論文名稱：以多元貝氏定理建構文件自動分類系統　　總頁數:86

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：97 學年度第 2 學期碩士學位論文提要

研究生：周政淇　　指導教授:李中彥

論文提要內容：

由於網際網路日益發達，使得數位化文件增加的速度越來越快，所以文件自動分類的重要性也慢慢增加，如何在更短的時間內將文件做最快與最正確的分類是文件分類領域中永遠的課題。

本系統是利用多元貝式定理，建構文件自動分類機制，可以有效的對文件進行分類。本系統的建置分為兩個階段，第一階段為訓練階段，透過建立字詞庫，再利用字詞庫針對訓練文件全文進行字詞的比對與過濾，最後在計算所有比對出字詞的 P_{ij} 值與 $\overline{P}_{ij}$ 值。第二階段為測試推論，將 P_{ij} 值與 $\overline{P}_{ij}$ 值代入多元貝式定理後便可計算出文件分別所屬類別的機率。多元貝式定理的優點在於機率的計算是可隨著已知文件的增加而逐次調整與修正，當有新樣本加入時只需局部調整某些機率值，即可得到新的分類模型，其分類模型的機動性相當高，在資料不斷的增加的情況下，可以得到較好的分類效能。實驗結果顯示本研究中使用的方法在文件分類上的正確率可以達到 95%。

關鍵字：多元貝氏(multimembership Bayesian)，文件分類(document classification)，文件自動分類(automatic document categorization)。

A Document Classification System Based on Multimembership Bayesian theorem

Student: Zeng Chi Chou Advisor: Prof. Chong-Yen Lee

Chinese Culture University

ABSTRACT

As a result of the development of Internet, it makes the increasing speed of digital documents faster. So the important of document automatic classification increases too. How to classify the documents quickly and correctly in shorter time is a very important question at the domain of document automatic classification. In this paper we establish the document automatic classification by Multimembership Bayesian. It could classify and manage documents usefully. In the training phrase, we establish the database of information word and to compare with the training documents. Finally, we compute the P_{ij} and $\overline{P}_{ij}$ to the Multi-membership Bayesian formula then we will get the probability of document belong the class. The merit of Multi-membership Bayesian is the value of probability will be modifying by the mount of document increase. We only need to modify some value of opportunity and we will get a new classification module when the new samples get in on. For this reason, the classification module has high mobility. The more documents increase, the more effective module we have.

Key Words: multimembership Bayesian, document classification, automatic document categorization

誌　謝　辭

總算將論文告了個段落，這一路走來所遇到的酸苦至今已化為煙雲，取而代之的是滿滿的感動、喜悅，以及對支持陪伴我的家人與朋友難以言喻的感激，在向未來勇敢邁步之前，我想對我的指導教授李中彥博士致上最高的感謝與敬意，李老師在本論文的寫作上給了我相當多的啟示與提點，讓我在遇到挫折時不至困於迷霧中，而其對完美的堅持及給予我們的肯定與信任，更讓我對人生有另一番的體認與想法，希望自己在未來的旅途上也能帶給身邊的人如此的溫暖與智慧。論文口試期間王美慈老師、郭乃文老師細心的指導與建議不僅使本論文更臻完滿，其對學生的提攜愛護之情亦讓人感動萬分。謹以本文獻給我最可愛也最親愛的家人，不善於情感表達的我想對你們說我愛你們，感謝你們所付出的一切。

內 容 目 錄

表　目　錄

圖　目　錄