# Application-of-Word2vec-in-Phoneme-Recognition

Build an attention-based model for speech recogntion.Use the Word2vec model to help to train the attention model.

Some code is quoted from https://github.com/thomasschmied/Speech\_Recognition\_with\_Tensorflow. But there are some mistakes in his file, which i have changed these mistakes.

Our model is built based on phoneme recognition. The datasets used are librispeech(http://www.openslr.org/12) and TIMIT.The pronunciation dictionary used is the 39 phoneme pronunciation dictionary of CMU.

We used a new method in the model. The word2vec model is used to initialize the embedding matrix in the attention model.This can make the distance between the embedding vectors larger, so it can improve the performance of the model.At the same time, in order to solve the overfitting problem of the attention dataset on the attention model.We use a new phoneme inverse mapping strategy to convert more 39 phoneme datasets to 61 phoneme datasets.

This is my paper (《Application of Word2vec in Phoneme Recognition》) about how to make this experiment and result.

The framework of our network:



Some mapping example:

|  |
| --- |
| **Table1: Mapping and inverse mapping 61-39 phonemes(some examples)** |
| |  |  |  |  | | --- | --- | --- | --- | | 61 to 39 | | 39 to 61 | | | aa,ao | aa | aa | aa/ao | | ah,ax,ax-h | ah | ah | ah/ax/ax-h | | bcl/dcl… | sil | d | dcl d | |

Result:

| **Table 2: Phoneme Error Rate (PER) on Librispeech and TIMIT in different experiments.** | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Index | Phoneme | Model set | Librispeech（PER） | TIMIT  (PER) | Cos | M-dist | Cov | P-dist |
| 1 | 39 | No-pretrain | 12.7 | 22.7 | 0.289 | 52.00 | 2.501 | 0.359 |
| 2 | Pretrain | 8.37 | 19.1 | 0.097 | 3.92 | 0.055 | 0.149 |
| 3 | 61 | Pretrain | 8.57 | 18.5 | 0.128 | 33.85 | 0.476 | 0.162 |
| 4 | Pretrain+all dataset | 9.41 | 18.1 | 0.142 | 44.24 | 0.631 | 0.176 |
| 5 | Pretrain+corrective training | 8.84 | **16.5** | 0.136 | 12.83 | 0.178 | 0.171 |