Weak Ratio Rules: 
A Generalized Boolean Association Rules

Baoqing Jiang, Henan University, China
Xiaohua Hu, Henan University, China, and Drexel University, USA
Qing Wei, Henan University of Economics and Law, China
Jingjing Song, Qingyuan Polytechnic, China
Chong Han, Henan University, China
Meng Liang, Henan University, China

ABSTRACT
This paper examines the problem of weak ratio rules between nonnegative real-valued data in a transactional database. The weak ratio rule is a weaker form than Flip Korn’s ratio rule. After analyzing the mathematical model of weak ratio rules problem, the authors conclude that it is a generalization of Boolean association rules problem and every weak ratio rule is supported by a Boolean association rule. Following the properties of weak ratio rules, the authors propose an algorithm for mining an important subset of weak ratio rules and construct a weak ratio rule uncertainty reasoning method. An example is given to show how to apply weak ratio rules to reconstruct lost data, and forecast and detect outliers.

Keywords: Association Rules, Data Mining, Ratio Rules, Uncertainty Reasoning, Weak Ratio Rules

INTRODUCTION
The problem of mining association rules from large databases has been subject of numerous studies. Some of them focus on developing faster algorithms for the classical method and/or adapting the algorithms to various situations, for example, distributed algorithm ODAM (Ashrafi, Taniar, & Smith, 2004), association rules mining in data warehouses (Tjioe & Taniar, 2005), multidimensional database mining (Yu et al., 2009; Casali et al., 2010), exception rules mining (Daly & Taniar, 2004; Taniar, Rahayu, Lee, & Daly, 2008) and redundant analysis (Ashrafi, Taniar, & Smith, 2007). Another direction is to define rules that modify some conditions of the classical rules to adapt to new applications (Marcus, Maletic, & Lin, 2001; Giannikopoulos, Varlamis, & Eirinaki, 2010). For instance, Srikant and Agrawal (1996) extended the categorical definition to include quantitative data and investigated the quantitative association rules problems.

For quantitative attributes, the general idea is partitioning the domain of a quantitative

DOI: 10.4018/jdwm.2011070103
attribute into intervals, and applying Boolean algorithms to the intervals. But there is a conflict between the minimum support problem and the minimum confidence problem, while existing partitioning methods cannot avoid the conflict (Tong et al., 2005).

Fuzzy method has successfully been used in data mining (Kuok, Fu, & Wong, 1998; Kwok, Smith, Lozano, & Taniar, 2002). Fuzzy association rules solve interval partition problem in some extent. It needs to do fuzzy partition to the domain of quantitative attribute, i.e., to consider fuzzy set on the domain. However, in some actual problem, to some test data, we can not propose meaningful fuzzy sets.

Flip Korn et al. (1998) focused on real-valued data such as dollar amounts spent by customers on products in transactional database and proposed the following ratio rule:

\[ \text{bread} : \text{butter} = 2 : 3 \]

which means that the ratio of the dollar amounts spent by a customer on bread to butter is 2 : 3 . Flip Korn also investigated the application of ratio rule on data cleaning, forecasting, “what-if” scenarios, outlier detection and visualization.

When quantitative attributes have no ratio relation, ratio rule can not better describe the quantitative association relation of quantitative attribute. At this time, there not exist real number \( a, b \) such that \( x : y = a : b \) for attributes \( x, y \), but there may exist real number \( a, b \) such that \( x : y \leq a : b \).

In this paper, we focus on the relationship between nonnegative real-valued data in transactional database and investigate the influence of the values, of some attributes called antecedent, to the values of other attributes called consequent. When antecedent and consequent only include one attribute, the quantitative relationship of antecedent and consequent can be expressed as \( x : y \leq a : b \). For example, \( \text{bread} : \text{butter} \leq 2 : 3 \) has such a reasoning meaning that if the dollar amount spent by a customer on bread is 2, then the dollar amounts spent by the customer on butter is at least 3.

It is precisely this reason that the rules proposed by this paper are called weak ratio rules.

In order to establish the mathematical model of weak ratio rules problem, we investigate the Boolean association rules problem deeply and propose the equivalent description of it.

By the equivalent description of Boolean association rules problem, we can generalize it naturally to weak ratio rules as an implication of the form \( A \Rightarrow B \), where \( A, B \) are nonnegative real-valued function defined on itemset \( I \). An attribute is also called an item in this paper, \( I \) is the set of all items.

We use \( 2^I \) to denote the set of all subset of \( I \), and \([0, +\infty)^I \) to denote the set of all nonnegative real-valued function on \( I \). For any subset \( X \) of \( I \), we consider \( X \) is the same as its characteristic function \( X_x \) , a nonnegative real-valued function on \( I \), whose value is 0 or 1. In this viewpoint, \( 2^I \) is a subset of \([0, +\infty)^I \).

By discussing the properties of weak ratio rule (simply as WRR in this paper sometimes), we come to that every WRR corresponds to a Boolean association rule (simply as BAR in this paper sometimes) as its support rule. The set of all WRRs can be classified by their support rules. In order to mine all WRRs, we need only mine all BARs, then for a single BAR \( (X \Rightarrow Y) \) to mine the set of all WRRs \( R_w(X \Rightarrow Y) \) supported by the BAR \( (X \Rightarrow Y) \).

By analyzing the structure of \( R_w(X \Rightarrow Y) \), we can find an important subset \( Q_w(X \Rightarrow Y) \) of \( R_w(X \Rightarrow Y) \). The subset can be expressed as an ideal of a chain-product. In this paper, we use a depth-first algorithm Boundary to mine the subset \( Q_w(X \Rightarrow Y) \).

By Remark 5, we come to that if \( (X \Rightarrow Y; ms, mc) \) is a BAR and a customer purchases all items in \( X \), then there is \( mc \) possibility that the customer purchases all items in \( Y \).

Similarly, if \( (A \Rightarrow B; ms, mc) \) is a WRR and a customer purchases \( F(x) \) (unit value) items \( x(x \in \text{supp} A) \), then how much unit
Related Content

Vertical Fragmentation in Databases Using Data-Mining Technique
[www.igi-global.com/article/vertical-fragmentation-databases-using-data/1812?camid=4v1a](www.igi-global.com/article/vertical-fragmentation-databases-using-data/1812?camid=4v1a)

A Survey of Open Source Tools for Business Intelligence
[www.igi-global.com/article/survey-open-source-tools-business/3896?camid=4v1a](www.igi-global.com/article/survey-open-source-tools-business/3896?camid=4v1a)
A Workload Assignment Strategy for Efficient ROLAP Data Cube Computation in Distributed Systems
Ilhyun Suh and Yon Dohn Chung (2016). *International Journal of Data Warehousing and Mining* (pp. 51-71).
[www.igi-global.com/article/a-workload-assignment-strategy-for-efficient-rolap-data-cube-computation-in-distributed-systems/168486?camid=4v1a](www.igi-global.com/article/a-workload-assignment-strategy-for-efficient-rolap-data-cube-computation-in-distributed-systems/168486?camid=4v1a)

Legal Framework for Data Mining and Privacy
[www.igi-global.com/chapter/legal-framework-data-mining-privacy/29150?camid=4v1a](www.igi-global.com/chapter/legal-framework-data-mining-privacy/29150?camid=4v1a)