Chapter XI
Meta-Knowledge Based Approach for an Interactive Visualization of Large Amounts of Association Rules

Sadok Ben Yahia
Faculty of Sciences of Tunis, Tunisia

Olivier Couturier
Centre de Recherche en Informatique de Lens (CRIL), France

Tarek Hamrouni
Faculty of Sciences of Tunis, Tunisia

Engelbert Mephu Nguifo
Centre de Recherche en Informatique de Lens (CRIL), France

ABSTRACT

Providing efficient and easy-to-use graphical tools to users is a promising challenge of data mining, especially in the case of association rules. These tools must be able to generate explicit knowledge and, then, to present it in an elegant way. Visualization techniques have shown to be an efficient solution to achieve such a goal. Even though considered as a key step in the mining process, the visualization step of association rules received much less attention than that paid to the extraction step. Nevertheless, some graphical tools have been developed to extract and visualize association rules. In those tools, various approaches are proposed to filter the huge number of association rules before the visualization step. However both data mining steps (association rule extraction and visualization) are treated separately in a one way process. Recently different approaches have been proposed that use meta-knowledge to guide the user during the mining process. Standing at the crossroads of Data Mining and Human-Computer
Meta-Knowledge Based Approach for an Interactive Visualization of Large Amounts of Association Rules

Interaction, those approaches present an integrated framework covering both steps of the data mining process. This chapter describes and discusses such approaches. Two approaches are described in details: the first one builds a roadmap of compact representation of association rules from which the user can explore generic bases of association rules and derive, if desired, redundant ones without information loss. The second approach clusters the set of association rules or its generic bases, and uses a fisheye view technique to help the user during the mining of association rules. Generic bases with their links or the associated clusters constitute the meta-knowledge used to guide the interactive and cooperative visualization of association rules.

INTRODUCTION

Data mining techniques have been proposed and studied to help users better understand and scrutinize huge amounts of collected and stored data. In this respect, extracting association rules has grasped the interest of the data mining community. Thus, the last decade has been marked by a determined algorithmic effort to reduce the computation time of the interesting itemset extraction step. The obtained success is primarily due to an important programming effort combined with strategies for compacting data structures in main memory. However, it seems obvious that this frenzied activity loses sight of the essential objective of this step, i.e., extracting a reliable knowledge, of exploitable size for users. Indeed, the unmanageably large association rule sets compounded with their low precision often make the perusal of knowledge ineffective, their exploitation time-consuming and frustrating for users. Moreover, unfortunately, this teenaged field seems to provide results in the opposite direction with the evolving “knowledge management” topic.

The commonly generated thousands and even millions of high-confidence rules – among which many are redundant (Bastide et al., 2000; Ben Yahia et al., 2006; Stumme et al., 2001; Zaki, 2004) – encouraged the development of more acute techniques to limit the number of reported rules, starting by basic pruning techniques based on thresholds for both the frequency of the represented pattern and the strength of the dependency between premise and conclusion. Moreover, this pruning can be based on patterns defined by the user (user-defined templates), on Boolean operators (Meo et al., 1996; Ng et al., 1998; Ohsaki et al., 2004; Srikant et al., 1997). The number of rules can be reduced through pruning based on additional information such as a taxonomy on items (Han, & Fu, 1995) or on a metric of specific interest (Brin et al., 1997) (e.g., Pearson’s correlation or \( \chi^2 \)-test). More advanced techniques that produce only lossless information limited number of the entire set of rules, called generic bases (Bastide et al., 2000). The generation of such generic bases heavily draws on a battery of results provided by formal concept analysis (FCA) (Ganter & Wille, 1999). This association rule reduction can be seen as a “sine qua non” condition to avoid that the visualization step comes up short in dealing with large amounts of rules. In fact, the most used kind of visualization categories in data mining is the use of visualization techniques to present the information caught out from the mining process. Graphical visualization tools became more appealing when handling large data sets with complex relationships, since information presented in the form of images is more direct and easily understood by humans (Buono & Costabile, 2004; Buono et al., 2001). Visualization tools allow users to work in an interactive environment with ease in understanding rules.

Consequently, data mining techniques gain in presenting discovered knowledge in an interactive, graphical form that often fosters new insights.