Introduction to Digital Speech Processing, Midterm

Nov. 19, 2016, 15:30-17:30

- OPEN Lecture Power Point (Printed Version) and Personal Notes
- You have to use CHINESE sentences to answer all the questions, but you can use English terminologies
- Total points: 135

---

1. (25) 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'
   Then Candy uses the following training algorithm to estimate the model parameters:

// Baum-Welch iterative training
Read in the observations (daily activities on Alices/Bobs blog)
for iter =1 to iteration_num do
  Clean all accumulators
  for sample = 1 to num_of_samples do
    T $\leftarrow$ length of the sample
    for t = 1 to T do
      calculate α_t (cold), α_t (warm) and α_t (hot)
      calculate β_t (cold), β_t (warm) and β_t (hot)
    end for
    calculate $\gamma_t(i)$, $\epsilon_t(i,j)$ iteratively where i, j = cold or warm or hot
    accumulate

$$\gamma_1(i), \sum_{t=1}^{T}\gamma_t(i), \sum_{t=1}^{T-1}\gamma_t(i), \sum_{o_t=baseball}\gamma_t(i), \sum_{o_t=movies}\gamma_t(i), \sum_{o_t=TV}\gamma_t(i), \sum_{o_t=study}\gamma_t(i), \sum_{t=1}^{T-1}\epsilon_t(i,j)$$

  end for
  update (A, B, π)
end for
Write out the new model

(a) (10) What should be read in for the algorithm to execute? (At least three answers, which were used in homework 1)

(b) (5) There is one more error in the pseudo code above, please point out.

(c) (10) Please use the following two models and **Viterbi algorithm** to classify (predict the author of) this observation sequence: (movies, TV, study)

Alice	Bob
start_probability = { 0.2, 0.3, 0.5}	start_probability = { 0.1, 0.4, 0.5}
transition_probability =	transition_probability =
{ 0.6, 0.4, 0.0	{ 0.3, 0.5 ,0.2
0.3, 0.5, 0.2	0.7, 0.3 ,0.0
0.1, 0.7, 0.2 }	0.0, 0.5 ,0.5 }
observation_probability =	observation_probability =
{ 0.0, 0.1, 0.3	{ 0.0, 0.1 ,0.2
0.5, 0.3, 0.4	0.1, 0.3 ,0.7
0.5, 0.5, 0.1	0.6, 0.6 ,0.1
0.0, 0.1, 0.2 }	0.3, 0.0 ,0.0 }

2. (10) Write down the procedures for LBG algorithm and discuss why and how it is better than the K-means algorithm.

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

4. (15) What is the perplexity of a language source? What is the perplexity of a language model with respect to a test corpus? How are they related to a "virtual vocabulary"?

5. (15) Please answer the following questions.

   (a) (5) Explain what a triphone is.

   (b) (10) Explain why and how the unseen triphones can be trained using decision trees.

6. (10) Explain how the MAP principle can be used to find a word sequence $\bar{W} = w_1 w_2 \cdots w_n$ given an observation sequence $\bar{O} = o_1 o_2 \cdots o_T$, how the hidden Markov model and language models can be used, and which the likelihood function and the prior probability are?

7. (5) Explain why and how **beam search** is useful in large vocabulary continuous speech recognition.

8. (10) What is the problem of coarticulation and context dependency considered in acoustic modeling?

9. (15) Answer the following questions about MFCC (without derivatives) extraction

   (a) (10) Why do we use a window to extract MFCC parameters ?

   (b) (5) After you obtain 13 MFCC parameters for 12 MFCC parameters plus a short-time energy, explain how to obtain the other 26 parameters and what they are.

10. (10) Explain how the tones in Mandarin (tones 1, 2, 3, 4) are reflected in speech signals.

11. (10) Explain what the Dynamic Time Warping (DTW) problem is and why and how it can be considered as the problem of finding a path on a two-dimensional plane.