A Novel Rule Ordering Approach in Classification Association Rule Mining

Yanbo J. Wang[1], Qin Xin[2], and Frans Coenen[1]

[1] Department of Computer Science, The University of Liverpool,
Ashton Building, Ashton Street, Liverpool, L69 3BX, UK
{jwang,frans}@csc.liv.ac.uk
[2] Department of Informatics, University of Bergen,
P.B.7800, N-5020 Bergen, Norway
Xin@ii.uib.no

Abstract. A Classification Association Rule (CAR), a common type of mined knowledge in Data Mining, describes an implicative co-occurring relationship between a set of binary-valued data-attributes (items) and a pre-defined class, expressed in the form of an "antecedent ⇒ consequent-class" rule. Classification Association Rule Mining (CARM) is a recent Classification Rule Mining (CRM) approach that builds an Association Rule Mining (ARM) based classifier using CARs. Regardless of which particular methodology is used to build it, a classifier is usually presented as an ordered CAR list, based on an applied rule ordering strategy. Five existing rule ordering mechanisms can be identified: (1) Confidence-Support-size_of_Antecedent (CSA), (2) size_of_Antecedent-Confidence-Support (ACS), (3) Weighted Relative Accuracy (WRA), (4) Laplace Accuracy, and (5) χ^2 Testing. In this paper, we divide the above mechanisms into two groups: (i) pure "support-confidence" framework like, and (ii) additive score assigning like. We consequently propose a hybrid rule ordering approach by combining one approach taken from (i) and another approach taken from (ii). The experimental results show that the proposed rule ordering approach performs well with respect to the accuracy of classification.

Keywords: Classification Association Rules, Classification Association Rule Mining, Data Mining, Rule Ordering.

1 Introduction

Classification Rule Mining (CRM) [15] is a well-known Data Mining technique for the extraction of hidden Classification Rules (CRs) from a given database that is coupled with a set of pre-defined classes, the objective being to build a classifier to classify "unseen" data records. One recent approach to CRM is to employ Association Rule Mining (ARM) [1] techniques to identify the desired CRs, i.e. Classification Association Rule Mining (CARM). In [9], Coenen *et al.* suggest that results presented in [13] and [14] show that CARM seems to offer greater accuracy of classification, in many cases, than other CRM methods such as C4.5 [15]. CARM mines a set of Classification Association Rules (CARs) from a class transaction database (the

P. Perner (Ed.): MLDM 2007, LNAI 4571, pp. 339–348, 2007.

well-established transaction database in a class fashion), where a CAR describes an implicative co-occurring relationship between a set of binary-valued data attributes (items in a transaction database) and a pre-defined class, expressed in the form of an "antecedent ⇒ consequent-class" rule. Regardless of which particular methodology is used to generate CARs, a classifier is usually presented as an ordered CAR list, based on an applied rule ordering mechanism. In [7] Coenen and Leng evaluated a number of alternative case satisfaction and rule ordering strategies. They indicate that (1) three common case satisfaction approaches are best first rule, best *K* rule, and all rules; and (2) five existing rule ordering mechanisms are Confidence-Support-size_of_Antecedent (CSA), size_of_Antecedent-Confidence-Support (ACS), Weighted Relative Accuracy (WRA), Laplace Accuracy, and χ^2 Testing. In this paper, we further divide (2) into two groups: (i) pure "support-confidence" framework like, and (ii) additive score assigning like. We consequently propose a hybrid rule ordering approach by combining one mechanism taken from (i) and another mechanism taken from (ii). The experimental results show good performance regarding the accuracy of classification when using the proposed rule ordering approach with the best first rule case satisfaction.

2 Related Work

2.1 An Overview of CARM Algorithms

The idea of CARM was first presented in [3]. Subsequently a number of alternative approaches have been described. Broadly CARM algorithms can be categorized into two groups according to the way that the CARs are generated:

- **Two stage algorithms** where a set of CARs are produced first (stage 1), which are then pruned and placed into a classifier (stage 2). Examples of this approach include CBA [14] and CMAR [13]. CBA (Classification Based Associations), developed by Liu *et al.* in 1998, is an Apriori [2] based CARM algorithm, which (1) applies its CBA-GR procedure for CAR generation; and (2) applies its CBA-CB procedure to build a classifier based on the generated CARs. CMAR (Classification based on Multiple Association Rules), introduced by Han and Jan in 2001, is similar to CBA but generates CARs through a FP-tree [11] based approach.
- **Integrated algorithms** where the classifier is produced in a single processing step. Examples of this approach include TFPC[1] [7] [9], and induction systems such as FOIL [16], PRM and CPAR [17]. TFPC (Total From Partial Classification), proposed by Coenen *et al.* in 2004, is a Apriori-TFP [8] based CARM algorithm, which generates CARs through efficiently constructing both P-tree and T-tree set enumeration tree structures. FOIL (First Order Inductive Learner) is an inductive learning algorithm for generating CARs developed by Quinlan and Cameron-Jones in 1993. This algorithm was later developed by Yin and Han to produce the PRM (Predictive Rule Mining) CAR generation

[1] TFPC may be obtained from http://www.csc.liv.ac.uk/~frans/ KDD/Software.