

# Deep Neural Network with Band Matrices for Speech Recognition

Hainan Xu  
hxu31@jhu.edu

## 1 Abstract

Deep Neural Networks (DNN) in recent years have been proven to be very successful in the Automatic Speech Recognition task, and in the speech recognition toolkit kaldi, DNN is consistently giving the best results for almost all the speech corpora.

In regular neural networks, each layer would be fully connected with the previous layer; In this project, I will investigate a special type of neural network with band matrices. Neural networks with band matrices, compared to regular neural networks, would hopefully better recognize correlations between adjacent entries in the feature vector, which might makes more sense in modeling speech signals.

## 2 Methods

According to Wikipedia, a band matrix is “a sparse matrix whose non-zero entries are confined to a diagonal band, comprising the main diagonal and zero or more diagonals on either side”. If we apply a band matrix into the neural network idea, we will get a neural that is not fully connected, and each node would only depend on nodes within a “neighborhood” in the previous layer.

I think this idea might work for speech because in speech signals, e.g. there might be stronger correlations between neighboring frequencies than just arbitrary frequencies.

In my previous research I already implemented a band matrix library, and the task here for this class would be implement on top of the matrix library a neural network component which supports forward/backward propagation algorithms that could be used in training and decoding for the speech recognition system kaldi.

## 3 Resources

I am going to use the open source speech recognition toolkit kaldi as my starting point. Kaldi supports training and testing of regular neural networks but there is not yet support for band matrices, and I am going to implement that.

The data resource needed for this project would mainly be the speech corpus used for training and testing. I will use the “Resource Management” corpus that is commonly used in the speech community for ASR system evaluations.

## 4 Milestones

### 4.1 Must achieve

Implementation of the neural network with band matrices, including all the functions that are needed for training and testing.

### 4.2 Expected to achieve

Scripts for training and testing the neural network code on the “RM” corpus.

### 4.3 Would like to achieve

Achieving comparable performance in terms of recognition accuracy on the “RM” corpus using the network.

## 5 Final Writeup

In the final writeup, I will include the detailed math behind band matrices and how it could be used for neural networks; I will also describe how to train such a neural network with band matrices.

I will also give numbers about how well the system will actually do compared to the regular neural network baseline, both in terms of accuracy and efficiency.

## 6 Bibliography

Dan Povey et al., “The Kaldi Speech Recognition Toolkit”

Geoffrey Hinton et al., “Deep Neural Networks for Acoustic Modeling in Speech Recognition”