

第五章 結論與建議

一個有效可靠的試題分類系統，可以簡化人工分類的手續。透過分類所提供分類建議，可以讓學習者更能有效的掌握學習的方向；教導者可以獲得能力項目參考，以了解應該準備的教材及教法；命題者更可以透過試題分類系統提昇命題的品質，並了解題庫中各能力項目的試題數量，是否有平均分佈。

第一節 結論

本研究延續文件分類的研究，轉化成試題分類機制，並開發一套試題分類系統 (QCS)，無論在字詞剖析工具、回歸測試及新增試題分類與三大部份，都可以得到可靠而精確的成果，關於題庫品質管控，也可以透過內因品質指數及外因品質指數的到試題品質修正的參考。成果說明如下：

一、有關於字詞剖析工具方面

Windows 98/2000 注音輸入法的相關字詞庫配合詞庫式斷詞法的中文字詞剖析工具，確實可以得到一份與 PreMOUS Word 相關的特徵關鍵字詞庫(FTB)，共得到字詞數有 415 筆。透過特徵關鍵字詞庫配合

IDF 的字詞權重函數(TWF)所構成的權重向量矩陣，運用在回歸測試上確實可以達到 100%的回歸率，可見此權重向量矩陣，確實具有 PreMOUS Word 的特徵。

二、有關於回歸測試方面

回歸測試共使用了 TF、IDF、TFIDF、WIDF、TF/WIDF 及 TFIDF² 六種字詞權重函數(TWF)，去針對 PreMOUS Word 的 91 道術科試題作回歸測試。其中除了 TF 有 73.6%的回歸率之外，其他的 TWF 都可以得到 90%以上的回歸率，而 IDF 與 WIDF 的表現更高達 100%，實驗結果如表 5-1 所示：

表 5-1 各 TWF 針對之回歸率比較表

TWF	回歸筆數	總筆數	回歸率
TF	67	91	73.63%
TFxIDF	82	91	90.11%
TF/WIDF	83	91	91.21%
TFxIDF ²	85	91	93.41%
WIDF	91	91	100.00%
IDF	91	91	100.00%

由表 5-1 的比較可以得知，使用 PreMOUS Word 所取樣的 FTB，並採用 IDF 的 TWF，可以對 PreMOUS Word 領域的術科試題得到很有效率及準確的分類。運用此 QCS，可以迅速提供給教育工作者、學習者及命題專家對任何新增的試題或既有的試題迅速的分類，並提高教

學、學習及命題效率。

三、 關鍵字詞數對回歸率的影響

由於 TF 回歸測試發現樣本試題多的能力項目，回歸率會偏低，為了解回歸率與關鍵字詞數量的關係，採用四組不同數量的樣本數去測驗其回歸率。得到的結果可以了解關鍵字詞數量增加對回歸率有提昇的效果，而其中 IDF 對於關鍵字詞的增加後所提昇的回歸率最為敏感，並且可以在較高樣本數的三組實驗中取得最高的回歸率。在第四組實驗中使用 64% 的試題作採樣，得到 588 筆關鍵字詞，IDF 亦可以達到 85.7%。

由此可以了解，IDF 在 PreMOUS Word 的術科試題樣本中，屬於表現最好的 TWF。

四、 新增試題測試部份

- (一) 本 QCS 採用 PreMOUS Word 術科試題作為樣本。使用於分析高複雜度、綜合能力項目類型的 MOCC Word 術科試題，且在能力項目數量多於 QCS 的情況之下，QCS 分類與專家分類的結果高達 72% 相同，具有中上的水準。使用於試題分類具有很高的可靠度，可以提供給教育工作者、學習者及命

題專家作為分類的參考。

(二) 試題分類錯誤可能由於下列三種因素之一分別是：

1. 能力項目定義不清或不周全
2. 試題本文的品質問題
3. TWF 的精密度不足

本實驗採用內因品質指標 Q_{in} 指出發生錯誤中缺判的程度，使用外因品質指標 Q_{out} 指出發生錯誤中誤判成分有多少。對於命題專家而言，指出錯誤 Q_{out} 會比 Q_{in} 的指標更具意義，因為內在因素的錯誤必須由改善 QCS 本身著手，命題專家無法處理這個部份，而 Q_{out} 所指出外在因素部份可以透過修改試題本文或刪除不合適的能力項目來達成。

藉此可以迅速回饋命題專家改善試題的品質建議。而如果權重向量矩陣的能力項目能夠完整定義，更可以的檢測出試題本文的內容是否正確。

第二節 建議

本研究礙於時間及許多現實的因素，仍有許多主題可提供給後續研究作為參考：

- 一、 未來研究者可利用試題分類系統作為命題品質管控的研究工具之一。
- 二、 本試題分類系統採用權重向量矩陣配合 TWF 來分類，函數本身已經具有考慮字詞出現率及普遍性的問題，而且在回歸率測試中已經可以證明具有很好的分類品質。且新增試題的分類測試也可以有 72% 的正確率。為了可以提昇到更高的分類的正確率，建議後續研究者可以透過機器學習的機制改變權重向量矩陣的權值，以衰減影響分類的關鍵字詞的權重，如不具 PreMOUS Word 特徵意義的字詞，可能因為試題需求，在試題的敘述內容中出現，而被擷取成關鍵字詞，以提昇分類的正確率。
- 三、 由圖 4-17 可以發現關鍵字詞數由 536 上升到 588 筆時，IDF 的上升曲線呈現趨緩的狀況，而 TFIDF 及 $TFIDF^2$ 卻有逐漸上升的趨勢，因此可以假設試題數量增加到某個程度，可讓此兩個函數超越 IDF。但本實驗無足夠的樣本，無法證明這個假設。建議可使

用灰色預測探討 TFIDF 與 $TFIDF^2$ 或其他的 TWF 是否可以超越 IDF。