方向性鏈結分析技術之開發與應用

黃勝偉、陳鴻文

E-mail: 9607621@mail.dyu.edu.tw

摘 要

由於涉及隱私、商業機密與社會環境背景等的特性資料取得不易，研究學者一般僅能針對少量的資料，運用社會網絡分析工具及技術，探討個體間的關係、群體結構特性及其行為特徵等意涵。然而過去社會網絡分析，大多偏重於網絡中心成員的辨識，甚少著墨於子群組間交互影響與關係上，也無法分析出群體群組間相互的作用關係。因此本研究主要目標是從社會網絡的潛在結構中，發掘出其間子群組之間的動態關係。 本研究在測試資料集方面，是以安隆公司電子郵件資料集及實際被起訴名單，來進行所設計技術的說明及效能驗證。首先是以內文主題分類(McCallum et al., 2005)的觀念進行前置處理，再利用所提出之方向性鏈結分析技術DLA的作法，對元件間橋樑路徑進行適當分類；接著再藉由所提出之網絡結構差異性分析DNSA技術，進行雜訊節點的過濾，以期發掘出可能涉及詐欺的員工。實驗結果顯示在預測方面的平均準確率可達83.07%，而平均所需執行時間為1.47秒。所以在圖形資料的自動分析方面，可達到節省大量人力時間的支出，和免除主觀上人為的判斷的目的。

關鍵詞 : 圖形理論(graph theory)、社會網絡分析(social network analysis)、方向性鏈結分析(directional link analysis)、安隆電子郵件分析(Enron corpus analysis)、弱鏈結(weak tie)

目錄

參考文獻

Buchanan, M.(2003)，連結:混沌、複雜之後，最具開創性的小世界理論，(胡守仁譯)，天下文化，(原文於2003年出版)。 楊境榮(2005)，自動資料鏈結分析技術之開發與應用，大葉大學資訊管理研究所未出版碩士論文。 熊瑞梅(1995)，社會網絡的資料蒐集、測量及分析，收於章英華、傅仰止、瞿海源編，社會調查與分析:社會科學研究方法檢討與前瞻(pp. 313-356)，台北:中央研究院民族學研究所。 二、英文部份 Adriaans, P., & Zantinge, D. (1999). Data Mining. Addi-son-Wesley Bekkerman, R., McCallum, A. & Huang, G.. (2004). Automatic Categorization of Email into Folders: Benchmark Ex-periments on Enron and SRI Corpora. CIIR Technical Report IR-418, Available: http://www.cs.umass.edu/~ghuang/foldering-tr05.pdf. Berry, M. J. A., & Linoff, G.. (1997). Data Mining Techniques: For Marketing Sale and Customer Support. John Wiley & Sons. Berry, M. W. & Browne M. (2005). Email Surveillance Using Nonnegative Matrix Factorization. Proceedings of Workshop on Link Analysis, Counterterrorism and Security, SIAM International Conference on Data Mining 2005. 45 – 54. Borgatti, S. P. (2004). The Key Player Problem. In: Dynamic So-cial Network Modeling and Analysis, Breiger, R., Carley, K.M., & Pattison, P., (Eds). National Acad-emies Press, 241-252. Brass, D. J., & Burkhardt, M.E. (1992). Centrality and Power in Organizations. In: Networks and Organizations, Nohria, N., & Eccles, R.G., (Eds). Boston:Harvard Business School Press, 191-215. Burt, R. S. (1980). Models of Network Structure. Annual Review of Sociology, 6, 79-141. Burt, R. S. (1992). Structural Holes:The Social Structure of Competition. Harvard University Press, 45-49. Chapanond, A., Krishnamoorthy, M. S. & Yener, B. (2005). Graph Theoretic and Spectral Analysis of Enron Email Data. Proceedings of Workshop on Link Analysis, Counterterrorism and Security, SIAM International Conference on Data Mining 2005. 15 – 22. Cohen, W. W. (2nd ed.), CALO, CMU.

[Online], Available: http://www-2.cs.cmu.edu/~enron/. Diesner, J., & Carley, K. M. (2005). Exploration of Communica-tion Networks from the

Enron Email Corpus. Pro-ceedings of Workshop on Link Analysis, 124-143. Diesner, J., Frantz, T. L., & Carley, K. M.(2005). Communication Networks from the Enron Email Corpus “ It's Always About the People. Enron is no Different ” . Computa-tional & Mathematical Organization Theory, 11(3), 201-228. Duan, Y., Wang, J., Kam, M. & Canny, J. (2002). A Secure Online Algorithm for Link Analysis on Weighted Graph. Proceedings of Workshop on Link Analysis, Counter-terrorism and Security, SIAM International Confer-ence on Data Mining 2005. 71 – 81. Emirbayer, M., & Goodwin, J. (1994). Network Analysis, Culture, and the Problem of Agency. American Journal of So-ciology, 99 (6), 1411-1454. Enron Dataset, Available: http://www.isi.edu/~adibi/Enron/Enron. htm, [2006, July 14]. Freeman, L. C. (1979). Centrality in Social Networks: Conceptual Clarification. Social Networks, 1, 215-239. Girvan, M., & Newman, M. E. J. (2002).Community structure in social and biological networks. Proceedings of the Na-tional Academy of Sciences, 99, 7821-7826. Granovetter, M. S. (1973).The strength of weak ties. Journal of American on Sociology, 78, 1360-1380. Hanneman, R. A. (2001). A.Introduction to social network meth-ods. California University Press, 87-105. Hansen, M. T. (1999). The search-transfer problem:the role of weak ties in sharing knowledge across organization subunits. Administrative Science Quarterly, 44, 82-111. Huberman, B. A., & Hogg, T. (1995). Communities of Practice: Performance and Evolution. Journal of Computational and Mathematical Organization Theory, 1(1), 73-92. Holme, P., Huss, M., & Jeong, H. (2003). Subnetwork hierarchies of biochemical pathways. Journal of the Bioinformat-ics, 19, 532-538. Johnson, M. P.; & Milardo, R. M. (1984). Network Interference in Pair Relationships: A Social Psychological Recasting of Slater's Theory of Social Regression Source. Jour-nal of Marriage and the Family, 46(4), 893-899. Keila, P. S. & Skillicorn, D. B. (2005). Structure in the Enron Email Dataset. Proceedings of Workshop on Link Analysis, Counterterrorism and Security, SIAM Inter-national Conference on Data Mining 2005., 55 – 64. Klimt, B., & Yang, Y. (2004). The Enron Corpus: A New Dataset for Email Classification Research. In Proceedings of ECML ’ 04, 15th European Conference on Machine Learning, 217-226. Klimt, B. & Yang, Y. (2004, a). Introducing the Enron Corpus. First Conference on Email and Anti-Spam (CEAS), Mountain View, CA, Available: http://www.ceas.cc/ papers-2004/168.pdf. Klimt, B. & Yang, Y. (2004, b). The Enron Corpus: A New Data-set for Email Classification Research. European Con-ference on Machine Learning, Pisa, Italy. Knoke, D., & Burt, R. S. (1983). Prominence. In: Applied Net-work Analysis: A Methodological Introduction, Burt, R.S., & Minor, M.J., (Eds), 195-222. Knoke, D., & Kuklinski, J. H. (1982). Network Analysis. Sage, Beverly Hills. Knoke, D., & Kuklinski, J. H. (1991). Network Analysis: Basic Concepts. In: Thompson, G., Frances, J. and Levacic, R. (Eds). Sage, London, 173-182. Krackhardt, D. (1992). The strength of strong ties: The impor-tance of philos in organizations. In N. Nohria, & Ec-cles, R.(Ed.), Networks and organizations:Structure, form, and action. Harvard Business School Press, 216-239. Krebs, V. E. (2002). Uncloaking Terrorist Networks. First Mon-day, 7(4), 549-560. Laumann, E. O., Galaskiewicz, J., & Marsden, P.V. (1978). Community structure as interorganizational linkages. Annual Review of Sociology, 4, 455-484. Lazarevic, A., Ertoz, L., Ozgur, A., Srivastava, J., Kumar, V. A Comparative Study of Anomaly Detection Schemes in Network Intrusion Detection. Proceedings of the 3rd SIAM Conference on Data Mining, 2003., Available: http://www.siam.org/meetings/sdm03/proceedings/sdm03_03.pdf Luo, J. D., Yen,G. L., & Hui, W. S.(2003). Social Network Struc-ture and Performance of Knowledge Team: A Case Study in the Chinese Cultural Settings. Proceedings of Academy on Management. Marsden, P. (1990). Network Data and Measurement. Annual Re-view of Sociology, 16, 435-463. Mcandrew, D. (1999). The structural analysis of criminal net-works. In: The Social Psychology of Crime: Groups, Teams, and Networks. Canter D., & Alison, L., (Eds). Dartmouth Publishing, Aldershot, UK, 53-94. McCallum, A., Corrada-Emanuel, A., & Wang, X. (2005a). Topic and role discovery in social networks. Proceedings of the Nineteenth International Joint Conference, 14, 786-791. McCallum, A., Corrada-Emmanuel, A., & Wang, X. (2005b). The Author-Recipient-Topic Model for Topic and Role Discovery in Social Networks, with Application to Enron and Academic Email. Proceedings of Workshop on Link Analysis, Counterterrorism and Security, SIAM International Conference on Data Mining 2005. 33-44. Milgram, S. (1967). The small world problem. Psychology Today, 2, 60-67. MIP. (2002). Fortune Magazine ’ s List of 10 Corporate Sins. Mok, E. (2004). Enron Email Corpus: Mapping names to email addresses and doing network analysis. Applied Natu-ral Language Processing. SIMS 290-2: Fall 2004, Prof. Marti Hearst, Assignment 4. Newman, M. E. J., & Girvan, M. (2003). Finding and evaluating community structure in networks. Michigan University press. Newman, M. E. J. (2004). Fast algorithm for detecting commu-nity structure in networks. Michigan University press. Ouchi, W. G.. (1980). Markets, Bureaucracies, and Clans. Admin-istrative Science Quarterly, 25, 129-141. Priebe, C. E., Conroy, J. M.., Marchette, D. J. & Park Y. (2005). Scan Statistics on Enron Graphs. Proceedings of Workshop on Link Analysis, Counterterrorism and Security, SIAM International Conference on Data Mining 2005. 23 – 32. Scott, J. (2000). Social Network Analysis: A Handbook. Sage, London. Scott, J. (2002). Social networks: critical concepts in sociology. Routledge, New York. Scott, W. R. (1992). Organizations: Rational, Natural, and Open Systems. Prentice-Hall. Seidam, S. B. (1983). Network Structure and Minimum Degree. Social Networks, 5, 269-287. Shetty, J., & Adibi, J. (2005a). Discovering Important Nodes through Graph Entropy: The Case of Enron Email Database. In Proceedings of ACM SIGKDD LinkKDD, 74-81. Shetty, J. and Adibi, J. (n.d., b), Ex employee status report. Re-trieved November 4, 2004, from http://www.isi.edu/ ~adibi/Enron/Enron_Employee_Status.xls. Shetty, J. and J. Adibi (n.d., c), The Enron Dataset Database Schema and Brief Statistical Report, Retrieved No-vember 4, 2004, from http://www.isi.edu/~adibi/ En-ron/Enron_Dataset_Report.pdf. Shetty, J. and J. Adibi (n.d., d), The Enron Dataset Mysql dump file, Retrieved November 4, 2004, from ftp://ftp.isi.edu/sims/philpot/data/enron-mysqldump.sql.gz. Sparrow, M. K. (1991). The application of network analysis to criminal intelligence: An assessment of the prospects. Social Networks, 13, 251-274. SRI International's Artificial Intelligence Center, Cognitive Agent that Learns and Organizes Project, April 2006,[Online], Available: http://www.ai.sri.com/people/gervasio. Tichy, N. M., Tushman, M. L., & Fombrun, C. (1979). Social network analysis for organizations. Academy of man-agement Review, 4(4), 507-519. Tyler, J. R., Wilkinson, D. M., & Huberman, B. A. (2005). Email as spectroscopy: Automated discovery of community structure within organizations. Journal of the Informa-tion Society, 21(2), 143-153. Wasserman, S., & Faust, K. (1997). Social network analysis: Methods and application. Cambridge University Press. Watts, D. J., & Strogatz, S. H. (1998). Collective Dynamics of Small-World. Networks, 1(393), 440-442. Wilkinson, D., & Huberman, H. (2002). A Method for Finding Communities of Related Genes. submitted for publica-tion, http://www.hpl.hp.com/shl/papers/ communi-ties/index.html. Westphal, C., & Blaxton, T. (1998). Data Mining Solutions. John Wiley & Sons. Weiss, M. A. (1993). Data Structures and Algorithm Analysis in C. Benjamin Cummings.