The New Stage of Data Mining Research : Variable Metric Data Mining

Ai Wang*, Xuedong Gao

Donglinks School of Economics and Management,
University of Science and Technology Beijing,
Beijing, China
Email: dddercwise@163.com
Email: gaoxuedong@manage.ustb.edu.cn
*Corresponding author

Leading the new trend of human development, artificial intelligence has drawn wide attention (Wang, 2015). While on the current research stage, there are essential differences between human intelligence and mechanical intelligence. It is effortless for human intelligence to conduct continuous learning and create things (Dennis et al., 2015). However, it is hard for mechanical intelligence to start qualitative change without depending on quantitative change, or to move from one concept (essence) directly to another.

The current data mining research has the following problems: (1) it takes the single top-down or down-top path (Wu et al., 2003); (2) it is an open-loop process, which means the research ends up with solving the data mining tasks without automatically learning and memorizing the mining results; (3) it requires the participation of people, for example the determination of the initial parameter values (Hosking et al., 1997).

For the problem (1), since everything is dialectical unity, the top-down and down-top thought should be combined together for solving problems. It can be seen that human thinking granularity switches between coarse granularity and fine granularity at any time, which makes it possible for human to get the satisfactory solution quickly. For the problem (2), one data mining process does not gain only one result (Han et al., 2012). As the same data mining task, different algorithms or different initial parameter values will lead to different results. Thus, an expert is needed to determine initial parameters and judge the quality of results nowadays, according to his (her) rich data mining experience no matter success or failure. So after finishing a data mining task, it is more important to analyze the whole process and accumulate experience, which could change the process to a close-loop structure. For the problem (3), if the close-loop structure is established, any computer can auto-determine optimal algorithm and initial parameter value via referring to history experience (memory) instead of people.

Human thinking ability contributes to solving complex problems, and it works through thinking activities which is a universal set of organic connected thinking sub-activities (Goldstein, 2015; Yin, 2007).

Definition 1: **Thinking Meta activity** is the independent thinking sub-activity, which can be understood as the activity of thinking activities.

Based on the background above, the framework of variable metric data mining research is put forward.

Definition 2: **Variable metric data mining** is the intelligent data mining process through simulating the regulatory mechanism among human thinking process when solving problems and the metric consists of the granularity and algorithm parameter. The intelligence of variable metric data mining is reflected in auto-choosing algorithm, auto-determining initial parameter value and auto-confirming the evaluation measurement and standard.

Thus, it can be seen that variable metric data mining pursues a genuine kind of artificial intelligence by simulating human thinking ability. The feasibility of this research can be verified by the principle of data mining.

Data mining starts its research from the perspective of human learning process, from the clustering to classification and association rules (Weiss et al., 1998). Obviously, data mining is one of human thinking Meta activities.

The research contents and their relations of variable metric data mining are shown in Fig.1.

The research field of artificial intelligence is quite large, including natural language processing, pattern recognition, genetic algorithm etc (Yang et al., 2009). Although variable metric data mining is just a small part of artificial intelligence research, it still has its feature. Variable metric data mining is able to establish the new problems found mechanism and let the computer gain the ability of epiphany, which other artificial intelligence methods do not have.

Variable metric data mining, as an intelligent process, has the characteristics of independently calculating, learning, and evaluating as well as finding new problems etc. These characteristics have been beyond the scope of original data mining problems. Variable metric data mining turns the research focus on the thinking regulatory mechanism instead of the data itself. Therefore, variable metric data mining is the ultimate problem of data mining research field.

The purpose of variable metric data mining is to establish an intelligent computing center with thinking ability, which could support the realization of human intelligence. What's more, the intelligent computing center could be trained by dynamic data sets of any industry, and is able to grow into an expert in this field through self-taught.
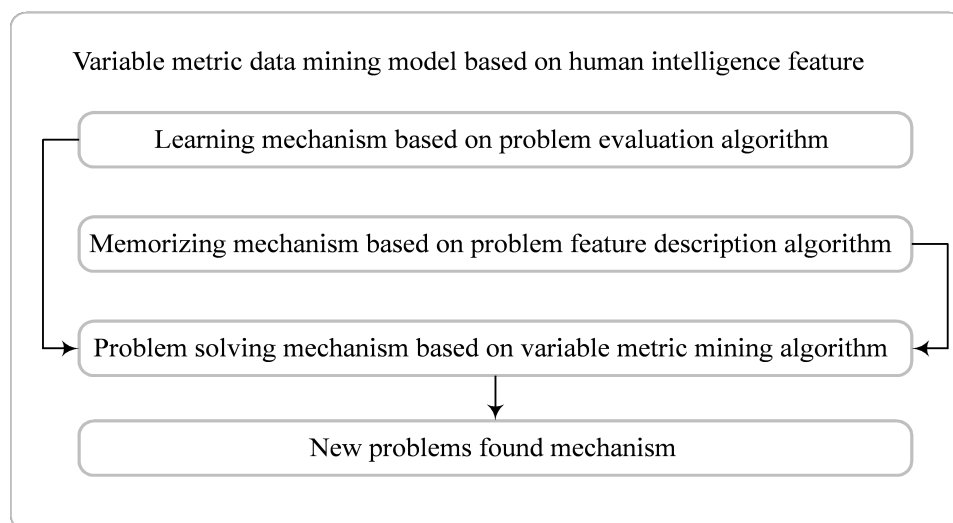
Fig.1. The research framework

References

Wu, S., Gao, X. and Bastian, M. (2003) *Knowledge Discovery of the High Dimensional Sparse Data*, Metallurgical Industry Press, China.

Han, J., Micheling K. and Pei, J. (2012) *Data Mining Concepts and Techniques*, Machine Press, China.

Hosking, J., Pednault, E. and Sudan, M. (1997) 'A statistical perspective on data mining', *Future Generation Computer Systems: Special Issue on Data Mining*, Vol. 3, No. 2, pp.117–134.

Dennis, C. and John, O. (2015) *Psychology: A Journey*, China Light Industry Press, Beijing.

Goldstein, E. (2015) *Cognitive Psychology*, China Light Industry Press, Beijing.

Yin, W. (2007) *Neuropsychological*, Science Press, Beijing.

Weiss, S. and Indurkhya, N. (1998) *Predictive Data Mining*, Morgan Kaugmann Press, San Francisco.

Yang, Z. and Lin, J. (2009) 'The present situation and development of artificial intelligence', *Science & Technology Information*, Vol. 4, No. 1, pp.524–525.

Wang, Y. (2015) 'Development and Application of Artificial Intelligence', *2015 4th International Conference on Mechatronics, Materials, Chemistry and Computer Engineering*.