Digital Speech Processing, Midterm

May. 16, 2012, 19:00-21:00

- OPEN Course PowerPoint, Course Reference, Personal Notes
- You have to use CHINESE sentences to answer all of the problems
- Total point: 120

1.

Alice and Bob are both interested in only four activities: playing baseball, going to the movies, watching television and studying. Their choices are influenced by the temperature of the cities they live, city A and city B, on a given day. One of their friends, Candy, has no definite information about the temperature in the cities, but she believes that they both operate as discrete Markov chains. Candy assumes that the weather conditions can be modeled as either "hot", "warm", or "cold", but she cannot observe them directly, that is, they are hidden because Candy lives in neither city A nor city B. Candy can see that on the blogs of Alice and Bob, they post their daily activities, which are the observations she can get. The entire systems are like two hidden Markov models (HMMs). Candy set the following model setting:

```
states = ' cold ', ' warm ', ' hot '
observations = 'baseball', 'movies', 'TV', 'study'
start_probability = ( P(cold), P(warm),P(hot) )
transition_probability = {
    P(cold|cold)   P(warm|cold)  P(hot|cold)
    P(cold|warm)  P(warm|warm)P(hot|warm)
    P(cold|hot)    P(warm|hot)   P(hot|hot)    }
observation_probability = {
    P(baseball|cold)    P(basball|warm)    P(baseball|hot)
    P(movies|cold)      P(movies|warm)     P(movies|hot)
    P(TV|cold)          P(TV|warm)         P(TV|hot)
    P(study|cold)       P(study|warm)      P(study|hot)
        }
```

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 $\alpha_t(\text{Rainy})$ and $\alpha_t(\text{Sunny})$

calculate $\beta_t(\text{Rainy})$ and $\beta_t(\text{Sunny})$

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

(10) (a)What should be read in for the algorithm to execute? (Hint: Three answers, which are used in homework 1)

(20) (b)Please use the following two models and **Viterbi algorithm** to classify(predict the author of) this collection: (movies, TV, study)

Alice

start_probability = (0.6, 0.2,0.2)

transition_probability = {

0.5 0.5 0.0

0.3 0.4 0.3

0.0 0.5 0.5}

observation_probability = {

bb 0.0 0.1 0.2

movie 0.0 0.2 0.4

TV 0.2 0.2 0.2

study 0.8 0.5 0.2 }

Bob

start_probability = (0.2,0.2,0.6)

transition_probability = {

0.5 0.5 0.0

0.3 0.4 0.3

0.0 0.5 0.5}

observation_probability = {

bbb 0.0 0.1 0.2

movie 0.2 0.3 0.4

TV 0.6 0.5 0.4

study 0.2 0.1 0.0 }

2. 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?

3. (15) Please briefly describe **LBG algorithm** and **K-means algorithm** respectively. Which one of the above two algorithms usually performs better in performance? Explain your answer.

4. Assume we have 3 different document sets (C1、C2、C3), and each document set has only 4 words (A、B、C、D). The count information of each document set is in Table

1. Please use Table 1 to answer the following questions.

Table 1

Count	A	B	C	D
C1	250	250	300	200
C2	200	200	200	400
C3	100	400	100	400

(10) (a) Compute the **perplexity** of the above 3 document sets respectively.
(log2=0.3010, log3=0.4771, log7=0.8451)

(10) (b) According to the perplexity in (a), what can we learn from these 3 document sets?

(10) (c) Now we have another document K. Use **Cross-Entropy** to determine to Which document set K belongs.

Table 2

Count	A	B	C	D
K	30	30	20	20

5. (10) Explain how the **tree lexicon** can be used in the search algorithm for large vocabulary continuous speech recognition and how it is helpful.

6. (15) Choose ONE of the problems to answer.

12-1. 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 added 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.

12-2. 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 pronunciation methods of plosive consonants and the resultant characteristics in signals.