新式探勘方法在關聯法則門檻值制定之研究

鄧安生、李德治

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

摘 要

由於在大型資料庫中所儲存的資料往往非常龐大，分析處理資料的工作因此愈加困難。為了有效的從資料庫中尋找出有用的資料，便孕育出資料探勘(Data Mining)技術的產生。在近幾年來有許多學者們從事資料探勘等相關技術的研究，其中一項被廣泛討論的議題就是從交易資料庫中挖掘關聯法則(Association Rules)。關聯法則主要是在協助尋找資料庫中資料與資料間的相互關係。關聯法則的產生過程大致可分為二個步驟：第一個步驟是產生大項目集合(Large Itemset)，第二個步驟則依據第一個步驟所產生的大項目集合來產生規則(Rules)。由於第一個步驟是演算法的瓶頸所在，已有許多學者針對此問題進行相關的研究，本研究主要是針對第一個步驟產生的規則時所制定的門檻值問題來探討。 由於產生的規則必須大於支持度(Support)及信度(Confidence)的門檻值，這樣的規則才具有它的意義，然而門檻值的訂定是人為所制定的，無一定的標準，太高太低都有相關的問題產生。有鑑於此，本研究嘗試以新的方法制定門檻值，使所導出的規則更具有意義及可靠性。本研究所提出之方法為平均項目集合分割法(Mean Itemset Divide Method)，並運用此方法訂定出較佳的門檻值。在進行實驗時，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，針對其產生的資料，經由平均項目集合分割法處理之後，以視覺化方式呈現其實驗結果，並分析其實驗數據之走勢，以利未來研究之進行。

關鍵詞 : 資料探勘、關聯法則、大項目集合、支持度、信度、平均項目集合分割法。

目錄

參考文獻

[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.