Chapter 9

Association Rule Mining in Collaborative Filtering

Carson K. Leung
University of Manitoba, Canada

Fan Jiang
University of Manitoba, Canada

Edson M. Dela Cruz
University of Manitoba, Canada

Vijay Sekar Elango
University of Manitoba, Canada

ABSTRACT

Collaborative filtering uses data mining and analysis to develop a system that helps users make appropriate decisions in real-life applications by removing redundant information and providing valuable to information users. Data mining aims to extract from data the implicit, previously unknown and potentially useful information such as association rules that reveals relationships between frequently co-occurring patterns in antecedent and consequent parts of association rules. This chapter presents an algorithm called CF-Miner for collaborative filtering with association rule miner. The CF-Miner algorithm first constructs bitwise data structures to capture important contents in the data. It then finds frequent patterns from the bitwise structures. Based on the mined frequent patterns, the algorithm forms association rules. Finally, the algorithm ranks the mined association rules to recommend appropriate merchandise products, goods or services to users. Evaluation results show the effectiveness of CF-Miner in using association rule mining in collaborative filtering.

INTRODUCTION

With the advances in technology, high volumes of high-veracity and valuable data can be easily generated or collected at a high velocity from large varieties of data sources in various real-life applications for both public services and private sectors. Embedded in these data—such as the financial data, retail data, social network data, and web data—are rich sets of useful information and knowledge. Hence, with rich sets of information, it is desirable to be able to remove redundant information and to be provided with valuable information in short periods of time. To this end, having a recommendation system would be helpful because such a recommendation system helps users make appropriate decisions in various real-life applications—such as e-commerce, e-learning, social networks, and web search—by remov-
ing redundant information and providing users with valuable information in short periods of time. For instance, a recommendation system helps business owners make business decisions by revealing their customers’ shopping behavior. Similarly, a recommendation system helps consumers by recommending goods, products, or services. As a third example, a recommendation system helps social network users connect to appropriate person or organizations by recommending friends or social communities (Lu & Lakshmanan, 2012; Maserrat & Pei, 2012; Leung, Medina & Tanbeer, 2013; Jiang & Leung, 2014; Leung, Tanbeer & Cameron, 2014).

With the exponential growth rate of data, collaborative filtering is commonly used to make recommendations. **Collaborative filtering**—which usually applies data mining techniques to develop a system with precise knowledge and accuracy for helping users—is a popular technique used for recommendation system. Here, **data mining** aims to search data for rich sets of implicit, previously unknown and useful information and knowledge embedded in the data. Common data mining tasks including classification, clustering, and association rule mining. **Classification** focuses on performing supervised learning from the historical/training data and categorizing future/test data with class labels describing the data. **Clustering** focuses on performing unsupervised learning and grouping similar data together. **Association rule mining** focuses on mining frequent patterns as well as forming rules in the form of A→C (where both A and C are frequent patterns) for revealing the associations or relationships between frequent patterns A and C. **Frequent pattern mining** (Aggarwal & Srikant, 1994; Leung et al., 2013; Leung, 2014b) focuses on discovering knowledge and useful information—in the form of patterns revealing frequently co-occurring items, events, or objects (e.g., frequently purchased merchandise items in shopper market basket, frequently co-located events)—from the data.

Since the introduction of the research problem of frequent pattern mining (Aggarwal & Srikant, 1994), numerous algorithms (Leung, 2013; Négrevergne et al., 2013; Cuzzocrea et al., 2014; Leung, MacKinnon & Jiang, 2014). Among them, the Apriori algorithm (Aggarwal & Srikant, 1994) is a notable one. However, as a breadth-first, level-wise bottom-up mining algorithm, Apriori requires many database scans—one scan for each level/cardinality of frequent patterns. The FP-growth algorithm (Han, Pei & Yin, 2000) improves efficiency by using a depth-first, tree-based approach. However, FP-growth improves efficiency at a price of requiring memory space to keep the following:

1. **The Global Frequent Pattern Tree (FP-Tree):** Captures the content of the original database.
2. **All Subsequent Sub-Trees:** Each captures the contents of subsequent projected databases.

Both the TD-FP-Growth algorithm (Wang et al., 2002) and the H-mine algorithm (Pei et al., 2001) avoid building and simultaneously keeping multiple FP-trees during the mining process. However, during the mining process, TD-FP-Growth keeps updating the global FP-tree by adjusting tree pointers, whereas H-mine keeps updating pointers/hyperlinks in the corresponding H-struct. As alternatives to the aforementioned “horizontal” frequent pattern mining algorithms (which use a transaction-centric approach to find which pattern is supported by or contained in a transaction), both the Eclat algorithm (Zaki, 2000) and the VIPER algorithm (Shenoy et al., 2000) mine frequent patterns “vertically” by using an item-centric approach to count the number of transactions supporting or containing the patterns. Take into account the advantages and disadvantages of the above algorithms, we present an algorithm that performs association rule mining for collaborative filtering. The objective of this chapter is to show how association rule mining can be applicable in collaborative filtering.