Digital Speech Processing, Midterm

May. 15, 2007, 9:10-11:10

- OPEN EVERYTHING
- 除專有名詞可用英文以外，所有文字説明一律以中文爲限，未用中文者不計分
- Total points: 170

---

1. (10) Describe what you know about the basic elements, operations and relevant research issues of conversational interfaces or spoken dialogue systems.

2. (10) Assume $\bar{\mathbf{X}} = (x_1, x_2)^t$ is a two-dimensional random vector with bi-variate Gaussian distribution, a mean vector $\bar{\mu} = (\mu_1, \mu_2)^t$ and a co-variance matrix Σ. x_1, x_2 are two random variables and "t" means transpose. Discuss how the distribution of $\bar{\mathbf{X}}$ depends on $\bar{\mu}$ and Σ.

3. (25) Given a HMM $\lambda = (A, B, \pi)$ with N states, an observation sequence $\bar{O} = o_1 o_2 \ldots o_t \ldots o_T$ and a state sequence $\bar{q} = q_1 q_2 \ldots q_t \ldots q_T$, define

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

   (a) (5) What is $\sum_{i=1}^{N} \alpha_t(i)\beta_t(i)$? Show your results.

   (b) (5) What is $\frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^{N} \alpha_t(j)\beta_t(j)}$? Show your results.

   (c) (5) What is $\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)$? Show your results.

   (d) (10) Formulate and describe the Viterbi algorithm to find the best state sequence $\bar{q}^* = q_1^* q_2^* \cdots q_t^* \cdots q_T^*$ giving the highest probability $\text{Prob}[\bar{O}, \bar{q}^*|\lambda]$. Explain how it works and why backtracking is necessary.

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

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

6. (10) In acoustic modeling the concept of "senones" is very useful. Explain what is a "senone" and how it can be used.

7. (10) Explain the basic principles in selecting the voice units for a language for hidden Markov modeling.

8. (10) Explain what the class-based language model is and why it is useful?

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

10. (10) Explain why the use of a window with finite length, $w(n), n = 0, 1, 2, \ldots, L-1$, is necessary for feature extraction in speech recognition.

11. (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.

12. (10) In large vocabulary continuous speech recognition, explain:

    (a) (5) What the "language model weight" is.

    (b) (5) Why the language model has the function as the penalty of inserting extra words.

13. (20) What is the maximum a posteriori (MAP) principle? How can it be used to integrate acoustic modeling and language modeling for large vocabulary speech recognition? Why and how this can be solved by a Viterbi algorithm over a series of lexicon trees?

14. (15) Under what kind of condition a heuristic search is admissible? Show or explain why?