ENGINEERING TRIPOS PART IIB

Tuesday 23 April 2013 9.30 to 11

Module 4F11

SPEECH AND LANGUAGE PROCESSING

*Answer not more than **three** questions.*

All questions carry the same number of marks.

*The **approximate** percentage of marks allocated to each part of a question is indicated in the right margin.*

There are no attachments.

STATIONERY REQUIREMENTS	SPECIAL REQUIREMENTS
Single-sided script paper	Engineering Data Book
	CUED approved calculator allowed

You may not start to read the questions printed on the subsequent pages of this question paper until instructed that you may do so by the Invigilator

pcw04

1 (a) What are the basic modelling assumptions in using hidden Markov models (HMMs) to recognise speech? [10%]

(b) Mel frequency cepstral coefficients (MFCCs) are often used as a front-end parameterisation in HMM-based speech recognition systems.

(i) What are the desirable properties of MFCCs? [10%]

(ii) Why is the HMM observation vector often extended to include first and second differential coefficients? [10%]

(c) HMMs are to be trained using Baum-Welch re-estimation for an isolated word recognition task. Each word is modelled by a single HMM with Gaussian state output distributions. A particular N-state HMM, with parameter set λ, is to be trained on an observation sequence $\mathbf{O} = \{\mathbf{o}_1, \mathbf{o}_2 \ldots \mathbf{o}_T\}$. The forward probability is defined as

$$\alpha_j(t) = p(\mathbf{o}_1, \mathbf{o}_2, \ldots, \mathbf{o}_t, x(t) = j \,|\, \lambda)$$

where $x(t)$ denotes the state occupied at time t.

(i) Define a corresponding backwards probability, $\beta_j(t)$, and show how it can be computed recursively. [20%]

(ii) How can $\alpha_j(t)$ and $\beta_j(t)$ be combined to find the posterior probability of state occupation, $L_j(t)$? [10%]

(iii) Using $\alpha_j(t)$, $\beta_j(t)$, and $L_j(t)$, show how a maximum likelihood estimate for the model transition probabilities can be found and suggest how the transition probabilities should be initialised. [20%]

(iv) Explain why computational underflow is a problem in computing the forward and backward probabilities, and describe **two** approaches for overcoming this problem. [20%]

2 Triphones are often used as the acoustic modelling units in large vocabulary speech recognition systems.

(a) Explain what is meant by triphone units and why they are appropriate models. What is the difference between word-internal and cross-word triphones? [15%]

(b) Describe methods of estimating triphone models including state-level parameter tying that use

(i) Phonetic decision trees [30%]

(ii) Bottom-up agglomerative clustering [15%]

In each case, describe the basic operation of the technique and list the advantages and disadvantages for estimating triphone models.

(c) A large vocabulary speech recognition system using a bigram language model is to use triphones as acoustic units.

(i) Describe the impact of using both word-internal and cross-word triphones on the model-level network used in speech recognition decoding. [20%]

(ii) Briefly describe **two** approaches to designing a decoder for a speech recognition system that uses cross-word triphones. [20%]

3 (a) Give **two** reasons why machine translation is difficult and briefly discuss them. [10%]

(b) A pair of sentences $e_1^I = e_1 \dots e_I$ and $f_1^J = f_1 \dots f_J$ are known to be translations. Their word-to-word alignment is described by the alignment process $a_1^J = a_1 \dots a_J$. By making simplifying conditional independence assumptions, describe the translation probability distribution $P(f_1^J, a_1^J, J|e_0^I)$ in terms of its three component distributions: the sentence length distribution, the word translation distribution, and the word alignment distribution. [20%]

(c) Give the formulae of the word alignment distribution under Model 1, Model 2, and the HMM alignment model. [20%]

(d) Derive expressions for the efficient calculation of $P(f_1^J, a_1^J|e_1^I)$ and $P(a_j = i|e_1^I, f_1^J)$ under Model 1. [20%]

(e) Describe an iterative estimation procedure for the word-to-word translation distribution $P_T(f|e)$. [20%]

(f) Models are often trained from parallel text using a 'flat start' procedure in which models are estimated in the following sequence: Model 1, Model 2, HMM alignment model. Provide a justification for this practice by discussing the differences between the models. [10%]

4 (a) What is a weighted finite state acceptor? Explain the differences between a weighted finite state acceptor and a weighted finite state transducer. [20%]

(b) An automatic speech recognition (ASR) system is to be implemented using weighted finite state automata. Describe the role of each of the following components of the ASR system and explain how each can be implemented with weighted finite state automata.

(i) A pronunciation lexicon. [15%]

(ii) A transducer that maps monophone sequences to triphone sequences. [15%]

(c) A bigram language model with back-off and discounting has the following form:

$$\hat{P}(w_j|w_i) = \begin{cases} d(f(w_i,w_j))\frac{f(w_i,w_j)}{f(w_i)} & f(w_i,w_j) > C \\ \alpha(w_i)\hat{P}(w_j) & \text{otherwise} \end{cases} \quad (1)$$

(i) Name the quantities $d(\cdot)$, $\alpha(\cdot)$, and C and describe their role. [20%]

(ii) Give an algorithm to construct a weighted finite state acceptor for an exact implementation of the bigram language model of Equation (1). [30%]

END OF PAPER