Speech Technology Mini Project 2

EE17B035 — EE17B047

May 2020

1 Problem Statement

This mini project involves building Hidden Markov Models for isolated digit recognition followed by continuous digit recognition using the isolated digit HMM through concatenation.

2 Implementation of Isolated Digit Recognition

Prior to training the Hidden Markov Models, firstly we must discretize the observations. In order to facilitate this, we perform a K-means clustering over the development data which allows us to convert them into given number of symbols.

Then, in order to build the Hidden Markov Models for each of the isolated digits, we made use of Baum-Welch Algorithm to train the parameters of the models namely, State Transition Matrix(A), Emission Probability Matrix(B) and Initial State Probability Vector(π).

Firstly, initialize all the three parameters accordingly. Then, we implement the forward-backward algorithm to obtain the alpha and beta matrices. The equations for the previously stated operation is as below:

$$\alpha_i(1) = \pi_i B_i(y_1)$$

$$\alpha_i(t+1) = B_i(y_{t+1}) \Sigma_{j=1}^{N} A_{ji} \alpha_j(t)$$

In the above equations, i refers to a specific state and t refers to the time instant in the observation sequence and y_t refers to the observation at time t. Here, α serves as a measure of the forward probabilities. Now, we measure β which gives us a measure of the backward probability. The equations related to it are as given below :

$$\beta_i(T) = 1$$

$$\beta_i(t) = \Sigma_{j=1}^{N} B_j(y_{t+1}) A_{ij} \beta_j(t+1)$$

Once we calculate α and β, we can now calculate the updated parameters for the HMM. The corresponding equations are as given below:

$$\gamma_i(t) = \frac{\alpha_i(t)\beta_i(t)}{\Sigma_{j=1}^{N}\alpha_j(t)\beta_j(t)}$$

$$\eta_{ij}(t) = \frac{\alpha_i(t)A_{ij}\beta_j(t+1)B_j(y_{t+1})}{\Sigma_{k=1}^{N}\Sigma_{w=1}^{N}\alpha_k(t)A_{kw}\beta_w(t+1)B_w(y_{t+1})}$$

The update A, B, π are as follows :

$$\pi_i^* = \gamma_i(1)$$

$$A_{ij}^* = \frac{\Sigma_1^{T-1}\eta_{ij}(t)}{\Sigma_1^{T-1}\gamma_i(t)}$$

$$B_j(v_k) = \frac{\Sigma_1^{T}1_{y_t=V_k}\gamma_i(t)}{\Sigma_1^{T}\gamma_i(t)}$$

Hence, by running the above for a set of iterations would give us the desired HMM model parameters.

3 Implementation of Multiple Digit Recognition

Below are the methods enlisted using which we proceeded to predict multiple digit numbers.

3.1 Concatenation of Isolated HMM's

In this method, we concatenated the A, B matrices of the isolated digits so as to make sure that we have a model for the corresponding number we want to recognize. Then, we ran the same test procedure to figure out which number produced the maximum probability of the observation sequences.

3.2 Dividing the observation sequence

In this method to recognize multiple digit number, we divided the observation sequence into equal parts of number of digits in the number. Post this, we tried to find out what digit each of the parts corresponded most to and thereby try to figure out the entire number.

4 Observations

4.1 Isolated Digit Recognition

4.1.1 Accuracy

The above section states the method of implementation of the Isolated digit HMM's. Our group was assigned the numbers {1,2,8,o,z}. The graph which shows the accuracy for various number of symbols in a given state for a specific state value is as below:

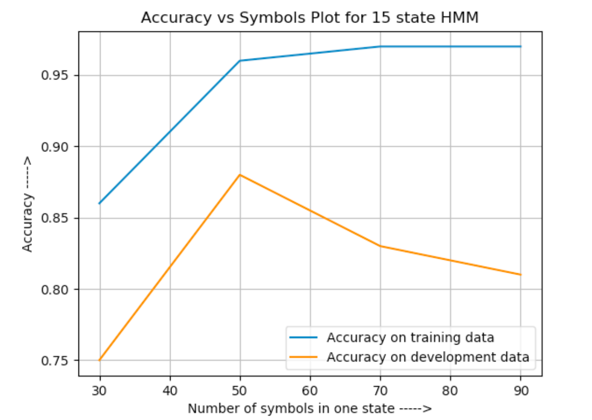

Figure 1: Accuracy for 15 state HMM

4.1.2 Confusion Matrix

Below is the confusion matrix for the development and train data respectively for various configurations of states and number of symbols in each state as mentioned below. All the models were trained for 30 iterations.

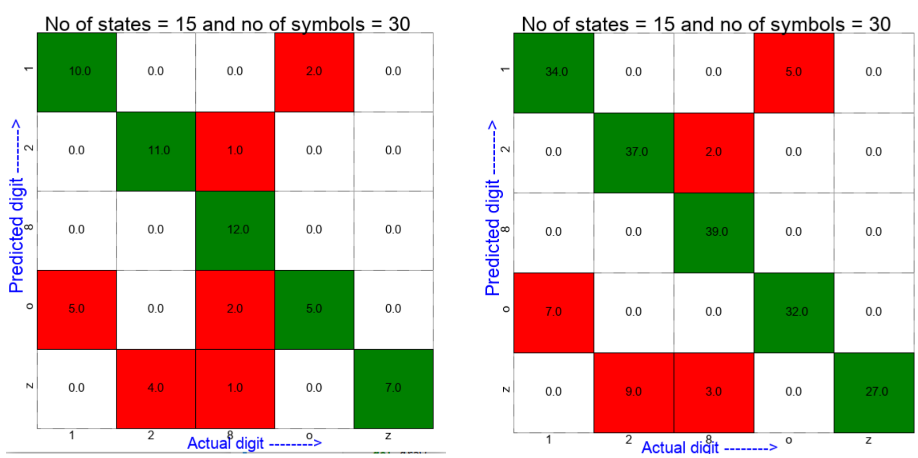

Figure 2: (a) Development data (b) Training data on 30 states

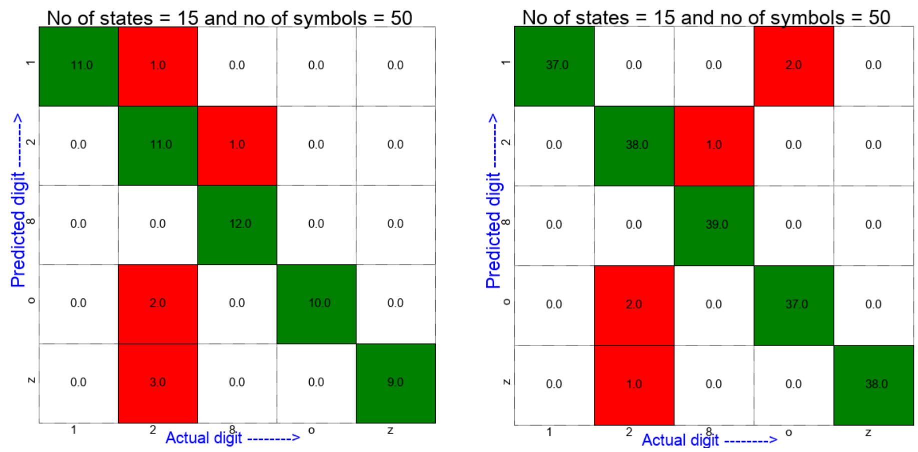

Figure 3: (a) Development data (b) Training data on 50 states

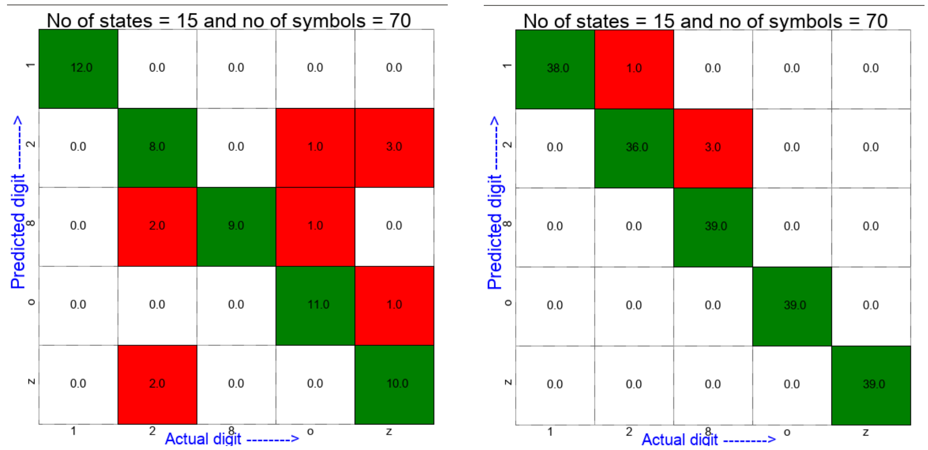

Figure 4: (a) Development data (b) Training data on 70 states

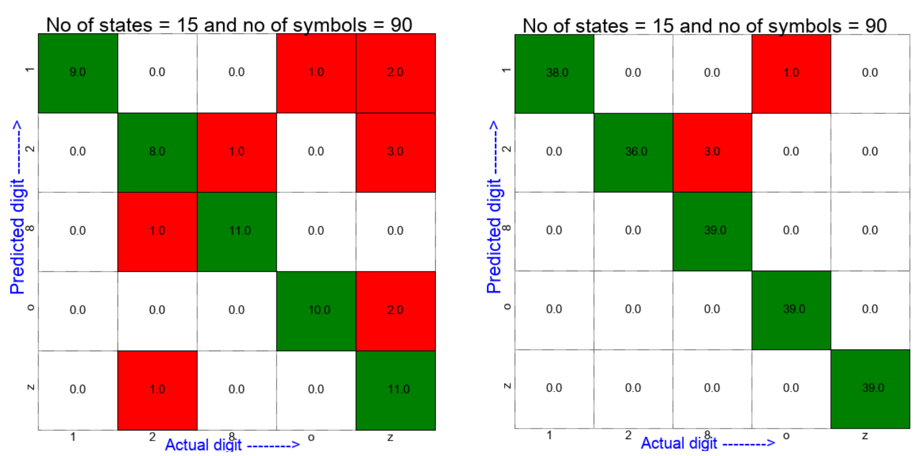

Figure 5: (a) Development data (b) Training data on 90 states

4.1.3 ROC curves

ROC curves are plotted with true positive rate on the y axis and false positive axis against the x axis. Below is the graph which represents the combined curves of various configuration of 15 state HMM's.

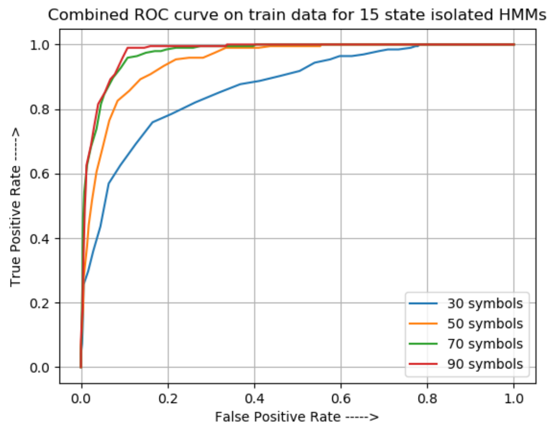

Figure 6: combined ROC curves on training data

Below are the individual ROC curves for 15 states isolated HMM digits for varied number of symbols in each state. The below ROC curves are obtained on training data sets.

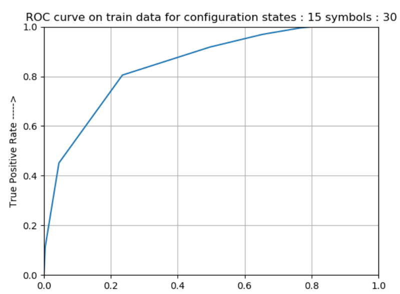

Figure 7: ROC for 15 state 30 symbol HMM

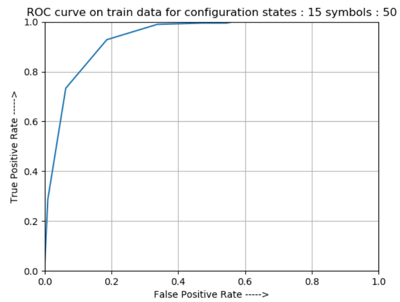

Figure 8: ROC for 15 state 50 symbol HMM

From the above, we can observe that the 50 symbol HMM has a better ROC curve over 30 symbol HMM and similar argument holds true for the 70 and 90 symbol plots mentioned below. Therefore, we can conclude that larger the number of symbols in each state of HMM, the better true positive rate we can get for smaller false positive rate.

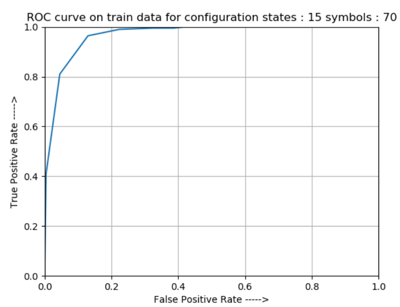

Figure 9: ROC for 15 state 70 symbol HMM

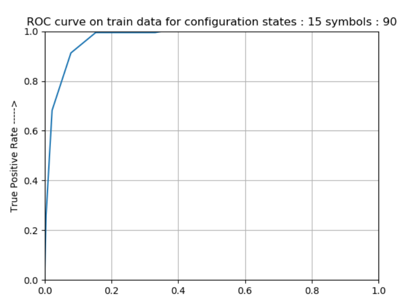

Figure 10: ROC for 15 state 90 symbol HMM

The below plotted graphs represent the ROC curves for the development data. We can observe that the trend of a better ROC curve with increase in number of symbols in a given state still holds. Firstly, we will plot the combined ROC curves for all the symbols which will make it easier for comparision followed by individual ROC curves for each symbol configuration.

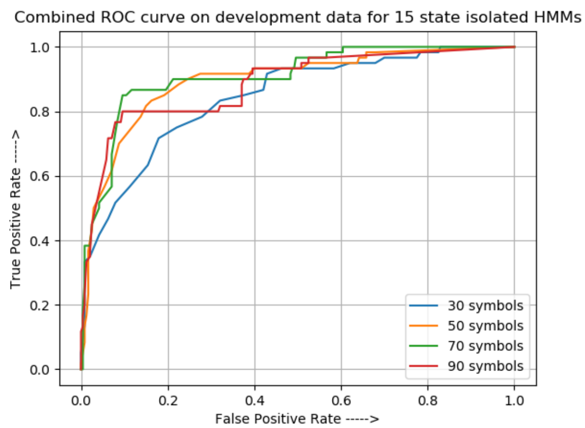

Figure 11: combined ROC for 15 state isolated digit HMM's

The following curves are the individual curves for each symbol configuration.

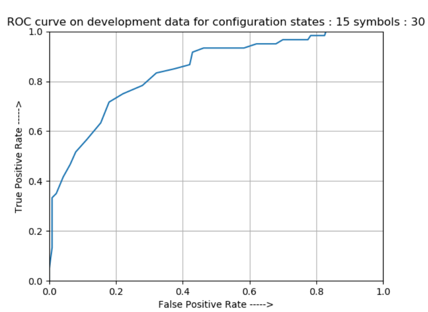

Figure 12: ROC for 15 state 30 symbol HMM

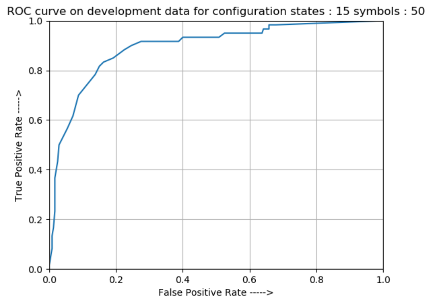

Figure 13: ROC for 15 state 50 symbol HMM

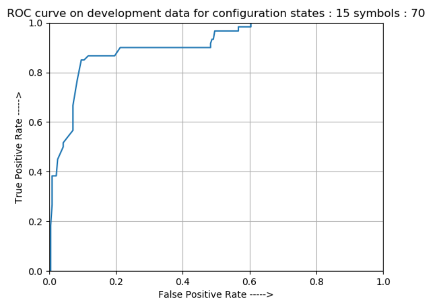

Figure 14: ROC for 15 state 70 symbol HMM

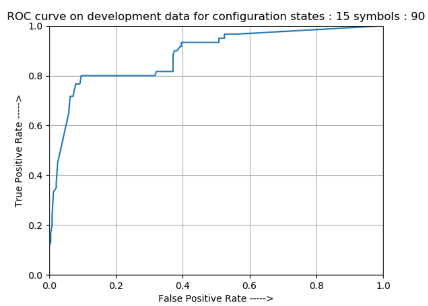

Figure 15: ROC for 15 state 90 symbol HMM

4.1.4 DET curves

Detection Error Tradeoff curves are log scaled curves plotted with false negative ratio on the y axis and false positive ratio on the x axis. Below is the figure representing the DET curves for 15 state isolated digit HMM's on development data.

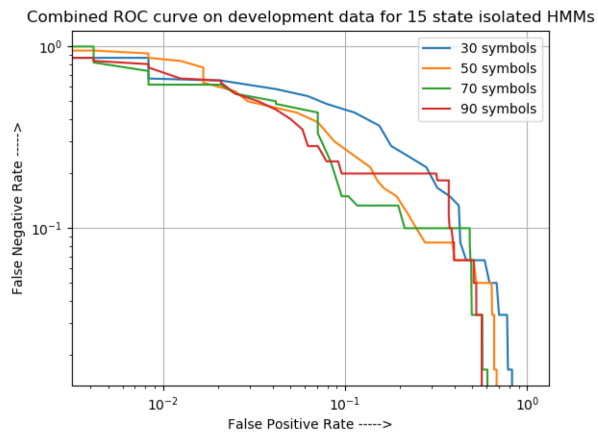

Figure 16: Log scale plotted DET for 15 state HMM's on development data

Below is the figure representing the combined DET curve for 15 state isolated HMM's on training data.

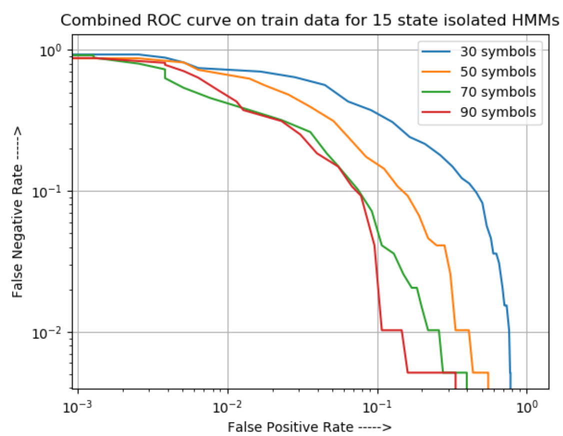

Figure 17: Log scale plotted DET for 15 state HMM's on train data

4.2 Multiple Digit Recognition

4.2.1 Observations on division of utterance

The below observations have been recorded using the division of utterances into equal parts and obtaining the corresponding scores and predicting the maximum scores thereon.

Below are the observation for 15 state 50 symbols HMM's.

Serial Number	Predicted Number
1	o18
2	zz
3	012
4	z8
5	o28

Below are the observation for 15 state 70 symbols HMM's.

Serial Number	Predicted Number
1	o18
2	zz
3	021
4	8o8
5	128

4.2.2 Observations on concatenation of isolated HMM models

The observations below are recorded for concatenation of isolated HMM's through joining the parameters : A,B,π.

Below are the observation for 15 state 50 symbols HMM's.

Serial Number	Best Prediction 1	Best Prediction 2	Best Prediction 3
1	o	1	00
2	z	2	z2
3	o	2	o21
4	z	z8	8
5	2	28	1

Below are the observation for 15 state 70 symbols HMM's.

Serial Number	Best Prediction 1	Best Prediction 2	Best Prediction 3
1	o	1	18
2	z	zz	z8
3	o	1	o21
4	z	z8	8
5	21	28	1

5 References

The link for the code is : Link