The Design and Application of an Automatic Link Analysis Technology

楊境榮、陳鴻文

E-mail: 9422529@mail.dyu.edu.tw

ABSTRACT

Each data mining technique is aimed to automatically analyze implied knowledge rules from data. However, the less frequent but high-value data association rules can not be extracted by simply setting a single proper threshold in algorithms of association rule analysis. For solving the kind of problems, data must be classified and then processed with different lowest support values. Unfortunately, the performance of handling many actual problems is still unacceptable. Besides, human experts need to use visualization tools to find out regular patterns and features with their eyes in traditional link analysis. An only exceptional case is Google, which takes Page Rank algorithm to automatically evaluate the weight of each network by hyperlink relations. In other words, an automatic technique for link analysis is eagerly needed for problems such as food chain, transportation network, etc. Some researches of social network analysis pointed out “ strong ties within a graph can group individuals of the same characteristic while weak ties can communicate and work as the bridge of different groups. ” Based on the concept of weak ties and group theory, we proposed to find out potential weak links beyond biconnected and strongly connected components and then form critical paths within a graph . The proposed algorithm can detect out association relations between rare critical data which is quite difficult to deal with in traditional association rule analysis. In order to verify and evaluate proposed automatic link analysis, the actual Enron Email Dataset announced by FERC(Federal Regulation and Oversight of Energy) was investigated. Experiments illustrated the efficiency of the algorithm in analyzing characteristics of a direct/undirected graph. Thus, it is highly recommended to solve problems such as detection of social group, organizational criminality, e-mail spam etc.

Keywords : data mining ; link analysis ; association analysis ; weak tie ; automatic link analysis

Table of Contents

REFERENCES

[1]吳寶秀，「台灣製造業員工個人社會網絡分析」，東海大學社會學研究所碩士論文，民國79年。

[2]林傑彬與劉明德，「資料採掘與OLAP理論與實務」，文魁資訊股份有限公司，民國91年。

[3]胡守仁譯，「連結:混沌、複雜之後，最具開創性的小世界理論」，天下文化，民國91年。

[4]鄭讚源，「社會網絡、社會整合與學習家庭:機會與挑戰」，載於教育部主辦《學習型家庭理論與實務研討會資料》，台灣師範大學，民國88年。

[5]蕭新煌與龔宜君，「東南亞台商與華人之商業網絡關係」，中央研究院東南亞區域研究所，民國87年。

[6]N.M. Adams, G. Blunt, D.J. Hand and M.G. Kelly, “ Data mining for fun and profit, ” Statistical Science, Vol. 15, No. 2, pp. 111-131, 2000.

[7]P. Adriaans and D. Zantinge, “ DATA MINING, ” ADDISON-WESLEY, 1999.

[8]R. Agrawal, T. Imilienski and A. Swami, “ 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.

[9]M. J. A. Berry and G. Linoff, Data Mining Techniques: For Marketing Sale and Customer Support, John Wiley & Sons, 1997.

[10]S. Brin and L. Page, “ The Anatomy of Large-Scale Hypertextual Web Search Engine, ” In Proceedings of the 7th InternationalWorldWideWeb Conference, pp. 107-117, 1998.

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

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

[13]F. L. Chung and C. L. Lui, " A post-analysis framework for mining generalized association rules with multiple minimum supports, " Workshop Notes of KDD'2000 Workshop on Post-Processing in Machine Learning and Data Mining, pp.9-14, 2000.

[14]U. Fayyad, S. G. Piatetsky and P. Smyth, " From data mining to knowledge discovery in database, " AI magazine, Vol. 17, pp. 37-54, 1996.

[15]L. Garton, C. Haythornthwaite and B. Wellman, " Studying Online Social Networks, " Journal of Computer-Medicated Communication, Vol. 3, No. 1, 1997.

[16]M. S. Granovetter, " The strength of weak ties, " American Journal of Sociology, Vol. 78, pp. 1360-1380, 1973.

[17]J. Han and M. Kamber, Data Mining : Concepts and Techniques, John Wiley & Son, 2001.

[18]M. R. Henzinger, " Hyperlink Analysis for the Web, " IEEE INTERNET COMPUTING, Vol. 5, pp. 1089-7801, 2001.

[19]M. P. Johson and R. M. Milardo, " Networkinterference in pair relationship : A social psychological recasting of Slater's (1963) theory of social regression, " Joural of Marriage and the Family, Vol. 46, pp. 893-899, 1984.

[20]C. Kleissner, " Data mining for the enterprise, " In Proceedings of the 35th Hawaii International Conference, Vol. 7, pp. 295-304, 1998.

[21]D. Knoke and J. H. Kuklinski, Network Analysis, Beverly Hills: Sage Publications, 1982.

[22]R. Lempel and A. Soffer, " PicASHOW: Pictorial Authority Search by Hyperlinks on the Web, " In Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 438-448, 2001.

[23]B. Liu, W. Hsu, and Y. Ma, " Mining Association Rules with Multiple Minimum Supports, " In Proceedings of the 1999 International Conference on Knowledge Discovery and Data Mining, pp. 337-341, 1999.

[24]P. Marsden, " Network Data and Measurement, " Annual Review of Sociology, Vol. 16, pp. 435-463, 1990.

[25]S. Milgram, " The small world problem, " Psychology Today, Vol. 2, pp. 60-67, 1967.

[26]S. Wasserman and K. Faust, Social network analysis: Methods and application, Cambridge University Press, 1997.

[27]D. J. Watts and S. H. Strogatz, " Collective Dynamics of Small-World, " Networks, Vol. 393, pp. 440-442, 1998.

[28]M. A. Weiss, Data Structures and Algorithm Analysis in C. The Benjamin / Cummings Publishing Company, 1993.

[29]P. C. Wang, " Visual Data Mining, " IEEE Computer Graphics and Applications, Vol. 19, No. 5, 1999.

[30]C. Westphal and T. Blaxton, Data Mining Solutions, John Wiley &Sons, 1998.

[31]H. Yun, D. Ha, B. Hwang and K. H. Ryu, " Mining Association Rules on Significant Rare Data Using Relative Support, " The Journal of Systems and Software, Vol. 67, pp. 181 – 191, 2003.

[32]Enron Email Dataset,　http://www-2.cs.cmu.edu/~enron/ [33]Enron Dataset,　http://www.isi.edu/~adibi/Enron/Enron.htm