

Digital Speech Processing, Midterm

May. 16, 2006, 10:10-12:10

- OPEN EVERYTHING
 - 除專有名詞可用英文以外，所有文字說明一律以中文為限，未用中文者不計分
 - Total points: 120, Time allocation: 1 point/min
-

1. (20) Given a HMM $\lambda = (A, B, \pi)$, an observation sequence $\bar{O} = o_1 o_2 \dots o_t \dots o_T$ and a state sequence $\bar{q} = q_1 q_2 \dots q_t \dots q_T$, define

$$\alpha_t(i) = \text{Prob}[o_1 o_2 \dots o_t, q_t = i | \lambda]$$
$$\beta_t(i) = \text{Prob}[o_{t+1} o_{t+2} \dots o_T | q_t = i, \lambda]$$

(a) (5) Show that $\text{Prob}(\bar{O} | \lambda) = \sum_{i=1}^N [\alpha_t(i) \beta_t(i)]$

(b) (5) Show that $\text{Prob}(q_t = i | \bar{O}, \lambda) = \frac{\alpha_t(i) \beta_t(i)}{\sum_{i=1}^N \alpha_t(i) \beta_t(i)}$

- (c) (10) Formulate and describe the procedures for Viterbi algorithm to find the best state sequence $\bar{q}^* = q_1^* q_2^* \dots q_t^* \dots q_T^*$.

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

$$\{p_i = \text{Prob}(X = x_i), i = 1, 2, 3, \dots, M\}, \quad \sum_{i=1}^M p_i = 1$$

Explain the meaning of $H(X) = - \sum_{i=1}^M p_i [\log(p_i)]$.

3. (10) What is coarticulation and context dependency for acoustic modeling? Why tri-phone models are difficult to train?
4. (10) For Chinese language models, the N-grams can be trained based on either characters or words. Discuss the considerations of choosing them.
5. (10) Explain the basic principles of back-off and interpolation to be used for language model smoothing.

6. (10) In feature extraction for speech recognition, after you obtain 12 MFCC parameters plus a short-time energy (a total of 13 parameters), explain how to obtain the other 26 parameters and what they are.
7. (10) Explain why the use of a window with finite length, $w(n), n = 0, 1, 2, \dots, L - 1$, is necessary for feature extraction in speech recognition.
8. (10) What do we mean by spoken document understanding and organization?
9. (30) Write down anything you learned about the following subjects that were NOT mentioned in the class. Don't write anything mentioned in the class.
 - (a) (15) classification and regression tree (CART)
 - (b) (15) search problem/algorithms for large vocabulary continuous speech recognition