論文名稱：應用詞頻以改良多元貝氏定理於文件分類之研究　　總頁數：88

校(院)所組別：中國文化大學商學院資訊管理研究所

畢業時間及提要別：100 學年度第 1 學期碩士學位論文提要

研究生：羅仁君　　指導教授：吳宏文 李中彥

論文提要內容：

多元貝氏定理(multimembership Bayesian，簡稱 MMB)近幾年曾運用於醫療、網站、郵件和文件的自動分類與推論，到目前 MMB 的在知識分類推論領域上的相關研究一直都持續進行著。而本研究是針對 MMB 文件自動分類提出以基因演算法為概念的改良方法，使用動態產生篩選門檻來達到真正完全的自動分類。

以基因演算法為概念的 MMB 改良方式「自動調適篩選門檻」來提取重要字詞進行 MMB 分類，最後也以「自動評估」取得最佳結果。經實驗發現，當類別彼此差異度大時，最佳分類準確率為 83.93%；類別彼此差異度小時，其分類準確率為 70.60%。

關鍵字：多元貝氏定理(multimembership Bayesian)、文件分類(document classification)、文件自動分類(automatic document classification)、基因演算法(genetic algorithm)。

A research on applying term frequency to improve
multimembership Bayesian theorem
on document classification

Student: Jen-Chun Lo Advisor: Prof . Homer Wu

Prof . Chong-Yen Lee

Chinese Culture University

ABSTRACT

In recent years, multimembership Bayesian (MMB) has had a wide application for medical, website, E-mail and other document processing use – the practices list above utilize the automatic classification and knowledge inference function of MMB to improve efficiency. Given MMB's practicality and popularity across all walks of lives, the research around MMB remains a constant focus academically. To further improve the strength of MMB's core automatic document classification function, our study proposes the additional application of genetic algorithm before traditional MMB.

Based on the law of probability, the extra step of genetic algorithm helps develop the "automatic adaptable screening threshold" mathematically, thus with more accuracy. Such calculation pin-points the significant, frequently-used words to form the threshold for further MMB classification. Since the application of genetic algorithm acts as the initial screen, consequently, the extracted leftovers are more precise for any document classification. Further, based on the mathematic results, the threshold leads to automatic assessment, which selects the most desirable word choice automatically.

The research presents significantly improved results. When class differences are relatively in large degree, its classification accuracy achieves 83.93%. Even when class differences are to a lesser extent, its classification accuracy rate of is able to reach 70.60%.

Key Words: multimembership Bayesian, document classification, automatic document classification, genetic algorithm

誌　謝　辭

原本以為研究所的日子很漫長，繁多的報告、可怕的英文、數不清且令人緊張的師生討論時間，最後還是得咬牙撐著。回頭一望才驚覺時間飛逝，終於可以結束這個煎熬。

一路走來要感謝的人很多，首先我要感謝的是我的兩位指導教授吳宏文老師與李中彥老師，有您們的支持幫助與寶貴意見，讓我在進行論文時，遇到任何問題都能迎刃而解，在此向兩位指導教授致上最高的感謝與敬意。

接著感謝的是論文口試期間陳武倚老師與王福星老師的細心指導與建議，不僅讓我受益良多，最重要的是讓我的論文能更趨完整。另外，還有百忙之中抽空前來指導論文的校外口試委員辜輝趁老師，感謝您的寶貴建議與鼓勵。

再來是研究所期間，感謝那群一同跟著老師作研究、相互扶持鼓勵的同伴們。最後，要感謝我最親愛的家人，你們的支持讓我可以沒有任何煩惱的從大學念到研究所。

承蒙諸位恩師的悉心指導，在您們的諄諄教誨下，研究所的課程與論文才能順利完成，師恩浩翰，無以回報，在此謹向諸位恩師致上最高的敬意！

內 容 目 錄

表　目　錄

圖　目　錄