



Chapter XIII

Visual Grouping of Association Rules by Clustering Conditional Probabilities for Categorical Data

Sasha Ivkovic, University of Ballarat, Australia

Ranadhir Ghosh, University of Ballarat, Australia

John Yearwood, University of Ballarat, Australia

Abstract

We demonstrate the use of a visual data-mining tool for non-technical domain experts within organizations to facilitate the extraction of meaningful information and knowledge from in-house databases. The tool is mainly based on the basic notion of grouping association rules. Association rules are useful in discovering items that are frequently found together. However in many applications, rules with lower frequencies are often interesting for the user. Grouping of association rules is one way to overcome the rare item problem. However some groups of association rules are too large for ease of understanding. In this chapter we propose a method for clustering categorical data

based on the conditional probabilities of association rules for data sets with large numbers of attributes. We argue that the proposed method provides non-technical users with a better understanding of discovered patterns in the data set.

Introduction

Traditional manual data analysis is becoming impractical in many domains as data volume grows exponentially. Depending on the type of analysis, several Knowledge Discovery in Databases (KDD) methods such as classification, regression, clustering and association rules use automated artificial intelligence, and mathematical and statistical techniques for the task. Frawley, Piatetsky-Shapiro, and Matheus (1992, p. 57) define KDD as “the non-trivial extraction of implicit, previously unknown and potentially useful information from the data.” The overall process of finding and interpreting patterns from data involves repeated application of the following steps — data selection, data pre-processing, data transformation, data mining, discovery interpretation / evaluation. KDD is a process involving human interactions. There are traditionally two human roles in any KDD process — a domain expert role and a data miner role. A data miner is someone who primarily uses sophisticated KDD technology in conjunction with existing data sources as the basis for discovering useful patterns in the data. A domain expert is a person with a comprehensive knowledge of a certain domain. However, some domain experts are non-data miners (e.g., data analyst and database administrator) and some are data miners (e.g., external KDD specialist).

A data miner can gather knowledge either from the domain expert or from a domain knowledge repository. However the domain expert is fully dependent on the data miner. One of the reasons that exploitation of KDD technology is not fully implemented within organizations is that the majority of KDD tools currently available are expensive and complex adjuncts to database management systems. Their operation typically requires specialist operators. Furthermore the countless data-mining techniques function in such different ways that even KDD experts cannot be expected to be proficient with all approaches. The specialist knowledge required and the cost of KDD tools militate against their use for non-technical domain experts. We argue that by implementing a balanced relationship between the domain expert and the data miner, the organization will benefit by exploiting its KDD technology at all organizational levels. To explain this further we consider the implementation of KDD technology in an organization.

Goebel and Gruenwald (1999) investigated the use of almost all mainstream commercial KDD products and reported that deploying KDD technology in an organization is traditionally implemented through the following three phases:

1. First, KDD studies are performed by the data-mining specialists (external consultants). (See Figure 1.)
2. Once the profitability of KDD is proven, data-analysis experts apply the KDD techniques (possibly with the help of a domain expert who has strong domain knowledge). (See Figure 2.)

17 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/visual-grouping-association-rules-clustering/6028

Related Content

A Model for Estimating the Benefits from Network Security Systems

Soumyo D. Moitra (2014). *International Journal of Business Analytics* (pp. 1-20).

www.irma-international.org/article/a-model-for-estimating-the-benefits-from-network-security-systems/117546

AI-Driven Financial Forecasting: The Power of Soft Computing

Nitendra Kumar, Priyanka Agarwal, Gauri Gupta, Sadhana Tiwari and Padmesh Tripathi (2024). *Intelligent Optimization Techniques for Business Analytics* (pp. 146-170).

www.irma-international.org/chapter/ai-driven-financial-forecasting/344520

A Comparative Study on Medical Diagnosis Using Predictive Data Mining: A Case Study

Seyed Jalaeddin Mousavirad and Hossein Ebrahimpour-Komleh (2016). *Business Intelligence: Concepts, Methodologies, Tools, and Applications* (pp. 923-954).

www.irma-international.org/chapter/a-comparative-study-on-medical-diagnosis-using-predictive-data-mining/142659

Online Advertisement Using Web Analytics Software: A Comparison Using AHP Method

Manu Sharma and Sudhanshu Joshi (2020). *International Journal of Business Analytics* (pp. 13-33).

www.irma-international.org/article/online-advertisement-using-web-analytics-software/246026

Innovation Capability in High-Tech Companies: Exploring the Role of Organizational Culture and Empowerment

Alper Ertürk (2012). *Managing Dynamic Technology-Oriented Businesses: High-Tech Organizations and Workplaces* (pp. 228-252).

www.irma-international.org/chapter/innovation-capability-high-tech-companies/67438