

# 主題分類關鍵詞之動態更新模式研究

林嵩富、陳振東

E-mail: 9222449@mail.dyu.edu.tw

## 摘要

隨著網際網路風潮席捲全球，愈來愈多使用者以網站作為資訊取得的來源。然而隨著網頁數量不斷地增加，使用者正面臨著資訊過載的問題。除了探討更為有效的資訊檢索技術外；藉由資訊推薦的概念，應用智慧型代理器所建立的個人化資訊推薦系統，也逐漸被用來解決資訊過載所產生的問題。資訊推薦方法的許多研究中，大多採用關鍵詞詞庫的分類策略。然而隨著時間的演變以及熱門主題的改變，分類主題關鍵詞也應適當調整，方能確保分類內的關鍵詞具有代表性。為此，本研究結合中文斷詞與相似度衡量的方法，提出一個關鍵詞詞庫動態更新模式，並據此開發動態更新系統，以進行詞庫中關鍵詞的自動更新，進而使得資訊推薦更為精準，藉以提高顧客對資訊服務的滿意度。從本研究的實驗結果可看出，本模式在關鍵詞擷取的準確度上，確實能達到預期的成效。另外，不同門檻值篩選的實驗也證明，本模式擷取關鍵詞之結果具有合理性與穩定性。

關鍵詞：資訊推薦、關鍵詞詞庫、相似度衡量、中文斷詞

## 目錄

第一章 緒論.....	1	第一節 研究背景與動機.....	1	第二節 研究目的.....	2	第三節 研究流程.....	3	第四節 研究範圍與限制.....	5
第二章 文獻探討.....	6	第一節 資訊推薦.....	6	第二節 中文斷詞.....	13	第三節 相似度衡量方法.....	18	第三章 關鍵詞動態更新系統架構.....	20
第一節 斷詞的處理.....	20	第二節 詞彙權重的計算方式.....	23	第三節 詞彙的相似度衡量.....	24	第四節 動態更新系統流程.....	26	第四章 實證分析.....	30
第一節 系統環境.....	30	第二節 系統架構與功能.....	30	第三節 實證分析.....	35	第四節 實驗結果的分析與討論.....	54	第五章 結論.....	57
第一節 結論.....	57	第二節 後續研究與建議.....	58	參考文獻.....	59				

## 參考文獻

- [1] 杜宜潔、翁頌舜，整合式客製化網站內容推薦系統，第七屆資訊管理研究暨實務研討會，民國90年。
- [2] 李雅萍，我國網際網路用戶數的調查統計分析，經濟部技術處Internet應用研究計畫/資策會ECRC-FIND，民國91年。
- [3] 陳克健、陳正佳、林隆基，中文語句的研究-斷詞與構詞，中央研究院技術報告，TR-86-006，民國75年。
- [4] 陳景揆，探勘中文新聞文件中的概念關聯及趨勢，雲林科技大學資訊管理研究所碩士論文，民國89年。
- [5] 陳稼興、謝佳倫、許芳誠，以遺傳演算法為基礎的中文斷詞研究，資訊管理研究，第二卷，第二期，民國91年，第27-44頁。
- [6] 曹又文，適地化與適性化之行動廣告資訊推薦機制，輔仁大學資訊管理研究所碩士論文，民國91年。
- [7] Atlam E. S., Fuketa M., Morita K. and Aoe J. I., "Similarity measurement using term negative weight and its application to word similarity," International Journal of Information Processing and Management, Vol. 36, Issue 5, 2000, pp. 717-736.
- [8] Belkin, N. J. and Croft, W. B., "Information Filtering and Information Retrieval: Two Sides of the Same Coin?" Communications of the ACM, Vol. 35, Issue 12, 1992, pp. 29-38.
- [9] Chen, C. T., "Extensions of the TOPSIS for group decision-making under fuzzy environment," Fuzzy Sets and Systems, Vol. 114, Issue 1, 2000, pp. 1-9.
- [10] Chen, K. J. and Kiu, S. H., "Word identification for Mandarin Chinese sentences," Fifth International Conference on Computational Linguistics, 1992, pp. 101-107.
- [11] Chen, S. M., "A new approach to handling fuzzy decision making problems," IEEE Trans. Systems, Man, Cybernetics, Vol. 18, 1988, pp. 1012-1016.
- [12] Chen, S. M., Horng, Y. J. and Lee, C. H., "Document Retrieval Using Fuzzy-Valued Concept Networks," IEEE Transactions on Systems, Man, and Cybernetics, Part B, Vol. 31, Issue 1, 2001, pp. 111-118.

- [13] Chen, S. M., Yeh M. S. and Hsiao P. Y., "A comparison of similarity measures of fuzzy values," *Fuzzy Sets and Systems*, Vol. 72, Issue 1, 1995, pp. 79-89.
- [14] Egghe, L. and Michel, C., "Strong similarity measures for ordered sets of documents in information retrieval," *International Journal of Information Processing and Management*, Vol. 38, Issue 6, 2002, pp. 823-848.
- [15] Fan, J. and Xie, W., "Some notes on similarity measure and proximity measure," *Fuzzy Sets and Systems*, Vol. 101, Issue 3, 1999, pp. 403-412.
- [16] Foo, S., Hui, S. C., Lim, H. K. and Li, H., "Automatic Thesaurus For Enhanced Chinese Text Retrieval," *Library Review*, Vol. 49, Issue 5, 2000, pp. 230-239.
- [17] Good, N., Schafer, B., Konstan, J. A., Borchers, A., Sarwar, B., Herlocker, J. and Riedl, J., "Combining Collaborative Filtering with Personal Agents for Better Recommendations," *Proceedings of the Sixteenth National Conference on Artificial Intelligence ( AAAI-99 )*, 1999.
- [18] Hong, D. H. and Kim, C., "A note on similarity measures between vague sets and between elements," *Information Sciences*, Vol. 115, Issue 1-4, 1999, pp. 83-96.
- [19] Horng, J. T. and Yeh, C. C., "Applying genetic algorithms to query optimization in document retrieval," *International Journal of Information Processing and Management*, Vol. 36, Issue 5, 2000, pp. 737-759.
- [20] Joentgen A., Mikenina L., Weber R. and Zimmermann H. J., "Dynamic fuzzy data analysis based on similarity between functions," *Fuzzy Sets and Systems*, Vol. 105, 1999, pp. 81-90.
- [21] Kang, B. J. and Choi, K. S., "Effective foreign word extraction for Korean information retrieval," *International Journal of Information Processing and Management*, Vol. 38, Issue 1, 2002, pp. 91-109.
- [22] Karacapilidis, N. and Pappis, C., "Computer-supported collaborative argumentation and fuzzy similarity measures in multiple criteria decision making," *Computers and Operations Research*, Vol. 27, Issue 7-8, 2000, pp. 653-671.
- [23] Karypis, G., "Evaluation of item-Based Top-N Recommendation Algorithms," *Proceedings of the Tenth International Conference on Information and Knowledge Management ( CIKM )*, Atlanta, 2001.
- [24] Kim, J. G. and Lee, E. S., "Intelligent Information Recommend System on the Internet," *International Wordshop on Industrial Applications on Network Computing ( INDAP'99 )*, 1999.
- [25] Kim, M. C. and Choi, K. S., "A comparison of collocation-based similarity measures in query expansion," *International Journal of Information Processing and Management*, Vol. 35, Issue 1, 1999, pp. 19-30.
- [26] Lanquillon, C., "Information Filtering in Changing Domains," *Workshop on Machine Learning for Information Filtering, International Joint Conference on Artificial Intelligence ( IJCAI' 99 )*, 1999, pp. 41-48.
- [27] Larsen, B. and Aone, C., "Fast and effective text mining using linear-time document clustering," *In Proc. of the Fifth ACM SIGKDD Int'l Conference on Knowledge Discovery and Data Mining*, 1999, pp. 16-22.
- [28] Mitra, P., Murthy, C. A. and Pal, S. K., "Unsupervised Feature Selection Using Feature Similarity," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 3, March 2002, pp. 301-312.
- [29] Motta, C. L. R. and Borges, M. R. S., "A cooperative approach for information recommendation and filtering," *Proceedings of the International Workshop on Groupware, IEEE Computer Society, Madeira, Portugal, October 2000*, pp. 42-49.
- [30] Nie, J. Y., Brisebois, M. and Ren, X., "On Chinese text retrieval," *Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval*, August 1996, pp. 225-233.
- [31] Nie, J. Y., Gao, J., Zhang, J. and Zhou, M., "On the Use of Words and N-grams for Chinese Information Retrieval," *Proceedings of the 15th international workshop on Information retrieval with Asian languages*, 2000, pp. 141-148.
- [32] Oard, D. W. and Marchionini, G., "A Conceptual Framework for Text Filtering," *Technical Report CS-TR3643, University of Maryland, College Park, MD, May 1996*.
- [33] Sarwar, B., Karypis, G., Konstan, J. and Riedl, J., "Analysis of Recommendation Algorithms for E-Commerce," *Proceedings of the 2nd. ACM conference on Electronic commerce*, 2000, pp. 158-167.
- [34] Schafer, J. B., Konstan, J. and Riedl, J., "Recommender Systems in E-Commerce," *Proceedings of the first ACM conference on Electronic commerce*, 1999, pp. 158-166.
- [35] Sproat, R. and Shih, C., "A Statistical Method for Finding Word Boundaries in Chinese Text," *Computer Processing of Chinese and Oriental Languages*, 1990, pp.336-351.
- [36] Wang, W. J., "New similarity measures on fuzzy sets and on elements," *Fuzzy Sets and Systems*, Vol. 85, Issue 3, 1997, pp. 305-309.
- [37] Wu, M., Fuller, M. and Wilkinson, R., "Using clustering and classification approaches in interactive," *International Journal of Information Processing and Management*, Vol. 37, Issue 3, 2001, pp. 459-484.
- [38] Yuan, S. T. and Liu, A., "Next-generation agent-enabled comparison shopping," *International Journal of Expert Systems with Applications*, Vol. 18, Issue 4, 2000, pp. 283-297.