


A Novel Incremental Framework for Building Classifier Using Constraint Class Association Rules

B. Subbulakshmi, Dept. of Computer Science and Engineering, Thiagarajar College of Engineering, Madurai, India

C. Deisy, Dept. of Information Technology, Thiagarajar College of Engineering, Madurai, India

S. Parthasarathy, Dept. of Applied Mathematics & Computational Science, Thiagarajar College of Engineering, Madurai, India

 <https://orcid.org/0000-0001-7439-6878>

ABSTRACT

Associative classification (AC) performs much better than other traditional classifiers. It generates a huge number of class association rules (CARs). Since users are interested in the subset of rules, constraints are introduced in the generation of CARs. Real-world databases are record-based in which data is continuously added which demands incremental mining. Hence, constraint class association rules (CCAR) is mined from incremental data. To limit the number of rules and to remove the duplicate rules, redundant rule pruning and duplicate rule pruning techniques are applied. To improve the accuracy of the classifier, the rule selection using principality metric has been applied and the classifier is constructed with rules possessing high principality. Then, classifier is evaluated using single rule and multiple rule prediction methods and the accuracy of the proposed classifier are measured. Experimental results show that the accuracy of the proposed classifier is relatively higher when compared to other algorithms.

KEYWORDS

Associative Classification, Class Association Rules, Constraints, Prediction, Pruning

1. INTRODUCTION

1.1 Associative Classification

Associative Classification (AC) (Vishwakarma et al. 2013; Phan-Luong, 2013) is an integration of association and classification methods aims to build a classifier describing the classes of the input training data set. Class Association Rules (CARs) (Mabu et al. 2011) are basically used to build a classification model for which prediction defines the relationship between the itemsets and the class

DOI: 10.4018/IJIRR.316125

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

labels. It is mainly used in the medical dataset. Based on the generated rules, it is possible to identify population at high risk for a particular disease.

Constructing an associative classifier involves two steps. Given the training dataset, class association rules are generated first. Then, a small subset of high- quality rules are obtained using pruning methods and an accurate classifier or model is built for the training data. However, in large or correlated data sets, rule mining may yield a huge number of classification rules. Hence, pruning techniques, in particular, support- based pruning, are exploited to reduce the complexity of the extraction task.

1.2 Motivation

A class association rule is a special case of association rule where the rule consequent contains only a class label (Nguyen et al., 2016). The basic objective of any user in mining class association rules (CARs) is to determine a complete set of CARs in such a way that it satisfies user-specified minimum support and minimum confidence thresholds. In literature there exists several methods and techniques have been proposed to manage this issue in CARs. In fact, the CAR mining has been widely used in various practical domains, namely tourism management ((Rong et al., 2012; Leung et al., 2013), social media security (Zhang et al., 2007), healthcare (Nahar et al., 2013), and education sectors (Luna et al., 2015). However, it is also observed by previous researchers that the results obtained as a result of CARs mining has redundancy.

Many real time applications are dynamic in nature such that the data is incrementally added to the database. The major problem with incremental mining is updating knowledge which was mined from the original database. Methods for mining frequent itemsets from incremental datasets have been proposed by (Cheung et al. 1996; Hong et al. 2001; Le et al. 2011, 2012 ; Vo et al. 2014). These methods reduce the execution time and memory usage compared to those obtained from rescanning the original dataset. Thus the techniques of incremental frequent itemset mining can be integrated into constraint CAR Mining. Hence, the incremental method generates all CARs satisfying the given constraint whenever the original dataset is added with new records. However, there exists the possibility of generation of duplicate and redundant CCARs. It can be addressed by applying post processing techniques such as pruning and selection of rules.

In the post-processing step, the Constraint Class Association Rules (CCARs) can be pruned, ranked and then quality rules are selected in order to build a compact and high quality classifier. The classifier is evaluated by measuring the number of correct classifications made for a given set of test instances. Finally, the prediction of test instances is made by selecting the single rule or multiple rules satisfying the antecedent of test instances.

1.3 Constraint Class Association Rule (CCAR) Mining

Class Association Rules (CARs) (Mabu et al. 2011) are basically used to build a classification model for prediction. These rules define the association between the itemsets and the class labels and it is useful in various application domains. However, the users are interested in only the subset of CARs, for instance, those rules that contain at least one itemset from a user-defined set of itemsets. Giving constraints on itemsets reduces the number of generated CARs and decrease the search space. Thereby, it improves the performance of the mining process. Moreover, constrained CARs help to discover interesting or useful rules specific to the end user. For example, while analyzing the patient data to identify those people who are at the high risk of HIV infection, analyst may focus on the rules that include information such as age, marital status, and the gender information in the rule antecedents. Similarly, biologists focus on rules involving new drugs to understand the effectiveness of new treatment strategies for cancer treatment applications.

Constraint-based mining (Nguyen & Vo 2014; Nguyen et al. 2015; 2016) provides

- Flexibility to user to express their interestingness using constraints

19 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/a-novel-incremental-framework-for-building-classifier-using-constraint-class-association-rules/316125

Related Content

A Unified Algorithm for Identification of Various Tabular Structures from Document Images

Sekhar Mandal, Amit K. Das, Partha Bhowmick and Bhabatosh Chanda (2013). *Modern Library Technologies for Data Storage, Retrieval, and Use* (pp. 1-28). www.irma-international.org/chapter/unified-algorithm-identification-various-tabular/73762

Ranking Documents Based on the Semantic Relations Using Analytical Hierarchy Process: Query Expansion and Ranking Process

Ali I. El-Dsouky, Hesham A. Ali and Rabab Samy Rashed (2017). *International Journal of Information Retrieval Research* (pp. 22-37). www.irma-international.org/article/ranking-documents-based-on-the-semantic-relations-using-analytical-hierarchy-process/181724

Data Mining Approach to Decision Support in Social Welfare

Ricardo Anderson and Gunjan Mansingh (2018). *Information Retrieval and Management: Concepts, Methodologies, Tools, and Applications* (pp. 1783-1802). www.irma-international.org/chapter/data-mining-approach-to-decision-support-in-social-welfare/198625

Nature-Inspired-Based Modified Multi-Objective BB-BC Algorithm to Find Near-OGRs for Optical WDM Systems and Its Performance Comparison

Shonak Bansal and Kuldeep Sharma (2018). *Handbook of Research on Biomimicry in Information Retrieval and Knowledge Management* (pp. 1-25). www.irma-international.org/chapter/nature-inspired-based-modified-multi-objective-bb-bc-algorithm-to-find-near-ogrs-for-optical-wdm-systems-and-its-performance-comparison/197692

A Noval Approach for Object Recognition Using Decision Tree Clustering by Incorporating Multi-Level BPNN Classifiers and Hybrid Texture Features

Upendra Kumar (2024). *International Journal of Information Retrieval Research* (pp. 1-31).

www.irma-international.org/article/a-noval-approach-for-object-recognition-using-decision-tree-clustering-by-incorporating-multi-level-bpnn-classifiers-and-hybrid-texture-features/338394