

本期要目

- | | |
|----------------------------|----------|
| 壹、ROCLING 2011 研討會剪影 | 第 2~3 頁 |
| 貳、第十二屆第一次會員大會剪影 | 第 4~5 頁 |
| 參、TAL 2012 Call for Papers | 第 6~7 頁 |
| 肆、專文-現階段鑑別式語言模型訓練研究之簡介 | 第 8~13 頁 |

ROCLING-2011 會議圓滿結束

由國立台北科技大學電子工程系及本會共同主辦的「第二十三屆自然語言與語音處理研討會」已於 100 年 9 月 8 日假台北科技大學科技大樓國際會議廳圓滿結束，參與此次盛會的人士分別來自新加坡及台灣，與會人數多達 180 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 12 篇口頭論文及 13 篇壁報論文。本屆最佳論文共選出兩篇，分別為：賴敏軒先生、黃邦烜先生、陳冠宇先生、及陳柏琳教授共同著作之「實證探究多種鑑別式語言模型於語音辨識之研究」；及李政儒先生、游基鑫先生、及陳信希教授共同著作之「廣義知網詞彙意見極性的預測」。兩篇論文分別於大會閉幕式中，獲頒獎金伍千元。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

第十二屆第一次會員大會圓滿結束

第十二屆第一次會員大會，已於 100 年 9 月 8 日假台北科技大學科技大樓國際會議廳圓滿結束，會中除了報告學會今年度工作會務、明年度工作計畫、及頒發博碩士論文獎外，並順利改選第十二屆理事、監事及海外理事，共選出理事 15 名、候補理事 3 名、監事 5 名、候補監事 1 名、及海外理事 3 名。同日下午五時隨即舉行第十二屆第一次理監事聯席會議，會中選出常務理事三席，三席常務理事分別為中研院資訊所王新民教授、中研院資訊所許聞廉教授及台灣師範大學資工系陳柏琳教授（以姓名筆劃順序排列）。許聞廉教授在所有理事的推選下當選理事長，王新民教授當選副理事長，監事召集人經由監事推舉，由成功大學資工系吳宗憲教授出任。理事及監事任期兩年，自 100 年 9 月 8 日起生效。

第十二屆理監事名冊一點閱

ROCLING 2011 研討會剪影 (1)

大會主席—台北科技大學廖元甫教授



會場—台北科技大學國際會議廳



專題演講—李海洲博士



專題演講—簡立峰博士



口頭論文報告(1)

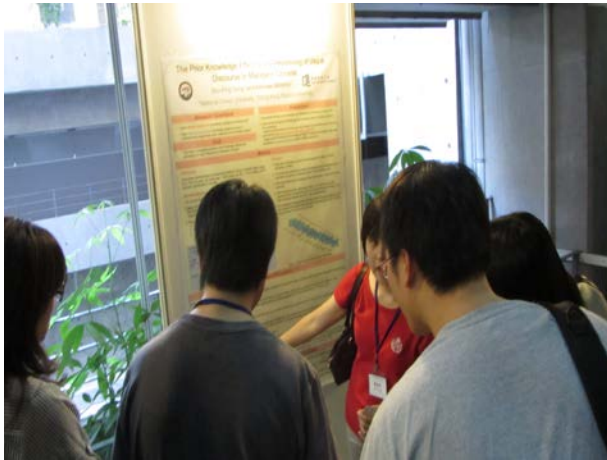


口頭論文報告(2)



ROCLING 2011 研討會剪影 (2)

壁報論文(1)



壁報論文(2)



座談會



攤位展示-賽微科技公司



中場休息



晚宴-美麗信花園酒店



第十二屆第一次會員大會剪影（1）

主席致詞—吳宗憲教授



會務工作報告



理監事改選—開票(1)



理監事改選—開票(2)



頒獎—博士論文優等獎



頒獎—博士論文佳作獎



第十二屆第一次會員大會剪影（2）

頒獎—碩士論文優等獎



頒獎—碩士論文佳作獎(1)



頒獎—碩士論文佳作獎(2)



頒獎—碩士論文佳作獎(3)



頒發感謝狀—理事長吳宗憲教授卸任



The Third International Symposium on Tonal Aspects of Languages (TAL 2012)

Nanjing, China, May 27-29, 2012

www.TAL2012.org

The Third International Symposium on Tonal Aspects of Languages (TAL 2012) will be held in Nanjing, China during May 27-29, 2012. This will be a long-awaited event following TAL 2004 (Beijing, China) and TAL 2006 (La Rochelle, France). As a satellite meeting of Speech Prosody 2012 (to be held in Shanghai, China during May 22-25, 2012), TAL 2012 will be organized by Nanjing Normal University under the joint supports of the International Speech Communication Association (especially by SProSIG and SIG-CSLP), the International Phonetic Association, the Phonetic Association of China, the Chinese Information Processing Society of China, the Acoustical Society of China, and the Chinese Dialect Society.

The theme of TAL 2012 is ‘Tonal aspects across tone and non-tone languages.’ Following the heritage of the past two symposia, TAL 2012 will continue to focus on tone languages, covering studies of tones from phonetic, phonological, psychological, technological, and pathological points of view; but will also welcome studies on tonal aspects of non-tone languages and singing. In particular, this symposium emphasizes the relationship between phonology and phonetics of tones, the relationship between production and perception of tones, the modeling of tones, and the practical utilization of tonal information in spoken language processing. We will make it a great opportunity for linguists, phoneticians, psychologists, language educators, speech pathologists, and speech engineers from all over the world to get together to share and deepen our understanding of tones.

Nanjing, which literally means ‘South Capital,’ has a prominent status in China’s long history and glorious culture. It had been the capital of China during ten dynasties including early Ming Dynasty and the Republic of China. With its beautiful nature and a variety of traditional architectures and historical sights, Nanjing is one of China’s most attractive cities. The transportation to Nanjing is very convenient. There are direct international flights between Nanjing and some cities in Europe, USA and Asian countries. Also, it takes only about 80-90 minutes to travel by train between Nanjing and Shanghai, where the airline connections to almost all main cities in the world are available.

Prospective authors are invited to submit papers (up to 6 pages and written in English) on original research related to tonal aspects of languages, including but not limited to:

- Phonology and phonetics of tones
- Typology of tones and tone languages
- Tone production and perception
- Tone language acquisition and learning
- Evolution of tones and tone change
- Tone, accent, and intonation
- Tonal variation in continuous speech
- Modeling of tonal aspects of languages
- Speech processing for tone languages
- Tonal processing in speech recognition and synthesis
- Psychological and neural mechanisms of tones
- Pathology and therapy for tonal aspects of languages
- Speech corpus and annotation for tones
- Cross-linguistic study of tone languages and non-tone languages
- Tonal information in forensic phonetics
- Tone in singing and traditional operas

We are also calling for exhibits and financial supports. Donations from corporate, academic, government, and individual sponsors would be greatly appreciated. Your support brings you various benefits as stated in our website.

Important Dates

Special session proposal:	December 20, 2011
Submission of full papers:	February 15, 2012
Notification of acceptance:	March 15, 2012
Camera-ready paper due:	March 25, 2012
Early registration deadline:	March 31, 2012

General Chair: Wentao Gu (Professor, Nanjing Normal University, China)

Contact Email: TAL2012nj@gmail.com

現階段鑑別式語言模型訓練研究之簡介

陳柏琳¹, 賴敏軒¹, 陳冠宇², 黃邦烜¹

¹ 國立台灣師範大學資訊工程系

{berlin, 698470623, 699470204}@ntnu.edu.tw

² 中央研究院資訊科學研究所

kychen@iis.sinica.edu.tw

一、前言

在人與人的互動當中，語音是最自然且直接的表達方式之一。透過語音，人們可以彼此溝通，傳達想法、感受以及情緒。因此，我們期望能讓電腦具備與人溝通的能力，能為生活帶來便利性。要達到此目標，我們必須先對使用者輸入的語音訊號進行辨識；待轉換成文字後，再對文字所欲表達的語意作理解，進而做出最適當的動作來回應使用者。將語音訊號轉換成文字的過程，可以透過自動語音辨識(Automatic Speech Recognition, ASR)技術來完成[1]。在自動語音辨識的過程中，我們必須先將語音訊號做特徵擷取(Feature Extraction)，保留語音訊號中的聲學特性(Acoustic Characteristics)，並轉換成能使電腦容易處理的聲學特徵向量(Acoustic Feature Vector)；利用這些聲學特徵向量，我們可以為不同的音素(Phoneme)分別建立聲學模型(Acoustic Model)，進而產生可能的候選詞序列(Candidate Word Sequences)。另一方面，我們也必須收集大量的文字訓練語料，用以統計自然語言中各種詞序列的出現情形，並藉此訓練語言模型(Language Model)。傳統語言模型是收集各種詞彙出現在自然語言中的詞頻數，經由最大化相似度估測(Maximum Likelihood Estimation, MLE)來建立語言模型。例如， N 連(N -gram)語言模型[2]是估測每一個詞在其前面緊鄰 $N-1$ 個歷史詞序列已知情況下的條件機率；它可協助語音辨識器從所產生的候選詞序列中，選取機率最高(最可能)的詞序列做為最後的語音辨識結果。

利用傳統語言模型(例如 N 連語言模型)所選出的語音辨識結果通常是發生機率最高的詞序列，但未必是最佳或是詞錯誤率(Word Error Rate, WER)最低的；換句話說，在候選詞序列(Hypothesis Word Sequences)中其實有可能存在著其它擁有較低詞錯誤率的詞序列可以做為語音辨識器的輸出。於是，我們期望透過探索更多其它語言特徵、使用候選詞序列本身所擷取出的資訊，並經適當設計與訓練的語言模型將所有候選詞序列做重新排序(Reranking)，以輸出擁有較低詞錯誤率的語音辨識結果。近年來，有許多學者提出使用鑑別式語言模型(Discriminative Language Model, DLM)[3, 4, 5, 6]來幫助重新排序，如圖 1 所示。鑑別式語言模型不同於傳統語言模型的是以最小化訓練語料的語音辨識錯誤率為目標，藉由一組

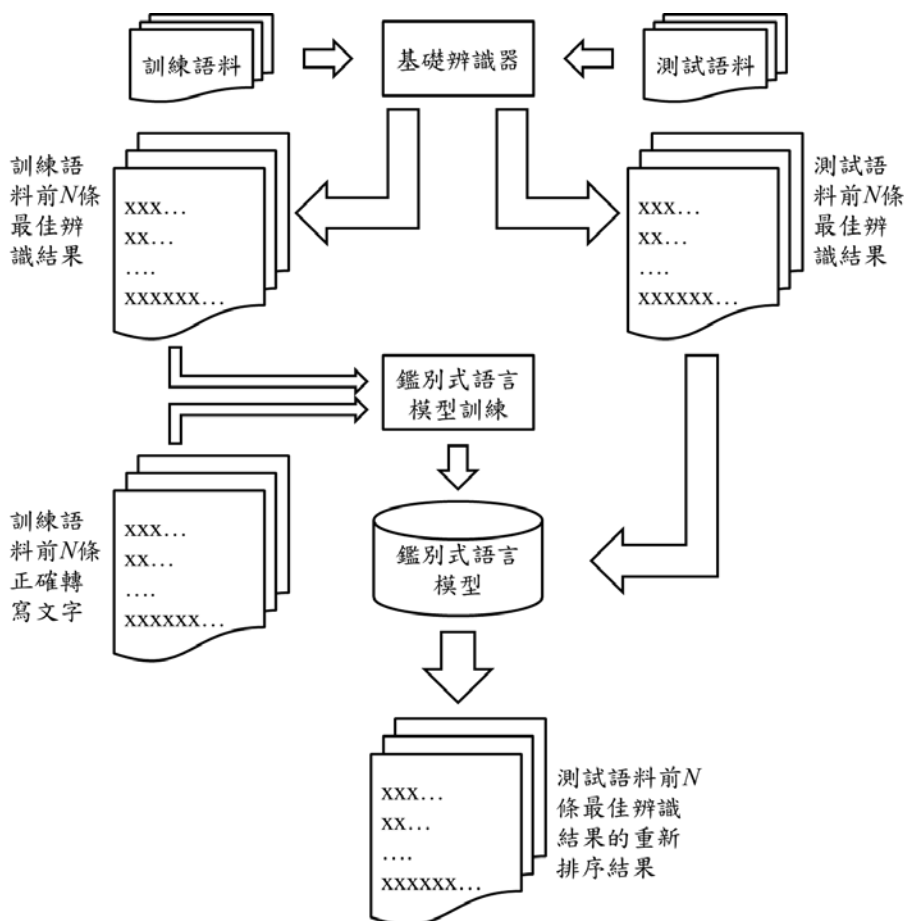


圖 1. 鑑別式語言模型使用於重新排序示意圖。

預先定義的語言特徵與其對應特徵權重參數的乘積之和，將所有存在於詞圖 (Word Graph) 或 M 條最佳序列 (M -best List) 的語音辨識候選詞序列重新計分 (Rescoring) 或重新排序 (Reranking)，期望使具有最低錯誤率的候選詞序列能擁有最高的分數 (或排序)，以做為最後的輸出結果，其流程圖如圖 1 所示 [7]。

本專文將簡介近幾年基於不同訓練準則所發展的鑑別式語言模型，並且說明它們的差異性。

二、常見的鑑別式語言模型訓練方法

早期的鑑別式語言模型大多都使用在其它的應用領域上：例如，機器翻譯 (Machine Translation, MT)、自然語言處理 (Natural Language Processing, NLP) 等。近十幾年來，陸續有許多學者將各種基於不同訓練準則的鑑別式語音模型介紹到大詞彙連續語音辨識 (Large Vocabulary Continuous Speech Recognition, LVCSR) 來使用。目前最常見的鑑別式語言模型依其訓練準則可區分成下四類：最小化平方誤差、最小化錯誤率期望值、最大化對數條件機率、以及考量候選詞序列彼此間之關係。

1. 最小化平方誤差：

感知器演算法(Perceptron)[8]：早期是被應用在人工類神經網路(Artificial Neural Network)領域中；在 2002 年，美國學者 Collins[9]將感知器演算法應用在自然語言處理領域中。感知器演算法可視為是最大化熵值法(Maximum-Entropy, ME)或條件式隨機域(Conditional Radom Fields, CRF)[10, 11]的一種變形。感知器演算法以最小平方誤差(Least Squared Error, LSE)[12]為觀念，透過最小化其訓練目標函數(Training Objective)[13]以求得最佳的特徵權重參數向量。

2. 最小化錯誤率期望值：

最小化錯誤率訓練(Minimum Error Rate Training, MERT)：是在 2003 年由學者 Och[14]提出，並且運用在機器翻譯(Machine Translation)領域中；分別在 2005 年及 2008 年由 Kuo 等學者[3]以及 Kobayashi 等學者[15]將最小化錯誤率訓練方法介紹到語音辨識中的語言模型訓練。當應用於語音辨識時，其訓練準則定義成最小化基礎語音辨識器所產生的 M 條候選詞序列之錯誤率期望值，藉此找出一個最合適的語言模型特徵權重向量。

3. 最大化對數條件機率：

(1) 全域條件式對數線性模型(Global Conditional Log-linear Model, GCLM)：早期被應用在自然語言處理領域中；2007 年 Roark 等學者[6]以有限狀態機(Weighted Finite State Automata, WFSA)實作全域條件式對數線性模型於語音辨識結果的重新排序上，並且與感知器演算法進行比較。全域條件式對數線性模型是希望在給定一句語音訊號與所對應的 M 條最佳候選詞序列時，其中擁有最低辨識錯誤率的詞序列其對數條件機率可以越大越好。

(2) 權重式全域條件式對數線性模型(Weighted Global Conditional Log-linear Model, WGCLM)：不同於全域條件式對數線性模型(GCLM)，Oba 等學者[16]在 2010 年提出將樣本權重加入全域條件式對數線性模型進行改良，為每一個候選詞序列的分數加上一個不同的權重，用來表示每一條候選詞序列不同的重要程度。換句話說，每一個候選詞序列都會有一個相對應的樣本權重；根據不同的樣本權重來表示每一個候選詞序列對於語言模型訓練的不同重要性。

4. 考量語句之間彼此之關係：

輪轉雙重鑑別式模型(Round-Robin Dual Discrimination Model, R2D2)[17]：全域條件式對數線性模型(GCLM)是期望最低錯誤率詞序列的對數條件機率能夠越大越好；Oba 等學者等針對全域條件式對數線性模型提出改良方法，在訓練目標函數中考慮了訓練語句所有候選詞序列

彼此之間的關係，因而有所謂的輪轉雙重鑑別式模型。輪轉雙重鑑別式模型可以視為是全域條件式對數線性模型(GCLM)的一種延伸；它因為考量了兩兩候選詞序列彼此之間的關係，使得其擁有較好的一般化能力。同時，類似於權重式全域條件式對數線性模型(WGCLM)，輪轉雙重鑑別式模型也使用了樣本權重。

此外，最近亦有許多學者針對鑑別式語言模型提出了不同的觀點與做法，例如：

1. 爲了讓鑑別式語言模型的訓練更有效率，在感知器演算法中融入了訓練語句不同錯誤率程度的資訊[18]。
2. 額外地將候選詞序列的文法結構與各種詞性的出現頻率等語言特徵加入鑑別式語言模型使用，讓鑑別式語言模型在對候選詞序列進行重新排序時，可以參考詞序列所含豐富的語言相關資訊[19, 20]。
3. 根據每一句測試語句本身的特性，或者是語句的相關資訊(Relevance Information)，來估測語句相依(Utterance-dependent)的鑑別式語言模型[7]。
4. 考慮邊際(Margin)資訊的概念[21, 22, 23]於鑑別式語言模型之訓練資料選取；對於每個訓練語句嘗試以其每一個候選詞序列各自的辨識錯誤率爲基礎，動態地來決定訓練資料(亦即候選詞序列)是否選取以用於模型訓練。

三、結論與未來展望

語言模型不論是在自然語言處理或語音辨識等領域中，都扮演一個不可或缺的重要角色。在語音辨識中，語言模型能輔助解決聲學模型的混淆，估計一段語句在自然語言中發生的可能性。 N 連語言模型是最常被使用的，但它僅能捕捉到短距離的詞彙規則資訊。近十幾年來，鑑別式語言模型陸續被提出，並且廣泛的使用在於各個領域之中；在語音辨識的領域中，它提供了一個新的視野，以直接降低語音辨識錯誤率爲訓練目標。未來，發展基於不同訓練準則之鑑別式語言模型與探討將各式語言特徵加入鑑別式語言模型使用將持續是重要的課題。另外，如何應用鑑別式語言模型至語音辨識相關領域，諸如語音摘要(Speech Summarization)與語音檢索(Speech Retrieval)，也將是十分吸引人的研究方向。

四、參考文獻

- [1] F. Jelinek, *Statistical Methods for Speech Recognition*, the MIT press, 1999.
- [2] R. Rosenfeld, "Two decades of statistical language modeling: Where do we go from here?," *Proceedings of the IEEE*, 88(8), pp. 1270-1278, 2000.

- [3] J.-W. Kuo and B. Chen, "Minimum word error based discriminative training of language models," *the Annual Conference of the International Speech Communication Association*, pp. 1277-1280, 2005.
- [4] M. Collins and T. Koo, "Discriminative reranking for natural language parsing," *Computational Linguistics*, 31(1), pp. 25-70, 2005.
- [5] J. Gao, H. Suzuki and W. Yuan, "An empirical study on language model adaptation," *ACM Transactions on Asian Language Information Processing*, 5(3), pp. 209-227, 2006.
- [6] B. Roark, M. Saraclar, M. Collins and M. Johnson, "Discriminative n -gram language modeling," *Computer Speech and Language*, 21(2), pp. 373-392, 2007.
- [7] B. Chen and C.-W. Liu, "Discriminative language modeling for speech recognition with relevance information," *the International Conference on Multimedia and Expo*, 2011.
- [8] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychological Review*, 6, pp. 386-408, 1958.
- [9] M. Collins, "Discriminative training methods for hidden markov models: theory and experiments with perceptron algorithms," *the Conference on Empirical Methods in Natural Language Processing*, pp. 1-8, 2002.
- [10] J. Lafferty, A. McCallum and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," *the International Conference on Machine Learning*, pp. 282-289, 2001.
- [11] F. Sha, F. Pereira, "Shallow parsing with conditional random fields," *the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, pp. 134-141, 2003.
- [12] T. M. Mitchell, *Machine Learning*. The McGraw-Hill Companies, 1997.
- [13] Z. Zhou, J. Gao, F.K. Soong and H. Meng, "A comparative study of discriminative methods for reranking LVCSR n -best hypotheses in domain adaptation and generalization," *the International Conference on Acoustics, Speech and Signal Processing*, pp. 141-144, 2006.
- [14] F.J. Och, "Minimum error rate training in statistical machine translation," *the Annual Meeting of the Association for Computational Linguistics on Computational Linguistics*, pp. 160-167, 2003.
- [15] A. Kobayashi, T. Oku, S. Homma, S. Sato, T. Imai and T. Takagi, "Discriminative rescoring based on minimization of word errors for transcribing broadcast news," *the Annual Conference of the International Speech Communication Association*, pp. 1574-1577, 2008.

- [16] T. Oba, T Hori and A. Nakamura, “A comparative study on methods of weighted language model training for reranking LVCSR n-best hypotheses,” *the International Conference on Acoustics, Speech and Signal Processing*, pp. 5126-5129, 2010.
- [17] T. Oba, T. Hori and A. Nakamura, “Round-robin discriminative model for reranking ASR hypotheses,” *the Annual Conference of the International Speech Communication Association*, pp. 2446-2449, 2010.
- [18] T. Oba, T. Hori and A. Nakamura, “An approach to efficient generation of high-accuracy and compact error-corrective models for speech recognition,” *the Annual Conference of the International Speech Communication Association*, pp. 1753-1756, 2007.
- [19] E. Arisoy, M. Saraclar, B. Roark and I. Shafran, “Syntactic and sub-lexical features for Turkish discriminative language models,” *the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 5538-5541, 2010.
- [20] E. Arisoy, B. Ramabhadran and H.-K. J. Kuo, “Feature combination approaches for discriminative language models,” *the Annual Conference of the International Speech Communication Association*, pp. 617-620, 2011.
- [21] V. Magdin and H. Jiang, “Large margin estimation of n -gram language models for speech recognition via linear programming,” *the International Conference on Acoustics, Speech and Signal Processing*, pp. 5398-5401, 2010.
- [22] M.-H. Lai, B.-X. Huang, K.-Y. Chen and B. Chen, "Empirical comparisons of various discriminative language models for speech recognition," *ROCLING XXIII: Conference on Computational Linguistics and Speech Processing*, pp. 4-20, 2011.
- [23] E. Dikici, M. Semerci, M. Saraclar, and E. Alpaydm, “Data sampling and dimensionality reduction approaches for reranking ASR outputs using discriminative language models,” *the Annual Conference of the International Speech Communication Association*, pp. 1461-1464, 2011.