

Chapter 12

Blending Association Rules for Knowledge Discovery in Big Data

Ali Anaissi

The University of Sydney, Australia

Madhu Goyal

The University of Technology Sydney, Australia

ABSTRACT

Data mining techniques have been widely applied in several domains to support a variety of business-related applications such as market basket analysis. For instance, basket market transaction accumulate large amounts of customer purchase data from their day-to-day operations. This paper delivers a strategy for the implementation of a systematic analysis framework built on the established principles used in data mining and machine learning areas. We employ Apriori and FP-growth algorithms coupled with support vector machine to implement our recommendation systems. Experiments are done using a real market dataset and the 0.632+ bootstrap method is used here in order to evaluate our framework. The obtained results suggest that the proposed framework will be able to generate benefits for grocery chain using a real-world grocery store data. FP-growth algorithm shows better performance over Apriori in terms of time complexity.

INTRODUCTION

The mining of data collection has received a lot of interests in several domains such as market, financial and biomedical (Rajendran & Madheswaran, 2010; Chen & Jaggi, 2001; Kamley, Jaloree, & Thakur, 2014). The aim of that is to discover knowledge patterns hidden in large data sets that can yield more understanding to the data holders and identify new opportunities for imperative tasks including strategic planning and decision making. One methodology of mining complex dataset is determining the association rules which mainly used in the analysis of the market basket data (Kamley, Jaloree, & Thakur, 2014; Kasthuri & Meyyappan, 2013). The main purpose of that is to find connections existing among

DOI: 10.4018/978-1-5225-6023-4.ch012

the items that can assist retailer with distinguishing new open doors for cross-offering their items to customers. This area of data mining i.e. association rules has received a great deal of interest in the field of market basket analysis. Figuring out what items clients are liable to purchase together could be extremely helpful for products arrangement and promotions (Shim, Choi, & Suh, 2012). The rationale behind that is to find hidden relationships between the frequent items in the presence baskets to generate association rules from these items.

Hence, basket market data needs to be systematically analyzed such that deriving the association rules and presented in a manner such that it will provide 'actionable knowledge' for the market's analyst. Association rule mining is an important task and a key issue in knowledge discovery and data mining. This has led to the growth of the tremendous research in data mining such as correlation mining, associative classification, and frequent pattern based clustering. It has proven to be quite important for handling product layout based business problems, such as goods promotion strategy and correlation product recommendation. For example, association rule mining is widely employed in retail industry to discover interesting association rules to help with better decision making. The objective of association mining is the elicitation of useful or interesting rules from which new knowledge can be derived. Initial research was largely motivated by the analysis of market basket data, which allowed companies to understand customers' buying behavior and to target more or right customers. Thus the association mining's rule in this process is to facilitate the discovery and enable the rules for subsequent interpretation by the user to determine their usefulness.

This chapter is an extension of the aforementioned work (Anaissi & Goyal, 2015). In combining with the Apriori algorithm (Agrawal & Srikant, 1994), this paper uses the Frequent Pattern-growth (FP-growth) algorithm (Han, Pei, & Yin, 2000) for generating the association rules. The rest of this paper is organized as follows. Section 2 provides a short overview about association rules and Section 3 introduces the related work of association rules. Section 4 presents the methods and algorithms used in this work to achieve the desired results. Algorithms of clusters generation, Apriori and support vector machine (SVM) are discussed in detail showing that how the association rules are generated in Apriori algorithm and how the classification model is built in SVM. Section 5 presents the experiments and discusses the obtained results and calculates the accuracy of classification performance. Section 6 draws a conclusion about the methods we applied and the results we achieved by our proposed framework.

Association Rules Technique

The association rules problem can be defined as per the following, let $I = i_1, i_2, i_3 \dots i_n$ is a set of items, and I is the dataset containing all the transactions. Each record in this dataset represents one transaction T having a set of objects such that $T \subseteq I$. Let A, B be a set of items such that $A, B \subseteq I$. A suggested association rule can be written in the structure $A \Rightarrow B$, where $A \subset I, B \subset I, A \cap B = \varphi$ (Agrawal & Srikant, 1994). A real typical example of that can be represented in the following statement:

{Peanut butter, Jelly} \Rightarrow {Bread}

This simple association rule says that Bread is likely to be bought if peanut butter and jelly are purchased. The items surrounded by brackets are called itemsets. With small datasets, human beings are able to find interesting connections in small datasets and can build association rules. However, this task becomes a significant challenge in the case of extremely complex and large transactional datasets with a

16 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/blending-association-rules-for-knowledge-discovery-in-big-data/214815

Related Content

Reliable Routing Protocols in VANETs

Gongjun Yan, Danda B. Rawat and Samy El-Tawab (2010). *Advances in Vehicular Ad-Hoc Networks: Developments and Challenges* (pp. 200-213).

www.irma-international.org/chapter/reliable-routing-protocols-vanets/43171

Certifications in Cybersecurity Workforce Development: A Case Study

Ping Wang and Hubert D'Cruze (2019). *International Journal of Hyperconnectivity and the Internet of Things* (pp. 38-57).

www.irma-international.org/article/certifications-in-cybersecurity-workforce-development/241804

Ventilator Production in Turkey Amid Pandemic Through Service-Dominant Logic Lenses

Aslisah Senak and Nuray Girginer (2022). *Handbook of Research on Digital Innovation and Networking in Post-COVID-19 Organizations* (pp. 115-136).

www.irma-international.org/chapter/ventilator-production-in-turkey-amid-pandemic-through-service-dominant-logic-lenses/307538

Comparison of Policies for Epidemic Broadcast in DTNs under Different Mobility Models

Francesco Giudici, Elena Pagani and Gian Paolo Rossi (2010). *Networking and Telecommunications: Concepts, Methodologies, Tools, and Applications* (pp. 805-819).

www.irma-international.org/chapter/comparison-policies-epidemic-broadcast-dtns/49777

Investigating the Performance of the TSS Scheme in Noisy MANETs

Hussein Al-Bahadili, Shakir M. Hussain, Ghassan F. Issa and Khaled El-Zayyat (2012). *Simulation in Computer Network Design and Modeling: Use and Analysis* (pp. 315-335).

www.irma-international.org/chapter/investigating-performance-tss-scheme-noisy/63290