Hierarchical Committee and Top-Down Attention for Robust Classification: Cases for Emotional Facial Expression and Noisy Speech Recognition

Bo-Kyeong Kim, Ho-Gyeong Kim, and *Soo-Young Lee

School of Electrical Engineering and Brain Science Research Center,

Korea Advanced Institute of Science and Technology

***E-mail: sylee@kaist.ac.kr**

This talk consists of two approaches to provide robust classification performance in real world applications, i.e., emotion recognition from facial expression and speech recognition in noisy environment.

The emotional facial expression recognition incorporating a hierarchical committee machine won the emotion-from-images competition at the third Emotion Recognition in the Wild (EmotiW2015) challenge. [1] We trained multiple deep convolutional neural networks (CNNs) as committee members and combined their decisions with two strategies: (1) in order to obtain diverse decisions from deep CNNs, we incorporated several different network architectures, input normalization, and random weight initializations for training these deep models, and (2) in order to form a better committee in structural and decisional aspects, we constructed a hierarchical architecture of the committee with exponentially-weighted decision fusion. For the recognition of seven emotional categories in the wild, we achieved a test accuracy of 61.6 %. Moreover, on other public databases, our hierarchical committee of deep CNNs yielded superior performance, outperforming or competing with the state-of-the-art results for these databases.

To achieve high accuracy for noisy speech recognition, we also incorporated top-down attention which automatically assigned attention gain on input and/or hidden variables for higher confidence on the classification decision. [2=4] Although it is basically similar to recent attentive networks, unlike image recognition tasks in big background, the segmentation of speech even in noisy environment is relatively easy task and we applied the top-down attention only at test phase. Also, several top candidate classes were attended and only the class with the mximum confidence was selected as the final decision. This approach successfully resulted in sequential recognition of superimposed patterns and continuous speech recognition in noisy environment.

Refereneces

[1] Bo-Kyeong Kim, Jihyeon Roh, Suh-Yeon Dong, Soo-Young Lee, “Hierarchical committee of deep convolutional neural networks for robust facial expression recognition,” Journal on Multimodal User Interfaces, June 2016, Volume 10, Issue 2, pp 173–189.

[2] Ho=Gyeong Kim and Soo-Young Lee, in preparation.

[3] Chang-Hoon Lee and Soo-Young Lee, “Noise-Robust Speech Recognition Using Top-Down Selective Attention with an HMM Classifier”, IEEE Signal Processing Letters, Vol. 14, Issue 7, pp. 489-491, 2007. 07.
[4] B.T. Kim and S.Y. Lee, "Sequential Recognition of Superimposed Patterns with Top-Down Selective Attention", Neurocomputing, Vol. 58-60, pp. 633-640, 2004. 06.