ASSOCIATION RULES ALGORITHM AND ITS APPLICATION IN THE MAINTENANCE OF THE TUNNEL

Xu, W.-X.[*]; Fan, Y.-H.[*] & Zhang, J.-X.[**]

[*] School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China

[**] School of Languages and Communication Studies, Beijing Jiaotong University, Beijing, China

E-Mail: wxxu@bjtu.edu.cn, yhfan@bjtu.edu.cn, jxzhang@bjtu.edu.cn

Abstract

In this paper, effective data mining methods are adopted for the tunnel management information system to deal with safety issue data and work out the relationship among these safety issues in order to estimate risk, establish intelligent decision support, provide basis of governance for railway maintenance departments and remedy the defects of the existed management information systems. In view of the bottleneck of Apriori algorithm, two new algorithms are proposed in this paper. The first is AprioriN algorithm based on arrays, which converts the operation on database to the operation on memory via coding. The second is a high performance association rule mining algorithm based on FP-tree, which accelerates the speed of traverse itemsets by adding an extra data structure. During the second scan of the database, a matrix is generated to save frequent 2-itemsets when the basic FP-tree is created. This paper attempts the improved algorithms to improve the efficiency.
(Received, processed and accepted by the Chinese Representative Office.)

Key Words: Data Mining, Association Rule, Frequent Itemsets, FP-Tree, Tunnel Safety Issue

1. INTRODUCTION

The first railway tunnel in China was constructed in 1888 and after more than one century of construction, there were 7153 railway tunnels in mainland China with a total length of 4601.836 km by the end of 2011. However, a very stunning fact is that most of them are suffering from safety defects originated from low designing standards, ill-considered situation, and poor quality of construction and shortage of maintenance management. Some of the defects endanger the traffic safety directly, while some others potentially, and the railway sectors annually invest a lot of manpower, funds and other resources for the maintenance and remediation of the tunnel safety, but they still couldn't make a fundamental turn for better conditions of the tunnel and the amount of data are idled. To ensure a safe operation of tunnels, equipment management department ought to master all the information which may manifest insecurity and take appropriate measures to eliminate the hidden dangers of the tunnel as far as possible. For a long time in the past years, the collection of tunnel insecurity information mainly relies on human eye-viewing, the measuring tapes and other simple and crude devices. Data management is no more than live recording data to fill in routine reports. Thus are the results to follow: an incomplete understanding of the security information, poor management, shortage of risk knowledge and inadequate prevention against accidents. Such status certainly calls for reforms in information management and renovation in the study of methodologies.

As one of the human-dominated computer systems, management information system utilizes computer hardware, software and other office equipment to collect, transmit, store, process and maintain information in terms of supporting the operation, management and decision-making. The existed railway tunnel management information systems have already achieved the goal of managing raw data, images and text reports dynamically, with establishment of models and improvement of visual degree of data. However, most of these systems only focus on a specific class of factors or engineering conditions. The studies on

will only exhaust the computer's memory and CPU consumption and it can't accelerate the speed of mining. Therefore, we need to find enhancement algorithms to deal with the dense data in the future. As geologic conditions differ greatly all over China, the diversity of the railway tunnel structures is also enormous. Association rule can be implemented not only to detect the relationship among the various safety issues but also extract the common characteristics of tunnel geological conditions.

6. ACKNOWLEDGEMENTS

REFERENCES

[1]   Wang, X. & Liu, M. G. (2010). Association rules mining, *Economic Research Guide*, Vol. 2010, No. 11, 198-199

[2]   Xu, W. X.; Xu, L.; Liu, X. M.; Jones, J. D. (2008). A new approach to decision-making with key constraint and its application in enterprise information systems, *Enterprise Information Systems*, Vol. 2, No. 3, 287-308, doi:10.1080/17517570802302341

[3]   Lee, N.; Jung, J. J. (2012). Profit-based association rule mining from commercial transactions, *Information – An International Interdisciplinary Journal*, Vol. 15, No. 8, 3469-3476

[4]   Abdullah, Z.; Herawan, T.; Deris, M. M. (2012). Mining highly-correlated least association rules using scalable trie-based algorithm, *Journal of the Chinese Institute of Engineers*, Vol. 35, No. 5, 547-554, doi:10.1080/02533839.2012.679064

[5]   Achar, A.; Laxman, S.; Sastry, P. S. (2012). A unified view of the apriori-based algorithms for frequent episode discovery, *Knowledge and Information Systems*, Vol. 31, No. 2, 223-250, doi:10.1007/s10115-011-0408-2

[6]   Chen, K.; Zhang, L.-J.; Li, S.; Ke, W. (2011). Research on association rules parallel algorithm based on FP-growth, Liu, C.; Chang, J.; Yang, A. (Eds.) *Information Computing and Applications*, Springer-Verlag, Berlin, 249-256, doi:10.1007/978-3-642-27452-7_33

[7]   Jiang, H.; Zhao, Y.-Y.; Dong, X.-J. (2008). Mining positive and negative weighted association rules from frequent itemsets based on internet, *International Symposium on Computational Intelligence and Design*, Wuhan, 242-245, doi:10.1109/ISCID.2008.172

[8]   Abdullah, Z.; Herawan, T.; Deris, M. M. (2010). Mining significant least association rules using fast SLP-growth algorithm, Kim, T. H.; Adeli, H. (Eds.) *Advances in Computer Science and Information Technology*, Springer-Verlag, Berlin, 324-336, doi:10.1007/978-3-642-13577-4_28

[9]   Amin, G. R.; Gattoufi, S.; & Seraji, E. R. (2011). A maximum discrimination DEA method for ranking association rules in data mining, *International Journal of Computer Mathematics*, Vol. 88, No. 11, 2233-2245, doi:10.1080/00207160.2010.543457

[10] Totad, S. G.; Geeta, R. B.; Prasad Reddy, P. V. G. D. (2012). Batch incremental processing for FP-tree construction using FP-Growth algorithm, *Knowledge and Information Systems*, Vol. 33, No. 2, 475-490, doi:10.1007/s10115-012-0514-9

[11] Lin, K.-C.; Liao, I.-E.; Chen, Z.-S. (2011). An improved frequent pattern growth method for mining association rules, *Expert Systems with Applications*, Vol. 38, No. 5, 5154-5161, doi:10.1016/j.eswa.2010.10.047

[12] Mustafa, M. D.; Nabila, N. F.; Evans, D. J.; Saman, M. Y.; Mamat, A. (2006). Association rules on significant rare data using second support, *International Journal of Computer Mathematics*, Vol. 83, No. 1, 69-80, doi:10.1080/00207160500113330

[13] Liu, Y.-M.; Guan, Y. (2009). Application in market basket research based on FP-Growth algorithm, *2009 WRI World Congress on Computer Science and Information Engineering*, Vol. 4, 112-115

[14] Li, H.; Wang, Y.; Zhang, D.; Zhang, M.; Chang, E. Y. (2008). Pfp: Parallel FP-Growth for query recommendation, *Proceedings of the 2008 ACM Conference on Recommender Systems*, 107-114