Knowledge-Based Systems 67 (2014) 16–25

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Principal Association Mining: An efficient classification approach

Fuzan Chen [a], Yanlan Wang [a], Minqiang Li [a,c], Harris Wu [b,*], Jin Tian [a]

[a] College of Management and Economics, Tianjin University, China
[b] Department of Information Technology and Decision Sciences, Old Dominion University, USA
[c] State Key Laboratory of Hydraulic Engineering Simulation and Safety, Tianjin University, China

ARTICLE INFO

Article history:
Received 28 June 2013
Received in revised form 13 May 2014
Accepted 11 June 2014
Available online 19 June 2014

Keywords:
Data mining
Classification
Associative classification
Association rule
Knowledge discovery

ABSTRACT

Classification is one of the key tasks in business intelligence, decision science, and machine learning. Associative classification has aroused significant research interest in recent years due to its superior accuracy. Traditional association rule mining algorithms often yield many redundant and sometimes conflicting class association rules. This paper presents a new, efficient associative classification approach. This new approach produces a compact classifier with a small number of association rules, yet with good classification performance. This approach is based on a novel rule quality metric, named as Principality, which measures an association rule's classification accuracy and coverage for a specific class. Heuristic methods utilizing the Principality metric are applied to rule pruning and associative classifier construction to produce a compact classifier. This Principal Association Mining (PAM) approach is confirmed to be effective at improving classification accuracy as well as decreasing classifier size by experiments conducted on 17 datasets.

1. Introduction

Classification is one of the most important data analysis tasks in business intelligence, machine learning and pattern recognition. Accurate and efficient classifiers for large scale datasets can help us better understand big data. Classification is frequently used in business decision-making, such as electronic commerce, financial markets, trend prediction and loan approval.

Classification is a well-studied problem. Different methods for classifier construction have been proposed, including decision trees, rule induction, Naive Bayesian model, neural networks, support vector machines, and statistical models such as linear/quadratic discriminant analysis [4,7,10]. In particular, rule-based methods, which induce minimal rule-based concept descriptions from training datasets, are a mainstay of research in classification because of various desirable properties, e.g., their expressiveness and intelligibility to human as well as their efficiency and effectiveness in classification.

Association mining was first proposed by Agrawal et al. [1], which aimed to discover association rules that determine implication or correlation among co-occurring elements within a dataset. The relationships in association mining are represented by frequent itemsets/patterns and association rules. The general form of association rule is the $X \Rightarrow Y$ implication, where X and Y are called *antecedent* and *consequent* respectively. Association rules try to answer questions such as "if a customer purchases product A, how likely is she to purchase product B?" or "What products will a customer buy if she buys products C and D?"

In the last decade classification based on association rule mining, also known as associative classification, has emerged as a powerful enhancement of traditional rule-based learning [18]. The basic intuition of associative classification is to substitute traditional rule induction with an association rule mining process. The resulting classification model, called associative classifier (AC), consists of class association rules (CAR). The *antecedent* of a CAR is co-occurrent attribute values, which frequently appear across the training data, while the *consequent* is the target class attribute, i.e. the class label. In general, associative classifiers yield better accuracy than decision trees and rule-based classifiers [23]. The reason is that CARs represent the correlations among different attributes simultaneously and provide confidence probability which can be used to address the uncertainty problem of classification. However, an associative classifier often consists of a huge number of association rules. Today's data gets larger and richer along with decreasing computing costs. The number and complexity of CARs increase exponentially as the number of underlying data attributes increases. The growth in number and complexity

* Corresponding author. Address: Department of Information Technology and Decision Sciences, Old Dominion University, 5105 Hampton Blvd, Norfolk, VA 23529, USA. Tel.: +1 757 683 4460.
E-mail address: hwu@odu.edu (H. Wu).

http://dx.doi.org/10.1016/j.knosys.2014.06.013
0950-7051/© 2014 Elsevier B.V. All rights reserved.

of rules results in vastly increased efforts to understand the rules and to resolve redundancy (when rules do not bring new information to user) and conflicts (when rules have the same antecedent but different consequent class labels). The trend of larger and more complex data sets triggers reconsideration of data classification options. The compactness of a classifier now deserves more attention. We aim to construct a compact associative classifier that performs well in terms of accuracy and yet is small in size.

In this paper, we present a new, efficient approach for constructing a compact associative classifier. We present a novel rule quality metric, named as Principality, which combines both classification accuracy and coverage of a class association rule. We propose an associative classifier construction method that derives class association rules from frequent patterns and then prunes the rules utilizing the principality metric. The presented method is shown to produce high classification accuracy with a reduced number of rules in the classifier according to the experiments on 17 University of California, Irvine (UCI) benchmark datasets [19].

The remainder of this paper is organized as follows. Section 2 gives a brief overview of related work. Section 3 describes basic concepts of associative classification and problem statements. Section 4 presents our associative classification approach and section 5 presents the experiments results. Section 6 concludes this study and outlines future research.

2. Related work

As a new classification method, associative classification has become popular in recent years. Several methods have been proposed to build a classifier with high quality class association rules. Such methods include CBA (Liu et al., 1998), CMAR [13], L^3 [3], CPAR [28], MCAR [22], as well as methods for a non-standard classification task such as CAEP [6]. These techniques use several different approaches to discover frequent patterns, extract rules, rank rules, prune redundant or harmful rules (rules that lead to incorrect classification) and classify new test objects.

CBA (Classification Based on Associations) introduced the idea of utilizing frequent patterns for classification and built an associative classifier to predict class labels according to the most confident classification rule. CBA employed the famous Apriori candidate generation method [1] to find the frequent patterns. Strong rules, whose confidences are not less than the user-defined confidence threshold, are then generated by these frequent patterns. After that, a pruning strategy is applied to discard useless and redundant rules. All of the CARs are ranked in a descending order of confidence, support and generated time. The classification for a new data object is based on the highest precedence rule which matches the object. An improved CBA algorithm named CBA(2) was proposed by Liu et al. [14], which overcomes the weaknesses of the single-support limitation and lengthy rule mining. The first weakness is addressed by using multiple minimum class supports, and the second is addressed by integrating CBA with decision tree and Naive Bayesian methods. Several other approaches have been proposed attempting to improve the performance of CBA in different ways. To improve the efficiency of the Apriori candidate generation step, for example, CMAR and L^3 use the FP-growth approach [9], which adopts a divide-and-conquer method to project and partition the dataset. Nguyen et al. [17] designed a structure called lattice of class rules for mining CARs efficiently, where each node contained attribute values. Rak et al. [21] organized the datasets in a tree-projection structure. The branches in these advanced structures represent only particular items instead of the whole itemsets, therefore the number of candidate tests is reduced.

The associative classification approaches have higher accuracy than decision tree classifiers because association rules explore highly confident associations among multiple variables at one time, whereas decision tree classifier examines one variable at once. Unfortunately, recent researches show that association classifiers may suffer from severe inherent limitations. Such weaknesses include large sets of CARs, problems with the support–confidence framework, and handling of continuous data. In order to tackle above problems, especially for the first two issues, and to achieve high classification accuracy, extensive research has been carried out to develop better methods for associative classification.

In many cases, the huge number of resulting CARs may potentially overfit the training dataset. Therefore there have been many attempts to reduce the size of CARs to construct a compact and accurate associative classifier. The challenging task in this phase is how to select a good criterion to evaluate the quality of the rules. Only "high quality" rules are selected to form a classifier. For example, redundant rules, i.e., those rules whose confidence is lower than the confidence of more general rules, are pruned in most approaches. E.g. R_1: $p_1p_2 \Rightarrow c$ will be marked as redundant in the presence of rule R_2: $p_1 \Rightarrow \mathrm{c}$ and $conf(R_1) < conf(R_2)$. In this case, R_1 is in fact a more specific version of R_2. It does not actually bring any new information to the user, as the information contained in R_1 is actually part of the information contained in R_2. Thus R_1 is redundant to rule R_2. Ashrafi et al. [2] propose a fixed antecedent and consequent method to remove redundant rules from the resultant rule set, where a rule is redundant when it finds a set of other rules that also convey the same knowledge. Liu et al. [15] employ closed sets to post-process the CARs and remove insignificant rules. The basis of this method is the dependency among rules, from which closed sets can be derived. Some special data structures are also adopted to reduce useless rules. Costa et al. [5] build a hierarchical classification framework, which combines associative rule learning and probabilistic smoothing. Therein a global rule-based classifier is refined from the local probabilistic models by performing a probabilistic analysis of the coverage of individual rules. CPAR inherits the basic idea of rule-based methods, such as FOIL/FFOIL [20] and RIPPER [26], and integrates the features of associative classification into predictive rule analysis, which can generate a smaller set of high-quality predictive rules and prevent redundant rule generation. GEAR [30] uses the information gain to determine the best attribute for each class and therefore generate a compact rule set. The database coverage method is also used to prune CARs. For example, the rule selection in CBA, CMAR, and MCAR is made using the database coverage heuristic, which evaluates the complete set of CARs on the training data set; and only CARs that cover a certain number of training data objects are considered. Since the ideal support threshold is not known in advance, the database coverage pruning method often discards some useful knowledge.

There is a trade-off between the size of the classifiers and the predictive accuracy: often slightly lower accuracy can be tolerated in exchange for a more compact set of concise rules. Over-pruning results in a small number of rules but the resulting classification may be inaccurate. Some associative classification techniques adopt a lazy pruning strategy, which limit its pruning to only negative or harmful rules. L^3 adopts a lazy pruning strategy in which only rules that yield wrong case classifications are discarded. Veloso et al. [25] also employs lazy pruning with an MDL (Minimum Description Length)-based entropy minimization method. In general, lazy pruning results in classifiers holding a very large number of spare or secondary rules, which is inefficient for classifying large datasets.

Many associative classification approaches work in a support–confidence framework, where a support threshold is used to