Chapter 7
User-Centered Interactive Data Mining

Yan Zhao
University of Regina, Canada

Yiyu Yao
University of Regina, Canada

ABSTRACT

While many data mining models concentrate on automation and efficiency, interactive data mining models focus on adaptive and effective communications between human users and computer systems. User requirements and preferences play an important role in human-machine interactions, and guide the selection of knowledge representations, knowledge discovery operations and measurements, combined with explanations of mined patterns. This chapter discusses these fundamental issues based on a user-centered three-layer framework of interactive data mining.

INTRODUCTION

Data mining is featured by applying computer technologies to carry out nontrivial calculations for many important tasks, such as description, prediction and explanation of data. Computer systems can maintain precise operations under heavy information load, and maintain steady performance. Without the aid of computer systems, it is very difficult for people to aware, extract, memorize, search and retrieve knowledge in large and separate datasets, to interpret and evaluate data and information that are constantly changing, and to make recommendations or predictions in the face of inconsistent and incomplete data. It is true that computer technologies have freed humans from many time-consuming and labour-intensive activities. However, full automation of cognitive functions such as decision making, planning, and creative thinking remains human’s job. Implementations and applications of computer systems reflect requests and preferences of human users, and contain certain human heuristics. Computer systems must rely on human users to set goals, select alternatives if original approach fails, participate in unanticipated
emergencies and novel situations, and develop innovations in order to preserve safety, avoid expensive failure, or increase product quality (Elm et al. 2004; Hancock and Scallen, 1996; Shneiderman, 1998).

According to the above observations, we believe that interactive systems are required for data mining tasks. Though human-machine interaction has been emphasized for many disciplines, it did not receive enough attention in the domain of data mining until recently (Brachmann and Anand, 1996; Han, Hu and Cercone, 2003; Zhao, 2007; Zhao and Yao, 2005). Generally, an interactive data mining system is an integration of a human user and a computer. They can communicate and exchange information and knowledge. Through interaction and communication, computers and users can divide the labours in order to achieve a good balance of automation and human control. Computers are used to retrieve and keep track of large volumes of data, and to carry out complex mathematical or logical operations. Users can avoid routinized, tedious, and error-prone tasks, concentrate on critical decisions, planning, and cope with unexpected situations (Elm et al. 2004; Shneiderman, 1998). Moreover, interactive data mining can encourage learning, improve insights and understandings of the domain, stimulate the exploration of creative possibilities, and help users to solve particular problems. Users’ feedback can be used to improve the system. The interaction is mutual beneficial. A foundation of human-computer interaction may be provided by cognitive informatics (Wang, 2002; Wang, 2003; Wang, 2004).

For conceptually modelling data mining, Yao (2003) proposed a three-layered framework consisting of the philosophy layer, the technique layer, and the application layer. The main objective of this chapter is to extend the framework for interactive data mining by introducing the notion of user preference and judgement. Within this new user-centered framework, we revisit and summarize our recent studies on data mining regarding the three layers. The study of different decision logic languages enables the definition of granules and concepts at the philosophy layer (Yao, 2003). The study of rule interestingness measures reveals the relationships among granules and concepts in the philosophy layer, and facilitates the discovery of interesting patterns in the technique layer (Yao, Chen and Yang, 2003; Yao and Zhong, 1999, Zhong, Yao and Ohshima, 2003; Zhong, Yao, Ohshima and Ohsuga, 2001). The study of different knowledge discovery strategies is essential for the technique layer (Yao, Wang, Wang and Zeng, 2005; Yao and Wong, 1992). Finally, the study of explanation-oriented data mining demonstrates the importance of having user involvement before and inside the application layer (Yao, Zhao and Maguire, 2003). The synthesis of the existing results leads to a high-level understanding of interactive data mining, as well as new insights to the potential of human-machine interaction in the design of interactive data mining systems.

A FRAMEWORK OF INTERACTIVE DATA MINING

A three-layered conceptual framework represents the understanding, discovery, and utilization of knowledge, and can be extended to a user-centered conceptual framework for interactive data mining by introducing the notion of user preference.

Modelling User Preference

User preference can be expressed in various forms. Quantitative judgement involves the assignment of different weights to different entities. Qualitative judgement is expressed as an ordering of entities. In many situations, user judgement is determined by semantic considerations. For example, it may be