600.676 Machine Learning Mid Report

Tianyi Chen, Huizhan Lu, Guanyi Wang, Mingyue Gao

1 Goal

- General Goal: Based on speech file, predict the most likely words.
- Achieved consequence: Construct training HMM model.

2 Related works

- *S.Young*, *A Review of Large Vocabulary Continuous Speech Recognition*,*IEEESignalProcessingMagazine*, *pp*45−57, *Sept*1996.
- *Rabiner*, *LA tutorial on hidden Markov models and selected applications in speech recognition*,*ProceedingsoftheIEE* 77, *Issue* : 2*Feb*1989
- *Bahl*, *L.R.Acoustic Markov models used in the Tangora speech recognition system*,*Acoustics*, *Speech*, *andSignalProces* 88., 1988*InternationalConferenceon*11 − 14*Apr*1988

3 Model

3.1 Fenonic Model

First, we should construct the acoustic model by computing the probability $P(A|W)$ for any acoustic data string $A = a_1, a_2, a_3, ..., a_m$ and hypothesized word string $W = w_1, w_2, ..., w_n$. Our aim is to find a desired transcribed word string $\hat{W}$ defined by

$$\hat{W} = \arg\max_{W} P(A|W)P(W)$$

Under finite acoustic data alphabets $\mathcal{A}$ with several hundreds of symbols. In our project, we will use 256 two-character-long fenemes to create a fenonic baseforms for each words as acoustic alphabets $\mathcal{A}$. Then a Hidden Markov Model is constructed on fenone, to simply the work, we just assume that there exists two or four hidden states, the meaning of these hidden states is not fixed, they may be consonant, vowel,etc.
The structure looks like Figure 1.

3.2 Create Silence Model

Except all acoustic symbols, there exists silence in one word. We should also incorporate the silence in our model. A similar HMM model is also constructed on silence symbol, the number of hidden states may be also two or four.

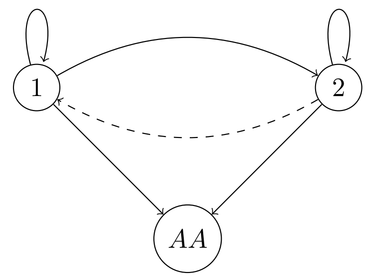

Figure 1: fenonic HMM, there exist two possible states, the dash line represents null transition that only change the state but exchange none information

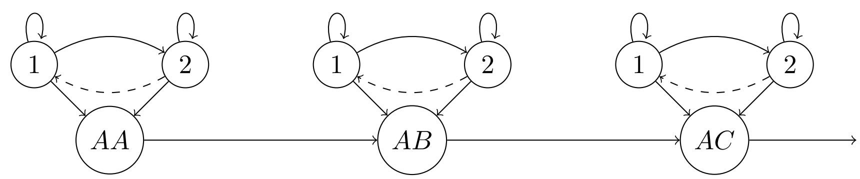

Figure 2: word HMM, constructed by combining the HMM model of each fenonic together

3.3 Establish fenonic bases for each training word

We can construct a fenonic bases by combining 256 fenemes with the silence symbol. The each word can map to some fenonic bases. In terms of the existence of accent, one word may have different fenonic bases, we can collect them to establish possible fenonic set for each word.
For example, for the word "hopkins", its corresponding fenonic base can be (AA, AB, sil, BC),etc.

3.4 Construct word HMM model

After establishing each word's fenonic bases, and HMM model for each fenonic bases, a word HMM model can be constructed by connecting HMM model of each fenonic symbol and silence symbol together,like Figure 2. $HMM_{word} = HMM_{fenonic_1} \bigcup HMM_{fenonic_2} \bigcup HMM_{fenonic_3} \bigcup ... \bigcup HMM_{fenonic_n}$.

3.5 Training the word HMM model

After establishing the word HMM model, the next work is to estimate the transition matrix. To train the word HMM, we can use EM algorithm to estimate the parameters, including forward-backward algorithm, Baum-Welch Algorithm. Given a observed sequence of fenonic symbols, we can also use Viterbi Search to find the most possible state transition. A python code has been provided in order to train the parameters of HMM model. It can be found in my github:

`https://github.com/motian12ps/machine-learning-data-to-model/tree/master/project`

3.6 Predict the word by using the word HMM system

Given a sequence of combination of fenonic symbols, we can predict the word by using the trained word HMM model. For details, we can calculate the forward-probability of the acoustics, and pick the word with the highest likelihood.

4 Inference Technique

We plan to use forward propagation, and backward propagation to inference the probability of each possible word given an observed fenonic symbol combination.

5 Dataset

Data set comes from information extraction course of Prof. Sanjeev. I audit it this semester in order to complete our machine learning project. I have uploaded them in our github.

`https://github.com/motian12ps/machine-learning-data-to-model/tree/master/project`

References

[1] S. Young, *A Review of Large Vocabulary Continuous Speech Recognition,* IEEE Signal Processing Magazine, pp 45-57, Sept 1996.