

利用緣集合與資料探勘於急診病患回診規則之研究

李宗鴻、陳郁文

E-mail: 9806125@mail.dyu.edu.tw

摘要

資料探勘能從資料中挖掘出所隱藏的實用價值，提供決策者作為參考；而且面對急診室如此忙碌的醫療環境下，如何提供醫療人員有效的參考依據以提升醫療品質降低回診頻率是一件相當重要的課題。本研究利用緣集合理論、倒傳遞類神經網路、粗略集合理論、支援向量機、決策樹及關聯規則等六種方法，依據各種模型分析資料後，經由ROC (Receiver Operating Characteristic)曲線比較找出預測急診病患回診規則能力佳的模型。研究結果發現，支援向量機的預測分類能力為最佳，其次則是緣集合理論，且均能達到80%，但是支援向量機無法對變數進行解釋的能力。因此本研究以選擇緣集合模型做為預測說明，提供醫生或醫療人員在診視病患時，以緣集合分析結果所產生的判中率較高之規則庫作為診斷時的參考依據，以增加正確診斷率及減少回診的發生降低醫療資源的浪費。

關鍵詞：資料探勘、緣集合理論、ROC曲線、醫療

目錄

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 2

1.3 研究範圍及限制 3

1.4 研究流程及內容 3

第二章 文獻探討 6

2.1 非計劃性回診 7

2.2 資料探勘(DATA MINING) 8

2.2.1 資料探勘的功能 9

2.2.2 資料探勘在醫療上的應用 10

2.3 資料探勘的技術 12

2.3.1 緣集合理論(Affinity Set) 12

2.3.2 倒傳遞類神經網路(BACK-PROPAGATION NEURAL NETWORK) 15

2.3.3 支援向量機(Support Vector Machine, SVM) 17

2.3.4 粗略集合(Rough Set) 18

2.3.5 決策樹(Decision Tree) 20

2.3.6 關聯規則 (Association Rules) 22

2.4 接受者操作特徵曲線 (Receiver Operating Characteristic, ROC)與應用 26

第三章 研究方法 29

3.1 研究流程架構 29

3.2 以緣集合模型建立資料探勘的建構 31

3.3 倒傳遞類神經網路模型的建構 35

3.4 支援向量機模型的建構 36

3.5 粗略集合模型的建構	37
3.6 決策樹模型的建構	38
3.7 關聯規則模型的建構	39
第四章 實例分析初步結果	41
4.1 研究對象與資料說明	41
4.1.1 傷患屬性與編碼	41
4.2 實際資料分析	43
4.2.1 緣集合分析結果	43
4.2.2 倒傳遞類神經網路分析結果	49
4.2.3 支援向量機分析結果	51
4.2.4 粗略集合理論分析結果	53
4.2.5 決策樹分析結果	55
4.2.6 關聯規則分析結果	57
4.3 分析結果比較	61
4.4 分析結果規則討論	64
第五章 結論與建議	68
5.1 結論	68
5.2 建議	69
參考文獻	71
附錄A	77
附錄B	78
附錄C-1	79
附錄C-2	80
附錄D	81
附錄E	85
附錄F	89

參考文獻

- 英文部分:[1]Agrawal, R. Imielinski, T. and A. Swami. (1993) 「Mining association rules between sets of items in large databases,」 Proceedings of the ACM SIGMOD Conference on Management of data, p.p. 207-216.
- [2]Berry M., J., A., Linoff, G. (1997) 「Data Mining Techniques: For Making Sales and Customer Supports,」 John Wiley and Sons, Inc..
- [3]Chae, Y.M., Kim, H.S., Tark, K. C., Park, H. J., Ho, S. H. (2003) 「Analysis of healthcare quality indicator using data mining and decision support system,」 Expert System with Applications, 24, pp.167-172.
- [4]Chen, Y. W ,and Larbani, M., (2006) 「Developing the Affinity Set and Its Applications,」 Proceeding of the Distinguished Scholar Workshop by National Science Council, Jul. pp.14-18.
- [5]Chen, Y. W., Moussa, L., Hsieh, C. Y., and Chen, C. W. (2009) 「Introduction of affinity set and its application in data-mining example of delayed diagnosis,」 Expert Systems with Applications, 36, pp. 10883 – 10889.
- [6]Christy, T., (1997), 「Neural Networks: Not Just a Black Box,」 Insurance & Technology, pp. 30-32.
- [7]Delen, D., and Walker, G., and Kadam, A., (2005) 「Predicting breast cancer survivability: a comparison of three data mining methods」, Artificial Intelligence in Medicine. Vol. 34, 2, pp.113-127[8]Ernest P.G. and R. Harish (1998) 「Survival Prediction in the Intensive Care Unit: a Comparison of Neural Network and Binary Logit Regression,」 Socio Economic Planning Science, pp. 189-198[9]Evaline A , Alessandrini MD, et al, (2004) 「Return visits to a pediatric emergency department,」 Pediatric Emergency Care, 20, pp. 166-171.
- [10]Han, J. and Kamber, M., (2001) 「Data Mining: Concepts and Techniques,」 Morgan Kaufmann.
- [11]Ho, D.Y.F., (1998) 「Interpersonal Relationships and Relationship Dominance: An Analysis Based on Methodological Relationism,」 Asian Journal of Social Psychology, Vol. 1, No.1, pp. 1-16.
- [12]Huang, Y.L., and Chen, D.R. (2005) 「Support vector machines in sonography: Application to decision making in the diagnosis of breast cancer,」 Journal of Clinical Imaging, pp. 179-184[13]Hwang, K. (1987) 「Face and Favor: The Chinese Power Game,」 The American Journal of Sociology, Vol. 92, No. 4, pp. 944-974.
- [14]Jerzy, M. (2008) 「Credit Scoring Model Base on The Affinity Sst,」 conference of ICEIS.
- [15]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, pp 311-327.
- [16]Larbani, M. and Chen, Y. W., (2007) 「Affinity Set and Its Applications,」 Proceeding of the International Workshop on Multiple Criteria

Decision Making, Apr.14-18, Poland.

- [17]Li, Y. C., Liu, L., Chiu, W. T., and Jian, W. S. (2000) 「Neural network modeling for surgical decisions on traumatic brain injury patients,」 International Journal of Medical Informatics, 57, pp. 1 – 9[18]Moussa, L. and Chen, Y. W. (2007), 「A fuzzy Set Based Framework for the Concept of Affinity」.
- [19]Pawlak Z. (1982) Rough sets. International Journal of Computer Science, 11, pp. 341 – 356.
- [20]Quinlan J.Ross., (1993) C4.5: Programs for machine learning, Morgan Kaufmann Publishers.
- [21]Rosetta Software: <http://rosetta.lcb.uu.se/general/>[22]Suykens, J. A. K., Van Gestel, T., De Brabanter, J., De Moor, B., and Vandewalle, J. (2002) 「Least Squares Support Vector Machines,」 World Scientific, Singapore.
- [23]Vapnik, V. N. (1998) 「Statistical Learning Theory. Adaptive and learning systems for signal processing, communications, and control,」 Wiley, New York.
- [24]Wang, H. Y., Chew, G., Kung, C. T., Chung, K. J., and Lee, W. H., (2007) 「The Use of Charlson Comorbidity Index for Patients Revisiting the Emergency Department within 72 Hours,」 Chang Gung Med J,30, pp 437-44[25]Yen, Z. S., Shih, F. Y., Tasi, K. C., Chen, S. C., and Chen, W. J., (2003) 「Revisits to the Emergency Department: Analysis of Risk Factors and Reasons,」 J Emerg Crit Care Med. Vol.14, 3, pp. 93-98中文部份:[26]丁一賢, 資料探勘, 滄海書局。
- [27]謝邦昌, 資料探勘在統計上的應用, 2001。
- [28]葉怡成, 「類神經網路模式應用與實作」, 儒林圖書股份有限公司, 臺北。
- [29]丁修文, 急診醫療結果面品質影響因素之探討 - 以某區域教學醫院兒科氣喘急診病患為例, 國立中山大學醫務管理研究所碩士論文, 2004。
- [30]余尚武、賴珮君, 「灰色系統、模糊理論與約略及合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, 第10卷4期, 2007。
- [31]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。
- [32]李雅雯, 類神經網路輔助醫療診斷分類模式之建構, 國立勤益科技大學工業工程與管理系碩士論文, 2006。
- [33]何韋蕓, 應用緣集合與資料探勘於判斷需要頭部電腦斷層掃描病患之研究, 大葉大學工業工程與管理學系碩士論文, 2008。
- [34]吳中倫, 結合粗略集合論及基因演算法於提昇頭部斷層掃描效益之研究, 大葉大學工業工程與管理學系碩士論文, 2007。
- [35]林志傑, 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。
- [36]林明昌, 利用資料探勘技術探討中央健康保險局中區分局資料庫門診高就診次者之特性, 國立中正大學資訊管理研究所碩士論文, 2007。
- [37]林瓊茹, 以醫師觀點探討急診病患再回診原因之質性研究, 義守大學醫務管理系碩士論文, 2008。
- [38]沈智敏, 「應用緣集合於延遲性診斷關鍵屬性之研究」, 大葉大學工業工程與管理學系碩士論文, 2007。
- [39]周啟凱, 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。
- [40]洪冠群, 多重最小支援度關聯規則探勘演算法之醫療檢驗應用:以血液透析病患之住院預測為例, 國立東華大學資訊工程學系碩士在職班論文, 2004。
- [41]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。
- [42]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。
- [43]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。
- [44]陳佳楨, 資料探勘應用於就診行為與醫師排班之研究 - 以埔裡基督教醫院為例, 國立暨南大學資訊管理研究所碩士論文, 2003。
- [45]黃昱銘, 有效率地探勘疾病和病症之複合項關連規則, 南台科技大學資訊管理學系碩士論文, 2004。
- [46]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。
- [47]黃勝崇, 資料探勘應用於醫療院所輔助病患看診指引之研究, 南華大學資訊管理學系碩士班 碩士論文, 2001。
- [48]黃仁貴, 以叢集分析技術探討病患就診屬性與看診時間之關係, 台北醫學院醫學資訊研究所碩士論文, 2002。
- [49]黃宗賢, 急診病患非計劃性返診之相關因素探討, 長庚大學醫務管理研究所碩士論文, 2006。
- [50]黃燕美, 支援向量機應用於醫學健檢之肝病分類, 國立高雄第一科技大學資訊管理研究所碩士論文, 2006。
- [51]黃堅展, 結合基因演算法及神經網路應用於急診病患回診次數原因之研究, 大葉大學工業工程與管理學系碩士論文, 2007。
- [52]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。
- [53]楊景晴, 整合決策樹與關聯規則之資料挖礦架構及其實證研究, 國立清華大學工業工程與工程管理學系碩士論文, 2004。
- [54]趙雲瀚, 以資料探勘分析氣候因素對蔬菜供給量之影響, 南華大學資訊管理學系碩士班 碩士論文, 2001。
- [55]應哲磊, 使用支援向量機預測台灣期貨指數, 華梵大學資訊管理學所碩士論文, 2006。
- [56]蕭柏宏, ROC 曲線之應用:以房貸之違約與提前清償為例, 國立高雄第一科技大學風險管理與保險所碩士論文, 2005[57]謝政彥, 利用緣集合在降低延遲性診斷之研究, 大葉大學工業工程與管理學系碩士論文, 2008。
- [58]台灣醫療品質指標計劃之急診結果面醫療品質指標[59]醫策會. (民96). 新制醫院評鑑基準及評分說明之第八章人力素質及品質促進 Retrieved 9/15, 2007, from <http://www.tjcha.org.tw/Public/HotNews/2007491443517607.pdf>