

Research on Vietnamese speech recognition method using deep learning HMM-DNN model

Pham Van Quang

Abstract: This study proposes a deep learning-based approach for Vietnamese speech recognition, utilizing a hybrid Hidden Markov Model-Deep Neural Network (HMM-DNN) architecture. While traditional HMM-Gaussian Mixture Model (GMM) systems have been widely used, they often struggle to model the complex acoustic variations inherent in Vietnamese, particularly its tonal characteristics. To address this limitation, we replace the GMM with a Deep Neural Network to estimate posterior probabilities of tied triphone states. The DNN is trained using a discriminative criterion, enabling it to learn high-level, non-linear feature representations directly from spectral features. The model is evaluated on a standard Vietnamese speech corpus, and experimental results demonstrate that the proposed HMM-DNN system significantly outperforms the conventional HMM-GMM baseline, achieving a relative reduction in Word Error Rate (WER) of over 15%. Furthermore, the system shows improved robustness in recognizing tonal syllables and co-articulated phonemes. These findings confirm that the HMM-DNN hybrid model is highly effective for large-vocabulary continuous Vietnamese speech recognition, offering a scalable and accurate solution for real-world applications.

Keywords: Vietnamese speech recognition, HMM-DNN, deep learning, acoustic modeling