雙門檻值制定應用於關聯法則之研究

陳政富、李德治

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

摘 要

近年來有關資料探勘的技術已成為相當熱門的研究議題之 一，其中有一項重要的議題就是如何從交易資料庫中挖掘出關聯法則。關聯法則的成立必須滿足使用者所制定的最小支持度 (minimum support)與最小信賴度(minimum confidence)。傳統的關 聯法則演算法僅制定一個最小支持度與最小信賴度，因此支持度 較低但信賴度較高的項目將不會被挖掘出來。為了有效防止這種 問題，有學者提出相關支持度Apriori 演算法(Relative support Apriori Algorithm; RSAA)以解決這類問題。 然而，傳統的關聯法則演算法對於門檻值的制定都是經由專 家根據經驗自行訂定，相關支持度Apriori 演算法的第一個最小支持度(1st support)與第二個最小支持度(2nd support)亦是如此。有鑑 於此，本研究嘗試以新的方法制定門檻值。先以平均項目集合分 割演算法訂定出第一個最小支持度，再根據第一個最小支持度的 制定情況以及資料庫項目特性，如商品的獲利，訂出第二個最小 支持度。 最後，我們利用程式以亂數的方式產生資料，做為本研究之 資料來源，以雙門檻值法實際挖掘出關聯法則，並與Apriori 演算 法比較資料挖掘在各方面的績效狀況，以利未來研究之進行。

關鍵詞 : 資料探勘、關聯法則、相關支持度Apriori演算法、門檻值、支持度、信賴度、平均項目集合分割、雙門檻值。

目錄

參考文獻

[1] 鄧安生，「新式探勘方法在關聯法則門檻值制定之研究」， 大葉大學資訊管理學系碩士論文，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.