

عنوان مقاله:

Persian Speech Recognition Through the Combination of ANN/HMM

محل انتشار:

فصلنامه پردازش سیگنال و انرژیهای تجدیدپذیر، دوره 7، شماره 2 (سال: 1402)

تعداد صفحات اصل مقاله: 22

نویسندگان:

Ladan Khosravani pour - Department of Electrical Engineering, South Tehran Branch, Islamic Azad University, Tehran, Iran

Ali Farrokhi - Department of Electrical Engineering, Islamic Azad University, South Tehran Branch, Tehran, Iran

خلاصه مقاله:

The goal is to create a speech recognition system that is able to recognize Persian speech. Pro-sodic speech is attributed to the hierarchical structure from speech rhythm and tonal expression to the smallest syllable components and provides important information about trans segmental features such as F_0 (fundamental frequency), intensity, and duration, which are crucial for natural sound. Prosodic features are highly language dependent, however, the relationship between linguistic features and prosodic data is not well understood in some languages. While relatively high-performance prosodic generators have been developed for many languages, very limited work has been done on prosodic generators in Farsi. In this article, we first use a simple four-layer RNN to extract prosodic information, then we investigate the hybrid ANN/HMM model for Persian speech recognition. ۲۱۰ samples of the speech of a male person were collected and after removing the noise, ۴۷ of the samples were manually labeled phonetically. Then, the remaining training samples were automatically labeled and new neural networks (ANN) were created for the final recognition of the three-layer MLP type. Four methods including MEL, MEL derivative, energy, and energy derivative were used to extract features, and the values of each of these four methods were combined and given to the neural network. Then we use the neural network to classify these feature vectors and get the most similar vowels. We give the order of vowels as "observations" to HMMs (which are created based on pronunciations) and then find the most probable HMM (or in other words, the most words) to the input sound and output it. By applying recognition on ۹۹.۴% of test data, we even reached ۱۰۰% accuracy in one case, which is a very favorable result considering the small number of speech data

کلمات کلیدی:

Artificial Neural Networks, hidden Markov models, Discrete Fourier transform, Vector Digitizer, Linear predictive coding, Viterbi Algorithm, Fuzzy Expectation Maximization, probabilistic neural networks, recurrent neural networks

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/1680958>



