English

Research Express@NCKU Volume 30 Issue 3 - March 4, 2016
[http://research.ncku.edu.tw/re/articles/e/20160304/4.html]

Code-Switching Event Detection by Using a Latent Language Space Model and the Delta-Bayesian Information Criterion

Chung-Hsien Wu*, Han-Ping Shen and Chun-Shan Hsu
Department of Computer Science and Information Engineering, National Cheng-Kung University
chunghsienwu@gmail.com

IEEE/ACM Trans. Audio, Speech, and Language Processing, Vol. 22, No. 10, October 2014, pp.1558-1570.

This study proposes a new paradigm for code-switching event detection based on latent language space models (LLSMs) and the delta-Bayesian information criterion (ΔBIC). In the proposed approach, acoustic feature and AF posterior probabilities for each senone segment are first extracted and then transformed to the principal components (eigenvectors) in the eigenspace by using principal component analysis (PCA). Latent semantic analysis (LSA) is then adopted to construct a matrix to model the importance of each principal component in the eigenspace for the senones and AFs in each language based on the training data. The spatial relationships among the senones (or AFs) represented by the PCA-transformed eigenvalues in the LSA-based matrix are employed to construct a latent language space model (LLSM) for characterizing a language. Based on this notion, in the detection phase, the acoustic features (or AFs) in the recognized senones of an input speech utterance are PCA-transformed to the eigenspace to form an input senone (or AF)-based sub-LLSM. The LLSM of each input speech utterance is compared with the LLSM of a target language for likelihood estimation. The ΔBIC is consequently adopted to compute the language transition score for each potential change point in the input speech utterance. To avoid an exhaustive search among all of the potential change points, the recognized phone boundaries from the automatic speech recognizer are regarded as the potential language change points. Finally, the dynamic programming algorithm is employed for identifying the most likely language sequence based on the similarities estimated from the LLSMs and ΔBIC.

Figure 1 illustrates the system framework of the proposed code-switching event detection mechanism.

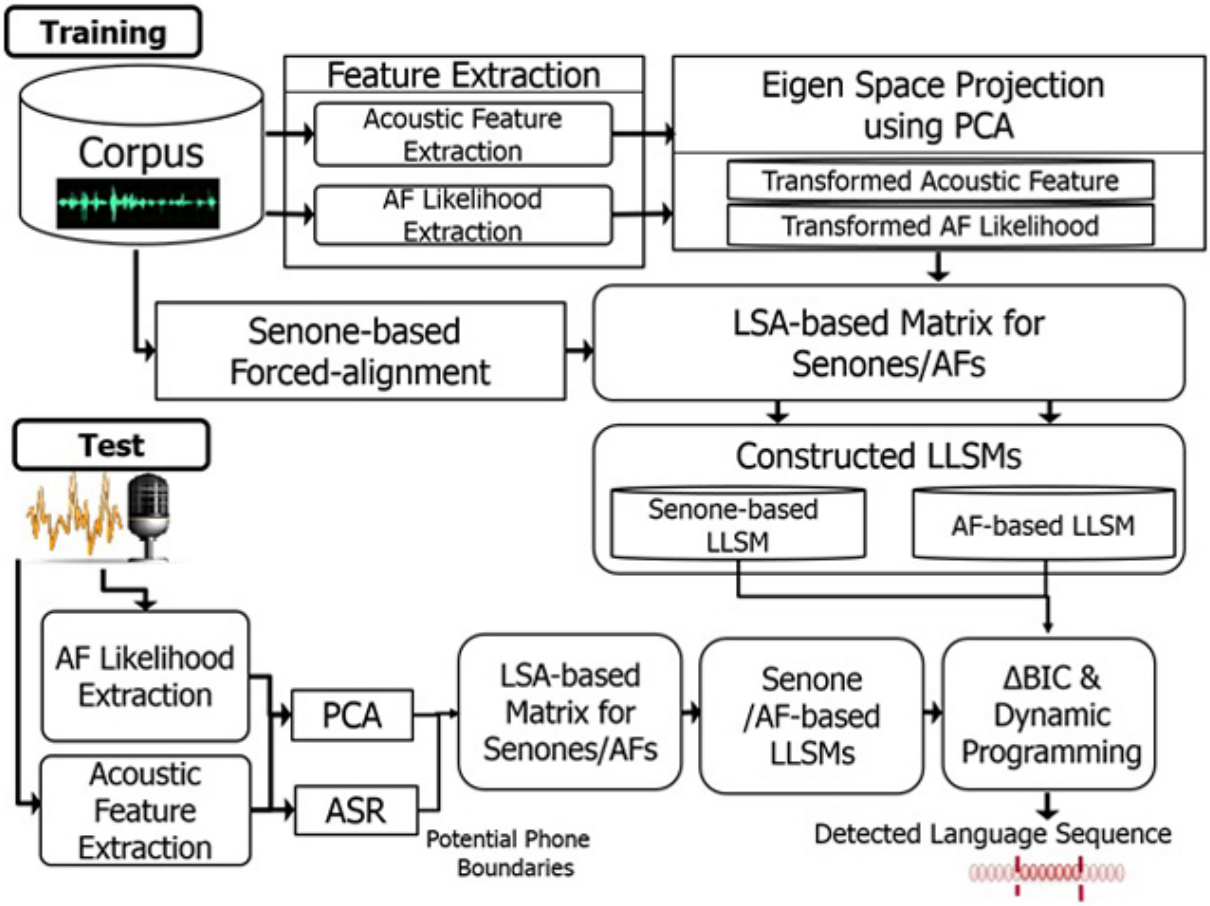

Fig.1 Illustration of the system framework of the proposed approach for code-switching event detection.

For evaluation, SVM-, GMM-, ANN-based and the proposed approaches were compared. These approaches were used to tokenize the incoming speech into language sequence. Fig. 2 presents all of the results obtained using these methods. The evaluation results indicated that the proposed method along with MFCC features mostly outperformed the other three methods using precision, recall, harmonic mean and duration accuracy.

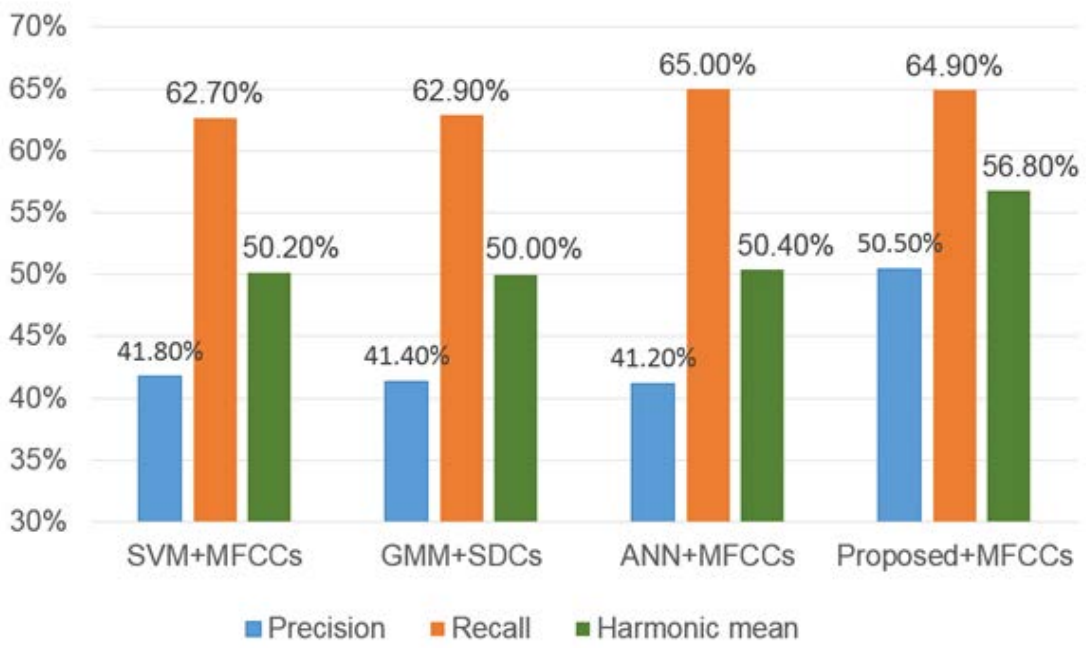

Fig. 2 Precision, recall and harmonic mean for the SVM-, GMM-, ANN-based and the proposed methods. The term following the approach is the features used.

[1] H. Y. Su, “Code-switching between Mandarin and Taiwanese in three telephone conversation: The negotiation of interpersonal relationships among bilingual speakers in Taiwan,” in *Proc. The Symposium about Language and Society*, Apr 2001.

[2] C. Chen, “Two types of code-switching in Taiwan,” in *Proc. Sociolinguistics Symposium 15* (SS15), Newcastle upon Tyne, United Kingdom., Apr 2004.

[3] H. Halmari, Government and Code-Switching: Explaining American Finnish. Amsterdam: John Benjamins, 1997.

[4] J. Weiner, N. T. Vu, D. Telaar, F. Metze, T. Schultz, D.-C. Lyu, E.-S. Chng and H. Li, “Integration Of Language Identification Into A Recognition System For Spoken Conversations Containing Code-Switches,” in Proceedings of The third International Workshop on Spoken Languages Technologies for Under-resourced Languages (SLTU'12), 2012.

[5] A. Hanani, “Human and Computer Recognition of Region Accents and Ethnic Groups from British Speech,” PhD Thesis, The University of Birmingham, March 2012.

[6] C.-H. Wu and C.-H. Hsieh, “Multiple change-point audio segmentation and classification using an MDL-based Gaussian model,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 2, pp. 647–657, Mar 2006.

[7] C.-H. Wu, H.-P. Shen and Y.-T. Yang, “Chinese-English Phone Set Construction for Code-Switching ASR Using Acoustic and DNN-Extracted Articulatory Features,” *IEEE/ACM Trans. Audio, Speech, and Language Processing*, Vol. 22, No. 4, April 2014, pp. 858-862.