The Study of Giving Double Threshold for Association Rules

E-mail: 9314358@mail.dyu.edu.tw

ABSTRACT

In recent years, the technique about data mining has become one of the most popular research subjects, and there is a popular issue to find the association rules from database transactions. The association rules that must happen only among the data that satisfy the minimum support and the minimum confidence is set by the user's. The traditional association rule algorithm establishes a minimum support and minimum confidence only, so it cannot discover the rule in which the support is low but the confidence is high. For preventing this problem effectively, the scholar putting forward Relative Support Apriori Algorithm to resolve this problem. But the threshold is given arbitrarily with no reason. The first support and the second support of the Relative Support Apriori Algorithm also is such. In this research, we will try to modify the threshold by a new method. The first support is given with the Mean Itemset Divide Method first, again the second support is given with the first support and database item characteristics, as the profit of the merchandise. Finally, we produce some of random data source for our research, and mine association rules with Double Threshold. We evaluate the performance of the proposed algorithm by comparing it with Apriori algorithm for discovering the association rules for the future work.

Keywords: Data Mining　Association Rules　Relative Support Apriori Algorithm　Threshold　Support　Confidence　Mean Itemset Divide Method　Double Threshold.

Table of Contents

REFERENCES

[1] 2003

[2] 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.

[3] 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.

[4] Alsabti, K., Ranka, S. and Singh, V., "An Efficient K-Means Clustering Algorithm," PPS/SPDP Workshop on High performance Data Mining, 1997.

[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] Cabena, P., Hadjinian, P., Stadler, R., Verhees, J. and Zanasi, A., " Discovering Data Mining From Concept to Implementation," Prentice-Hall Inc., 1997.

[7] Chen, M.S., Park, J.S. and Yu, P.S., " Efficient Data Mining for Path Traversal Patterns," IEEE Transactions on Knowledge and Data Engineering, Vol. 10, No. 2, pp. 209-221, April 1998.

[8] Cheung, D.W., Han, J., Ng, V.T., Fu, A.W. and Fu, Y., " A Fast Distributed Algorithm for Mining Association Rules," In Proc. of 1996 Int 'l Conf. on PDIS' 96, Miami Beach, Florida, USA, Dec. 1996.

[9] 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.

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

[11] Frawley, W.J., Piatetsky-Shapiro, G. and Matheus, C., " Knowledge Discovery in Databases: An Overview," AI Magazine, pp. 213-228, 1992.

[12] Han, J. and Kamber, M., " Data Mining: Concepts and Techniques," Morgan Kaufmann Publisher, 2000.

[13] Yun, H., Ha, D., Hwang, B. and Ryu, K.H., " Mining association rules on significant rare data using relative support," The Journal of Systems and Software 67, pp. 181– 191, 2003.

[14] Kaufman, L. and Rousseeuw, P.J., " Finding Groups in Data: an Introduction to Cluster Analysis," John Wiley & Sons, 1990.

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

[16] Liu, B., Hsu, W. and Ma, Y., " Mining Association Rules with Multiple Minimum Supports," Proc. 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 337-341, May 1999.

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

[18] Ng, R.T. and Han, J., " Efficient and Effective Clustering Methods for Spatial Data Mining," Proc. of the 20th Int' l Conf. on Very Large Databases, Santiago, Chile, pp. 144-155, 1994.

[19] Weiss, S.M. and Kulikowski, C.A., " Computer System that Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert System," Morgan Kaufman, 1991.

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