive rules induced by decision trees. We investigate why decision trees miss certain rules, why they tend to have lower confidence and the possibility of improving them to match constrained association rules. Based on our experimental results, we show association rules, compared to decision trees, tend to have higher confidence, they involve larger subsets of the data set, they are better for multiple target attributes, they work better with user-defined binning and they are easier to interpret.

Keywords: association rule, decision tree, classification

1 Introduction

Association rules are a popular and powerful data mining technique, which have triggered an important body of research [15, 16]. Association rules have been successfully applied on basket, census, geographic and financial data [15]. On the other hand, medical data sets have been generally analyzed with decision trees [15], clustering [15], regression [16] or statistical tests [16], but less commonly with association rules. We believe combinatorial search techniques like association rules and OLAP processing have promise in the analysis of medical data sets, which are small, but have complex information content. This work studies association rule discovery in medical records to improve disease diagnosis when there exist multiple target attributes.

Association rules exhaustively look for hidden patterns, making them ideal to discover predictive rules involving subsets of the medical data set attributes [26]. Nevertheless, there exist three main issues. First, in general, given a medical data set a significant fraction of association rules is irrelevant because they are trivial or do not make medical sense. Second, many significant rules having high quality metrics tend to appear only at low support values. Third and most importantly, the number of discovered rules becomes extremely large at low support. Therefore, it is necessary to introduce search constraints [26, 21] to reduce the number of association rules and accelerate search. On the other hand, decision trees represent a well-known machine learning technique used to find predictive rules combining numeric and categorical attributes, which raises

the question of how predictive rules mined by constrained association rules compare to rules induced by a decision tree. With that motivation in mind, we compare association rules and decision trees with respect to accuracy, applicability, interpretability and time efficiency to predict several related target attributes, in the context of heart disease prediction. Our constrained association rule algorithm was programmed and optimized with SQL queries, like other data mining techniques [15, 23].

The article is organized as follows. Section 2 provides a discussion on related research work and motivates our approach. Section 3 introduces basic definitions on association rules and decision trees. Section 4 explains how to transform a medical data set into a binary data set appropriate for association rule mining, introduces search constraints to accelerate the discovery process and identifies important differences between both techniques, in the context of disease prediction. Section 5 presents experiments with a medical data set, comparing constrained association rules and decision trees from several angles, including reliability, overfit, number of rules, simplicity and time performance. Section 6 presents conclusions and directions for future work.

2 Related Work

Literature on association rules is extensive [15]. Most research has concentrated on accelerating algorithms to mine association rules since they work in a combinatorial search space. Although there has been interest in applying association rules in classification problems, it has been done mostly considering a single binary target attribute [14, 15]. The application of association rules in problems that have multiple target attributes, possibly non-binary, remains poorly understood. An issue that complicates matters is that multiple association rules generally refer to overlapping subsets of the data set. From a performance perspective, most research has ignored how processing time behaves when the data set is small, but is highly dimensional. Therefore, our goal is to understand well how association rules can discover rules involving multiple attributes, by comparing them to a standard classification technique like decision trees. As a secondary aspect, we consider high dimensionality rather than data set size as the main performance bottleneck. Such problem setting is precisely the case of a medical data set having heart disease measurements and risk factor attributes, which is the basic data set analyzed in our work. In the following paragraphs we will review related work on efficient approaches to mine association rules, we will discuss important research on applying data mining on medical data and we will provide a focused review on using association rules to improve heart disease diagnosis.

Association rules represent a fundamental data mining technique [15]. The A-priori algorithm remains a basic search algorithm framework as of today. There exist many efficient techniques to discover association rules. Most approaches concentrate on speeding up the frequent itemset generation phase [15]. Some of them use data structures that can help frequency counting for itemsets like the hash-tree, the FP-tree [15] or heaps [17]. Reference [17] proposes an efficient algorithm based on a heap than can efficiently mine frequent itemsets with new records or decreasing support. In [30] global association support is bounded and approximated for data streams with the support of recent and old itemsets; this approach relies on discrete algorithms for efficient frequency computation. It is possible to approximate support and confidence from a clustering model or a correlation matrix on binary data [24]. This model-based approach can be combined with constraints to produce more reliable rules. Our group constraint represents a user-defined way to prune the search space, instead of being automatic. Both [31] and [19] use different approaches to automatically bin numeric attributes. Instead, in our approach it was preferred to use well-known medical cutoffs for binning numeric attributes, to improve result interpretation and validation. Our search constraints share some similarities with previous work incorporating constraints into association rule mining [32]. There are algorithms that can incorporate constraints to include or exclude certain items in the association generation phase [15, 32]; but most approaches focus only on two types of constraints: items constrained by a certain hierarchy or associations which include certain items. It is well-known that simple constraints on support can be used for pruning the search space in frequent itemset search [32]. Association rules and prediction rules

from decision trees are contrasted in [14], emphasizing classification is a broad and ill-defined problem, whereas association rules are a specific, simpler and well-defined problem. The main ingredient to apply association rules for classification is to test them on an independent data set, to get a reliable measure of their accuracy [14, 21]. A lift measure for association rules, which helps us rank rules, was introduced in [3]. Rule covers [18] and basis [6] are alternative mechanisms to build condensed representations of association rules. Our SQL implementation of constrained association rules is related to developing data mining algorithms in SQL [23, 27]. K-means clustering was programmed with SQL queries introducing three variants [23]: standard, optimized and incremental. Later, this proposal was extended to develop Bayesian classification models exploiting K-means to perform class decomposition [27]. Generating frequent itemsets and searching for constrained association rules in SQL represent a complementary (discrete, lattice-based) approach to building (continuous, multidimensional) clustering and Bayesian classification models.

We now give an overview of important related work on exploiting data mining techniques in medicine. Important issues in medical data [29] include distributed and uncoordinated data collection, strong privacy concerns, diverse data types (image, numeric, categorical, missing information), complex hierarchies behind attributes and a comprehensive knowledge base. A well-known program to help heart disease diagnosis based on Bayesian networks is described in [13, 20]. Association rules have been used to help infection detection and monitoring [4, 5], to understand what drugs are co-prescribed with antacids [7], to discover frequent patterns in gene data [1, 9], to understand interaction between proteins [28] and to detect common risk factors in pediatric diseases [11]. Fuzzy sets have been used to extend association rules [10]. Reference [26] studies the impact of search constraints on the number of discovered rules and algorithm running time and also proposes a summarization of a large number of rules having the same consequent. This work shows constraints are essential to reduce the number of rules and running time.

Our work is a continuation of previous studies using data mining, statistical analysis and machine learning techniques on medical data sets to improve heart disease diagnosis and treatment [8, 21]. In [2], neural networks were used to predict heart response based on exercise stress and heart muscle thickening images. Search constraints for association rules are presented in [26]. Reference [21] introduces techniques to learn predictive association rules for disease prediction by testing and filtering them on independent subsets of a data set, eliminating rules that are particular to the data set. More recently, [25] studies how to combine OLAP cube [15] processing with parametric statistical tests to isolate risk factors which may trigger heart disease. This is accomplished by building pairs of subsets which have a significant difference in heart disease and differ on one characteristic (e.g. risk factor). Such approach is complementary to constrained association rules. Advantages of such approach are the fact that heart artery measurements do not need to be binned and negation of attributes is automatically considered. But such approach cannot explain why two or more risk factors lead to heart disease and may not give good results when target attributes do not follow a Gaussian distribution. Our work has common research goals with [21] in the sense that we explore a large search space to find interesting results. However, this article represents a first attempt to compare constrained association rules with a well-known classification technique. Our choice of decision trees has several reasons. Decision trees work well with low dimensional data sets having several numeric attributes. Decision trees can isolate good attributes to classify a data set. Rules induced by decision trees are highly similar to association rules: both have a conjunction of predicates (comparisons) on the left hand side of the rule, both have a measure of reliability (confidence factor and confidence), both can have one target attribute on the right hand side of the rule (although association rules can have more attributes) and both refer to specific subsets of the data set. From a processing perspective both techniques require multiple passes over the data set to gradually grow rules, but decision trees tend to be faster.

This article is a significant extension of [22], where association rules and decision trees are compared for the first time, also in the context of heart disease prediction. Reference [22] presents preliminary evidence that association rules are more powerful to predict multiple target attributes. This work also showed decision trees where numeric attributes are automatically split do not produce much better rules than trees on binary attributes. In this new article we present a more comprehensive experimental evaluation comparing rules

from a qualitative and quantitative perspective, we analyze how good predictive attributes are isolated by both techniques and we study in more depth why decision trees miss specific rules found by a constrained association rule algorithm.

3 Definitions

3.1 Association Rules

The standard definition of association rules [15, 24] is as follows. Let $D = \{T_1, T_2, \ldots, T_n\}$ be a set of n transactions and let $\mathcal{I}$ be a set of items, $\mathcal{I} = \{i_1, i_2 \ldots i_m\}$. Each transaction is a set of items, i.e. $T_i \subseteq \mathcal{I}$. An association rule is an implication of the form $X \Rightarrow Y$, where $X, Y \subset \mathcal{I}$, and $X \cap Y = \emptyset$; X is called the antecedent and Y is called the consequent of the rule. Basically, the rule states that if X happens then Y happens. In general, a set of items, such as X or Y, is called an itemset. In this work, a transaction is a patient record transformed into a binary format where only positive binary values are included as items. Therefore, each item corresponds to one numeric range or one categorical value. In medical terms, an association rule relates medical measurements and patient risk factors (X) to the existence of disease in a subset of patients (Y).

Let $P(X)$ be the probability of appearance of itemset X in D and let $P(Y|X)$ be the conditional probability of appearance of itemset Y given itemset X appears. For an itemset $X \subseteq \mathcal{I}$, $s(X)$ is defined as the fraction of transactions $T_i \in D$ such that $X \subseteq T_i$. That is, $P(X) = s(X)$. The support of a rule $X \Rightarrow Y$ is defined as $s(X \Rightarrow Y) = P(X \cup Y)$. An association rule $X \Rightarrow Y$ has a measure of reliability called $c(X \Rightarrow Y)$ defined as $P(Y|X) = P(X \cup Y)/P(X) = s(X \cup Y)/s(X)$. The standard problem of mining association rules is to find all rules whose metrics are equal to or greater than some specified minimum support τ and minimum confidence ψ thresholds, respectively [15, 24]. A k-itemset with support above the minimum threshold is called frequent. We use a third significance metric for association rules called *lift* [26]: $l(X \Rightarrow Y) = P(Y|X)/P(Y) = c(X \Rightarrow Y)/s(Y)$. Lift quantifies the predictive power of $X \Rightarrow Y$; we are interested in rules such that $l(X \Rightarrow Y) > 1$.

3.2 Decision Trees

In decision trees (DT) [12] the input data set has one attribute called class $\mathcal{C}$ that takes a value from K discrete values $1, \ldots, K$, and a set of numeric and categorical attributes $A_1, \ldots, A_p$. The goal is to predict $\mathcal{C}$ given $A_1, \ldots, A_p$. Decision tree algorithms automatically split numeric attributes A_i into two ranges and they split categorical attributes A_j into two subsets at each node. The basic goal is to maximize class prediction accuracy $P(\mathcal{C} = c)$ at a terminal node (also called node purity) where most points are in class c and $c \in \{1, \ldots, K\}$. Splitting on numeric attributes is generally based on the information gain ratio (an entropy-based measure) or the Gini index [12]. The splitting process is recursively repeated until no improvement in prediction accuracy is achieved with a new split. The final step involves pruning nodes to make the tree smaller and to avoid model overfit. The output is a set of rules that go from the root to each terminal node consisting of a conjunction of inequalities for numeric variables ($A_i <= x, A_i > x$) and set containment for categorical variables ($A_j \in \{x, y, z\}$) and a predicted value c for class $\mathcal{C}$. In general decision trees have reasonable accuracy and are easy to interpret if the tree has a few nodes. Detailed discussion on decision trees can be found in [15, 16].

4 Comparing Constrained Association Rules and Decision Trees

We start by introducing a transformation process of a data set with categorical and numerical attributes to transaction (sparse binary) format. We then introduce search constraints to find only predictive association

rules and to accelerate the search process. We then discuss important differences between association and decision trees. We call our improved association rule mining technique constrained association rules (CAR).

4.1 Transforming Medical Data Set

A medical data set with numeric and categorical attributes must be transformed to binary dimensions, in order to use association rules on the data set D, defined in Section 3. Numeric attributes are binned into intervals and each interval is mapped to an item. Categorical attributes are transformed by mapping each categorical value to one item. Our first constraint is the negation of an attribute, which makes search more exhaustive. If an attribute has negation then additional items are created, corresponding to each negated categorical value or each negated interval. Missing values are assigned to additional items, but they are not used. 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.

4.2 Constrained Association Rules

We now introduce our constrained association rules (CAR) technique. Search constraints were studied before in [26, 21], but here we provide a summary to make exposition clear. A more detailed theoretical discussion on the properties of constraints is available in [26]. Our discussion is based on the standard association rule search algorithm [15], which has two phases. The first phase finds all itemsets having minimum support, proceeding bottom-up, generating frequent 1-itemsets, 2-itemsets and so on, until there are no frequent itemsets. The second phase produces all rules whose support and confidence are above user-specified thresholds. Two of our constraints work on the first phase and the other one works on the second one. Constraints are generally specified on attributes and not on items. Let $\mathcal{I} = \{i_1, i_2, \ldots i_m\}$ be the set of items to be mined, obtained by the transformation process from the attributes $\mathcal{A} = \{A_1, \ldots, A_p\}$. Let $attribute()$ be a function that returns the mapping between one attribute and one item.

The first constraint is κ, the user-specified maximum itemset size. This constraint prunes the search space for k-itemsets of size such that $k > \kappa$. This constraint reduces the combinatorial explosion of large itemsets and helps finding simple rules. Each predictive rule will have at most κ attributes (items).

The second constraint involves selecting those items relevant to the predictive task at hand. Generally speaking, there are many items whose appearance is irrelevant depending on the predictive task at hand. For instance, if certain risk factors are considered unimportant to predict disease in a specific artery they can be eliminated before the algorithm starts computing frequent itemsets.

We now introduce the third constraint. Let $\mathcal{C} = \{c_1, c_2, \ldots c_p\}$ be a set antecedent and consequent constraints for each attribute A_j. Each c_j can take two values: 1 if attribute A_j can only appear in the antecedent of a rule and 2 if A_j can only appear in the consequent. 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 satisfy the antecedent constraint if for all $i_j \in X$ then $ac(attribute(i_j)) = 1$; X satisfies the consequent constraint if for all $i_j \in X$ then $ac(attribute(i_j)) = 2$. This constraint ensures we only find predictive rules with disease attributes in the consequent. Intuitively, this constraint acts as a template to filter rules.

The fourth and last constraint is called the group attribute constraint. Let $\mathcal{G} = \{g_1, g_2, \ldots g_p\}$ be a set of p group constraints corresponding to 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 $group : \mathcal{A} \rightarrow \mathcal{G}$ as $group(A_j) = g_j$. 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 satisfies the group constraint if for each item pair $\{a, b\}$ s.t. $a, b \in \mathcal{I}$ it is true $group(attribute(a)) \neq group(attribute(b))$. The group constraint avoids finding trivial or redundant rules as well as reducing the number of frequent itemsets to accelerate rule generation.

Input: raw data set S with p attributes and n records.
Output: predictive association rules of the form $A \Rightarrow B$, satisfying thresholds τ, ψ, λ and constraints $group()$ and $ac()$.

1. Data set preprocessing:
   Transform data set S with n patient records and p attributes $A_1, \ldots, A_p$ into a transaction data set $D = \{T_1, \ldots, T_n\}$, based on input cutoffs for bins and negation for numeric attributes.

2. Frequent itemset search:
   Search for frequent itemsets X of size k, $k = 1 \ldots \kappa$, such that $s(X) \geq \tau$ and which satisfy the $group()$ constraint.

3. Rule generation:
   Generate predictive rules in sets of rules having the same consequent, filtering them with minimum confidence ψ, minimum lift λ and the $ac()$ constraint. Each set of rules is of the form: $\mathcal{R} = \{A_1 \Rightarrow B, A_2 \Rightarrow B, \ldots, A_d \Rightarrow B\}$, where A, B are itemsets.

Figure 1: Constrained association rule algorithm.

4.3 Constrained Association Rule Algorithm

We join the data set transformation and search constraints into an algorithm that goes from transforming medical records into transaction to getting predictive rules. The transformation process using the given cutoffs for numeric attributes and desired negated attributes, produces the input data set for frequent itemset search. Each patient record becomes a transaction T_i (see Section 3). After the medical data set is transformed, items are further filtered out depending on the prediction goal: predicting absence or existence of heart disease. Items can only be filtered after attributes are transformed because they depend on the numeric cutoffs and negation. That is, it is not possible to filter items based on raw attributes. This is explained in more detail in Section 5. In frequent itemset search we use the $group()$ constraint to avoid searching for trivial itemsets. Frequent itemset search finds all frequent itemsets from size 1 up to size κ. Rule generation produces predictive rules satisfying the $ac()$ constraint. Our algorithm main input parameters are κ (maximum itemset length), τ (minimum support) and ψ (minimum confidence). Section 3 provides further details. The constrained association rule algorithm input, output and major steps are shown in Figure 1.

It is necessary to run the algorithm twice: one to mine "absence" of disease and another one to mine "existence" of disease; items are filtered before generating 1-itemsets, respectively. Preprocessing is semi-manual, requiring the user to specify cutoffs, negation, constraints and items to be discarded. Frequent itemset search and rule generation are automatic, getting minimum support τ, confidence ψ and lift λ thresholds as input parameters.

4.4 Similarities and Differences between Association Rules and Decision Trees

We start by explaining how rules from both techniques can be compared with each other. Recall that we use the acronym CAR to refer to constrained association rules and DT for decision trees. Assume DT is built with binary attributes. In other words, splits on numeric attributes are predefined. Notice that when a decision tree automatically splits numeric attributes it becomes impossible to find an equivalence with association rules on manually binned attributes. Consider a predictive rule $X \Rightarrow Y$. An association rule can have multiple attributes on the consequent Y, whereas decision tree rules have only one attribute on Y. In the decision tree the target attribute appears only in a terminal node (leaf) and it corresponds to one predicted value. Therefore, the path from root to leaf determines all attributes in the antecedent X. A leaf determines only one target attribute value of Y. In a rule $X \Rightarrow Y$ induced by a decision

tree CF is called its confidence factor and N refers to the number of records complying with $X \cup Y$. Assuming a rule from a decision tree involves exactly the same binary attributes we have identified the following similarities and differences. We base our discussion on confidence and support, which are the two fundamental reliability metrics in association rules. The rule confidence $c(X \Rightarrow Y)$ is equal to the CF, with a nice and straightforward equivalence. Support, on the other hand, requires more careful interpretation. In an association rule $s(X \Rightarrow Y)$ is equivalent to $N * CF$ in a decision tree, where N is the number of records belonging to the terminal node. This is a consequence of not having a pure node, where all records belong to the same class. We now discuss some interesting aspects about metric thresholds. The minimum support threshold τ corresponds to the minimum number of records λ, required to split a node ($\lambda = \tau n$). By the induction process, a decision tree on binary attributes can only obtain rules whose CF is at least 50%, whereas association rules can obtain rules with confidence below 50%, if needed. Thus given an association rule, it may not be possible to find an equivalent rule in the decision tree. On the other hand, given a rule in the decision tree it is always possible to find an equivalent association rule setting sufficiently low thresholds τ and ψ.

We now discuss interesting algorithmic aspects. Association rules perform an exhaustive combinatorial search, whereas decision trees recursively partition the data set using one attribute at a time to find hyper-rectangles with most records in one class. Therefore, association rules aim to find all rules above the given thresholds involving overlapping subsets of records, whereas decision trees find regions in space where most records belong to the same class. Association rules analyze item combinations corresponding to input attributes and target attributes together, whereas decision trees select only one input attribute at one time and they analyze the interaction of that attribute with the target attribute. In other words, decision trees resemble an algorithm that analyzes 2-itemsets at each level. In general, association rules work on previously binned attributes (although it is possible to perform automated binning to maximize confidence [15]), whereas decision trees can handle a combination of binned attributes and raw numeric attributes. Association rules algorithms can be slow, despite many optimizations proposed in the literature because they work on a combinatorial space, whereas decision trees can be comparatively much faster because each split obtains successively smaller subsets of records.

Based on the similarities and differences explained above, we summarize advantages and disadvantages for each technique. Association rules can find all existing predictive rules from a data set, given sufficiently low thresholds τ and ψ and no constraints. However, they depend on binned or binary attributes, which is fortunately the case for our medical data set. On the other hand, decision trees can miss many predictive rules found by association rules because they successively partition into smaller subsets. When a rule found by a decision tree is not found by association rules it is either because a constraint pruned the search space or because τ or ψ were too high. When there exist multiple target attributes, separate decision trees must be computed, one for each attribute. Even further, if it is necessary to analyze combinations of target attributes it is necessary to extend the data set with new attributes for each target combination (in general as a conjunction). When a decision tree is built an unbounded depth in the induction process can lead to small groups of records, decreasing rule generality and reliability. In a decision tree, internal nodes do not produce any rules (although tentative rules can be derived), which leads to increasingly longer and more complex rules until an acceptable node purity is reached. On the other hand, association rules can indeed produce rules corresponding to internal nodes corresponding to multiple trees, but they require careful interpretation since any two rules may refer to overlapping data subsets. Decision trees tend to be overfit for a particular data set, which may affect their applicability. Post-processing pruning techniques can reduce overfit, but unfortunately they also reduce rule confidence. Another issue is that decision trees can repeat the same attribute multiple times for the same rule because such attribute is a good discriminator. This is not a big issue since rules are conjunctions and therefore the rule can be simplified to one interval for the attribute, but such interval will be generally small and the rule too specific. In the experimental section we will study the differences and issues explained above.

5 Experiments

Our experiments focus on comparing the significance, accuracy and usefulness of predictive rules obtained by constrained association rules and decision trees. We used standard decision trees (CN4.5 [15]) with two variants: automatically deciding split points with gain and manually splitting attributes into ranges based on domain knowledge. We did not use bagging or boosting, combining multiple decision trees, which could improve the accuracy of induced rules. The parameter settings for both techniques are explained in more detail below. Recall that we use the acronym CAR to refer to constrained association rules and DT for decision trees. Our experiments were conducted on a database server running at 3.2 GHz with 4 GB of main memory and 256 GB of disk space. We used the Teradata DBMS V2R6 running on the Windows operating system. The constrained association rule (CAR) and the decision tree (DT) algorithms were implemented with SQL code generated by a Java program. Explaining the technical details on implementing these two data mining techniques in SQL falls outside the scope of this article, but the reader can consult [23, 27] for a brief introduction on related approaches. Time measurements are given in seconds.

5.1 Medical Data Set Description

There are three different kind of attributes for analysis: perfusion defect measurements, risk factors and coronary stenosis. Perfusion measurements are simplified imaging data comparing the muscle from the patient's heart to the muscle in a healthy "average" heart. The medical data set contains the profiles of $n = 655$ patients and has $p = 25$ medical attributes corresponding to the numeric and categorical attributes listed in Table 1. The data set has personal information such as age, race, gender and smoking habits. There are additional medical measurements such as weight, heart rate, blood pressure. There are also historical attributes describing pre-existence of related diseases. Finally, the data set contains the degree of artery narrowing (stenosis) for the four heart arteries. Such narrowing is associated to the risk of developing a heart attack when the artery blood flow gets interrupted due to a clot.

For decision trees we considered two versions of the data set. One version where independent numeric attributes were not binned and a second version where all independent attributes were binned. In the first version the decision tree automatically splits numeric attributes choosing the best cutoff point given the current data subset. Our first experiments use both versions for rule reliability analysis purposes, but we use the binary data set in the remaining experiments in order to produce subsets of rules that are equal on both techniques.

5.2 Default Parameter Settings

This section explains default settings for algorithm parameters, that were based on medical opinion and previous research work [26]. Table 1 contains a summary of medical attributes and default CAR search constraints.

Transformation parameters

To set the transformation parameters default values we must discuss attributes corresponding to heart vessels. The LAD, RCA, LCX and LM numbers represent the percentage of vessel narrowing (stenosis) compared to a healthy artery, where narrowing is 0%. Attributes LAD, LCX and RCA were binned at 50% and 70%. In cardiology a 70% value or higher indicates significant stenosis and a 50% value indicates borderline disease. Stenosis below 50% generally indicates the patient is considered healthy since it is unlikely the artery may get blocked. The LM artery has a lower (more stringent) cutoff because it poses a higher health risk than the other three arteries. The fundamental reason is LAD and LCX arteries branch from LM. Therefore, a defect in LM is likely to trigger more severe disease. Attribute LM was binned at 30% and 50%. 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,

Table 1: Medical data set.

Attribute	Description	Constraints		
		neg	group H D	ac
AGE	Age of patient	N	0 0	1
LM	Left Main narrowing	Y	0 0	2
LAD	Left Anterior Desc. artery narrowing	Y	0 0	2
LCX	Left Circumflex artery narrowing	Y	0 0	2
RCA	Right Coronary artery narrowing	Y	0 0	2
AL	Antero-Lateral	N	1 1	1
AS	Antero-Septal	N	1 1	1
SA	Septo-Anterior	N	1 1	1
SI	Septo-Inferior	N	1 1	1
IS	Infero-Septal	N	1 1	1
IL	Infero-Lateral	N	1 1	1
LI	Latero-Inferior	N	1 1	1
LA	Latero-Anterior	N	1 1	1
AP	Apical	N	1 1	1
SEX	Gender	N	0 0	1
HTA	Hyper-tension Y/N	N	2 0	1
DIAB	Diabetes Y/N	N	2 0	1
HYPLD	Hyperloipidemia Y/N	N	2 0	1
FHCAD	Family hist. of disease	N	2 0	1
SMOKE	Patient smokes Y/N	N	0 0	1
CLAUDI	Claudication Y/N	N	2 0	1
PANGIO	Previous angina Y/N	N	3 0	1
PSTROKE	Prior stroke Y/N	N	3 0	1
PCARSUR	Prior carot surg Y/N	N	3 0	1
CHOL	Cholesterol level	N	0 0	1

meaning a perfusion measurement greater or equal than 0.2 indicated a severe defect. CHOL was binned at 200 (warning) and 250 (high). AGE was binned at 40 (adult) and 60 (old). Finally, only the four artery attributes (LAD, RCA, LCX, LM) had negation to find rules referring to healthy patients and sick patients. The other attributes did not have negation. The data set had nulls values. We applied a common missing value imputation solution: for numeric attributes we used the mean and missing categorical values were substituted by the mode of the attribute.

Constrained Association Rules Parameters

The maximum association size κ was varied from 2 to 6. Minimum support, confidence and lift were used as the main filtering parameters. Minimum lift in this case was 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 power. Confidence, combined with lift, was used to evaluate the significance of each rule. The minimum support was fixed at $1\% \approx 7$. That is, rules referring to 6 or less patients were eliminated. Such threshold eliminated rules that were probably particular for our data set. From a medical point of view, rules with high confidence are desirable, but unfortunately, they are infrequent. Based on the domain expert opinion, the minimum confidence was set at 70%, which provides a balance between sensitivity (identifying sick patients) and specificity (identifying healthy patients) [26]. Minimum lift was set slightly higher than 1 to filter out rules where X and Y are very likely to be independent. Finally, we use a high lift threshold (1.2) to get rules where there is a stronger implication dependence between X and Y.

In our case, we selected items corresponding to the lower ranges of each artery to predict healthy arteries. On the other hand, when we are after rules predicting diseased arteries we select items representing the upper ranges. Similarly, since high perfusion measurements point to abnormal heart muscle we select items corresponding to the upper ranges to predict existence of heart disease.

The group constraint and the antecedent/consequent constraint had the following settings. Since we are trying to predict likelihood of heart disease, the 4 main coronary arteries LM, LAD, LCX and RCA are 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, risk factors (medical history and measurements) and perfusion measurements (9 heart regions) appear in the antecedent, whereas the four artery measurements appear in the consequent of a rule. From a medical perspective, determining the likelihood of presenting a risk factor based on artery disease is irrelevant. The default constraints are summarized in Table 1. Under column "group", the H subcolumn presents the group constraint to predict healthy arteries and the D subcolumn has the group constraint to predict diseased arteries.

It was necessary to perform separate runs for healthy and diseased arteries. We explain parameter settings for healthy arteries. 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 medically important. Risk factors HTA, DIAB, HYPLD, 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 unselected items (third constraint) involved arteries with values in the higher (not healthy) ranges (e.g. $[30, 100]$, $[50, 100]$, $[70, 100]$), perfusion measurements in $[-1, 0.2)$ (no perfusion defect), and risk factors equal to "y" for the patient (person presenting risk factor). Minimum support was $\tau = 1\%$ and minimum confidence was $\psi = 70\%$. On the other hand, the parameter settings to mine rules for diseased arteries were as follows. The four arteries (LAD, LCX, RCA, LM) had negation (which was not used for healthy arteries). Rules relating presence of risk factors (equal to "y") with diseased arteries were considered interesting. Group constraints were different. 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 patterns. 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 having "n" for the patient (person not presenting risk factor). Minimum support was 1% and minimum confidence was 70%.

Decision Trees Parameters

We now explain the decision tree construction algorithm and its main parameters. We used the CN4.5 decision tree [12] algorithm using gain ratio for splitting and pruning nodes. Such algorithm is widely used and it is one of the most accurate decision tree algorithms available [15, 16]. Pruning is essential to reduce overfitting, reduce data set fragmentation and to get more general rules [16]. In some experiments the depth of trees had a threshold to produce simpler rules. We show some classification rules with the percentage of patients (LS) they involve and their confidence factor (CF). The confidence factor has a similar interpretation to association rule confidence, but the percentage refers to the fraction of patients where the antecedent appears (i.e. support of antecedent itemset). For instance, if CF is less than 100% and $LS = 10\%$ then the actual support of the rule is less than 10%.

5.3 Comparing Association Rules and Decision Trees

The goal is to link perfusion measurements and risk factors to artery disease. Some rules were expected, confirming valid medical knowledge, and some rules were surprising, having the potential to enrich medical knowledge. We show some of the most important discovered rules. Predictive rules were grouped in two sets: (1) if there is a low perfusion measurement or no risk factor then the arteries are healthy; (2) if there exists a risk factor or a high perfusion measurement then the arteries are diseased.

Comparing Most Significant Rules

We used the following guidelines to select best rules. Rules with confidence $\geq 90\%$, with lift $>= 1.5$, and with two or more items in the consequent were considered medically significant and thus they were the most important. Rules with high support, only risk factors, low lift or borderline confidence were considered interesting, but not significant. Rules with artery figures in wide intervals (more than 70% of the attribute range) were not considered significant, such as rules having a measurement in the 30-100 range for the LM artery because they provide imprecise patient profiles. Figure 2 and Figure 3 show a sample of discovered association rules predicting healthy and diseased arteries, respectively.

Table 2 and Table 3 compare the total number of rules discovered by each technique and the total running time. For DT we computed trees with manual binning of attributes (based on the same cutoff points as association rules) and with automatic splits (cutoff point was picked by DT). We can observe CAR finds many rules than DT. In fact, the number of rules is two orders of magnitude larger. On the breakdown we can observe CAR always finds more rules in each rule category. DT with automatic splits produces more high confidence rules than the manual DT (as expected), but the number of rules is much smaller than association rules anyway. Another drawback is that automatic splits make medical interpretation more difficult [26, 21]. In the high confidence range DT finds very few rules. Rules with three attributes in the consequent are particularly hard to find; in this case CAR can find two rules but DT none. Both CAR and DT find an increasing number of rules as confidence or support go down. It is noteworthy DT tends to find many more rules having low support, which indicates the data set gets fragmented into small subsets.

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

Figure 2: Association rules for healthy arteries.

confidence = 1:
IF $0.2 <= SA < 1.0$ $HYPLPD = y$ $PANGIO = y$
THEN $70 <= LAD < 100$, s=0.01 c=1.00 l=3.2
IF $60 <= AGE < 100$ $0.2 <= SA < 1.0$ $FHCAD = y$
THEN $not(0 <= LAD < 50$, s=0.02 c=1.00 l=1.9
IF $0.2 <= IS < 1.0$ $CLAUDI = y$ $PSTROKE = y$
THEN $not(0 <= RCA < 50)$, s=0.02 c=1.00 l=2.3
IF $60 <= AGE < 100.0$ $0.2 <= IS < 1.0$ $250 <= CHOL < 500$
THEN $70 <= RCA < 100$, s=0.02 c=1.00 l=3.2
IF $0.2 <= IS < 1.0$ $SEX = F$ $250 <= CHOL < 500$
THEN $70 <= RCA < 100$, s=0.01 c=1.00 l=3.2
IF $0.2 <= IS < 1.0$ $HTA = y$ $250 <= CHOL < 500$
THEN $70 <= RCA < 100$, s=0.011 c=1.00 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=0.80 l=4.5
IF $0.2 <= AP < 1.1$ $PCARSUR = y$
THEN $not(0 <= LAD < 50)$ $not(0 <= RCA < 50)$, s=0.01 c=0.80 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 $60 <= AGE < 100.0$ $0.2 <= IL < 1.1$ $250 <= CHOL < 500$
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

Figure 3: Association rules for diseased arteries.

Table 2: Comparison between CAR and DT (healthy; time in secs).

Classification	CAR	DT manual	DT auto
Total associations	28313	-	-
Total rules	3826	26	22
Confidence 0.9-1.0	213	1	5
Confidence 0.8-0.9	1308	2	3
Confidence 0.7-0.8	2305	11	7
one item in consequent	3020	24	19
Two items in consequent	804	2	3
Three items in consequent	2	0	0
Support>=0.2	122	3	3
support 0.1-0.2	715	3	4
support<0.1	2989	20	15
Time	242	205	220

Table 3: Comparison between CAR and DT (diseased; time in secs).

Classification	CAR	manual DT	auto DT
Total associations	28976	-	-
Total rules	1459	17	22
Confidence 0.9-1.0	134	1	6
Confidence 0.8-0.9	464	3	6
Confidence 0.7-0.8	861	5	7
One item in consequent	1301	17	19
Two items in consequent	152	0	2
three items in consequent	6	0	1
Support>=0.2	3	3	3
support 0.1-0.2	62	3	0
support<0.1	1394	11	19
Time	82	194	188

Table 4: Analysis of attributes in DT.

	Attributes at different tree depth (healthy and diseased).			
Target attribute	1	2	3	4
Healthy				
LAD	AP	AGE IL	IL HYPLPD	LI IS
LCX	LI	SEX	HTA CHOL	AGE LA IS
RCA	IL	CLAUDI HYPLPD	AGE IS	AL SI
LM	-	-	-	-
Diseased				
LAD	SA	AP	AGE	AS
LCX	LI	SEX	HTA	PSTROKE SI
RCA	IL	HYPLPD	PCARSUR	CHOL
LM	-	-	-	-

Table 5: Analysis of rules in CAR but not in DT (healthy and diseased).

Reason	Number of rules	
	Healthy	Diseased
root attribute not in given CAR rule	1249	326
no internal node attribute in given CAR rule	16	0
no leaf attribute in given CAR rule	4	0
one CAR attribute never selected in DT	1105	304
two CAR attributes never selected in DT	376	48
three CAR attributes never selected in DT	22	0
two CAR attributes appear in different rules	146	149
three CAR attributes appear in different rules	1372	584

Analyzing DT structure

We now analyze which attributes DT picks at each level based on the automatic DT (the one producing better rules). Table 4 analyzes which attributes tend to appear closer to the root of the tree. Such attributes, in theory, discriminate healthy from sick patients. The first column shows a specific artery as the target attribute and the right columns show the attributes at different depths. We can see perfusion measurements are the best discriminator; risk factors tend to appear at deeper levels. We can also observe it is difficult to understand which risk factors are relevant for all arteries: HYPLPD and AGE appear only twice. DT cannot find any good rules involving LM.

Table 5 provides an interesting analysis of why DT does not find rules. We created a program that stored rules in the database and created queries to explain why a rule did not appear. We now summarize the main reasons rules were not found by DT. The attribute selected at the root of DT determines all rules derived from the tree. It turned out there were many rules in which the preferred "root" attribute selected by DT did not appear in many rules. Such finding highlights how differently each technique works. Another important reason is that DT discarded one attribute as a good discriminator, whereas CAR did include it. Finally, we can see that in many cases the attributes participating in a rule appeared in rules from DT, but they appeared in different tree branches. That is, they ended being disconnected. In general, they were very few association rules that did not involve attributes selected somewhere by DT. That is, both techniques made their best effort to exploit all attributes, but CAR was better at linking attribute combinations.

As explained in Section 4, CAR is guaranteed to find all predictive rules given sufficiently low thresholds. For the data set with binary attributes $\psi = 50\%$ and τ is the same as the minimum number of records needed to continue splitting. However, constraints eliminate some rules found by DT. Table 6 analyzes the reasons why rules are filtered out by CAR. As we can see confidence alone is not the reason rules are discarded. In most cases, a combination of constraints eliminates rules, highlighting their interaction.

Attribute Importance

Table 7 and Table 8 provide an analysis of attribute importance for CAR. This table shows which attributes tend to appear more frequently in rules of different length. The numbers indicate the percentage of rules where the given attribute appears. Notice these percentages do not add up to 100% because a combination of attributes may appear on each rule. We build one decision tree for each rule length in order to make a fair comparison. We do not show combinations of target attributes (arteries), but results are worse for DT. The left column shows each predictive attribute and the right column counts the number of attributes in the antecedent of the rule (i.e. the IF part of the rule). A general observation is that there are many more zeroes for DT than for CAR. CAR provides a more comprehensive coverage of attributes: in general more perfusion measurements and more risk factors participate in rules. On the other hand, DT isolates a

Table 6: Analysis of rules in DT but not in CAR.

Reason	# of attribs. in anteced. 1	2	3	4
Healthy:				
group constraint only	1	1	1	1
item selection constraint only	3	1	4	3
confidence threshold only	0	2	3	2
combination: any	3	4	5	6
Diseased:				
group constraint only	3	0	2	2
item selection constraint only	2	2	3	3
confidence threshold only	1	3	3	1
combination: any	4	5	6	5

Table 7: Percentage of rules where attribute appears (healthy).

	1		2		3		4	
Attribute	CAR	DT	CAR	DT	CAR	DT	CAR	DT
AGE	74	0	0	30	0	20	0	20
AL	37	0	42	0	0	0	0	100
AS	30	0	38	0	0	0	0	0
SA	25	0	36	0	0	0	0	0
SI	32	0	36	0	0	0	0	50
IS	35	0	32	0	0	50	0	50
IL	36	45	30	23	0	9	0	0
LI	46	47	34	0	0	6	0	12
LA	46	0	37	0	0	0	0	100
AP	37	57	26	0	0	0	0	0
SEX	10	0	47	58	19	0	0	0
HTA	1	0	19	0	35	43	6	0
DIAB	2	0	26	0	39	0	5	0
HYPLD	2	0	25	47	30	27	4	0
FHCAD	1	0	23	0	48	0	5	0
SMOKE	1	0	20	0	42	0	17	0
CLAUDI	1	0	12	100	40	0	12	0
PANGIO	0	0	5	0	42	0	51	0
PSTROKE	0	0	4	0	34	0	40	0
PCARSUR	0	0	5	0	36	0	44	0
CHOL	0	0	1	0	23	67	49	0

Table 8: Percentage of rules where attribute appears (diseased).

	1		2		3		4	
Attribute	CAR	DT	CAR	DT	CAR	DT	CAR	DT
AGE	27	0	0	0	0	30	0	0
AL	10	0	11	0	0	0	0	0
AS	16	0	16	0	0	0	0	100
SA	19	100	20	0	0	0	0	0
SI	14	0	18	0	0	0	0	50
IS	19	0	14	0	0	0	0	0
IL	22	23	12	0	0	0	0	0
LI	10	35	10	0	0	0	0	0
LA	8	0	9	0	0	0	0	0
AP	24	0	13	43	0	0	0	0
SEX	1	0	16	42	5	0	0	0
HTA	1	0	17	0	18	57	3	0
DIAB	3	0	8	0	13	0	4	0
HYPLD	1	0	10	27	20	0	8	0
FHCAD	0	0	4	0	16	0	4	0
SMOKE	0	0	3	0	10	0	7	0
CLAUDI	0	0	4	0	18	0	14	0
PANGIO	0	0	1	0	1	0	1	0
PSTROKE	0	0	7	0	12	0	3	100
PCARSUR	0	0	5	0	8	100	3	0
CHOL	0	0	2	0	3	0	1	33

few perfusion measurements at each rule lengths and less than three risk factors. These results state decision trees constitute a succinct model, capable of identifying good discriminating attributes, at the price of getting lower confidence rules. On the other hand, association rules do not represent a model, but a collection of high quality patterns hidden in the data set.

5.4 Time Efficiency

We now compare time efficiency between both techniques. Notice that for DT a rule of length $k-1$ is related to an association rule of length k, given a consequent Y with one attribute. Table 9 compares times with rules of increasing length. For each length we set κ (first constraint) for CAR and an equivalent tree depth for DT. Short length rules from DT have low reliability, but we compare both techniques producing similar rules. As can be seen CAR is significantly faster than DT for short rules. However, as rule length grows CAR becomes increasingly slower, which can be explained by the growing number of itemsets found. On the other hand, DT time growth is almost independent from the length of rules, showing a small growth as k increases. For diseased artery prediction CAR matches the time required by DT at maximum k. On the other hand, for healthy artery prediction CAR is twice as slow. Despite CAR being slower than DT there is not a significant difference for short length rules, which are the kind of rules we are after. CAR is competitive because it finds many more rules than DT. In fact, for $k = 6$ CAR finds a number of rules that is two orders of magnitude larger than those from DT.

We explore why DT is not much faster than CAR. Table 10 provides a breakdown of time for DT. We include the times needed to predict multiple arteries: a separate decision tree is needed for each combination. As can be seen, time grows slowly as there are more arteries in each combination. The time required to build

Table 9: Time: Analysis of rules with different number of antecedent items (time in secs).

Items	Itemsets	Rules CAR	Rules DT	Time CAR	Time DT
Healthy:					
k=2	351	6	6	13	183
k=3	2368	120	12	68	188
k=4	10006	793	19	166	199
k=5	28313	3826	26	242	205
k=6	56342	6798	35	363	204
Diseased:					
k=2	565	2	3	6	168
k=3	3688	70	7	36	175
k=4	13011	615	11	69	189
k=5	28976	1459	17	82	194
k=6	46504	1636	26	195	196

each decision tree is small, but all times add up to a significant amount of time, which in this case is in the range of minutes. Another important observation is that the time to predict healthy or diseased arteries is quite similar.

5.5 Discussion

Our experiments provide evidence that decision trees are not as powerful as association rules to exploit a set of numeric attributes manually binned and categorical attributes to predict several related target attributes. Decision trees do not work well with combinations of several target variables (arteries), which requires defining one class attribute for each values combination. Decision trees fail to identify many medically relevant combinations of independent numeric variable ranges and categorical values (i.e. perfusion measurements and risk factors). When given the ability to build height-unrestricted trees decision trees tend to find complex and long rules, making rule applicability and interpretation difficult. Also, in such case decision trees find few predictive rules with reasonably sized ($> 1\%$) sets of patients; this is a well-known drawback known as data set fragmentation [16]. To complicate matters, rules sometimes repeat the same attribute several times creating a long sequence of splits that needs to be simplified.

We have provided evidence, that for the purpose of predicting disease with several related target attributes, association rules are more effective. However, our constraints for association rules may be adapted to decision trees, but that is subject of future work. Decision trees do have advantages over association rules. A decision tree partitions the data set, whereas association rules on the same target attribute may refer to overlapping subsets; sometimes this makes result interpretation difficult. A decision tree represents a predictive model of the data set, whereas association rules are disconnected among themselves. In fact, the large number of discovered association rules requires rule summarization or ranking. A decision tree is guaranteed to have at least 50% prediction accuracy and generally above 70% accuracy for binary target variables, whereas association rules specifically require the confidence threshold ψ to filter out low confidence rules.

6 Conclusions

We compared constrained association rules and decision trees to find predictive rules on a data set having multiple target attributes. For association rules we used an A-priori style algorithm using constraints to prune the itemset search space and reduce the number of rules. On the other hand, we used the well-known

Table 10: Time: Breakdown of running time for DT (time in secs).

Arteries	Healthy time	Diseased time
LM	11	12
LAD	11	12
LCX	12	11
RCA	12	11
LM LAD	12	13
LM LCX	14	11
LM RCA	13	12
LAD LCX	14	13
LAD RCA	14	13
LCX RCA	15	13
LM LAD LCX	16	14
LM LAD RCA	15	14
LM LCX RCA	15	15
LAD LCX RCA	15	14
LM LAD LCX RCA	16	15

CN4.5 decision tree algorithm. We did not use bagging or boosting, combining multiple decision trees, which could improve the accuracy of induced rules. We focused on solving a medical problem, where the objective was to find rules predicting absence or existence of artery disease on several arteries, given patient risk factors and medical measurements as predictive attributes. We introduced search constraints to produce only medically useful rules and to reduce running time. We identified important differences and similarities between constrained association rules and decision trees in the context of heart disease prediction. Experiments with a medical data set, obtained from a hospital, compare predictive constrained association rules with rules induced by decision trees. Our experiments show constrained association rules find many high confidence rules, whereas decision trees only find a few. Decision trees are shown to miss many important predictive rules found by constrained association rules. In general, decision trees tend to favor a few attributes to initially partition the data set and such choice cannot be changed later. Important attributes are considered irrelevant by decision trees, despite the fact that they participate in high confidence rules. Rules from decision trees tend to refer to very small sets of records because the data set gets fragmented. Given an association rule it is often the case that its predictive attributes appear scattered on different branches of the tree and therefore they also appear in different rules. The only rules that are found by decision trees that were not found association rules are those that are filtered by combinations of constraints. Therefore, they are medically irrelevant. From an efficiency point of view, constrained association rules are faster and better than decision trees to find short length rules. On the other hand, they become slower than decision trees to find long rules, but they find two orders of magnitude more rules than decision trees. In summary, to predict multiple target attributes, as is the case for artery stenosis, constrained association rules find more rules, there are more rules with high reliability and the additional time in processing is reasonable.

Our work suggests several directions to improve decision trees and association rules. A mixed set of attributes may produce higher confidence rules, where some attributes are automatically binned by the decision tree, while other attributes are manually binned by the user. A set of small decision trees may be an alternative to using a large number of association rules. Decision trees may be used to pre-process a large data set to partition it into focused subsets, where association rules may be applied in a second phase.

Acknowledgments

We thank Dr. Cesar Santana from the Emory University Hospital for his medical opinion to validate results. We want to thank the Emory University Hospital for providing the medical data set used in this article. This research work was partially supported by US National Science Foundation grants CCF 0937562 and IIS 0914861.

References

[1] 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.

[2] L. Braal, N. Ezquerra, 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.

[3] 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.

[4] 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.

[5] 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.

[6] A. Bykowski and C. Rigotti. DBC: a condensed representation of frequent patterns for efficient mining. *Information Systems*, 28(8):949–977, 2003.

[7] 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.

[8] D. Cooke, C. Ordonez, E.V. Garcia, E. Omiecinski, E. Krawczynska, R. Folks, C. Santana, L. de Braal, and N. Ezquerra. Data mining of large myocardial perfusion SPECT (MPS) databases to improve diagnostic decision making. *Journal of Nuclear Medicine*, 40(5), 1999.

[9] C. Creighton and S. Hanash. Mining gene expression databases for association rules. *Bioinformatics*, 19(1):79–86, 2003.

[10] 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.

[11] 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.

[12] U. Fayyad and G. Piateski-Shapiro. *From Data Mining to Knowledge Discovery*. MIT Press, 1995.

[13] 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.

[14] A. Freitas. Understanding the crucial differences between classification and association rules - a position paper. *SIGKDD Explorations*, 2(1):65–69, 2000.

[15] J. Han and M. Kamber. *Data Mining: Concepts and Techniques*. Morgan Kaufmann, San Francisco, 1st edition, 2001.

[16] T. Hastie, R. Tibshirani, and J.H. Friedman. *The Elements of Statistical Learning*. Springer, New York, 1st edition, 2001.

[17] J. Huang, S. Chen, and H. Kuo. An efficient incremental mining algorithm-QSD. *Intelligent Data Analysis*, 11(3):265–278, 2007.

[18] M. Kryszkiewicz. Concise representation of frequent patterns based on disjunction-free generators. In *Proc. IEEE ICDM Conference*, pages 305–312, 2001.

[19] B. Lent, A. Swami, and J. Widom. Clustering association rules. In *Proc. IEEE ICDE Conference*, pages 220–231, 1997.

[20] 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.

[21] C. Ordonez. Association rule discovery with the train and test approach for heart disease prediction. *IEEE Transactions on Information Technology in Biomedicine (TITB)*, 10(2):334–343, 2006.

[22] C. Ordonez. Comparing association rules and decision trees for disease prediction. In *Proc. ACM HIKM Workshop*, pages 17–24, 2006.

[23] C. Ordonez. Integrating K-means clustering with a relational DBMS using SQL. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 18(2):188–201, 2006.

[24] C. Ordonez. Models for association rules based on clustering and correlation. *Intelligent Data Analysis*, 13(2):337–358, 2009.

[25] C. Ordonez and Z. Chen. Evaluating statistical tests on OLAP cubes to compare degree of disease. *IEEE Transactions on Information Technology in Biomedicine (TITB)*, 13(5):756–765, 2009.

[26] C. Ordonez, N. Ezquerra, and C.A. Santana. Constraining and summarizing association rules in medical data. *Knowledge and Information Systems (KAIS)*, 9(3):259–283, 2006.

[27] C. Ordonez and S.K. Pitchaimalai. Bayesian classifiers programmed in SQL. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 22(1):139–144, 2010.

[28] 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.

[29] J.F. Roddick, P. Fule, and W.J. Graco. Exploratory medical knowledge discovery: Experiences and issues. *SIGKDD Explorations*, 5(1):94–99, 2003.

[30] C. Silvestri and S. Orlando. Approximate mining of frequent patterns on streams. *Intelligent Data Analysis*, 11(1):49–73, 2007.

[31] R. Srikant and R. Agrawal. Mining quantitative association rules in large relational tables. In *Proc. ACM SIGMOD Conference*, pages 1–12, 1996.

[32] 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.