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. In this article, we have presented a meta-learning approach to support users automatically selecting most suitable algorithms during data mining model building process. The article discusses the meta-learning method in details and presents some empirical results that show the improvement we can achieve with the hybrid model by combining meta-learning method and Rough Set feature reduction. The redundant properties of the dataset can be found. Thus, we can speed up the ranking process and increase the accuracy by using the reduct of the properties of the dataset. With the reduced searching space, users’ cognitive load is reduced.

Keywords: algorithm selection; cognitive overload; meta-learning; rough set

INTRODUCTION

The exploding amount of available data has led researchers to the area of knowledge discovery and data mining. 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

Copyright © 2008, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
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 percent 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, 2007a, 2007b) proposed by professor Yingxu Wang investigates the internal information processing mechanisms and process of the natural intelligence. Woods et al (2002) (Peng, Flach, Brazdil, & Soares, 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 is a process of learning at meta-level (Dzeroski & Zenko, 2002). The objective of meta-learning is to generate a user support system for selection of the most appropriate supervised learning algorithms for such tasks. The meta-learning framework is usually based on a classification on the meta-level often disregarding a large amount of information gained during the induction process. Considering only a small subset of meta-attributes may significantly reduce both the time and effort applied for the corresponding process (Peng et al., 2002). The use of meta-knowledge will greatly reduce the amount of experience to be carried out. This also will enhance the performance of data mining and encourage users to work in a comfortable and effective way. Cognitive overload will be reduced. An excellent review of different aspects of meta-learning is given by Vilalta and Drissi (2002).

Several related algorithm selection systems and strategies have been proposed. An expert system called “Consultant” (Sleeman, Rissakis, Craw, Graner, & Sharma, 1995), determines the characters of data by asking users to answer some questions. The system just depends on the users’ subjective experiences and their answers to the system’s questionnaire about the data. Rendell and Seshu in VBMS (Rendell, Seshu, & Tcheng, 1987) give a classification of the problem to let the users to select, and then give the prediction about the corresponding algorithms. It considers the new tasks as new classification tasks, and this slows down the process, particularly if the user is a novice. STATLOG (Michie & Spiegelhalter, 1996) extracts various characteristics from a set of registered datasets, and then combines these characteristics with the performance of the algorithms. It generates rules to guide inducer selection based on the characteristics of the dataset. This method just bases on the morphological similarity between the new dataset and the existing collection. Besides, it does not classify these datasets. When a new dataset comes, it compares the characteristics of the new dataset to all the old datasets. This takes a lot of time. Data Mining Advisor (DMA) (Alexandros & Melanie, 2001) is a system that already has a set of candidate algorithms and a group of training datasets. The performance of the candidate algorithms for each subset in the training datasets is known. When the user gives a new dataset, DMA first uses K-NN algorithm to find a similar subset in the training datasets. Based on this subset, it retrieves information about the performance of candidate algorithms, then ranks the candidate algorithms and gives the appropriate recommendation. DCT tool (Data Characterization Tool) is used to find the
On Abstract Intelligence and Brain Informatics: Mapping Cognitive Functions of the Brain onto its Neural Structures


[www.igi-global.com/article/abstract-intelligence-brain-informatics/75451?camid=4v1a](www.igi-global.com/article/abstract-intelligence-brain-informatics/75451?camid=4v1a)