Chapter 4

An Analytical Survey of Current Approaches to Mining Logical Rules from Data

Xenia Naidenova
Military Medical Academy, the Russian Federation

ABSTRACT

An analytical survey of some efficient current approaches to mining all kind of logical rules is presented including implicative and functional dependencies, association and classification rules. The interconnection between these approaches is analyzed. It is demonstrated that all the approaches are equivalent with respect to using the same key concepts of frequent itemsets (maximally redundant or closed itemset, generator, non-redundant or minimal generator, classification test) and the same procedures of their lattice structure construction. The main current tendencies in developing these approaches are considered.

INTRODUCTION

Our objectives, in this chapter, are the following ones:

1. To give an analytical survey and comparison of existing and most effective approaches for mining all kinds of logical rules (implicative, association rules and functional dependencies) in the following frameworks: Apriori-like search, Formal Concept Analysis, closure operations of Galois connections, and Diagnostic Test Approach.

2. To show that all these approaches use the equivalent definitions of the key concepts in mining all kinds of logical rules: item, itemset, frequent itemset, maximal itemset, maximally redundant itemset, generator, minimal generator (non-redundant or irredundant itemset), closed itemset, support, and confidence.

3. To consider all these approaches on the base of the same mathematical language (the lattice theory) and to analyze the interconnections between them.
4. To present the Diagnostic Test Approach (DTA) to mining logical rules. This approach is an integrated system of operations and methods capable to solve any kind of supervised symbolic machine learning problems including mining implications, association rules, and functional dependences both in incremental and non-incremental manner.

NOTATIONS AND BASIC CONCEPTS

Mining itemsets of different properties (as a basis of logical rule mining) is a core problem for several data mining applications as inferring association rules, implicative and functional dependencies, correlations, document classification and analysis, and many others, which are extensively studied. Moreover, databases are becoming increasingly larger, thus requiring a higher computing power to mine different itemsets in reasonable time.

We begin with the definitions of the main concepts of itemset mining: item, itemset, transaction, *tid*, and *tid*-set or *tid*-list. The definitions of these concepts go from database system applications.

By Lal, & Mahanti, 2010, the set $I = \{ i_1, i_2, \ldots i_m \}$ is a set of m distinct literals called items. Transaction is a set of items over I. Items may be products, special equipments, service options, objects, properties of objects, etc.

Any subset X of I is called an itemset. As an example of the itemset, it may be considered a set of products that can be bought (together). An example of customer purchase data as a set of itemsets is given in Table 1.

Huge amounts of customer purchase data are collected daily at the checkout counters of some supermarket. The data in Table 1 is commonly known as market basket transactions.

Each row corresponds to a transaction and each column corresponds to an item.

Traditionally, transaction is an itemset which is a record in a database. Transactions need not to be pair wise different. A transaction has an associated unique identifier called *tid*.

A transaction over an item set I can be considered as a pair $t = (tid, X)$, where *tid* is a unique transaction identifier and $X \subseteq I$ is an item set.

In general case, we can consider the set I of items as a set of all attributes' values that can appear in descriptions of some objects or situations, consequently, a transaction is a collection of attribute values composing a description of some object or situation. Table 2 gives an example of object descriptions.

More formally, let $I = \{i_1, i_2, \ldots, i_N\}$ be a set of distinct values of some object properties, called items.

A transaction database (TDB) is a set of transactions, where transaction $<tid, X>$ contains a set of items (i.e., $X \subseteq I$) and is associated with a unique transaction identifier *tid*.

A non-empty itemset $Y \subseteq I$ is called l-itemset if it contains l items.

An itemset $\{x_1, x_2, \ldots, x_n\}$ is also denoted as $x_1, x_2, \ldots, x_n$.

A transaction $<tid, X>$ is said to contain itemset Y if $Y \subseteq X$. Transaction X is said to support an itemset Y if $Y \subseteq X$.

The number of transactions in TDB containing itemset X is called the support of itemset X,

Table 1. An example of market basket transactions

TID	Itemsets (Transactions)
1	Bread, Milk
2	Bread, Diapers, Beer, Eggs
3	Milk, Diapers, Cola, Beer
4	Bread, Milk, Diapers, Beer
5	Bread, Milk, Coffee, Cheese
6	Bread, Milk, Butter, Coffee, Cakes

Related Content

Discovering Spatio-Textual Association Rules in Document Images
Donato Malerba, Margherita Berardiand Michelangelo Ceci (2008). *Data Mining Patterns: New Methods and Applications (pp. 176-197).*
www.irma-international.org/chapter/discovering-spatio-textual-association-rules/7565

Discovery of Emergent Sorting Behavior using Swarm Intelligence and Grid-Enabled Genetic Algorithms
Dimitris Kalles, Alexis Kaporis, Vassiliki Mperoukliand Anthony Chatzinouskas (2014). *Biologically-Inspired Techniques for Knowledge Discovery and Data Mining (pp. 249-271).*
www.irma-international.org/chapter/discovery-of-emergent-sorting-behavior-using-swarm-intelligence-and-grid-enabled-genetic-algorithms/110463

A Blind Signature-Based Approach for Cross-Domain Authentication in the Cloud Environment
Aniello Castiglione, Francesco Palmieri, Chin-Ling Chenand Yao-Chung Chang (2016). *International Journal of Data Warehousing and Mining (pp. 34-48).*
www.irma-international.org/article/a-blind-signature-based-approach-for-cross-domain-authentication-in-the-cloud-environment/143714

A Novel Approach Using Non-Synonymous Materialized Queries for Data Warehousing
Sonali Ashish Chakraborty (2021). *International Journal of Data Warehousing and Mining (pp. 22-43).*
www.irma-international.org/article/a-novel-approach-using-non-synonymous-materialized-queries-for-data-warehousing/286614

Resource Constrained Data Stream Clustering with Concept Drifting for Processing Sensor Data
Gansen Zhao, Zhongjie Ba, Jiahua Du, Xinming Wang, Ziliu Li, Chunming Rongand Changqin Huang (2015). *International Journal of Data Warehousing and Mining (pp. 49-67).*
www.irma-international.org/article/resource-constrained-data-stream-clustering-with-concept-drifting-for-processing-sensor-data/129524