表示法學習技術於節錄式語音文件摘要之研究

A Study on Representation Learning Techniques for Extractive Spoken Document Summarization

施凱文　Kai-Wun Shih, 陳柏琳　Berlin Chen
國立臺灣師範大學資訊工程學系
Department of Computer Science & Information Engineering
National Taiwan Normal University
{60247065S, berlin}@ntnu.edu.tw

陳冠宇　Kuan-Yu Chen, 劉士弘　Shih-Hung Liu, 王新民　Hsin-Min Wang
中央研究院資訊科學研究所
Institute of Information Science
Academia Sinica
{kychen, journey, whm}@iis.sinica.edu.tw

摘要

隨著網路科技的蓬勃發展，大量含有語音資訊的多媒體內容(像是電視新聞、課程演講、會議錄音等)快速地傳遞並分享於全球各地，進而促使自動語音文件摘要成為一項重要的研究議題。其中，長久以來一直最為被廣泛地探究的是節錄式語音文件摘要(Extractive Spoken Document Summarization)[1-4]；其目標在於根據一定的摘要比例，從語音文件中選取重要語句並組合成摘要，以期能夠扼要的表示語音文件主要的主題或語意資訊。藉此，使用者能迅速地瀏覽大量多媒體內容並能充分理解原始語音文件的主題或語意資訊。另一方面，表示法學習(Representation Learning)是近期相當熱門的一個研究議題[5-7]，多數的研究成果也證明了這項技術在許多自然語言處理(Natural Language Proceeding, NLP)的相關任務上，可以進一步地獲得優良的成效。有鑑於此，本論文首先探討使用不同的詞表示法(Word Representations)及語句表示法(Sentence Representations)，包括了連續型詞袋模型(Continuous Bag-of-Words, CBOW)、跳躍式模型(Skip-Gram, SG)、分散式儲存模型(Distributed Memory Model of Paragraph Vector, PV-DM)以及分散式詞袋模型(Distributed Bag-of-Words of Paragraph Vector, PV-DBOW)[8, 9]，於節錄式中文廣播新聞語音文件摘要之應用。其次，基於詞表示法及語句表示法，本論文提出使用三種簡單且有效的排序模型(Ranking Models)，包括了餘弦相似度(Cosine Similarity)、馬可夫隨機漫步(Markov Random Walk, MRW)以及文件相似度量值(Document Likelihood Measure, DLM)[10]，來選取重要語句以形成摘要。再者，除了使用文件中的文字資訊外，本論文更進一步地結合語音文件上的各式聲學特徵，諸如韻律特徵(Prosodic Features)等[11]，以期望能獲得更好的摘要成效。在實驗設定上，本論文的語音文件摘要實驗語料是採用公視廣播新聞(Mandarin Chinese Broadcast News Corpus, MATBN)[12]；一系列的實驗結果顯示，不論是在使用含有錯誤資訊的語音辨識轉寫(Speech Recognition Transcripts)或者是使用正確參考轉寫(Reference Transcripts)的情況下，相較於其它現有的摘要方法，

我們所提出的新穎式摘要方法的確都能夠獲得供顯著的摘要效能增進。

關鍵字: 語音文件、節錄式摘要、詞表示法、語句表示法、韻律特徵

致謝

參考文獻

[1] H. P. Luhn, "The automatic creation of literature abstracts," *IBM Journal of Research and Development*, Vol. 2, No. 2, pp. 159-165, 1958.

[2] B. Chen and S.-H. Lin, "A risk-aware modeling framework for speech summarization," *IEEE Transactions on Audio, Speech and Language Processing*, Vol. 20, No. 1, pp. 199-210, 2012.

[3] K.-Y. Chen, S.-H. Liu, B. Chen, H.-M. Wang, E.-E. Jan, W.-L. Hsu and H.-H. Chen, "Extractive broadcast news summarization leveraging recurrent neural network language modeling techniques," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 23, No. 8, pp. 1322-1334, 2015.

[4] S.-H. Liu, K.-Y. Chen, B. Chen, H.-M. Wang, H.-C. Yen and W.-L. Hsu, "Combining relevance language modeling and clarity measure for extractive speech summarization," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 23, No. 6, pp. 957-969, 2015.

[5] G. E. Hinton, "Learning distributed representations of concepts", in *Proc. the Cognitive Science Society*, pp. 1-12, 1986.

[6] Y. Bengio, R. Ducharme, P. Vincent and C. Jauvin, "A neural probabilistic language model," *Journal of Machine Learning Research*, Vol. 3, pp. 1137-1155, 2003.

[7] T. Mikolov, K. Chen, G. Corrado and J. Dean, "Efficient estimation of word representations in vector space," in *Proc. the International Conference on Learning Representations*, pp. 1-12, 2013.

[8] T. Mikolov, I. Sutskever, K. Chen, G. Corrado and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. the International Conference on Learning Representations*, pp. 1-9, 2013.

[9] Q. V. Le and T. Mikolov, "Distributed representations of sentences and documents," in *Proc. the International Conference on Machine Learning*, 2014.

[10] S.-H. Lin, Y.-M. Yeh and B. Chen, "Leveraging Kullback-Leibler divergence measures and information-rich cues for speech summarization," *IEEE Transactions on Audio, Speech and Language Processing*, Vol. 19, No. 4, pp. 871-882, 2011.

[11] B. Chen, S.-H. Lin, Y.-M. Chang, J.-W. Liu, "Extractive speech summarization using evaluation metric-related training criteria," *Information Processing & Management*, Vol. 49, No. 1, pp. 1-12, 2013.

[12] H. M. Wang, B. Chen, J. W. Kuo, and S. S. Cheng, "MATBN: A Mandarin Chinese broadcast news corpus," *Journal of Computational Linguistics and Chinese Language Processing*, Vol. 10, No. 2, pp. 219-236, 2005.