Accepted Manuscript

Speech Bottleneck Feature Extraction Method Based on Overlapping Group Lasso Sparse Deep Neural Network

Yuan Luo , Yu Liu , Yi Zhang , Congcong Yue

PII: S0167-6393(17)30128-0
DOI: 10.1016/j.specom.2018.02.005
Reference: SPECOM 2537

To appear in: *Speech Communication*

Received date: 9 April 2017
Revised date: 28 November 2017
Accepted date: 10 February 2018

Please cite this article as: Yuan Luo , Yu Liu , Yi Zhang , Congcong Yue , Speech Bottleneck Feature Extraction Method Based on Overlapping Group Lasso Sparse Deep Neural Network, *Speech Communication* (2018), doi: 10.1016/j.specom.2018.02.005

Speech Bottleneck Feature Extraction Method Based on Overlapping Group Lasso Sparse Deep Neural Network

Yuan Luo[a], Yu Liu[a,]*, Yi Zhang[b], Congcong Yue[a]

[a]*School of Optoelectronic Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China.*

[b]*School of Advanced Manufacture Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China.*

Abstract: In this paper, a novel speech bottleneck feature extraction method based on overlapping group lasso sparse deep neural network is proposed. This method extracts the speech bottleneck features which contain supervised class information and adjacent voice frames related information by changing the architecture of deep neural network (DNN), to improve of the performance speech recognition. Firstly, in order to construct the sparse deep neural network, the sparse regularization of the overlapping group lasso algorithm is added to the single DNN object function which is regarded as the penalty item. Second, the speech bottleneck features can be extracted from the bottleneck layer of the sparse bottleneck DNN (BN-DNN). Finally, the speech bottleneck features are used as the input features of the deep neural network-hidden Markov model (DNN-HMM) speech recognition system. The large vocabulary continuous speech recognition experiment results on the Switchboard database indicate that the algorithm proposed in this paper can extract the speech bottleneck features with priori information, and reduce the word error rate in continuous speech recognition.

Key words: automatic speech recognition, deep neural network, overlapping group lasso, speech bottleneck feature

1 Introduction

Automatic speech recognition (ASR) technology is one of the key technologies of human-computer interaction. Automatic speech recognition system consists of five parts, which are speech signal preprocessing, feature extraction, acoustic model, language model and decoder. As for feature extraction, we mainly extract partial speech features which are more conducive to speech recognition from a large number of voice data, and try to eliminate a large amount of redundant information which is not useful for speech recognition. There are many types of speech features currently used in speech recognition system, such as mel-frequency cepstral coefficient (MFCC), liner prediction cepstral coefficient (LPCC) [1]. But these features are susceptible to speaker accent and other factors in the actual environment, leading to poor robustness [2]. In addition, there are two disadvantages: the influence of adjacent speech frames is neglected during feature extraction; the voice label information used for forced alignment to obtain a frame level labeling for fine-tuning the DNN is also ignored, and some of the features related to the target label are not extracted, so we cannot effectively use the