Visual Analysis of User-Driven Association Rule Mining

Wei Chen, Cong Xie, Pingping Shang, Qunsheng Peng

PII: S1045-926X(17)30007-1
DOI: 10.1016/j.jvlc.2017.08.007
Reference: YJVLC 803

To appear in: *Journal of Visual Languages and Computing*

Received date: 14 January 2017
Revised date: 27 August 2017
Accepted date: 28 August 2017

Please cite this article as: Wei Chen, Cong Xie, Pingping Shang, Qunsheng Peng, Visual Analysis of User-Driven Association Rule Mining, *Journal of Visual Languages and Computing* (2017), doi: 10.1016/j.jvlc.2017.08.007

Visual Analysis of User-Driven Association Rule Mining

Wei Chen[a,1], Cong Xie[b,1,*], Pingping Shang[a], Qunsheng Peng[a]

[a]State Key Lab of CAD& CG Zhejiang University, Hangzhou, China
[b]Stony Brook University, New York, USA

Abstract

Association rules have been widely used for detecting relations between attribute-value pairs of categorical datasets. Existing solutions of mining interesting association rules are based on the support-confidence theory. However, it is non-trivial for the user to understand and modify the rules or the results of intermediate steps in the mining process, because the interestingness of rules might differ largely for various tasks and users. In this paper we reinforce conventional association rule mining process by mapping the entire process into a visualization assisted loop, with which the user workload for modulating parameters and mining rules is reduced, and the mining efficiency is greatly improved. A hierarchical matrix-based visualization technique is proposed for the user to explore the measure value and the intermediate results of association rules. We also design a set of visual exploration tools to support interactively inspection and manipulation of mining process. The effectiveness and usability of our approach is demonstrated with two scenarios.

Keywords:
Categorical Data, Association Rules, Visual Analysis

1. Introduction

1.1. Background

The association rules mining [1] is an effective method to find relationships between different attribute-value pairs in different datasets. For example, an online commercial company records a set of commodities in a customer's basket. It is critical for a sales manager to learn the co-occurred purchased items, such as a customer who buys bread is also likely to purchase butter. The sales manager can use the rules for decision making, recommendation and other marketing activities.

However, strong rules found by the automatic rule mining methods (e.g., support-confidence framework) are not necessarily interesting [2]. The judgment of whether a given rule is interesting may vary based on different factors:

- Interestingness is based on domain-specific knowledge and mining tasks. The interesting patterns of rules can differ largely under different usage scenarios. For example, diamond can have a very low sales amount, but analysts are still interested because of its high price. However, the traditional support-confidence framework will exclude this item due to its low purchasing frequency.

- Interestingness is based on structure of the data. Association may involve specific structures and sequences. For example, customers tend to buy headsets after buying a digital audio player. Rules with a desirable sequence are interesting in such cases.

- Interestingness can be subjective. The user assesses a rule based on his or her knowledge, experiences, and understanding of their tasks, which may differ from one to another.

To identify interesting rules, previous mining methods either allow the user to extend the existing association rules by developing different measures [3], or narrow the interesting rules by setting various constraints for certain attributes, itemsets, and rules [4] (e.g., specifying the interesting subset in database). However, the automatic mining process is actually hard to understand [5], which makes it challenging for defining measures and constraints. Meanwhile, most of these works do not provide methods for modulating the mining process interactively.

Existing visualization methods [6] [7] for rule mining only support basic interactions, such as searching and deleting rules. They are incapable of various tasks that require user expertise and domain knowledge, such as defining measures or modulating different types of constraints.

To deal with the above problems, a user-driven visual analysis pipeline [8] is proposed. It employs a visual-assisted rule mining pipeline that reinforces the iterative mining loop for understanding the learning process, which also enables the modulation of the process of mining interesting rules. However, scalability is one of the major challenges because it is difficult to visualize a large amount of rules in the screen space. Another problem is its usability since this method only deal with online transaction dataset.

In this paper, we extend the existing visual analysis of rule

*Corresponding author
Email addresses: chenwei@cad.zju.edu.cn (Wei Chen), coxie@cs.stonybrook.edu (Cong Xie)
[1]These authors contributed equally to this work.