

行政院國家科學委員會專題研究計畫 成果報告

文件主題分段的探討

計畫類別：個別型計畫

計畫編號：NSC93-2213-E-009-069-

執行期間：93年08月01日至94年07月31日

執行單位：國立交通大學資訊科學學系(所)

計畫主持人：李嘉晃

計畫參與人員：張道行 張佑銘 蔡沛言 林信宏

報告類型：精簡報告

報告附件：出席國際會議研究心得報告及發表論文

處理方式：本計畫可公開查詢

中 華 民 國 94 年 12 月 21 日

行政院國家科學委員會補助專題研究計畫

成果報告
 期中進度報告

文件主題分段的探討

計畫類別： 個別型計畫 整合型計畫

計畫編號：NSC 93-2213-E-009-069

執行期間：2004年8月1日至2005年7月31日

計畫主持人：李嘉晃

共同主持人：

計畫參與人員：張道行 張佑銘 蔡沛言 林信宏

成果報告類型(依經費核定清單規定繳交)： 精簡報告 完整報告

本成果報告包括以下應繳交之附件：

- 赴國外出差或研習心得報告一份
- 赴大陸地區出差或研習心得報告一份
- 出席國際學術會議心得報告及發表之論文各一份
- 國際合作研究計畫國外研究報告書一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢

涉及專利或其他智慧財產權， 一年 二年後可公開查詢

執行單位：國立交通大學資訊科學所

中 華 民 國 年 月 日

成果報告

一、 中文摘要

本論文探討是否能夠以機器學習的方式來實行中文自動作文評分以協助作文的評分工作，此項評分限於修辭方面。主要基礎在於設計及擷取修辭技法來做為機器學習所需要的特徵，這些特徵包含詞數、形容詞數、成語數、譬喻手法、非口語化的喻詞、排比手法。由於不同老師批改同一篇作文也會有誤差出現，所以我們認為系統判斷出的分數及作文實際修辭分數的差距在一分(含)以內為可容許的誤差，因此以「避免產生過大誤差，但容許微小的差異」為主要概念建立改良式的 ID3 演算法做為機器學習的方式。訓練和測試作文都由老師先行評分，每篇作文的修辭分數由低到高為一到六分。利用改良式的 ID3 演算法產生決策樹，決策樹會將測試的作文分類到一分到六分其中的一個等級。以上述特徵及機器學習方式建立的修辭評分系統，其評量出的分數與真實分數的誤差比兩名沒有受過標準化作文評分訓練的老師間之誤差來得低。

Abstract:

This thesis explores the possibility of designing an Automated Chinese essay scoring system using machine learning method. Especially, we are interested in rhetorical aspect of the writing. The System focuses on designing and retrieving features related to rhetoric. Since human grader might assign different score to the same essay at different times, the system treats the error of one point as tolerable. Based on it, a modified ID3 algorithm is constructed.

二、 緣由與目的

現今學生語文能力普遍低落，其部份原因在於網路及手機通訊的發達，傳達訊息變得迅速且即時，使得一般學生慣於以簡短且口語化的文字互相溝通，結構嚴謹、詞藻洗鍊的傳統書信因而逐漸被取代，這樣一來便減少了鍛鍊文筆的機會。其次由於升學考試取消作文測驗，學生因而不重視國語文能力的培養。台積電董事長張忠謀先生曾提出「中文優勢論」[10]，在大中國市場興起的時代，中文能力已成為台灣吸引外資的人才優勢。但是事實上隨著非華語系國家學習中文的人數不斷增加，台灣的中文優勢正逐步消失。為了改善這個情形，教育部在民國九十三年十月宣佈，國中基本學力測驗將在民國九十五年試辦加考作文，民國九十六年正式實施。

作文可以使學生明瞭本國語言之特質、加強組織及表達概念的能力、強化寫作技巧、亦可增進文藝欣賞及創作之能力。不斷練習寫作可以整理學生從教學中零星片斷所獲得的寫作方法和知識，學生自然能增強寫作能力[2]。

一般而言，老師批改作文時，主要是以該文是否符合主題、句法是否通順、文章完整性、組織架構、修辭用字等來做為評分的標準。判斷文章是否符合主題通常需要分析句子及段落對於文章主題在語意上的關聯性。判斷文章句法是否通順需要檢查句子是否符合文法。評估文章完整性及組織架構需要對作文各段落的主要概念之間做語意及連貫性分析。而評估修辭用字則需要觀察各種修辭手法及詞彙的使用技巧。

作文要寫得動人，必然要注重修辭，修辭學探討表答情思的技巧與規則，以期達到精確明瞭、優美生動的境界[1]。本論文針對修辭用字方面，以機器學習的方式來學習老師在修辭學上的評分準則，並且發展自動

化的評閱系統來協助閱卷及訓練老師的工作。

三、系統設計

自動修辭評分系統的主要流程，第一步先處理中文斷詞及標記詞性，然後擷取修辭特徵，接著使用改良式的 ID3 演算法生成決策樹，最後以決策樹判斷測試作文的分數。自動修辭評分系統設計的兩個重點為擷取修辭特徵以及設計機器學習方法。

作文中的特徵可分為間接特徵與直接特徵。間接特徵(或稱為表面特徵)的出現頻率雖然與作文的分數有相關性，但是老師在批改作文時，並不是以間接特徵作為主要的評分依據，因此這些特徵比較缺乏教育上的反饋功能。一般老師閱卷時觀察的特徵我們稱之為直接特徵，這類特徵能夠更直接地評量文章的品質，因此本研究在評分系統中除了間接特徵外加入直接特徵，以期能更有效估計作文的修辭水準。

間接特徵

間接特徵包含詞數、形容詞數、及成語數。表達感情與思想需要透過形容詞的修飾。另外，使用成語代表寫作者受過較深入的國語文訓練，修辭能力會有一定的水準。兩者的數量又與文章長度有所關聯，因此我們加入了詞數這一項特徵。

圖 2、3、4 分別是詞數、形容詞數、成語數在所有作文樣本中的分布。很明顯隨著修辭分數的增加，平均詞數、平均形容詞數、平均成語數的數量也都會上升，也就是說這些特徵的統計數量與修辭能力有正相關性。

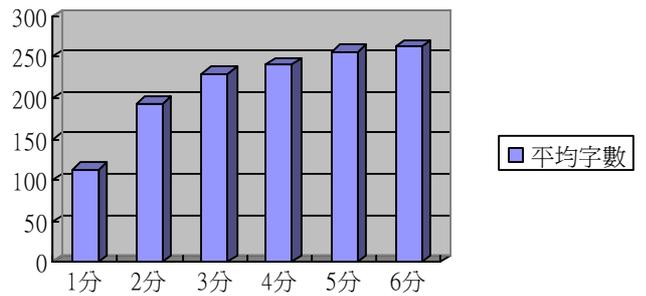


圖 2 各分數作文平均詞數統計圖。

1 分：114、2 分：194、3 分：230、4 分：241、5 分：257、6 分：263。

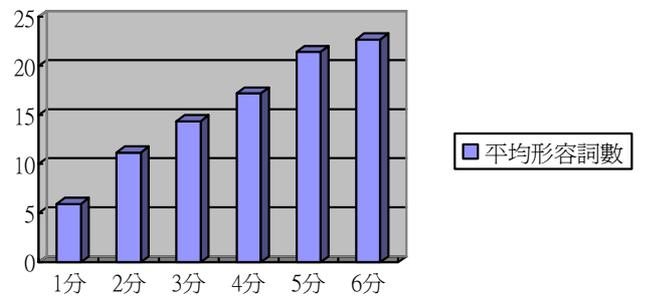


圖 3 各分數作文平均形容詞數統計圖。

1 分：5.95、2 分：11.19、3 分：14.4、4 分：17.2、5 分：21.41、6 分：22.67。

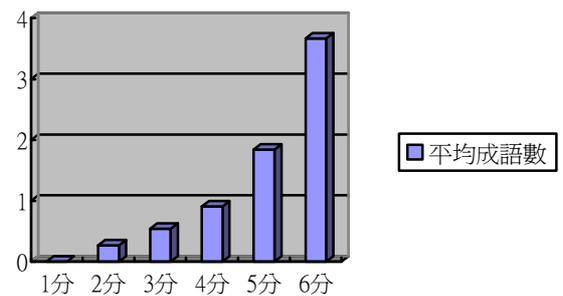


圖 4 各分數作文平均成語數統計圖。

1 分：0、2 分：0.27、3 分：0.54、4 分：0.91、5 分：1.84、6 分：3.67。

直接特徵

本研究使用的直接特徵包含譬喻手法、非口語化的喻詞以及排比手法。

譬喻手法

譬喻就是用已知喻未知[1]，利用已知的知識描繪未知的情境。一個好的譬喻，往往能夠給人留下深刻的印象。例如：「下課十分鐘時，學校就像一個菜市場。」這樣的一個譬喻手法是借用菜市場很吵雜的已知知識來描述下課時學校的情境。

譬喻手法包含明喻、隱喻、略喻、借喻等[1]。因為略喻、借喻需要語意的分析，所以本研究只使用明喻及隱喻手法。明喻及隱喻主要是由本體、喻體、及喻詞所構成。舉例來說：「下課十分鐘時，學校就像一個菜市場。」這句話中「學校」是本體、「像」是喻詞、「菜市場」是喻體。本論文採用的明喻及隱喻手法基本形式有兩種，一種是本體、喻詞、喻體皆在同一句；另一種是喻詞及喻體在同一句，本體在前面一句的句子組合，如：「原本寧靜的校園，霎時變成了熱鬧的菜市場。」底下是從作文中找出明喻及隱喻手法的步驟及實例說明。參照圖 5，首先要找出所有可能的喻詞，再搜尋符合譬喻手法形式的句子。這個演算法需要使用辭典，我們使用的辭典為「中文詞庫」，其中記載了 78219 個詞以及各個詞的詞性。該演算法是半自動化的，有些步驟需要語意及文法分析，因此我們以人工處理。

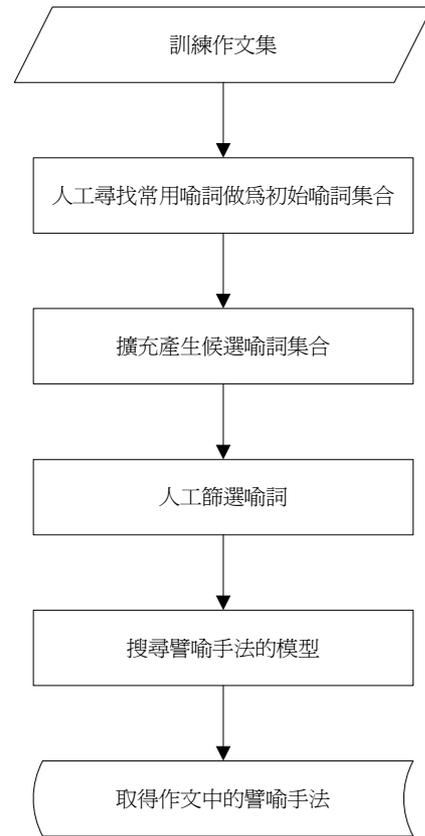


圖 5 擷取譬喻手法流程圖

四、 實驗結果

我們使用的作文來自台北市立敦化國中，這些作文輸入成電子檔時保留所有的錯字以及標點符號，不加以修改，以維持學生作文的原貌。修辭分數從 1 分到 6 分，1 分為最低，6 分為最高。每篇由二到三名老師評分，取平均分數。實驗資料包括 693 篇作文。修辭分數 1 分到 6 分的文章篇數分別為 40 篇、166 篇、234 篇、177 篇、70 篇以及 6 篇。每次實驗時隨機取出 394 篇訓練作文，其中 197 篇用來建立決策樹，197 篇用來作為 prune 時的檢驗資料，299 篇測試作文則用來評估整個系統的效能。實驗結果詳見下表：

	ID3	Modified ID3
--	-----	--------------

實驗一	0.85	0.93
實驗二	0.857	0.89
實驗三	0.884	0.913
平均	0.864	0.911

表 1 Adjacent value 統計表

	ID3	Modified ID3
實驗一	0.345	0.375
實驗二	0.395	0.408
實驗三	0.362	0.385
平均	0.367	0.389

表 2 Exact value 統計表

	Exact	Adjacent
Two teachers	0.297	0.749
Modified ID3	0.389	0.911

表 3 系統效能評估表

五、 參考文獻

- [1] 黃麗貞。《實用修辭學「增訂本」》台北，文津出版社，2004。
- [2] 陳品卿。《中學作文教學指導》國立臺灣師範大學中等教育輔導委員會，1989。
- [3] Jill Burstein. The E-rater Scoring Engine: Automated Essay Scoring With Natural Language Processing. *Automated Essay Scoring: A Cross-Disciplinary Perspective* (2003), pp. 113-121
- [4] Thomas K Landauer, Darrell Laham, Peter W. Foltz. Automated Scoring and Annotation of Essays with the

Intelligent Essay Assessor. *Automated Essay Scoring: A Cross-Disciplinary Perspective* (2003), pp. 87-112

- [5] Thomas K. Landauer, Darrell Laham, Bob Rehder, and M. E. Schreiner. How Well Can Passage Meaning be Derived without Using Word Order? A Comparison of Latent Semantic Analysis and Humans. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the 19th annual meeting of the Cognitive Science Society* (1997), pp. 412-417.
- [6] Berry, M. W., Dumais, S. T. and O'Brien, G. W.. Using linear algebra for intelligent information retrieval. *SIAM: Review* (1995), 37(4), 573-595.
- [7] Hearst, M.. The debate on automated essay grading. *IEEE Intelligent Systems* (2003), 15(5), 22-37, IEEE CS Press.
- [8] Richard Sproat, Thomas Emerson. The First International Chinese Word Segmentation Bakeoff. Which was held as part of the *Second Meeting of SIGHAN*, July 11-12, 2003 in Sapporo, Japan.
- [9] Tom Mitchell, McGraw Hill. *Machine Learning* (1997).
- [10] 92.10.22 經濟日報
http://mag.udn.com/mag/life/storypage.jsp?f_ART_ID=6732
- [11] <http://163.26.9.12/noise/hcjh-ca/1-12.htm>