

Using Affinity Set and Data Mining to Judge the Necessity of Head Computer Tomography for Patients

何韋蓁、黃開義、陳郁文

E-mail: 9707306@mail.dyu.edu.tw

ABSTRACT

According to statistics from the Department of Health, cerebral vascular disease had been the second-largest cause of death in Taiwan over the last few years, and head injuries leading to death is significant in fatal accidents. In order to diagnose a patient's brain trauma damage correctly, a head-computed tomography scan (head CT scan) is necessary. In the emergency room, doctors have to determine quickly and correctly whether a patient needs a head CT scan to save his or her life. Therefore, this research employs patients from the Emergency Trauma Department at Chung-Ho Memorial Hospital at Kaohsiung Medical University who have had a head CT scan before as research subjects. This research adopts Affinity Set Theory to construct a model for data mining to deduce important attributes and rules, and to support doctors make judgments. The judgment accuracy rate of the affinity set model proposed in this research is high as 99.4%, higher than those by neural network, logistic regression model, rough set model or support vector machine (SVM) model. Therefore, the model based on the basis of the affinity set can provide doctors decision support to assess whether a head-injury patient needs a head CT scan, so that the diagnosis efficiency of head CT scans can be improved.

Keywords : Data Mining ; Affinity Set ; Neural Network ; Logistic Regression ; Rough Set

Table of Contents

封面內頁	簽名頁	博碩士論文暨電子檔案上網授權書	iii	中文摘要	iv	ABSTRACT	v	誌謝	vi	目錄	vii	圖目錄	x	表目錄	xi																																																																																																																																																																																																		
第一章 緒論	1	1.1 研究背景與動機	1	1.2 研究目的	3	1.3 研究範圍	3	1.4 研究流程及內容	4	第二章 文獻探討	7	2.1 電腦斷層掃描	7	2.2 資料探勘	8	2.2.1 資料探勘的功能	9	2.2.2 資料探勘的應用	10	2.2.3 資料探勘在醫學上的應用	12	2.3 緣集合	15	2.3.1 緣集合的應用	18	2.4 類神經網路	19	2.4.1 類神經網路在醫療上的應用	21	2.5 羅吉斯迴歸	22	2.5.1 羅吉斯迴歸在醫療上的應用	22	2.6 粗略集合	23	2.6.1 粗略集合的定義	23	2.6.2 粗略集合的應用	25	2.7 支援向量機	26	2.7.1 SVM基本表示法	26	2.7.2 圖形分類	27	2.7.3 特徵空間學習	28	2.7.4 支援向量機之應用	29	2.8 接受者特徵操作曲線	30	2.8.1 ROC曲線的應用	32	第三章 研究方法	33	3.1 研究流程與架構	33	3.2 緣集合模型的建構	35	3.3 倒傳遞類神經網路模型的建構	38	3.4 羅吉斯迴歸的模型建構	40	3.5 粗略集合的模型建構	41	3.6 支援向量機的模型建構	43	第四章 實例分析	45	4.1 研究對象與資料收集	45	4.1.1 傷患屬性與編碼	45	4.2 實際分析結果	47	4.2.1 緣集合分析結果	47	4.2.2 倒傳遞類神經網路分析結果	55	4.2.3 羅吉斯迴歸分析結果	63	4.2.4 粗略集合分析結果	69	4.3 支援向量機分析結果	73	4.4 分析結果比較	77	4.5 規則解釋	78	第五章 結論與建議	82	5.1 結論	82	5.2 後續研究及建議	83	參考文獻	85	附錄A	92	附錄B	94	附錄C	96	附錄D	99	圖目錄	圖1.1 研究流程圖	6	圖2.1 緣集合模型	圖17	圖2.2 類神經網路架構圖	圖20	圖2.3 SVM說明圖	圖27	圖2.4 空間轉換圖	圖29	圖2.5 ROC曲線圖	圖31	圖3.1 研究分析流程圖	圖34	圖3.2 nntool之主要視窗	圖39	圖3.3 羅吉斯迴歸之視窗	圖41	圖3.4 粗略集合之視窗	圖42	圖3.5 SVM之視窗	圖44	圖4.1 不同變數組合下的最佳緣度(80/20)	圖48	圖4.2 不同變數組合下的最佳緣度(70/30)	圖51	圖4.3 不同變數組合下的最佳緣度(60/40)	圖53	圖4.4 五種分析方法之ROC曲線	圖78	表目錄	表2.1 資料探勘技術的應用領域	表11	表2.2 資料探勘應用於醫療產業論文整理	表12	表2.3 混沌矩陣	表31	表3.1 假設病患資料	表37	表3.2 猜想規則	表37	表4.1 屬性編碼	表46	表4.2 緣集合規則(80/20)	表48	表4.3 緣集合理論之混沌矩陣(80/20)	表49	表4.4 緣集合規則(70/30)	表51	表4.5 緣集合混沌矩陣70/30	表52	表4.6 緣集合規則(60/40)	表54	表4.7 緣集合混沌矩陣60/40	表54	表4.8 4個神經元倒傳遞類神經網路之混沌矩陣(80/20)	表56	表4.9 8個神經元倒傳遞類神經網路之混沌矩陣(80/20)	表57	表4.10 16個神經元倒傳遞類神經網路之混沌矩陣(80/20)	表57	表4.11 4個神經元倒傳遞類神經網路之混沌矩陣(70/30)	表59	表4.12 8個神經元倒傳遞類神經網路之混沌矩陣(70/30)	表59	表4.13 16個神經元倒傳遞類神經網路之混沌矩陣(70/30)	表60	表4.14 4個神經元倒傳遞類神經網路之混沌矩陣(60/40)	表61	表4.15 8個神經元倒傳遞類神經網路之混沌矩陣(60/40)	表62	表4.16 16個神經元倒傳遞類神經網路之混沌矩陣(60/40)	表62	表4.17 羅吉斯迴歸逐一選擇屬性說明表(80/20)	表64	表4.18 羅吉斯迴歸之混沌矩陣(80/20)	表65	表4.19 羅吉斯迴歸逐一選擇屬性說明表(70/30)	表66	表4.20 羅吉斯迴歸之混沌矩陣(70/30)	表66	表4.21 羅吉斯迴歸逐一選擇屬性說明表(60/40)	表68	表4.22 羅吉斯迴歸之混沌矩陣(60/40)	表68	表4.23 粗略集合之混沌矩陣(80/20)	表70	表4.24 粗略集合之混沌矩陣(70/30)	表71	表4.25 粗略集合之混沌矩陣(60/40)	表72	表4.26 支援向量機之混沌矩陣(80/20)	表73	表4.27 支援向量機之混沌矩陣(70/30)	表74	表4.28 支援向量機之混沌矩陣(60/40)	表76	表4.29 五種分析方法的預測正確率比較	表77	表4.30 規則縮減範例	表79	表4.31 決策屬性為1之規則	表79

REFERENCES

- 中文部份 [1] 鄭國順, 透視人體的利器 電腦斷層掃描系統, 科學發展, 382期, 52-61, 2004。
- [2] 行政院衛生署: http://www.doh.gov.tw/CHT2006/index_populace.aspx [3] 黃勝堅, 輕度頭部外傷與腦震盪後症候群, 台大醫療網, 24期, 11月, 2004。
- [4] 紀煥庭, 輕度頭部外傷-常見, 但可能致死的傷害, 台中榮民總醫院急診通訊第七卷第十一期, 2006。
- [5] 劉斯顯, 淺談頭部外傷, 高雄榮民總醫院醫訊第五卷第十一期, 2002。
- [6] 中央健保局: <http://210.69.214.131/index.asp> [7] 霍爾(E.J.Hall)和布倫納(D.J.Brenner), 新英格蘭醫學期刊, 2007 [8] 廖興國, 頭部外傷治療準則, 教您如何緊急處理頭部外傷, KungNet 國家網路醫院, 2007。
- [9] 廖元麟、孫永年, 透視人體的奧秘, 行政院國家科學委員會科普知識, 2005。
- [10] 行政院衛生署疾病管制局: http://www.cdc.gov.tw/file/38980_6723263889119-電腦斷層掃描.doc [11] 許明暉, 以頭部外傷病人之臨床變數預測頭部斷層掃描結果, 台北醫學院醫學資訊研究所碩士論文, 1999。
- [12] 楊旭峰, 以類神經網路預測輕度頭部外傷後電腦斷層是否出現異常之分析, 台北醫學大學傷害防治學研究所碩士論文, 2006。
- [13] 丁一賢, 資料探勘, 滄海書局。
- [14] 謝邦昌, 資料探勘在統計上的應用, 2001。
- [15] 藍中賢, 結合模糊集合理論與貝氏分類法之資料探勘技術--應用於健保局醫療費用審查作業, 元智大學資訊管理研究所碩士論文, 2000。
- [16] 林志杰, 應用資料探勘技術建立中西醫腦中風模型之研究, 長庚大學資訊管理研究所碩士論文, 2006。
- [17] 姚志成, 運用資料探勘技術建構脂肪肝預測模式, 中原大學資訊管理研究所碩士論文, 2005。
- [18] 吳充平, 資料探勘技術於台灣地區國人健康狀況之研究, 南台科技大學碩士論文, 2005。
- [19] 左涵瀧, 資料探勘在心血管疾病預測之應用, 南台科技大學碩士論文, 2005。
- [20] 周欽凱, 利用『資料探勘技術』探討急診高資源耗用者之特性, 國立台灣大學醫療機構管理研究所碩士論文, 2004。
- [21] 陳迪祥, 以資料探勘技術發掘疾病隱藏關係之研究, 國立暨南大學資訊管理研究所碩士論文, 2003。
- [22] 陳佳楨, 資料探勘應用於就診行為與醫師排班之研究 以埔里基督教醫院為例, 國立暨南大學資訊管理研究所碩士論文, 2003。
- [23] 陳瑞杰, 探討外傷系統之設計-資料探勘分析法於健康資料庫之應用, 台北醫學大學醫學資訊研究所碩士論文, 2003。
- [24] 廖雅郁, 應用資料探勘於我國西藥行銷之研究, 國立交通大學經營管理研究所碩士論文, 2001。
- [25] 吳智誠, 資料探勘於影像資訊之應用-以乳房微鈣化特徵處理為案例, 大葉大學工業工程研究所碩士論文, 2001。
- [26] 施庭芳, 醫學影像與健康, 國立台灣大學醫學院, 通識課程-醫學與生活(02), 單元一, 2006。
- [27] 黃肇偉, 整合粗略及不與緣集合理論於資料探勘在生技產品生命週期與銷售之研究, 大葉大學工業工程與科技管理研究所碩士論文, 2007。
- [28] 沈智敏, 應用緣集合理論於延遲性診斷關鍵屬性之研究, 大葉大學工業工程與科技管理研究所碩士論文, 2007。
- [29] 彭士晏, 應用類神經網路建構手術後噁心嘔吐之預測模型, 國立陽明大學衛生資訊與決策研究所碩士論文, 2006。
- [30] 葉怡成, 應用類神經網路, 儒林圖書公司, 2001。
- [31] 黃亞軒, 類神經網路之規則萃取應用於中醫線上診療系統, 中原大學資訊工程研究所碩士論文, 2000。
- [32] 王健亞, 類神經網路在臨床心理醫療費用精算模型之應用, 逢甲大學統計與精算研究所碩士論文, 2000。
- [33] 李建誠, 互動式之遠距腹腔醫療診斷系統, 國立成功大學電機工程學系博士論文, 2003。
- [34] 康惠雯, 邏輯斯迴歸分析在資料探勘上的應用, 東海大學統計研究所碩士論文, 2002 [35] 呂奇傑, 演化式類神經網路分類技術於資料探勘上之應用, 輔仁大學應用統計研究所碩士論文, 2001。
- [36] 陳鴻星, 台灣地區醫生治療氣喘之現象討論, 台北大學統計系碩士論文, 2003。
- [37] 林伯聲, 影響病患置中央健保局聯合門診中心就醫行為之研究, 高雄醫學大學公共衛生研究所碩士論文, 2002。
- [38] 林信成, 應用全民健保資料庫分析台灣地區糖尿病患者之醫療服務利用, 國立交通大學管理學院經營管理組碩士論文, 2005。
- [39] 陳利鈺, 以約略集合理論進行知識擷取, 國立台灣海洋大學航運管理研究所碩士論文, 2002。
- [40] 黃俊霖, 應用約略集合理論與模糊理論於台股指數期貨漲跌幅預測之實證研究, 台灣科技大學資訊管理研究所碩士論文, 2004。
- [41] 賴威利, 利用約略集合理論預測燒燙傷患者死亡率, 南台科技大學國際企業研究所碩士論文, 2005。
- [42] 雷賀君, 前十字韌帶傷害快速診斷系統-以粗略集合、基因演算法與倒傳遞網路為工具, 大葉大學工業工程預科技管理研究所碩士論文, 2004。
- [43] 李佩玲等, 以基因演算法建置不同風險接受度之投資組合, 電子商務與數位生活研討會, 2006。
- [44] 黃啟信, 應用資料修補機制技術於醫療環境之研究, 東海大學工業工程與經營資訊研究所, 2003。
- [45] 陳文山, 結合SVM技術與統計方法於骨質密度資料之分析及預測, 大葉大學工業工程與科技管理研究所碩士論文, 2005。
- [46] 蘇育成, 使用支援向量機和極端值偵測方法進行乳房腫瘤影像分類, 長庚大學資訊管理研究所碩士論文, 2006。
- [47] 吳忠原, 結合粗略集合論、支援向量機及最佳化演算法於財務系統之應用, 大葉大學工業工程與科技管理研究所碩士論文, 2006。
- [48] 黃玉櫻, 結合粗略集合論、支援向量機及最佳化演算法於製造系統之應用, 大葉大學工業工程與科技管理研究所碩士論文, 2006。
- [49] 施毓森, 銀行放款訂價模型之探討-ROC分析法之應用, 東吳大學會計系碩士論文, 2007。

- [50] 張緯翎，類神經網路與羅吉斯迴歸分析方法對於車體損失險出險預測能力之比較，逢甲大學保險學系碩士論文，2005。
- [51] 林佳慧，全民健保醫療審查費用之研究，朝陽科技大學保險金融管理系碩士論文，2004。
- [52] 陳麗琴，中文版五級急診檢傷分類電腦化系統之建構與臨床應用評估，台北醫學大學護理研究所碩士論文，2005 [53] 吳中倫，陳郁文，陳昭文，應用緣集合理論於提昇頭部斷層掃描效益之研究，2007。
- [54] 余尚武、賴珮君灰色系統、模糊理論與約略集理論於權變投資組合保險策略之應用，中華管理評論國際學報，Vol. 10, No. 4, 2007。
- [55] 吳明隆，SPSS問卷統計分析與實務，五南圖書出版公司，2007。英文部分: [56] Berry, M. J. A., and Linoff, G., Data Mining Techniques: for Marketing, Sales, and Customer Support, New York, John Wiley and Sons Inc., 1997。
- [57] Chen, Y. W. and Larbani, M., " Developing the Affinity Set and Its Applications, " Proceeding of the Distinguished Scholar Workshop by National Science Council, Jul. 14-18, 2006, Taiwan.
- [58] Ho, D.Y.F., " Interpersonal Relationships and Relationship Dominance: An Analysis Based on Methodological Relationism, " Asian Journal of Social Psychology, Vol. 1, No. 1, pp. 1-16, 1998.
- [59] Hwang, K, " Face and Favor: The Chinese Power Game, " The American Journal of Sociology, Vol. 92, No. 4, pp. 944-974, 1987.
- [60] Larbani, M. and Chen, Y. W., " Affinity Set and Its Applications, " Proceeding of the International Workshop on Multiple Criteria Decision Making, Apr. 14-18, 2007, Poland.
- [61] Pawlak, Z., " Rough Sets " , International Journal of Computer and Information Science, Vol. 11, pp. 341-356, 1982 [62] Pawlak Z. (1982). Rough sets. International Journal of Computer Science, 11,341 – 356.
- [63] Richard, J. R., and W. G.. Michael, "Data Mining a Tutorial-based Primer," 1st ed., Addison Wesley, New York, 2002, pp.4-27.
- [64] Woods ,K. and Bowyer, K. W.(1997), " Generating ROC curves for artificial neural networks, " IEEE Trans. Medical Imaging, vol. 16, no. 3, pp.329-337.
- [65] Moussa, L. and Chen, Y. W., A fuzzy Set Based Framework for the Concept of Affinity.
- [66] Moussa, L. and Chen, Y. W., Developing the Affinity Set (or Guanxi Set) Theory and Its Applications.
- [67] Kusiak, A. Dixon, B. and, Shah ,S. "Predicting survival time for kidney dialysis patients: a data mining approach" Elsevier journal of Computers in Biology and Medicine,35,311-327 [68] Kusiak, A., Caldarone , C. A. Kelleher , M.C. Lamb, F. S. Persoon , T. J. and Burns, A. "Hypoplastic left heart syndrome :knowledge discovery with a data mining approach" Elsevier journal of Computers in Biology and Medicine,36, 21-40 [69] Vapnik , V., "The Nature of Statistical Learning Theory" New York, Springer- Verlag (1995) [70] Vapnik, V.N., " An overview of statistical learning theory, " IEEE Transactions on Neural Networks, 10(5), 988 -999 (1999)