Chapter 3
Reducing Cognitive Overload by Meta-Learning Assisted Algorithm Selection

Lisa Fan
University of Regina, Canada

Minxiao Lei
University of Regina, Canada

ABSTRACT

With the explosion of available data mining algorithms, a method for helping user to select the most appropriate algorithm or combination of algorithms to solve a given problem and reducing users’ cognitive overload due to the overloaded data mining algorithms is becoming increasingly important. This chapter presents a meta-learning approach to support users automatically selecting most suitable algorithms during data mining model building process. The authors discuss the meta-learning method in detail and present some empirical results that show the improvement that can be achieved with the hybrid model by combining meta-learning method and Rough Set feature reduction. The redundant properties of the dataset can be found. Thus, the ranking process can be sped up and accuracy can be increased by using the reduct of the properties of the dataset. With the reduced searching space, users’ cognitive load is reduced.

The explosion in the amount of available data on any given subject has led researchers to the area of knowledge discovery and data mining. The main motivation of these research areas is that humans are not capable of analyzing the current size of the available data either manually or with basic statistical methods. As a result, the technological challenge of performing everything automatically has dominated the interests of researchers and developers. Thus, data mining was established as a methodology for extracting potentially useful information from very large amounts of data.

To deal with different complicated data, scientists have developed numerous data mining algorithms. These different data mining algorithms work well on different kinds of data. For example, Neural Network Algorithms (Wasserman, 1989) can deal with discrete data, Data Association Rule Algorithms
Reducing Cognitive Overload by Meta-Learning Assisted Algorithm Selection

(Kotsiantis & Kanellopoulos, 2006) can find groups of common items in transactions, and Clustering Algorithms (Kaufman & Rousseeuw, 1900) can both group similar items and deal with discrete data. Not a single algorithm can perform well on different data. As well, when different data mining algorithms are performed on the same data, the outputs are different. Thus, choosing the right algorithm to use for a specific type of data can be a challenge.

The Data Mining algorithm is the mechanism that creates mining models. Extensive research has been performed to develop appropriate machine learning techniques for different data mining problems, and has led to a proliferation of different learning algorithms (Bernstein & Provost, 2001). With the explosion of available data mining algorithms, select appropriate algorithms or combination of algorithms to solve a given problem becomes more important than its availability.

Ideally, there exists a single algorithm that can solve all the problems, or we can try all the algorithms to the problems to find the best algorithms, which can obtain the best accuracy and efficiency balance. Unfortunately, scientists have to develop different algorithms to satisfy different situations. Thus there are large amount of algorithms and models developed. Facing the enormous algorithms, the users can hardly handle all of them, and it is difficult to decide which one or combination of some algorithms is the most fitted for their problems. It shows that for a novice user, the data mining process space is overwhelming. Many novice users simply use the algorithms that they are familiar with (Kirsh, 2000). Consider the 2000 KDDCUP, in which 30 teams of data mining researchers and professionals competed to mine knowledge from electronic-commerce data. Most algorithms were tried by only a small fraction of participants. The only algorithm that was tried by more than 20% of the participants was decision-tree induction.

We can view data mining algorithms overload as information overload or data overload. Cognitive overload is the result of excessive demands made on the cognitive processes, in particular memory (Woods, Patterson, & Roth, 2002). Cognitive load increases with the amount of information to process. People feel information anxiety and suffer. Cognition in principal is mainly a process of information manipulation, according to the point of view of cognitive science. Cognitive Informatics (Wang, 2003) proposed by professor Y.X. Wang investigates the internal information processing mechanisms and process of the natural intelligence. Woods et al (2002) examined three different characterizations capturing the nature of the data overload problem and how they lead to different proposed solutions. Our approach to solve the data mining algorithms overload problem is to use machine intelligence (meta-learning) to cooperatively aid human users in selecting the most appropriate algorithms and assist the users to make decisions.

META-LEARNING BASIC CONCEPT

Meta-learning is an approach to select the appropriate learning algorithms for data mining, and the task of meta-learning is to find functions that can map datasets to the appropriate data mining algorithms. It is a process of learning at meta-level (Dzeroski & Zenko, 2002). It is similar to other machine learning methods in its ability to capture valuable information or knowledge from data. However, it is especially designed to enhance the machine learning process. Meta-learning is acquiring knowledge that can guide users in the application of data mining. It is a way to learn how a data mining system can work in a more effective way. Essentially, it is learning how to learn. The objective of meta-learning is to generate a user support system for selection of the most appropriate supervised learning algorithms for such tasks.