Chapter III
Combining Data-Driven and User-Driven Evaluation Measures to Identify Interesting Rules

Solang Oliveira Rezende
University of São Paulo, Brazil

Edson Augusto Melanda
Federal University of São Carlos, Brazil

Magaly Lika Fujimoto
University of São Paulo, Brazil

Roberta Akemi Sinoara
University of São Paulo, Brazil

Veronica Oliveira de Carvalho
University of Oeste Paulista, Brazil

ABSTRACT

Association rule mining is a data mining task that is applied in several real problems. However, due to the huge number of association rules that can be generated, the knowledge post-processing phase becomes very complex and challenging. There are several evaluation measures that can be used in this phase to assist users in finding interesting rules. These measures, which can be divided into data-driven (or objective measures) and user-driven (or subjective measures), are first discussed and then analyzed for their pros and cons. A new methodology that combines them, aiming to use the advantages of each kind of measure and to make user’s participation easier, is presented. In this way, data-driven measures
can be used to select some potentially interesting rules for the user’s evaluation. These rules and the knowledge obtained during the evaluation can be used to calculate user-driven measures, which are used to aid the user in identifying interesting rules. In order to identify interesting rules that use our methodology, an approach is described, as well as an exploratory environment and a case study to show that the proposed methodology is feasible. Interesting results were obtained. In the end of the chapter tendencies related to the subject are discussed.

**INTRODUCTION**

It can be said that the objective of the data mining process is to find knowledge from a group of data to be used in a decision making process. Therefore, it is important that the discovered knowledge is comprehensible and interesting to the final users. However, from the user’s point of view, one of the main problems is the difficulty of understanding the extracted models. These models can be very complex or difficult to understand by the domain experts. In the case of association rules, the fundamental issue in the analysis and interpretation of the extracted knowledge is the great number of patterns that makes the manual interpretation infeasible. Besides, few of these patterns are really interesting to the user.

One of the most used techniques in the evaluation of rules and search for interesting ones is the application of knowledge evaluation measures (Natarajan & Shekar, 2005). These measures are usually classified as data-driven or user-driven. The data-driven measures (or objective measures) depend exclusively on the pattern structure and the data used in the process of knowledge extraction, while the user-driven measures (or subjective measures) depend fundamentally on the final user’s interest and/or needs. Therefore, the data-driven measures are more general and independent of the domain in which the data mining process is carried out. These measures can be insufficient to identify the interesting rules, because the objectives and the specialists’ knowledge are not considered. Although the user-driven measures consider these factors, there can be some limitations in their use. A specialist can have difficulty in expressing himself/herself when supplying a system with his/her knowledge and interests that uses the user-driven measures. A good alternative to aid in the identification of interesting rules can be the combined use of data-driven and user-driven measures, exploiting the advantages of each type. This combination can facilitate the domain’s (or the final user’s) participation in the knowledge post-processing phase and allow a more efficient identification of interesting association rules than with the use of just one measure type.

In this context, the objective of this chapter is to discuss the evaluation of rules using data-driven and user-driven measures by presenting the advantages, difficulties and limitations in their use. Then, a methodology to identify interesting association rules is presented. This methodology combines the use of data-driven and user-driven measures, focusing strongly on the interaction with the expert user. Using this methodology it is possible to overcome some of the difficulties and limitations found in the post-processing of association rules.

Considering the established objective, a literature review is initially presented in the background section to discuss the challenges of using data mining and the association technique in real databases. Then the evaluation of rules is covered by evaluating the interestingness of these rules, considering the objective and subjective aspects. Some of the main techniques used in the identification of interesting rules are introduced, among them the application of knowledge evaluation