

Chapter XXVII

Mining Association Rules

Mihai Gabroveau
University of Craiova, Romania

ABSTRACT

During the last years the amount of data stored in databases has grown very fast. Data mining, also known as knowledge discovery in databases, represents the discovery process of potentially useful hidden knowledge or relations among data from large databases. An important task in the data mining process is the discovery of the association rules. An association rule describes an interesting relationship between different attributes. There are different kinds of association rules: Boolean (crisp) association rules, quantitative association rules, fuzzy association rules, etc. In this chapter, we present the basic concepts of Boolean and the fuzzy association rules, and describe the methods used to discover the association rules by presenting the most important algorithms.

INTRODUCTION

The progress of Information Technology has determined companies to record a large quantity of information in huge databases. Due to the fact that a lot of useful knowledge is hidden in these databases, the companies need to extract this knowledge in order to help the decision-making process.

Data mining, also known as knowledge discovery in databases (Chen, Han, & Yu, 1996),

provides efficient automated techniques for discovering potentially useful, hidden knowledge or relations among data from large databases. Data mining functions include classification, clustering, prediction, regression, and link analysis (associations), etc.

Mining association rules represent an unsupervised data mining method that allows identifying interesting associations, correlations between items, and frequent patterns from large transactional databases. This problem was first

introduced by Agrawal et al. in (Agrawal, Imielinski, & Swami, 1993).

The original motivation for searching association rules came from the need to analyze the *market-basket data* that stores items purchased on a per-transaction basis, in order to identify customer behaviors by finding different items often purchased together. A typical example of an association rule has the following statement:

$$popcorn \Rightarrow beer \text{ [support} = 7\%, \text{ confidence} = 74\%] \quad (1)$$

This rule expresses a relation between *popcorn* and *beer* namely “customers who purchase popcorn also purchase beer in the same transaction”. The *support* and *confidence* are two measures used to show the rule interestingness. The support of 7% for rule presented above states that *popcorn* and *beer* appeared together in 7% of all recorded transactions. The confidence measure describes the chance that there is beer in a transaction provided that there is also *popcorn*. In this case, 74% of all transactions involving *popcorn* also involved *beer*.

Such rules provide a source of information for retailers to develop marketing strategies, catalog design and store layout. Thus, the company can decide that items frequently purchased together to be placed in proximity in order to encourage the sale of such items together. Other strategy is placing the items frequently purchased together at opposite ends of the store in order to attract customers that purchase such items to pick up other items along the way.

The association rules can be categorized in different ways as follows (Han & Kamber, 2006):

1. Based on the types of values handled in the rule:
 - a. *Boolean (crisp) association rules*: the association is between the presence or absence of items. Rule (1) is an example of such a Boolean association rule.

- b. *Quantitative association rules*: the rules describe associations between quantitative items or attributes. Here the attribute values are partitioned into intervals. An example of such rule is the following:

$$\begin{aligned} &age(X, "20...30") \wedge income(X, "2000...3000") \\ &\Rightarrow cars(X, "0..1") \end{aligned} \quad (2)$$

- c. *Categorical association rules*: the attributes referred by rules are categorical, such as *occupation* or *sex*. An example of such a rule is the following:

$$\begin{aligned} &occupation(X, "student") \wedge sex(X, "male") \\ &\Rightarrow plays(X, "basket") \end{aligned} \quad (3)$$

2. Based on the number of data dimension involved in the rule:

- a. *Single-dimensional rules*: the items or attributes from rules refer only one dimension. An example of such rule is the following:

$$\begin{aligned} &buys(X, "beer") \wedge buys(X, "cola") \\ &\Rightarrow buys(X, "popcorn") \end{aligned} \quad (4)$$

The dimension referred is buys.

- b. *Multidimensional rules*: the rules refer to two or more attributes. The rule (2) is an example of multidimensional rule. The dimensions referred to are age, income and cars.

3. Based on the abstraction levels of data involved in the rule:

- a. *Single-level rules*: the rules refer attributes or items at the same level of abstraction, i.e., at the lowest level in

25 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/chapter/mining-association-rules/35878

Related Content

Mining Association Rules

Mihai Gabroveanu (2009). *Handbook of Research on Emerging Rule-Based Languages and Technologies: Open Solutions and Approaches* (pp. 647-673).

www.irma-international.org/chapter/mining-association-rules/35878

Using Device Detection Techniques in M-Learning Scenarios

Ricardo Queirós and Mário Pinto (2013). *Innovations in XML Applications and Metadata Management: Advancing Technologies* (pp. 118-134).

www.irma-international.org/chapter/using-device-detection-techniques-learning/73176

Modeling Temporal Information With JSON

Zhangbing Hu and Li Yan (2019). *Emerging Technologies and Applications in Data Processing and Management* (pp. 134-153).

www.irma-international.org/chapter/modeling-temporal-information-with-json/230687

Version Control of Software Models

Marcus Alanen and Ivan Porres (2005). *Advances in UML and XML-Based Software Evolution* (pp. 47-70).

www.irma-international.org/chapter/version-control-software-models/4930

Specification of Business Components Using Temporal OCL

Stefan Conrad and Klaus Turowski (2003). *UML and the Unified Process* (pp. 48-65).

www.irma-international.org/chapter/specification-business-components-using-temporal/30537