

本期要目

- | | |
|--------------------------|----------|
| 壹、ROCLING 2012 研討會剪影 | 第 2~3 頁 |
| 貳、第十二屆第二次會員大會會議記錄及剪影 | 第 4~8 頁 |
| 參、專文-自動化中文寫作分級技術之發展(張道行) | 第 9~16 頁 |

ROCLING 2012 圓滿結束

由元智大學資訊管理學系、資訊工程學系及本會共同主辦的「第二十四屆自然語言與語音處理研討會」已於 9 月 22 日假中壢市元智大學元智五館彥公廳圓滿結束，本次會議專題演講部分，分別邀請到 ACL 主席 Dr. Kenneth Church 及 Microsoft Research 首席研究員也是現任 IEEE Trans. Audio, Speech & Language Processing 期刊主編 Dr. Li Deng。大會也規劃了一個「Research & Application of Speech & Language Technology for Orange Computing」的座談會，由大仁科技大學校長王駿發教授主持，座談會成員包含：成功大學資訊工程系吳宗憲教授、工研院資通所郭志忠博士及國立成功大學陳柏煒博士。座談會融合了學界與產業界的觀點，提供與會者對於 Orange Computing 有更深一層的認識，同時也開啓了語音與語言處理技術一個全新的應用領域。參與此次盛會的人士除了國內學者外，還包含來自世界各地之學者—美國、加拿大、土耳其、印度、馬來西亞、越南等十餘國，與會人數多達 200 人次。

本次會議廣邀學界與產業界投稿、經嚴謹同儕評審之審稿程序，共收錄了 15 篇口頭論文及 19 篇壁報論文。黃邦烜先生、郝柏翰先生、陳冠宇先生及陳柏琳教授共同著作之「遞迴式類神經網路語言模型應用額外資訊於語音辨識之研究」獲最佳論文獎，於大會閉幕式中，獲頒獎金伍仟元。會議論文及專題演講投影片皆已公開建置在大會網站上。

- 專題演講投影片一點閱
- 會議論文檢索一點閱

第十二屆第二次會員大會圓滿結束

本會第十二屆第二次會員大會，已於 9 月 21 日(星期五)假元智大學元智五館彥公廳舉行，大會首先由許理事長聞廉報告過去一年來會務工作概況，以及未來一年經營會務之方針；接著由各工作小組、委員會及期刊總編輯進行年度工作報告，隨後，提請通過兩個提案，分別為追認一〇〇年度財務報表及追認一〇一年度收支預算書，最後，頒發本年度博碩士論文獎。大會內容請詳見本刊第四頁—會員大會會議紀錄。

ROCLING 2012 研討會剪影 (1)

大會主席—元智大學禹良治教授



大會主席—元智大學蔡宗翰教授



專題演講—Dr. Kenneth Church



專題演講—Dr. Li Deng



口頭論文報告

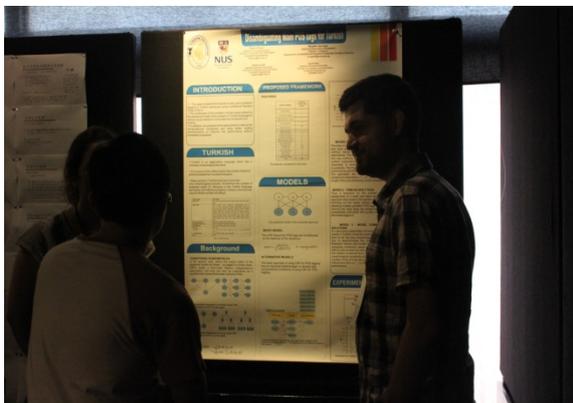


口頭論文報告



ROCLING 2012 研討會剪影 (2)

壁報論文



壁報論文



座談會



工作人員合影



中華民國計算語言學學會

第十二屆第二次會員大會紀錄

- 一、會議時間：中華民國一〇一年九月二十一日十三時十五分
- 二、會議地點：桃園縣中壢市元智大學元智五館彥公廳
- 三、出席人員：應出席 213 人，實際出席 109 人（親自出席 95 人，委託出席 14 名）
- 四、缺席人數：104 人
- 五、主席：許聞廉
記錄：黃琪
- 六、主席致詞：

- 今年的 ACL 會議台灣大約有 10 多位學者參加，反觀中國大陸學者的參與非常踴躍，ACL 的主席 Dr. Kenneth Church，也是今天的主講者，希望台灣學者多參與 ACL，也應積極的參與「亞洲自然語言處理聯盟(AFNLP)」的活動，AFNLP 每年定期舉行 NLP 國際會議；今晚的理監事聯席會議將討論爭取在 AFNLP 成立一個 Knowledge 及 Semantic 的 SIG(Special Interest Group)，希望大家支持這個活動。
- 學會期刊（中文計算語言學期刊）編輯助理將與 JISE 之編輯群合併，以節省學會人力開支，兩份刊物還是獨立發行。
- 中研院資訊所每年定期舉辦 Distinguished Lecture Series，今年學會即是與中研院資訊所合作，邀請會議期間受邀來訪的 Dr. Kenneth Church 擔任 Rocling 專題演講主講人，學會因此省下了一筆費用，所以希望往後 Rocling 在邀請專題演講之主講人時，主辦單位可與中研院資訊所聯繫。
- 今年上半年中研院資訊所的 Distinguished Lecture Series 邀請到 Prof. Eduard Hovy，並與學會共同在 Prof. Hovy 來訪期間合辦一場 Workshop，日後學會可多多利用這樣的機會，與中研院資訊所合作，以促進國際學術交流。

七、報告事項

(一) 秘書處報告

財務報告：截至 101/7/31 日止，經費收入 NT\$1,291,175 元，經費支出 NT\$825,965 元，結餘 NT\$465,210 元，流動資產 NT\$14,114,060 元。

(二) 學術委員會報告(陳柏琳)

1. 學術活動：自去年 Rocling 至目前為止，學會舉辦了三場研討會及協辦兩場國際知名學者在中研院資訊所的講座，總參加人數多達 500 人。依照往例，

預計今年 12 月在中研院資訊所舉辦資訊檢索研討會，歡迎踴躍參加。

2. 獎助出席國際會議：今年共獎助兩位學生出席 ACL，兩位同學發表的是 Short paper；學會獎助的會議包含 ACL、ACM-SIGIR、ICASSP 及 COLING，希望各位同學踴躍投稿及申請獎助。
3. 博碩士論文獎申請：今年博碩士論文獎共收到 6 件博士論文及 8 件碩士論文申請，今年收件期限延後一個月，本屆共頒發一位博士論文優等獎、兩位博士論文佳作獎、一位碩士論文優等獎及三位碩士論文佳作獎，明年歡迎碩博士畢業生踴躍申請。
4. 今明兩天將決定明年 Rocling 的主辦單位，屆時歡迎各位師長及同學踴躍參加及投稿。

(三) 會員委員會報告(洪志偉)

會員委員會主要工作之一是會員通訊專文徵稿，專文的主題只要是與語言或語音相關之領域皆可投稿，專文的形式可為研究成果，或是新的文獻整理，與一般論文投稿形式不同，目前稿源非常短缺，希望大家踴躍投稿，有興趣投稿者，請將文稿寄給本人或學會秘書。

(四) 資訊檢索組報告(盧文祥)

本人今年是最後一年擔任此工作小組召集人，六年的任期中，若有不盡理想之處，還望各位包涵。去年資訊檢索研討會稍微改變會議形式，會中除了邀請專家學者進行專題演講外，另外增加邀請出席國際會議發表論文之學生進行報告，去年邀請了三位同學進行報告，成效不錯，參加人數稍有成長；今年的研討會將延續此形式舉行，並將安排實驗室的研究報告，有興趣的同學，歡迎與本人聯絡。

(五) 口語處理組報告(廖元甫)

1. 學術活動：本組每年定期籌畫語音訊號處理研討會，今年是由交通大學冀泰石教授負責主辦，會議相當成功。明年主辦單位尚未決定，這兩日將徵求主辦單位，希望被徵詢到老師不要推辭。
2. 語音資料庫：口語小組今年並無收錄新的語料，在座各位老師若計畫收集語料，並願意提供學會發行，口語小組將可提供錄音所需經費。
3. 本人目前正在錄製「語音合成語料庫」，內容是中英夾雜的句子，待整理完成後將授權學會發行。

(六) 期刊出刊進度報告(曾元顯)

1. 出刊報告：期刊一年出刊四期，分別在 3 月、6 月、9 月、12 月，截至目前為止已編稿完畢至 2012 年 9 月，會員近日將會陸續收到去年及今年發行的期刊，在此，特別感謝期刊總編輯之一的陳光華教授，將近兩年拖稿的刊數

補齊。

2. 明年 6 月及 12 月兩期之特刊將分別由謝舒凱教授及曾淑娟博士負責，主題為語料分析，歡迎踴躍投稿，詳細內容請參閱[期刊網站](#)。
3. 期刊編委會：期刊編輯委員會每半年召開一次，針對審稿的時間，已達成共識，收到稿件後儘可能在 6 週內完成；審稿意見方面，建議主編或審稿老師多給作者正面的評語，並能提供具體修改意見，希望大家踴躍投稿。

八、討論提案

提案 1

案由：提請追認一〇〇年度「經費收支決算表」、「資產負債表」、「現金出納表」、「基金收支表」。

提案人：理事會

說明：業經第十二屆第二次理監事聯席會議(101/3/22)通過。

決議：照案通過。

提案 2

案由：提請追認一〇一一年度「經費收支預算書」。

提案人：理事會

說明：業經第十二屆第一次理監事聯席會議(100/9/8)通過。

決議：照案通過。

九、頒獎-第十二屆博碩士論文獎

十、臨時動議

十一、散會

會員大會剪影(1)

主席致詞-許聞廉教授



頒獎-博士論文優等獎



頒獎-博士論文佳作獎



頒獎-博士論文佳作獎



頒獎-碩士論文優等獎



頒獎-碩士論佳作獎



會員大會剪影(2)

頒獎-碩士論文佳作獎



頒獎-碩士論文佳作獎



自動化中文寫作分級技術之發展

張道行

國立高雄應用科技大學 資訊工程系

1. 前言

寫作測驗對語言教學、教育評量及心理計量等領域是非常重要的工具，然而閱卷的困難使得寫作測驗的實施規模與實施次數受到極大限制。舉例來說，測驗機構在舉行寫作測驗時必須考量閱卷的速度、成本、評分者信度三個因素。首先，測驗機構與受試者都希望能於施測後儘快取得測驗結果，然而由於傳統寫作測驗要加快閱卷速度只能增加評分者，但增加評分者人數將大幅增加測驗成本，更何況評分者必須經過嚴格的訓練及認證。再者，要提高測驗的評分者信度，則必須增加每份試卷的評分者人數及提高評分者品質。然而兩者都將導致測驗成本大幅增加。這些成本對經費有限的研究而言難以負荷，即使是商業測驗，雖然可以轉嫁成本，但募集及訓練足夠數量的合格評分者還是相當困難的工作。因此如何快速、有效、低成本進行閱卷成爲重要的研究課題。

自動寫作分級(Automated Essay Scoring, AES)技術是上述問題的解決方案。早在 1960 年 Page 即開始發展相關技術，但直到在 1990 年代初期，由於自然語言處理及資訊擷取技術的蓬勃發展，才使得 AES 研究有長足進步，並在 2000 年後陸續發表多個 AES 系統(Elliot, 2001; Larkey & Croft, 2003; Shermis & Burstein, 2003; Valenti, Neri, & Cucchiarelli, 2003)。這些系統至今仍不斷被改進並已廣泛使用於國際大型語文測驗、例如 GMAT、TOEFL、SAT 等。部分系統更進一步發展成可用以輔助學生學習寫作的系統，例如 ETS 的 Criterion 及 Vantage Learning 公司的 MY Access，國內也有部分大學於學生英文寫作教學課程使用。而一些研究用的系統如 BETSY 等也協助教育行政單位成功進行學童寫作能力變化的評估。這些應用證實了 AES 技術確實可行，在後續應用的可能性及重要性均不言可喻。

雖然理想上我們希望 AES 系統能具有完全模擬人工閱卷的能力與品質，但事實上並不是所有的需求都需要 AES 達到如此理想的境界，不同能力的 AES 系統能滿足不同層次的需求。這些需求可分爲三個層次：第一是當系統與人工分級的整體結果相當接近，就可以應用在整體寫作能力分析上，例如分析國民教育學生的寫作能力趨勢變化。由於這類研究樣本的量很大卻只需要整體的平均數據，因此可以容忍系統的少量錯誤。第二是能作爲需具有高評分者信度之測驗的分級輔助工具。例如目前與入學有關的寫作測驗，每份試卷都安排至少兩名評分者評分，若兩者評分結果不同則再交由資深評分者複閱。合格的 AES 可以做爲第三評分者，降低兩名評分者同時發生錯誤的可能，甚至更進一步取代其中一名評分者以節省大量成本。第三是能提供寫作學習者自學的工具，能提供改善的建議。

隨著國內外對華語文能力的日漸重視及大型寫作測驗的實施，自動化中文寫作分級也已經成為重要的研究課題。然而由於語言的差異性、對寫作評量設定的目標不同、以及在相關基礎研究的質與量均有差距，直接使用現有的 AES 技術在中文寫作上將產生相當大的困難。本文將介紹一個針對中文語言特性發展的 AES 技術。

2. 國外 AES 研究回顧

AES 研究可以區分為三個時期。第一時期是 1960-1980 年早期。1960 年 Page 發現文章中某些統計特徵與作品等級具有極高的關連性，例如字數、段落數等。利用統計文章的這些特徵得出的分級結果與人具有相當高的相關性。因此利用文章長度、平均詞長、逗號數量、介詞數量、非常用詞數量等特徵以多線性迴歸找出特徵的線性組合函數，提出名為 Project Essay Grader (PEG) 的分級系統，其系統分級結果與人工判斷結果達 .78 相關。這些特徵是透過統計結果間接推測作品的等級，並非直接偵測寫作中表現的能力，因此稱為間接特徵(或表面特徵)。由於 PEG 只使用寫作的間接特徵，無法偵測寫作的內容、組織、文法等直接特徵，這使得 PEG 無法應用在真實的測驗中，因為這些簡單的統計特徵很容易被受試者掌握，進而寫出人工一眼即可判斷明顯不佳、系統卻給予高評價的作品，例如完全不知所云、但字數及罕用字特別多的作品。

AES 研究的第二時期是 1980-1990 中期。1980 年初期，McDonald 等人(1982)發表 Writer's Workbench (WWB) 的系統，主要功用為修正文章的拼字、措辭及可讀性。這開啓了以判別寫作品質為基礎的研究方向。1990 年早期 AES 系統開始大量使用作品的語法及語意特徵。由於這些特徵代表作者展現的寫作能力的一部分，因此稱為直接特徵(或內在特徵)。例如 Page (1994) 爲了提高系統的效能，改採用句子完備性等特徵做為評分依據。Burstein 等人(1998)利用文法分析、詞彙內容分析等工具找出語法正確性、主題符合度、意見組織等寫作的直接特徵，並使用 Vector Space Model 處理這些特徵值。Landauer 及 Dumais (2000)除了使用間接特徵與文法特徵外，還利用作品與樣本詞彙的潛在語意關係(LSA)來推論兩份寫作間的語意關係，並提出以此語意特徵為核心的 AES 系統 IEA。

AES 研究的第三時期是 1990 後期迄今。由於 AES 技術已經廣泛應用，研究者將目標轉向發掘更多的語意特徵、例如 Attali & Burstein (2006) 在 e-rater 的第二版使用更多的語意特徵提高系統的有效性及信度，該文也探討如何提供資訊回饋給受試者，甚至與受試者進行互動教學。這些研究目前仍在持續進行。

雖然 AES 研究在英語的發展有長足進步，但因為幾個主要的困難使得應用在中文上效能不佳。第一、現有的 AES 系統採用許多與語法結構相關的特徵，例如句型、文法、邏輯性等。由於英語已經有效能相當好的文法剖析器，因此可以擷取這些特徵並進一步分析這些特徵與等級的相關性。然而雖然目前已經有些中文文法剖析工具(例如 Stanford parser)可供使用，然而對中文 AES 的需求而言現有工具的效能仍然不足。第二、中文在文字處理上遭遇比較多的困難。中文需要進行斷詞處理，且中文在未知詞擷取及詞性標記

的正確性要達到與英文同樣水準也不大容易。這些前處理工作所造成的錯誤會影響下一個分析階段的正確性，因此導致特徵擷取的準確性較差，進而影響分級正確性。第三、中文句子的定義及標點符號的使用較為模糊，這使得以句子為單位的特徵擷取方法使用於中文會產生相當大的困難。

3. 中文 AES 的發展現況

目前大多數 AES 的系統架構都是以特徵擷取作為基礎，以一個機器學習的方法整合這些特徵值，並進一步學習特徵與人工結果間的關連性，用以預測新作品的等級。現有 AES 系統間不同之處在於特徵種類、擷取方法以及預測模型的設計都各不相同。中文 AES 也可採用類似架構，圖 1 說明一個同樣採用特徵為本之中文 AES 系統的架構。

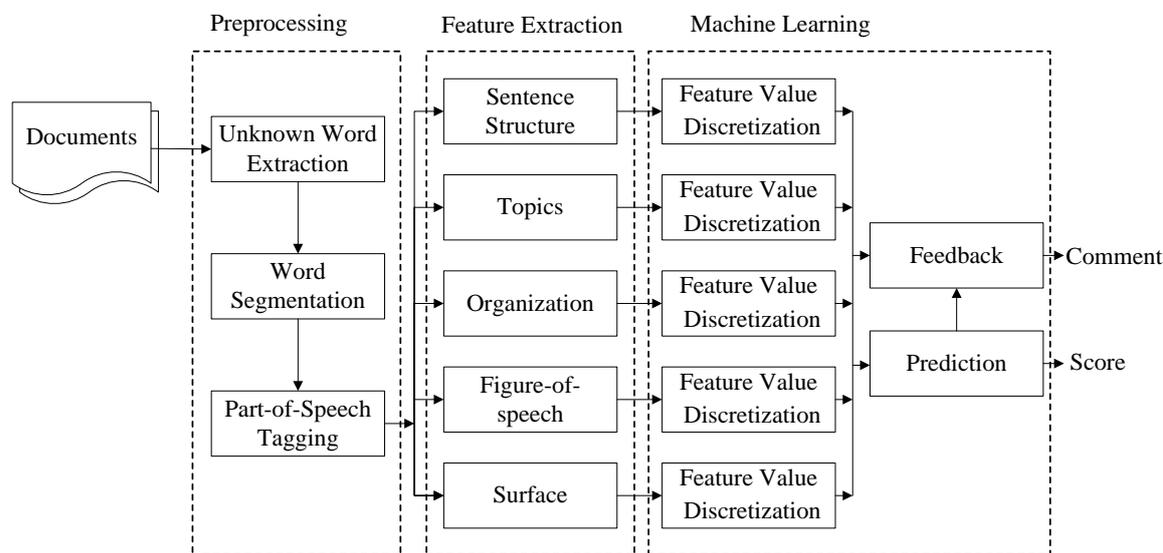


圖 1 特徵為本之中文 AES 的基本架構

到目前為止，中文 AES 的研究多集中在圖 1 架構中個別問題的討論，例如某一特徵的擷取方式、較適合的預測模型設計等等。而能整合圖 1 中大多數模組進行實際分級、並經由大量資料驗證的系統，是由台灣師大心測中心、交通大學及高雄應用科大所共同發展出的 ACES (Automatic Chinese Essay Scoring) 系統。本文接下來將介紹 ACES 的運作原理。其中前處理部分為自然語言處理領域長期關注且已有卓越研究成果與工具的研究題目，ACES 也是應用這些研究成果及工具進行前處理，因篇幅限制不再贅述。

3.1 特徵擷取

目前的 ACES 系統只使用三類直接特徵：取材、修辭與結構。而文法特徵如先前所述，由於目前尚無較適合用於寫作分析的文法剖析工具，故尚未納入文法特徵。這些特徵的使用方式主要可分為鑑別法與比對法。鑑別法是指若某現象在高等級作品經常出現而很少在低等級作品出現，那麼偵測這個現象出現與否就能推測作品等級的可能性。例如在「下課十分鐘」為題的寫作中可以發現「植物」這個概念只會在高等級作品中被提及，卻很少出現在低等級作品，這樣的觀念會用不同的詞彙出現，例如紅花、大樹、小草都屬於植物，但將這些詞彙透過知網(HowNet)轉換為語意類別（或稱為義原）後，都屬於「植物」這個類別，因此我們稱「植物」為這個題目的鑑別義原。這個現象的發生可能是因為高等級作品作者較有能力觀察到且使用鑑別義原描寫特定取材。ACES 便使用鑑別義原出現情形估計該篇作品在各等級的機率值(Chang et al., 2009)。

修辭特徵也有類似的現象。表 1 為先前的研究(Chang, Lee & Chang, 2006)的發現，顯示語料中寫作各等級分布比例、以及該語料中有出現譬喻及排比之子集合其各等級分布比例。很明顯地，有出現譬喻與排比的作品比起其他作品有更高比例屬於高等級作品。因此 ACES 使用數條規則來判斷作品中是否具有譬喻與排比修辭格，若出現則推測該作品有較高的機率是高等級作品。

表 1 所有寫作與出現修辭格之寫作在等級分布比例的差異

	低等級群			高等級群		
	1	2	3	4	5	6
所有作品	0.07	0.19	0.31	0.30	0.13	0.01
出現譬喻作品	0.03	0.15	0.28	0.35	0.19	0.01
出現排比作品	0.03	0.10	0.22	0.47	0.16	0.03

對於結構特徵，ACES 則是採用比對類似結構方式決定等級，也就是某篇作品的結構類似於高等級作品，則該作品傾向高等級的機率較高。而對結構的擷取方法，ACES 則是提出一個稱為 C-L 結構的觀念，也就是寫作中的結構可分為段落內與段落間的概念連結關係，連結的順序表示作者鋪陳段落或篇章的手法。若有分屬不同作品的兩個段落其內部概念連結順序相似，則表示兩個作者對一個段落次主題有類似的表達方式；若兩篇寫作其段落間的概念連結順序很相似，則表示兩位作者對題目的看法有類似的表達方式。因此 ACES 發展出一個比對 C-L 結構相似性的方法，將待分級寫作與已知等級的語料庫中寫作比對結構相似性，並經由比對結果給定結構面向的各等級機率值。詳細作法可參考 Chang & Lee (2009)的論文。

3.2 預測模型

對於一篇作品，經由上述步驟後可得該作品在各特徵所得之各等級機率值。ACES 使用 multi-variate Bernoulli model (MBM) 整合不同特徵所得之不同等級機率值，並成為預測模型。傳統 MBM 模型定義如下：

$$P(d_i | c_j) = \prod_{t=1}^V \left[B_{it} P(w_t | c_j) + (1 - B_{it})(1 - P(w_t | c_j)) \right]$$

其中 d_i 表示待分級寫作 i ； c_j 表示等級 j ； $B_{it} \in \{0,1\}$ ，表示第 t 個特徵是否出現在寫作 i ； V 表示特徵的數量； $P(w_t | c_j)$ 表示特徵 w_t 出現在等級為 c_j 之作品的機率值；而 $P(w_t | c_j)$ 可被定義如下：

$$P(w_t | c_j) = \frac{1 + \sum_{i=1}^{D_j} B_{it}}{J + D_j}$$

其中 D_j 是訓練語料中等級 c_j 的作品數量； J 則是一個常數項。

ACES 即是使用改良自上述 MBM 模型的修正模型計算待分級作品在每個等級的機率值，並取最高者作為該作品的預測等級。

4. 效能分析

本文使用兩個題目分別為「可貴的合作經驗」與「用餐時刻」的寫作語料集分析 ACES 效能。這些寫作都是台灣九年級學生的作品。寫作集「可貴的合作經驗」包含 1,200 篇短文寫作，並採取六等級制。此寫作集由大量已分級寫作中，隨機挑選各等級 200 篇寫作，並且每篇寫作均由二位合格閱卷者給予相同等級。寫作集「用餐時刻」包含 1,104 篇短文寫作，亦採取六級分制。該寫作集由大量未評分文件中隨機選出 2,000 篇作品，再經由二位閱卷者進行評閱，並從中保留二位閱卷者給予相同分數的作品。此寫作集的等級組成接近全體學生的能力分佈，各等級所佔有的篇數為一級 17 篇、二級 39 篇、三級 202 篇、四級 519 篇、五級 257 篇以及六級 70 篇等不均勻的分布。

為了比較效能，我們定義 Accuracy rate (AR)、Exact rate (ER)、Average accuracy rate (AAR) 以及 Average exact rate (AER) 四個指標。Accuracy rate 與 Exact rate 分別定義如下：

$$\text{Accuracy rate} = \frac{\sum_{i=1}^n \sum_{j=1}^n a_{i,j}}{\sum_{i=1}^n \sum_{j=1}^n c_{i,j}}, a_{i,j} = \begin{cases} c_{i,j}, & \text{if } |i-j| \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

$$\text{Exact rate} = \frac{\sum_{i=1}^n \sum_{j=1}^n e_{i,j}}{\sum_{i=1}^n \sum_{j=1}^n c_{i,j}}, e_{i,j} = \begin{cases} c_{i,j}, & \text{if } |i-j| = 0 \\ 0, & \text{otherwise} \end{cases}$$

其中 n 表示等級範圍,在此 n 值為 6; i 代表人工判斷等級且 j 代表 ACES 預測等級; $c_{i,j}$ 表示人工判斷等級為 i 且 ACES 預測等級為 j 的作品篇數; $a_{i,j}$ 代表允許一級誤差之作品篇數,而 $e_{i,j}$ 代表預測與人工等級完全相同之作品篇數。

由於 AR 與 ER 兩指標的計算方法以整個寫作集為基準,因此當寫作集在各等級的數量不同時,有些預測模型會傾向放棄量少等級的正確率換取對寫作數量較多之等級的更高正確率,因此設計平均正確率(AAR)以及平均精確率(AER)兩個指標判斷系統是否在各等級的預測能力接近一致。AAR 與 AER 定義如下:

$$\text{Average accuracy rate} = \frac{1}{n} \sum_{i=1}^n \frac{\sum_{j=1}^n a_{i,j}}{\sum_{j=1}^n c_{i,j}}, a_{i,j} = \begin{cases} c_{i,j}, & \text{if } |i-j| \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

$$\text{Average exact rate} = \frac{1}{n} \sum_{i=1}^n \frac{\sum_{j=1}^n e_{i,j}}{\sum_{j=1}^n c_{i,j}}, e_{i,j} = \begin{cases} c_{i,j}, & \text{if } |i-j| = 0 \\ 0, & \text{otherwise} \end{cases}$$

表 2 說明 ACES 在兩個寫作集的表現。在寫作集 1 中,ACES 對 98%的寫作預測等級與人工判斷等級在 1 級之內,64%與人工判斷等級完全相同。AAR 及 AER 的數據顯示 ACES 在各等級的表現相當一致。但在寫作集 2 中,ACES 的表現明顯下降。經過進一步實驗發現,差距並非是題目差異造成,而是在較高等級與較低等級的數量較少造成預測模型的判斷誤差較大。雖然寫作集 2 的效能不如寫作集 1 的表現,但這個表現已經與英文 AES 的預測能力相當接近。整體來說,若訓練語料足夠,自動中文寫作分級預測能力非常優異,而即使在訓練語料不足的實際應用環境下,ACES 仍有與英文 AES 相當、接近人工判斷水準的預測能力。

表 2 ACES 在兩個語料集的效能

	AR	ER	AAR	AER
寫作集 1	0.98	0.64	0.98	0.64
寫作集 2	0.89	0.46	0.87	0.46

5. 進一步的研究

由測試結果顯示，ACES 可以滿足第一級需求層次的工作。在訓練語料較多的情況下，也可滿足第二級需求層次的工作。然而不論英文 AES 或 ACES，都有可持續研究的問題。下面介紹目前學術社群進行中的幾個研究方向。

第一是尋找更多的直接特徵。當直接特徵越多、能辨識的特徵能力層次越高，就越能精確的判斷作品的等級。其中特別是語法特徵與語意特徵的發展。第二是發展可提供教學回饋的 ACES。目前 ACES 可對應用於總結性評量，但無法提出診斷，因此也無法應用於形成性評量。第三是發展非監督式機器學習預測模型。包括 ACES 在內，現有的 AES 必須對單一題目有足夠的訓練語料才能取得夠好的預測能力。先前 Chen 等人(2010)由先前語料在各特徵的分布作為對未人工分級之新語料的初始假設，提出一個疊代演算法預測新語料中的寫作等級。這類的研究若成功發展，將使 AES 更具實用及便利性。第四是針對第二語言作者的寫作分析。第二語言寫作者由於受母語的影響，其寫作發生錯誤的類型與表現與母語作者不盡相同。如何找尋適合中文為第二外語寫作者的寫作特徵並設計回饋模式，在全球華語熱的現在更顯得迫切。

參考文獻

- Attali, Y. & Burstein, J. (2006). *Automated Scoring Using With e-raterV2*. The Journal of Technology, Learning and Assessment, 4(3).
- Burstein, J., Kukich, K., Wolff, S., Lut, C., Chodorow, M., Braden-Harder, L. & Dee Harri, M. (1998). *Automated Scoring Using A Hybrid Feature Identification Technique*. Paper presented at the 36th Annual Meeting of the Association of Computational Linguistics, Montreal, Canada.
- Chang, T. H. & Lee, C. H. (2009). *Automatic Chinese Essay Scoring Using Connections between Concepts in Paragraphs*. Paper presented at the International Conference on Asian Language Processing, 265-268, Singapore.

- Chang, T. H., Lee, C. H. & Chang, Y. M. (2006). *Enhancing Automatic Chinese Essay Scoring System from Figures-of-Speech*. Paper presented at the 20th Pacific Asia Conference on Language, Information and Computation, Wuhan, China.
- Chang, T. H., Lee, C. H. & Tam, H. P. (2010). *Corpus-based Subtopic Segmentation Using Concept Segment Method*. *Information: An International Interdisciplinary Journal*, 13(3B), 975-982.
- Chang, T. H., Lee, C. H., Tsai, P. Y. & Tam, H. P. (2009). *Automated essay scoring using set of literary sememes*. *Information: An International Interdisciplinary Journal*, 12(2), 351-357.
- Chen, Y. Y., Liu, C. L., Lee, C. H. & Chang, T. H. (2010). *An Unsupervised Automated Essay Scoring System*. *IEEE Intelligent System*, 25(5), 61-67.
- Elliot, S.M., (2001) *IntelliMetric: From Here to Validity*. Paper presented at the annual meeting of the American Educational Research Association Seattle, WA, USA.
- Hearst, M. (2003). *The Debate on Automated Essay Grading*. *IEEE Intelligent Systems*, 15(5), 22-37.
- Landauer, T.K., Laham, D. & Foltz, P.W. (2000). *The Intelligent Essay Assessor*. *IEEE Intelligent System*, 15, 27-31.
- Larkey, L. S. and Croft, W.B. (2003). *A Text Categorization Approach to Automated Essay Grading*. *Automated Essay Scoring: A Cross-Disciplinary Perspective*. Lawrence Erlbaum Associates Inc., Mahwah New Jersey, 55-69.
- Ma, W. Y. & Chen, K. J. (2003). *Introduction to CKIP Chinese Word Segmentation System for the First International Chinese Word Segmentation Bakeoff*. Paper presented at 2nd SIGHAN Workshop on Chinese Language Processing, Sapporo, Japan.
- MacDonald, N., Frase, L, Gingrich, P, & Keenan, S. (1982). *The Writer's Workbench: Computer Aids for Text Analysis*. *IEEE Transactions on Communications*. 30(1), 105-110.
- Page, E. B. (1994). *Computer Grading of Student Prose, Using Modern Concepts and Software*. *Journal of Experimental Education*, 67, 127-142.
- Shermis, M. D., and Burstein, J. C. (Eds). (2003). *Automated Essay Scoring: A Cross-Disciplinary Perspective*. Mahwah New Jersey: Lawrence Erlbaum Associates Inc.
- Valenti, S., Neri, F. and Cucchiarelli, A. (2003). *An overview of current research on automated essay grading*. *Journal of Information Technology Education*, 2, 319-330.