

Using Affinity Set on Reducing Delayed Diagnosis

謝政彥、陳郁文

E-mail: 9707307@mail.dyu.edu.tw

ABSTRACT

According to "Institute of Medical investigation report", a half above percentage(58%) is a mistake which is caused by doctors can avoid in the medical injury, because medical care is risky. In Taiwan, medical events with delayed diagnosis occur increasingly. When facing the busy hospital environment, the examining time that a doctor has is limited. Thus, how to face patients to make the correct diagnosis is a quite important thing to doctors. This research uses Affinity Set, Rough Set, Back-propagation Neural Network, Logistic Regression, and SVM as tools of data mining to classify and analyze the relation within medical data. This research uses Receiver Operating Characteristic(ROC) curve to compare forecasting abilities with these five methods. Studying results indicate the forecast ability of SVM is the best, and the second best is Affinity. However, SVM cannot explain variables, this research chooses Affinity Set to be the final forecasting model. Affinity Set reaches 90%. Blood pressure and pulse are the frequent attributions. Therefore, when doctors diagnose patients, they should take these two attributions first. Affinity Set with high hit rate generated rules to offer doctors a reference when doctors make a diagnosis to increase accuracy and decrease delayed diagnosis.

Keywords : delayed diagnosis ; data mining ; affinity set ; rough set ; back-propagation neural network ; logistic regression ; SVM ; ROC curve

Table of Contents

封面內頁 簽名頁 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 研究範圍及限制 3 1.4 研究流程 3 1.5 研究方法 6 第二章 文獻探討 8 2.1 資料探勘在醫療上的應用 8 2.2 緣集合理論(Affinity Set) 11 2.2.1 以緣集合建立資料探勘模型 14 2.3 粗略集合論(Rough Set) 16 2.3.1 粗略集合論的基本概念 16 2.4 倒傳遞類神經網路(Back-propagation Neural Network) 18 2.5 羅吉斯迴歸(Logistic Regression) 20 2.6 支援向量機(Support Vector Machine) 20 2.7 接受者特徵操作曲線(Receiver Operating Characteristic) 21 第三章 研究方法 24 3.1 研究流程及架構 24 3.2 緣集合模型的架構 26 3.3 粗略集合論的架構 28 3.4 倒傳遞類神經網路的架構 29 3.5 羅吉斯迴歸的架構 30 3.6 支援向量機的架構 32 第四章 實例驗證 33 4.1 研究資料說明 33 4.2 實際資料分析 34 4.2.1 緣集合分析結果 34 4.2.2 粗略集合理論分析結果 40 4.2.3 倒傳遞類神經網路分析結果 43 4.2.4 羅吉斯迴歸分析結果 50 4.2.5 支援向量機分析結果 54 4.3 分析方法結果比較 55 4.4 規則之討論 56 第五章 結論與建議 59 5.1 結論 59 5.2 建議 60 參考文獻 61 附錄A 64 附錄B 67

REFERENCES

- 英文部分: [1] Berry M., J., A., Linoff, G. (1997), "Data Mining Techniques: For Making Sales and Customer Supports," John Wiley and Sons, Inc..
- [2] Baxt, W.G., "Using an Artificial Neural Network for Data Analysis in Clinical Decision-Making: The Diagnosis for Acute Coronary Occlusion," Neural Computayion, Vol. 2, pp. 480-489 [3] Ernest P.G. and R. Harish (1998), "Survial Prediction in the Intensive Care Unit: a Comparision of Neural Network and Binary Logit Regression," Socio Economic Pianning Science, pp.189-198.
- [4] Hung, Y.I., and D.R. Chen (2005), "Support vector machines in sonography: Application to decision making in the diagnosis of breast cancer," Journal of Clinical Imaging, pp.179-184.
- [5] Jerzy M. (2008), "Credit Scoring Model Base on The Affinity Sst," conference of ICEIS.
- [6] Kohn, L. T., J. M. Corrigan and M. S. Donaldson (eds) (2000), To Err is Human: Building a Safer Health System. Institute of Medicine Report. Washington, DC: National Academy Press.
- [7] Moussa, L. and Chen, Y. W. (2007), A fuzzy Set Based Framework for the Concept of Affinity.
- [8] Moussa, L. and Chen, Y. W. (2007), Developing the Affinity Set (or Guanxi Set) Theory and Its Applications.
- [9] Pawlak, Z. (1982), "Rough Sets", International Journal of Computer and Information Science, Vol. 11, pp. 341-356.
- [10] Rosetta Software: <http://rosetta.lcb.uu.se/general/>.
- [11] Vapnik, V. N. (1998), "Statistical Learning Theory. Adaptive and learning systems for signal processing, communications, and control." Wiley, New York.
- [12] Woods K. and K. W. Bowyer, (1997), "Generating ROC curves for artificial neural networks," IEEE Trans. Medical Imaging, vol. 16, no.

- [13] WD(Wrong Diagnosis): <http://www.wrongdiagnosis.com/>. 中文部份: [14] 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班, 2007。
- [15] 余尚武、賴珮君, 「灰色系統、模糊理論與約略及合理論於權變阻呵保險策略之應用」, 中華管理評論國際學報, 第10卷4期, 2007。
- [16] 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。
- [17] 許夙君, 「從臨床治理觀點探討醫學中心病人安全的風險因素」國立成功大學高階管理碩士論文, 2005。
- [18] 陳世源, 「資料探礦技術在病例與藥品關聯性之研究」, 國立中山大學資訊管理研究所碩士論文, 2000。
- [19] 陳佩欣, 「結合資料探勘與約略集合論建構以顧客問題為導向之動態常見問題集」, 中華大學資訊管理系碩士班碩士論文, 2004。
- [20] 陳郁文、Moussa Larbani、吳中倫、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。
- [21] 郭乃文, 「菸、酒、檳榔使用行為與憂鬱症狀之相關性探討」, 台北醫學大學醫務管理研究所碩士論文, 2007。
- [22] 張惠真, 「醫院藥局藥品調劑疏失之探討 以南部某醫學中心為例」, 國立成功大學碩士論文, 2005。
- [23] 許瑞文, 「地理資訊系統 類神經網路土石流潛勢判定方法」, 國立交通大學土木工程系碩士班碩士論文, 2003。
- [24] 黃昱銘, 「有效率地探勘疾病和病症之複合項關聯規則」, 南台科技大學資訊管理研究所碩士論文, 2004。
- [25] 黃育正, 「營建公司違約機率預測模型之比較研究」, 國立交通大學土木工程學系碩士論文, 2007。
- [26] 黃肇偉, 「整合粗略集合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管理碩士論文, 2007。
- [27] 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, 2006。
- [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] 羅均緯, 「應用約略集合理論預測交通事故嚴重程度之研究」, 南台科技大學碩士論文, 2005。
- [30] 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。
- [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] 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」, 台北醫學大學護理學研究所碩士論文, 2005。
- [33] 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, 2004。
- [34] 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」, 國立暨南國際大學資訊管理研究所碩士論文, 2003。
- [35] 吳建興, 「以約略集合與決策樹萃取危險因子-以逆流性食道炎為例」, 華梵大學資訊管理學系碩士班, 2002。