A New Method for Giving Association Rules Threshold

鄧安生、李德治

E-mail: 9222615@mail.dyu.edu.tw

ABSTRACT

The data store in a large database is usually vast, and the work of data analysis becomes more and more hard. Data mining technique was explored in order to find the useful data efficiently in a database. Now, many researchers study the relevant technique in data mining, and there is a popularly issue to find the association rule from database transactions. The association rule is to find interrelationship between data in a database. Association rule processes usually have two steps: The first, to produce large itemset. The second, according to the large itemset produce rules from the first step. The first step is the bottleneck of algorithm usually. There are many researchers have studied the relevant research about this problem. Now in this paper, we focus on the threshold of the rule from the first step. In order to the rule has meaningful, so the rule must be greater than the threshold of support and confidence. But the threshold is given arbitrarily. It is no any reason, and it is an invalid value if the threshold higher or lower. In this research we will try to modify the threshold by a new method. We expect the new method can make the important rules more meaningful. It cannot only clear the confusion of making the threshold, but also the rule can be meaningfully and reliably. In this paper, we invent Mean Itemset Divide Method, and use this method to get optimal threshold. We produce some of random data source for our research. And aimed at the produce data and deal with the Mean Itemset Divide Method, we present the result by a visualized method to analyze the trend of experimentation for the future work.

Keywords : Data Mining、Association Rules、Large Itemset、Support、Confidence、Mean Itemset Divide Method.

Table of Contents

REFERENCES

[1] Adriaans, P. and Zantinge, D., Data Mining, Addison Wesley Longman, 1996.

[2] Agrawal, R., Imilienski, T. and Swami, A., "Mining Association Rules between Sets of Items in Large Databases," In Proceedings of ACM SIGMOD International Conference on Management of Data, pp. 207-216, 1993.

[3] Agrawal, R. and Srikant, R., "Mining Sequential Patterns," In Proceedings of the IEEE Conference on Data Engineering, pp. 3-14, 1995.

[4] Agrawal, R. and Srikant, R., "Fast Algorithm for Mining Association Rules," In Proceedings of the 20th International Conference on Very Large Databases, pp. 487-499, 1994.

[5] Brin, S., Motwani R. and Silverstein, C., "Beyond market baskets: Generalizing association rules to correlations," In Proceedings of ACM SIGMOD Conference on Management of Data, pp. 265-276, 1997.

[6] Chen, M.S., Han J. and Yu, P.S., "Data Mining: An Overview from Database Perspective," IEEE Transactions on Knowledge and Data Engineering, Volume 8, Number 6, pp. 866-883, 1996.

[7] Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P., "The KDD Process for Extracting Useful Knowledge from Volumes of Data," Communications of The ACM, Volume 39, Number 11, pp. 27-34, 1996.

[8] Fayyad, U.M., "Data Mining and knowledge Discovery: Making Sense Out of data," IEEE Expert, Volume 11, Issue 5, pp. 20-25, 1996.

[9] Han, J. and Kamber, M., Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, San Francisco, 2000.

[10] Kleissner, C., "Data mining for the enterprise," In Proceedings of the Thirty-First Hawaii International Conference on, Volume 7, pp. 295-304, 1998.

[11] Michael, J.A. and Linoff, G., Data Mining Technique: for Marketing, Sales and Customer Support, Wiley Computer Publishing, New York, 1997.

[12] Olaru, C. and Wehenkel, L., "Data mining," IEEE Computer Applications in Power, Volume 12, Issue 3, pp. 19-25, 1999.

[13] Park, J.S., Chen, M.S. and Yu, P.S., "An Effective Hash-Based Algorithm for Mining Association Rules," In Proceedings of ACM SIGMOD International Conference on Management of Data, pp. 175-186, 1995.

[14] Park, J.S., Chen, M.S. and Yu, P.S., "Using a Hash-Based Method with Transaction Trimming and Database Scan Reduction for Mining Association Rules," IEEE Transactions on Knowledge and Data Engineering, Volume 9, Number 5, pp. 813-825, 1997.

[15] Savasere, A., Omiecinski, E. and Navathe, S., "An Efficient Algorithm for Mining Association Rules in Large Databases," In Proceedings of the 21st International Conference on Very Large Databases, pp. 432-443, 1995.

[16] Savasere, A., Omiecinski, E. and Navathe, S., "Mining for Strong Negative Associations in a Large Database of Customer Transactions," In Proceedings of the 14th International Conference on Data Engineering, pp. 494-502, 1998.

[17] Simoudis, E., "Reality check for data mining," IEEE Expert, Volume 11, Issue 5, pp. 26-33, 1996.

[18] Srikant, R., Vu, Q. and Agrawal, R., "Mining Association Rules with Item Constraints," In Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining, pp. 67-73, 1997.

[19] Srikant, R. and Agrawal, R., "Mining generalized association rules," In Proceedings of the 21st International Conference on Very Large Databases, pp. 407-419, 1995.

[20] Srikant, R. and Agrawal, R., "Mining Quantitative Association Rules in Large Relational Tables," In Proceedings of the ACM SIGMOD International Conference on Management of Data, pp. 1-12, 1996.

[21] Srikant, R. and Agrawal, R., Mining generalized association rules, Future Generation Computer Systems, 1997.

[22] Fu, Y., "Discovery of Multiple - Level Rules from Large Databases," In Ph. D. dissertation, School of Computing Science, Faculty of Applied Sciences, Simon Fraser University, 1996.

[23] Fu, Y., "Data mining Tasks, techniques and applications," IEEE Potentials, Volume 16, Issue 4, pp. 18-20, 1997.

[24] Zaki, M.J., Parthasarathy, S., Ogihara, M. and Li, W., "New Algorithms for Fast Discovery of Association Rules," In Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining, pp. 283-286, 1997.

[25] Zhang, C. and Zhang, S., Association Rule Mining: Model and Algorithms, Springer-Verlag Berlin Heidelberg, New York, 2002.