Chapter V

Mining Frequent Patterns Using Self-Organizing Map

Fedja Hadzic, University of Technology Sydney, Australia
Tharam Dillon, University of Technology Sydney, Australia
Henry Tan, University of Technology Sydney, Australia
Ling Feng, University of Twente, The Netherlands
Elizabeth Chang, Curtin University of Technology, Australia

Abstract

Association rule mining is one of the most popular pattern discovery methods used in data mining. Frequent pattern extraction is an essential step in association rule mining. Most of the proposed algorithms for extracting frequent patterns are based on the downward closure lemma concept utilizing the support and confidence framework. In this chapter we investigate an alternative method for mining frequent patterns in a transactional database. Self-Organizing Map (SOM) is an unsupervised neural network that effectively creates spatially organized internal representations of the features and abstractions detected in the input space. It is one of the most popular clustering techniques, and it reveals existing similarities in the input space by per-
forming a topology-preserving mapping. These promising properties indicate that such a clustering technique can be used to detect frequent patterns in a top-down manner as opposed to the traditional approach that employs a bottom-up lattice search. Issues that are frequently raised when using clustering technique for the purpose of finding association rules are: (i) the completeness of association rule set, (ii) the support level for the rules generated, and (iii) the confidence level for the rules generated. We present some case studies analyzing the relationships between the SOM approach and the traditional association rule framework, and propose a way to constrain the clustering technique so that the traditional support constraint can be approximated. Throughout our experiments, we have demonstrated how a clustering approach can be used for discovering frequent patterns.

Introduction

With large amounts of data continuously collected and stored, organizations are interested in discovering associations within their databases for different industrial, commercial, or scientific purposes. The discovery of interesting associations can aid the decision-making process and provide the domain with new and invaluable knowledge. Association rule mining is one of the most popular data mining techniques used to discover interesting relationships among data objects present in a database. A large amount of research has gone toward the development of efficient algorithms for association rule mining (Agrawal, Imielinski, & Swami, 1993; Agrawal, Mannila, Srikant, Toivonen, & Verkamo, 1996; Feng, Dillon, Weigana, & Chang, 2003; Park, Chen, & Yu, 1997; Tan, Dillon, Feng, Chang, & Hadzic, 2005; Zaki, 2001; Zaki, Pathasarthy, Ogihara, & Li, 1997). The algorithms developed have different advantages and disadvantages when applied to different domains with varying complexity and data characteristics. Frequent pattern discovery is an essential step in association rule mining. Most of the developed algorithms are a variant of Apriori (Agrawal et al., 1996) and are based on the downward closure lemma concept with the support framework. Besides the fact that the problem is approached differently, commonality that remains in most of the current algorithms is that all possible candidate combinations are first enumerated, and then their frequency is determined by scanning of the transactional database. These two steps, candidate enumeration and testing, are known to be the major bottlenecks in the Apriori-like approaches, which inspired some work towards methods that avoid this performance bottleneck (Han, Pei, & Yin, 2000; Park, Chen, & Yu, 1995). Depending on the aim of the application, rules may be generated from the frequent patterns discovered, and the support and confidence of each rule can be indicated. Rules with high confidence often cover only small fractions of the total number of rules generated, which makes their isolation more challenging and costly. In general, the task of frequent patterns discovery can have problems with complex real-world data, as for a transaction
Mining for Mutually Exclusive Items in Transaction Databases
www.igi-global.com/article/mining-mutually-exclusive-items-transaction/1789?camid=4v1a