Sign Language Recognition

April 2017

0.1 HMM Algorithm

0.1.1 Elements of an HMM

An HMM is characterized by

1. N, the number of states in the model. We denote the individual states as $S = S_1, S_2, ...S_N$ and the state at time t as q_t

2. The state transition probability distribution $A = a_{ij}$, where

$$a_{ij} = P(q_{t+1} = S_j | q_t = S_i) \tag{1}$$

3. The observation symbol probability distribution in state j, B = $\{b_j(O)\}$ where

$$b_j(k) = P(O|q_t = S_i), \quad 1 \leq j \leq N \tag{2}$$

4. The initial distribution $\pi = \pi_i$ where

$$\pi_i = P(q_1 = S_i), \quad 1 \leq i \leq N \tag{3}$$

Complete specification of HMM requires specification of N, specification of observation symbols, and the specification of the three probability measures A, B, and π. For convenience, we use the notation

$$\lambda = (A, B, \pi) \tag{4}$$

to indicate the complete parameter set of the model.

0.1.2 Problems in HMM

1. Given the observation sequence $O = O_1, O_2...., O_T$, and a model $\lambda = (A, B, \pi)$, how do we efficiently compute $P(O|\lambda)$, the probability of the observation sequence, given the model?

2. How do we adjust the model parameters $\lambda = (A, B, \pi)$ to maximize $P(O|\lambda)$

0.1.3 Solutions to the problem

A. Solution to problem 1

The Forward-Backward Procedure:

Consider the forward variable

$$\alpha_t(i) = P(O_1O_2....O_t, q_t = S_i|\lambda) \tag{5}$$

i.e, the probability of the partial observation sequence, $O_1, O_2..., O_t$ and state S_i at time t, given the model λ. We can solve for $\alpha_t(i)$ inductively as, follows:

1. **Initialization:**

$$\alpha_1(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq N. \tag{6}$$

2. **Induction:**

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i)a_{ij}\right] b_j(O_{t+1}), \quad 1 \leq t \leq \textit{T-1}, \quad 1 \leq j \leq N \tag{7}$$

3. **Termination:**

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i) \tag{8}$$

In a similar manner we define the backward variable $\beta_t(i)$ as

$$\beta_t(i) = P(O_{t+1}O_{t+2}....O_T|q_t = S_i, \lambda) \tag{9}$$

i.e., the probability of the partial observation sequence from $t + 1$ to the end, given state S_i at time t and the model λ. Again we can solve for $\beta_t(i)$ inductively, as follows:

1. **Initialization:**

$$\beta_T(i) = 1, \quad 1 \leq i \leq N. \tag{10}$$

2. **Induction:**

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), \quad t = T-1, T-2, ..., 1 \quad 1 \leq i \leq N \tag{11}$$

B. Solution to problem 2

For the continuous observations we use a Gaussian pdf to insure that the parameters of the pdf can be re-estimated in a consistent way.

$$b_j(O) = G(O, \mu_j, U_j) \quad 1 \leq j \leq N \tag{12}$$

We define $\gamma_t(i)$ as

$$\gamma_t(i) = P(q_t = S_i | O, \lambda) \tag{13}$$

i.e., the probability of being in state S_i at time t, given the observation sequence O, and the model λ. Equation (13) can be expressed in terms of forward-backward variables, i.e.,

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{P(O|\lambda)} = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{N} \alpha_t(i)\beta_t(i)} \tag{14}$$

We also define $\xi_t(i,j)$ as

$$\xi_t(i,j) = P(q_t = S_i, q_{t+1} = S_j | O, \lambda) \tag{15}$$

i.e., the probability of being in state S_i at time t, and state S_j at time $t+1$, given the model and the observation sequence.

From the definition of forward and backward variables, we can write $\xi_t(i,j)$ in the form

$$\xi_t(i,j) = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{P(O|\lambda)} = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)} \tag{16}$$

We can relate $\gamma_t(i)$ to $\xi_t(i,j)$ by summing over j, giving

$$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i,j) \tag{17}$$

$$\sum_{t=1}^{T-1} \gamma_t(i) = \text{ expected number of transitions from } S_i \tag{18a}$$

$$\sum_{t=1}^{T-1} \xi_t(i,j) = \text{expected number of transitions from } S_i \text{ to } S_j. \tag{18b}$$

A set of reasonable re-estimation formulas for π, A and B are

$$\bar{\pi}_i = \text{expected frequency (number of times) in state } S_i \text{ at time } (t=1) = \gamma_1(i) \tag{19a}$$

$$\bar{a}_{ij} = \frac{\text{expected number of transitions from state } S_i \text{ to state} S_j}{\text{expected number of transitions from state } S_i} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \tag{19b}$$

$$\bar{\mu_j} = \frac{\sum_{t=1}^{T} \gamma_t(j).O_t}{\sum_{t=1}^{T} \gamma_t(j)} \tag{19c}$$

$$\bar{U_j} = \frac{\sum_{t=1}^{T} \gamma_t(j).(O_t - \mu j)(O_t - \mu_j)'}{\sum_{t=1}^{T} \gamma_t(j)} \tag{19d}$$

0.1.4 Initialization of the model

1. π is initialized uniformly for all states
2. A is initialized using Bakis model.
3. Co-variance matrix corresponding to Gaussian pdf of each state is initialized with random diagonal matrices.
4. The observation sequence is divided uniformly among the states. Mean corresponding to Gaussian pdf of each state is initialized with mean of the observation sequence corresponding to the state.

0.1.5 Training

Create 95 HMM's, one for each word.

1. Initialize $\pi, A, \mu, .$
2. Calculate $\alpha, \beta, \gamma, \xi$
3. Update the model $\lambda = (\pi, A, \mu,)$ using the update rules (19)
   Continue the above steps till it converges.

0.1.6 Prediction

Calculate the likelihood of each observed sequence with 95 models and predict the one with the highest probability.

0.1.7 Problems faced

1. The value of probability$b_j(O)$ is very low and it is creating floating point overflows.

0.1.8 References

1. L. R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. pages 267–296, 1990
2. http://people.csail.mit.edu/yingyin/resources/doc/projects/6867term-project.pdf