Digital Speech Processing, Midterm

Dec. 14. 2011, 10:20-12:10

- OPEN Printed Course PowerPoint, Course References, Personal Notes
- You have to use **CHINESE** sentences to answer all of the problems
- Total point: 115

1. (10) In order to recognize L isolated words $w_1, w_2, \ldots, w_L$ each with an HMM respectively, $\lambda_1, \lambda_2, \ldots, \lambda_L$, it is well known that one can use either the forward algorithm (the left) or the Viterbi algorithm (the right),

   $$\arg\max_k P(\bar{O}|\lambda_k) \sim \arg\max_k P(\bar{q}^*, \bar{O}|\lambda_k)$$

   Explain why and discuss the difference between them.

2. (10) In training HMM models for isolated word recognition, do you think the more number of iterations you perform, the higher recognition accuracy you'll get? Note that it is guaranteed that the likelihood function will be increased in each iteration. Explain your answer.

3. (30) Alice, Bob and Cindy live in city A, B, and C respectively. Alice and Bob are both interested in only 3 activities: walking in the park, shopping, and cleaning her/his apartment. Their choices are influenced by the weather on a given day. Cindy has no definite information about the weather in city A or city B, but she believes that the weather both operate as discrete Markov chains. Cindy assumes that the weather is either "Rainy" or "Sunny", but she cannot observe them directly, that is, they are hidden. Cindy can see that Alice and Bob post their daily activities on blogs, those activities are the observations. The entire systems are then two HMMs. Cindy set the following model setting:

```
states = 'Rainy', 'Sunny'.        observations = 'walk', 'shop', 'clean'
start_probability = ( P(Rainy), P(Sunny) )
transition_probability = {        P(Rainy| Rainy), P(Sunny| Rainy),
                                  P(Rainy| Sunny), P(Sunny| Sunny) }
observation_probability = {       P(walk| Rainy), P(walk| Sunny), P(shop| Rainy),
                                  P(shop| Sunny), P(clean| Rainy), P(clean| Sunny)}
```

Then Cindy uses the following training algorithm to estimate the model parameters:

```
// Baum-Welch iterative training
Read in the observations (daily activities on Alice's/Bob's blog)
for iter =1 to iteration_num do
    Clean all accumulators
    for sample = 1 to num_of_samples do
        T ← length of the sample
```

for t = 1 to T **do**

calculate α_t(Rainy) and α_t(Sunny)

calculate β_t(Rainy) and β_t(Sunny)

end for

calculate $\gamma_t(i)$, $\varepsilon_t(i,j)$ iteratively where i, j = Rainy or Sunny

accumulate

$$\gamma_1(i), \sum_{t=1}^{T} \gamma_t(i), \sum_{t=1}^{T-1} \gamma_t(i), \sum_{o_t=\text{walk}} \gamma_t(i), \sum_{o_t=\text{shop}} \gamma_t(i), \sum_{o_t=\text{clean}} \gamma_t(i), \sum_{t=1}^{T-1} \varepsilon_t(i,j)$$

end for

update (A, B, π)

end for

Write out the new model

Please answer the following questions:

(a) (5) Besides observations (daily activities on Alice's/Bob's blog), what should also be read in for the algorithm to execute?

(b) (5) What is $\sum_{o_t=\text{walk}} \gamma_t(i)$? i = Rainy or Sunny

Use the observation (shop, walk, clean, walk, walk, clean) to explain.

Finally Cindy gets the two models.

Alice:

```
start_probability = ( 0.6, 0.4 )
transition_probability = {
    0.5, 0.5
    0.5, 0.5 }
observation_probability = {
    0.2 0.5
    0.4 0.4
    0.4 0.1 }
```

Bob:

```
start_probability = ( 0.4, 0.6 )
transition_probability = {
    0.5, 0.5
    0.5, 0.5 }
observation_probability = {
    0.1 0.5
    0.4 0.3
    0.5 0.2 }
```

(c) (20) Cindy collects blog articles of Alice and Bob, but two collections have their authors missing. Please use the above models and Viterbi algorithm to classify:

(shop, walk) (clean, clean)

4. (10) Explain the principles and procedures of estimating the probabilities for unseen events in **Katz smoothing**.

5. (20)The following are the procedures of MFCC (without derivatives) extraction
   (1) Pre-emphasis
   (2) Windowing
   (3) Discrete Fourier Transform
   (4) Mel filter-bank processing
   (5) Logarithmic operation
   (6) Inverse discrete Fourier transform(IDFT)

   Briefly answer the following question
   (10) (a)Why do we use a **window** to extract MFCC parameters ?
   (10) (b)Why **pre-emphasis** is performed?

6. (10)
   (a) (5) Given a discrete-valued random variable X with probability distribution

$$\{p_i = \mathrm{Prob}(X = x_i), i = 1,2,3,\ldots,M\}, \sum_{i=1}^{M} p_i = 1$$

   Explain the meaning of $H(X) = -\sum_{i=1}^{M} p_i[\log(p_i)]$.
   (b) (5) Explain why and how H(x) above can be used as the criterion to split a node into two in developing a decision tree.

7. (10) What is **LBG** algorithm and why is it better than **K-means** algorithm?

8. (15) Choose **ONE** of the problems to answer.

   8-1. HW2-1

   (a) (5) In homework 2-1, we build and train digit models, "sp model" and "sil model". What does 'sp' and 'sil' stand for seperately? How can they be used in digit recognition?
   (b) (5) What the following means?

   MU +2 {er.state[2-9].mix}

   If I add it into HHEd, will the accuracy increase? Why?
   (c) (5) Write down two methods (except (b)) in HW2-1 which can increase the accuracy of recognition and explain the reasons.

   8-2. HW2-2

   (a) (5) What are the voiced/unvoiced speech signals and their time domain waveform characteristics?
   (b) (5) What are the fricative consonants and their frequency characteristics compared with voiced signals?
   (c) (5) What are the plosive consonants (or stop consonants)? Describe the way of the plosive consonants are produced and the resultant characteristics in signals.