

2019「中技社AI創意競賽」

2019 CTCI Foundation AI Innovation Competition

AI 與教育 AI & Education

中文手寫作文自動辨識、批改與教學回饋系統



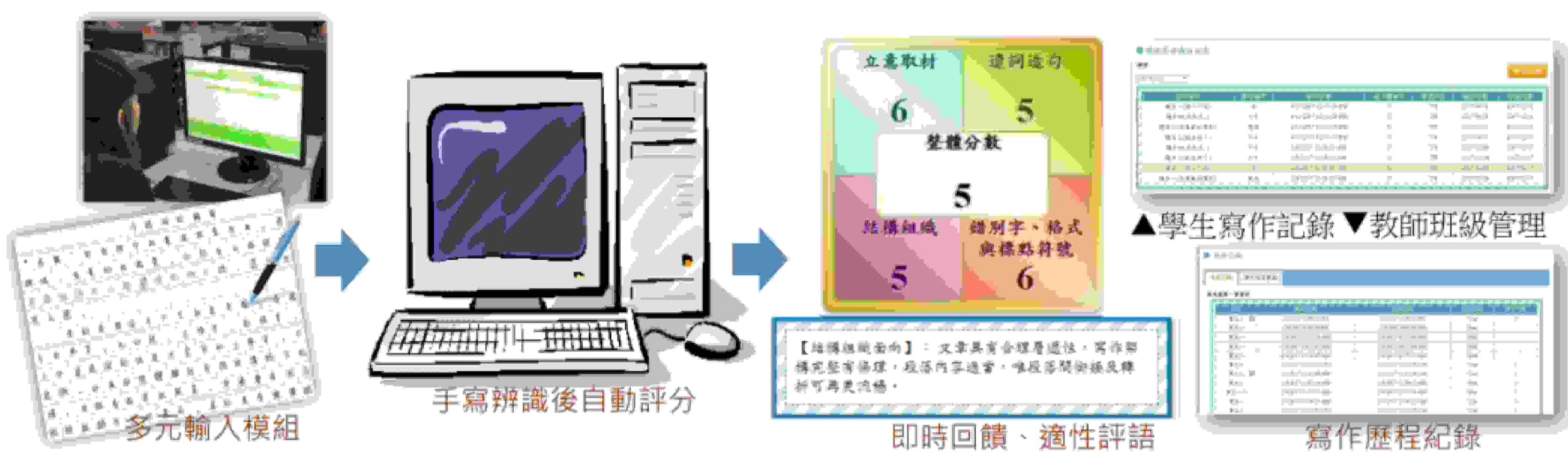
國立臺灣師範大學
教育心理與輔導學系 資訊工程學系
宋曜廷 副校長 曾厚強 博士候選人



國立高雄科技大學
資訊工程學系
張道行 副教授 李少榮 碩士生

作品概述

寫作是語言學習中最難以掌握及精通的一項能力。受限於教學時數不足及一對多的教學現況等因素，致使學生不但缺乏練習，更無法立即得到回饋，大幅降低學生的學習意願和成效。本團隊為不同族群開發出線上中文寫作自動評分系統，並提供全新的自動手寫辨識功能、精準的寫作分數批改、即時的適性化教學及完整記錄寫作歷程，以解決傳統寫作教學上的困難。



語意空間建立與相似度計算技術

透過13億詞的語料，利用word embedding語言模型建構中文語意空間，並提出使用語意空間計算文本語意相似度的演算法。此項技術可用於剖析詞彙關係、句子的語意及篇章主題。

中文斷詞、詞性標記與句法剖析工具

為滿足寫作需求並提高系統效能及技術自主性，團隊獨立研發了基於混合長詞優先及CRF技術中文斷詞與詞性標記元件，也發展了可容錯的中文文法剖析器。

文句偵錯與校正模組

藉由機器從偏誤語料庫中找尋可能的錯誤樣態，再經由基於貝式機率模型的機器學習方法建立偏誤樣態庫。藉由偵測句子中是否存在偏誤樣態，提供給評估模型參考。由於此模組具有校正能力，在回饋中也扮演提供學習者建議的重要角色。

創意及核心技術

為求能貼合教學現場的需求與基礎理論的實踐。本作品自主研发多項核心技術，完全掌握演算法與原始碼，使得後續的產品應用及衍生產品開發都能獨立自主。

中文OCR手寫自動辨識

因應傳統寫作中的紙本寫作方式，能夠將稿紙的手寫文字自動數位化。透過多個卷積神經網路(CNN)模型來進行手寫字的辨識，並利用語言模型來提高辨識的正確率。



| 寫作題目 | 篇數 | 正確率 |
|---------------|------|-----|
| 當一天的老師 | 1200 | 98% |
| 可貴的合作經驗 | 1200 | 99% |
| 當我和別人意見不同的時候 | 1200 | 97% |
| 影響生活的一項發明 | 1200 | 96% |
| 面對未來，我應該具備的能力 | 1200 | 98% |

▲臺灣地區，繁體文本

| 寫作題目 | 篇數 | 正確率 |
|-----------------|-----|-------|
| 我敬愛的唐老師 | 250 | 97.6% |
| 當我和別人發生衝突時 | 205 | 95.1% |
| 上大學前，假如有一年的休學時間 | 280 | 98.9% |

▲香港地區，繁體文本

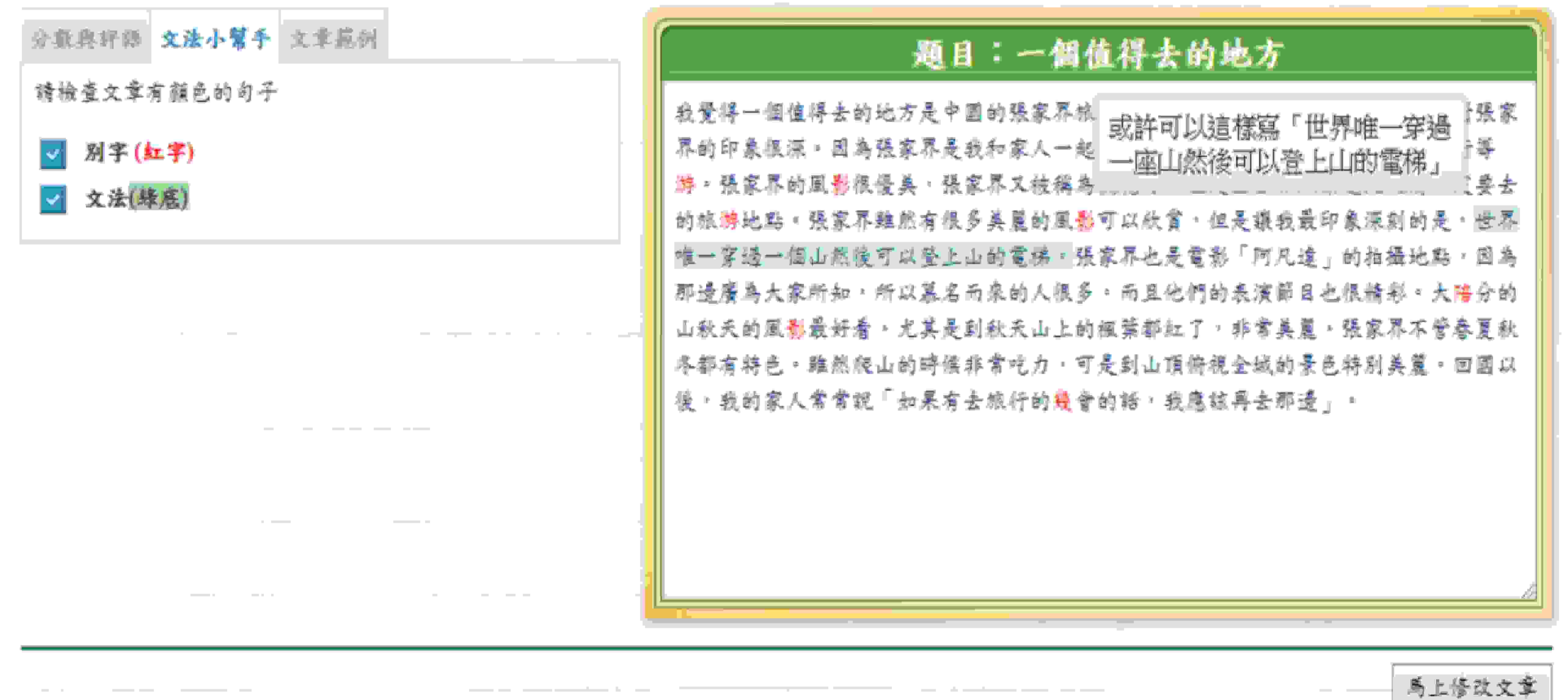
| 寫作題目 | 篇數 | 正確率 |
|-----------|-----|-----|
| 有一種溫暖來自自身 | 461 | 95% |

▲大陸地區，簡體文本

※備註：「正確率」指機器與專業評分員的評分結果，其誤差值在正負1分之內。

寫作能力評估模型

由於中文特性不適用拼音語言慣用的MBM分析模型，而文件分類流行的end-to-end模型不具備可解釋性，因此我們自力發展了以模糊推論引擎為核心的寫作評估模型。這個模型的效能不但能達到與其他模型及真人表現無異的水準，還具有訓練資料少、可解釋性、適用中文特性的優點，在此領域是領先全球的核心技術。



產業及社會貢獻

全球華語產業

成為華語產業指標型產品，強化我國在全球華語教育市場競爭力。

國民語文素養

翻轉中文寫作教學方法，有效提升國民語文素養，保持我國在華語文化圈的軟實力優勢。

AI產品典範

成為人工智慧導入數位教育產品的典範，提高教育產業引入AI技術的質與量。

融入台灣社會

提供新住民有效學習工具，提升華語學習成效。



財團法人 中技社
CTCI FOUNDATION