Interactive Visual Data Mining

Shouhong Wang  
University of Massachusetts Dartmouth, USA

Hai Wang  
Saint Mary’s University, Canada

INTRODUCTION

In the data mining field, people have no doubt that high level information (or knowledge) can be extracted from the database through the use of algorithms. However, a one-shot knowledge deduction is based on the assumption that the model developer knows the structure of knowledge to be deducted. This assumption may not be invalid in general. Hence, a general proposition for data mining is that, without human-computer interaction, any knowledge discovery algorithm (or program) will fail to meet the needs from a data miner who has a novel goal (Wang, S. & Wang, H., 2002). Recently, interactive visual data mining techniques have opened new avenues in the data mining field (Chen, Zhu, & Chen, 2001; de Oliveira & Levkowitz, 2003; Han, Hu & Cercone, 2003; Shneiderman, 2002; Yang, 2003).

Interactive visual data mining differs from traditional data mining, standalone knowledge deduction algorithms, and one-way data visualization in many ways. Briefly, interactive visual data mining is human centered, and is implemented through knowledge discovery loops coupled with human-computer interaction and visual representations. Interactive visual data mining attempts to extract unsuspected and potentially useful patterns from the data for the data miners with novel goals, rather than to use the data to derive certain information based on a priori human knowledge structure.

BACKGROUND

A single generic knowledge deduction algorithm is insufficient to handle a variety of goals of data mining since a goal of data mining is often related to its specific problem domain. In fact, knowledge discovery in databases is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns of data mining (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). By this definition, two aspects of knowledge discovery are important to meaningful data mining. First, the criteria of validity, novelty, usefulness of knowledge to be discovered could be subjective. That is, the usefulness of a data pattern depends on the data miner and does not solely depend on the statistical strength of the pattern. Second, heuristic search in combinatorial spaces built on computer and human interaction is useful for effective knowledge discovery. One strategy for effective knowledge discovery is the use of human-computer collaboration.

One technique used for human-computer collaboration in the business information systems field is data visualization (Bell, 1991; Montazami & Wang, 1988) which is particularly relevant to data mining (Keim & Kriegel, 1996; Wang, 2002). From the human side of data visualization, graphics cognition and problem solving are the two major concepts of data visualization. It is a commonly accepted principle that visual perception is compounded out of processes in a way which is adaptive to the visual presentation and the particular problem to be solved (Kosslyn, 1980; Newell & Simon, 1972).

MAIN THRUST

Major components of interactive visual data mining and their functions that make data mining more effective are the current research theme in this field. Wang, S. and Wang, H. (2002) have developed a model of interactive visual data mining for human-computer collaboration knowledge discovery. According to this model, an interactive visual data mining system has three components on the computer side, besides the database: data visualization instrument, data and model assembly, and human-computer interface.

Data Visualization Instrument

Data visualization instruments are tools for presenting data in human understandable graphics, images, or animation. While there have been many techniques for data visualization, such as various statistical charts with colors and animations, the self-organizing maps (SOM) method based on Kohonen neural network (Kohonen, 1989) has become one of the promising techniques of data visualization in data mining. SOM is a dynamic system that can learn the topological relations and abstract struc-