MP5: Hidden Markov Model

ECE 417 Fall 2017

Yuchen Liang

Zixu Zhang

November 7, 2017

1 Introduction

This MP performs audiovisual speech recognition on digit 2 and 5 using the Hidden Markov Model (HMM). In the MP, the audio feature is the 24 dimensional MFCC of the utterance of each digit, while the visual feature is a three dimensional vector [w,h1,h2], recording the distance between mouth corners, between upper and lower lips to the center line, respectively. Each digit has ten audio and visual features provided. The audiovisual (AV) feature is a simple concatenation of these two into a 27-dimensional data. The audio features are useful as shown in MP3, while the concatenation of visual features can potentially be useful because it also provides some information for digit classificaton.

The classification uses HMM, a structure that well takes into account the transitions between states over time. From the training set, we use Baum-Welch algorithm to learn the parameters (i.e. transition matrix A and observation Gaussian distribution μ_i and σ_{di}^2) of two HMMs (for each digit), and use the forward-backward procedure to calculate the probability of a test data, and finally compare results from both HMM and choose the more probable one as well as the corresponding digit.

2 Method

2.1 Hidden Markov Model

Hidden Markov Model (HMM) is a statistical Markov model whose system is assumed to be a Markov process with unobserved states S_i. We assume that our model is in Gaussian distribution with diagonal covariance matrix. In this case we can define following parameters for out HMM.

$$\pi_i = \Pr\{q_1 = S - i\}, \quad a_{ij} = \Pr\{q_{t+1} = j | q_t = S_i\}, \quad b_i(\vec{x}_t) = p(\vec{x}_t | q_t = i)$$

$$\mu_{di} = E[x_{dt} | q_t = S_i] \quad \sigma_{di}^2 = E\left[(x_{dt} - \mu_{di})^2 | q_t = S_i\right]$$

2.2 Forward Variable

If we define a forward variable $\alpha_j(t)$ as the probability of first time t observation with the state at time t $q_t = S_j$ given the current model, we will have

$$\alpha_j(t) = \Pr\{O_1 O_2 \cdots O_t, q_t = S_j | \lambda\} \tag{1}$$

In this case, we will have probability of first time t observation given current model as

$$\Pr(O_1 O_2 \cdots O_t|\lambda) = \sum_{j=1}^{N} \alpha_j(t) \tag{2}$$

The forward variable $\alpha_j(t)$ can be solve by induction as we have

$$\alpha_j(1) = \Pr\{O_1, q_1 = S_j|\lambda\} = \Pr\{O_1|q_1 = S_j, \lambda\} \Pr\{q_1 = S_j|\lambda\} = b_{j1}\pi_j \tag{3}$$

$$\alpha_j(t+1) = \sum_{i=1}^{N} (\alpha_i(t) a_{ij}) b_{jt+1} \tag{4}$$

Moreover, in order to avoid the α go below machine ϵ, we will use scaled forward variable $\tilde{\alpha}_i(t)$, as it is defined:

$$\tilde{\alpha}_j(1) = \frac{b_{j1}\pi_j}{\sum_{j=1}^{N} b_{j1}\pi_j}$$

$$\hat{\alpha}_j(t) = \sum_{i=1}^{N} (\tilde{\alpha}_i(t) a_{ij}) b_{jt+1}, \qquad \tilde{\alpha}_j(t) = \frac{\hat{\alpha}_j(t)}{\sum_{i=1}^{N} \hat{\alpha}_i(t)} = \frac{1}{\prod_{s=1}^{t} (\sum_{i=1}^{N} \hat{\alpha}_i(s))} \alpha_j(t)$$

$$\sum_{j=1}^{N} \tilde{\alpha}_j(T) = \frac{1}{\prod_{s=1}^{T} (\sum_{i=1}^{N} \hat{\alpha}_i(s))} \sum_{j=1}^{N} \alpha_j(T) = 1 \Rightarrow \Pr(O_1 \cdots O_T|\lambda) = \prod_{s=1}^{T} (\sum_{i=1}^{N} \hat{\alpha}_i(s))$$

In this case, we can get log likelihood of a test sequence given trained model as

$$\ln \Pr(O_1 \cdots O_T|\lambda) = \sum_{s=1}^{T} \ln(\sum_{i=1}^{N} \hat{\alpha}_i(s))$$

The forward procedure is performanced by function `gmihmm_fwd.m`. First, we assume that the observation probability B has Gaussian distribution, whereby we can get $B = \{b_{jt}\} \in \mathbb{R}^{N \times T}$. We name B matrix as `Pys`. Following equation 2 and 3, We calculated our first iteration from line 28 to 30. Then, we calculated the rest $T - 1$ iterations in a loop from line 32 to 39 by equation 4 and 5.

2.3 Backward Variable

If we define the backward variable $\beta_i(t)$ as the probability of the partial observation seq. after time t, given state S_i at time t.

$$\beta_i(t) = \Pr(O_{t+1} \cdots O_T|q_t = S_i, \lambda)$$

Thus, $\beta_i(t)$ can also be calculated inductively as:

$$\beta_i(T) = 1, \quad \beta_i(t-1) = \sum_{j=1}^{N} a_{ij} b_{jt} \beta_j(t)$$

Since we have scaled our forward parameters, we also introduce scaled backward variable $\tilde{\beta}_i(T)$ as

$$\tilde{\beta}_i(T) = \frac{1}{\sum_{j=1}^{N} \hat{\alpha}_j(T)}, \quad \tilde{\beta}_i(t-1) = \frac{1}{\sum_{j=1}^{N} \hat{\alpha}_j(t-1)} \sum_{j=1}^{N} a_{ij} b_{jt} \tilde{\beta}_j(t)$$

2.4 Forward-Backward Procedure

2.4.1 Calculating the Probability of an Observation Sequence

With both forward and backward variables calculated above, and initial parameters for HMM, we are able to calculate the probability of a given sequence:

$$\Pr(O_1 \cdots O_T | \lambda) = \exp(\sum_{s=1}^{T} \ln(\sum_{i=1}^{N} \hat{\alpha}_i(s)))$$

2.4.2 Training an HMM

Besides, we can also train and re-estimate HMM parameters to get their optimal values:

$$\gamma_t(i) = \Pr\{q_t = S_i | O_1 \cdots O_T\} = \frac{\alpha_i(t)\beta_i(t)}{\sum_{j=1}^{N} \alpha_j(t)\beta_j(t)} = \frac{\tilde{\alpha}_i(t)\tilde{\beta}_i(t)}{\sum_{j=1}^{N} \tilde{\alpha}_j(t)\tilde{\beta}_j(t)}$$

$$\xi_t(i,j) = \Pr\{q_t = S_i, q_{t+1} = S_j | O_1 \cdots O_T\} = \frac{\alpha_i(t) a_{ij} b_{jt+1} \beta_j(t+1)}{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \alpha_i(t) a_{ij} b_{jt+1} \beta_j(t+1)}$$

$$= \frac{\tilde{\alpha}_i(t) a_{ij} b_{jt+1} \tilde{\beta}_j(t+1)}{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \tilde{\alpha}_i(t) a_{ij} b_{jt+1} \tilde{\beta}_j(t+1)}$$

By assuming that we are training this HMM with L different sequence, with the probability we calculated above, we are able to re-estimate the HMM parameters with following steps:

$$a_{ij} = \Pr\{q_{t+1} = j | q_t = i\} \approx \frac{\sum_\ell \sum_t \xi_{t\ell}(i,j)}{\sum_\ell \sum_t \gamma_{t\ell}(i)}$$

$$\vec{\mu}_i = E[\vec{x}_t | q_t = i] \approx \frac{\sum_\ell \sum_t \gamma_{t\ell}(i) \vec{x}_t}{\sum_\ell \sum_t \gamma_{t\ell}(i)}$$

$$\sigma_{di}^2 = E\left[(x_{dt} - \mu_{di})^2 | q_t = i\right] \approx \frac{\sum_\ell \sum_t \gamma_{t\ell}(i)(x_{dt} - \mu_{di})^2}{\sum_\ell \sum_t \gamma_{t\ell}(i)}$$

3 Result

The accuracy table for HMM recognition is listed below.

Table 1: Accuracy of HMM for Audio, Visual and Audio-Visual Recognition

Digit	Audio	Visual	AV
2	100	60	100
5	100	100	100
Ave	100	80	100

4 Discussion

The accuracy of audio and AV recognition is higher than that of purely visual recognition. This is due to the higher dimensionality of audio and AV data. With data of higher dimensions, a larger number of Gaussians are multiplied, and thus the gap produced by the correct and incorrect HMM increases. Hence, the accuracy increases for higher dimension data. Note that curse of dimensionality is not the case here because the number of available states and the volume of training data samples remain small.