User-Centered Interactive Data Mining

Yan Zhao, University of Regina, Canada
Yaohua Chen, 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 the most important roles in human-machine interactions, and guide the selection of target knowledge representations, operations, and measurements. Practically, user requirements and preferences also decide strategies of abnormal situation handling, and explanations of mined patterns. In this article, we discuss these fundamental issues based on a user-centered three-layer framework of interactive data mining.

Keywords: concept formation; data mining; human-machine interaction; objective; subjective rule-interest measures; user preference

INTRODUCTION

Exploring and extracting knowledge from data is one of the fundamental problems in science. Many methods have been proposed and extensively studied, such as database management, statistics, machine learning and data mining. Particularly, data mining takes up many important tasks, such as description, prediction and explanation of data.

Data mining is featured by applying computer technologies to carry out nontrivial calculations. 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, 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 & Cook, Greitzer, Hoffman, Moon, & Hutchins, 2004; Hancock & 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 & Anand, 1996; Han, Hu & Cercone, 2003; Zhao, 2007; Zhao & 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. A foundation of human-computer interaction may be provided by cognitive informatics (Wang, 2002, 2004; Wang & Liu, 2003).

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 & Cook, 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.

For conceptually modeling 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 article is to extend the framework for interactive data mining. In particular, we introduce 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 philosophy layer, the technique layer, and the application layer, respectively. 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 & Yang, 2003; Yao & Zhong, 1999, Zhong, Yao & Ohshima, 2003; Zhong, Yao, Ohshima & Ohsuga, 2001). The study of user preferences provides a formal model for involving user’s judgement into the whole data mining process. The study of different strategies for abnormal situation handling is essential for the technique layer (Yao, Wang, Wang & Zeng, 2005; Yao & 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 & Maguire, 2003). This 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 viable interactive data mining systems.

**A FRAMEWORK OF INTERACTIVE DATA MINING**

A three-layered conceptual framework of data mining is proposed by Yao (2003), which consists of the philosophy layer, the technique layer and the application layer. The layered framework represents the understanding, discovery, and utilization of knowledge. The philosophy layer investigates the essentials of knowledge. One attempts to answer the fundamental question, namely, what is knowledge? There are many related issues to this question, such as the representation of knowledge, the expression and communication of knowledge in languages, the relationship between knowledge in the mind and in the external real world, and the classification and organization of knowledge. The technique layer is the study of knowledge discovery by
Related Content

Cognitive Memory for Semantic Agents Architecture in Robotic Interaction
Sébastien Dourlens and Amar Ramdane-Cherif (2013). *Cognitive Informatics for Revealing Human Cognition: Knowledge Manipulations in Natural Intelligence* (pp. 82-97).
[www.igi-global.com/chapter/cognitive-memory-semantic-agents-architecture/72284?camid=4v1a](www.igi-global.com/chapter/cognitive-memory-semantic-agents-architecture/72284?camid=4v1a)

Approximations in Rough Sets vs Granular Computing for Coverings
[www.igi-global.com/article/approximations-rough-sets-granular-computing/43878?camid=4v1a](www.igi-global.com/article/approximations-rough-sets-granular-computing/43878?camid=4v1a)

Technosocial Space: Connecting People and Places
Anne Sofie Laegran (2009). *Exploration of Space, Technology, and Spatiality: Interdisciplinary Perspectives* (pp. 54-69).
[www.igi-global.com/chapter/technosocial-space-connecting-people-places/18676?camid=4v1a](www.igi-global.com/chapter/technosocial-space-connecting-people-places/18676?camid=4v1a)
Interactive Classification Using a Granule Network
Yan Zhao and Yiyu Yao (2009). *Novel Approaches in Cognitive Informatics and Natural Intelligence* (pp. 235-245).
www.igi-global.com/chapter/interactive-classification-using-granule-network/27311?camid=4v1a