

運用概念模型化技術於中文大詞彙連續語音辨識之語言模型調適

Leveraging Concept Modeling Techniques for Language Model Adaptation in Mandarin Large Vocabulary Continuous Speech Recognition

郝柏翰 Po-Han Hao
國立臺灣師範大學資訊工程學系
60047082s@ntnu.edu.tw

陳思澄 Ssu-Cheng Chen
國立臺灣師範大學資訊工程學系
60247071s@ntnu.edu.tw

陳柏林 Berlin Chen
國立臺灣師範大學資訊工程學系
berlin@ntnu.edu.tw

摘要

語言模型已被廣泛地使用在各種自然語言相關的研究議題上；譬如，在語音辨識上，語言模型是關鍵的組成，其主要的功能通常是藉由已解碼的歷史詞序列資訊來預測下一個詞彙為何的可能性最大，以協助語音辨識系統從眾多混淆的候選詞序列假設中找出最有可能的結果[1, 2]。本論文旨在於發展新穎動態語言模型調適技術，用以輔助並彌補傳統 N 連(N -gram)語言模型不足之處。為此，我們本提出所謂的概念語言模型(Concept Language Model, CLM)，其主要假設是認為每一句的語句都是用來代表語者內心隱含而欲傳達的概念，並藉由語言(及語音)來具體表達相對應的語意內容。而概念模型最主要的目的則是希望能夠獲取使用者欲表達的概念，並假設在語音辨識過程中，同一概念之中歷史詞序列中所有詞彙以及待預測詞彙具有共同出現的關係，進而藉此關係達到預測待預測詞彙出現機率的目的。

在實作上，概念模型會使用(搜尋)與初步語音辨識結果相關的同領域文件(或調適語料)內表述的若干概念，用以近似語者內心欲傳達的真正含意，並基於此來建立概念語言模型。而概念語言模型的建立是分兩個面向來探討，它們分別是「詞彙」面向與「文件群聚」面向。首先，在實作上，概念模型會使用(搜尋)與初步語音辨識結果近似同領域文件(或調適語料)內表述的若干概念，用以近似語者內心欲傳達的真正含意，並基於此來建立概念語言模型。而概念語言模型的建立是分兩個面向來探討，它們分別是「詞彙」面向與「文件群聚」面向。首先，我們發展所謂的詞概念語言模型(Word-based Concept Language Model)，並應用於語言模型調適。在建構詞概念語言模型時，我們期望能夠針對每一語句不同的語意內容(第一階段語音辨識結果，以詞圖[3]表示)，在調適語料的若干相關的文件中挑選一組具有代表性的概念關鍵詞組，藉以描述任一對歷史詞序列中所有詞

彙與待預測詞彙之間的相依關係。其次，我們亦發展所謂的群聚概念語言模型(Cluster-based Concept Language Model)，假設在調適語料的文件集內之文件可以由一組概念類別度來獲得語句可能的概念分布，並做為語言模型預測的根據。實作上，概念類別的求取可透過一般分群演算法諸如 *K-Means* 演算法[4]而求得；而根據每一語句的語意內容與各個概念類別的相關性，挑選一組具有代表性的概念類別組，藉以描述任一對歷史詞序列中所有詞彙與待預測詞彙之間之共同出現關係，並獲得對於待測詞彙之預測機率。再者，我們嘗試以不同方式來估測此種概念語言模型，並將不同程度的鄰近資訊(Proximity Information)融入概念語言模型以放寬其既有詞袋(Bag-of-Words)假設[3]的限制。

本論文是基於公視電視新聞語料庫來進行大詞彙連續語音辨識(Large Vocabulary Continuous Speech Recognition, LVCSR)實驗，以比較本論文所提出語言模型調適技術與其它當今常用技術之效能，包括了觸發對模型(Trigger-Pair Model)、機率式潛藏語意分析(Probabilistic Latent Semantic Analysis)以及狄利克里分配(Latent Dirichlet Allocation)[5]等。實驗結果顯示我們的語言模型調適技在以字錯誤率(Character Error Rate, CER)評估標準之下，其它當今常用的語言模型調適技術相較，有相當具競爭性的表現，並且對於僅使用 N 連語言模型的基礎語音辨識系統能有明顯的效能提升。

關鍵詞：語音辨識、語言模型、概念模型、鄰近資訊

致謝：本論文之研究承蒙教育部-國立臺灣師範大學邁向頂尖大學計畫(102J1A0800)與行政院科技部研究計畫(MOST 103-2221-E-003-016-MY2, NSC 103-2911-I-003-301, NSC 101-2221-E-003-024-MY3、NSC 101-2511-S-003-057-MY3、NSC 101-2511-S-003-047-MY3 和 NSC 102-2221-E-003-014-MY3)之經費支持，謹此致謝。

參考文獻：

- [1] R. Rosenfeld, "Two decades of statistical language modeling: Where do we go from here," *Proceedings of IEEE*, vol. 88, no. 8, 2000, pp. 1270–1278, 2000.
- [2] J. R. Bellegarda, "Statistical language model adaptation: review and perspectives," *Speech Communication*, vol. 42, no. 11, pp. 93–108, 2004.
- [3] S. Ortmanns, H. Ney and X. Aubert, "A word graph algorithm for large vocabulary continuous speech recognition," *Computer Speech and Language*, vol. 11, 43-72, 1997.
- [4] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval: the Concepts and Technology behind Search*, Addison-Wesley Professional, 2011.
- [5] D. Blei and J. Lafferty, "Topic models," in A. Srivastava and M. Sahami, (eds.), *Text Mining: Theory and Applications*, Taylor and Francis, 2009.