ENGINEERING TRIPOS PART IIB

Friday 23 April 2010 2.30 to 4.00

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
Single-sided script paper

SPECIAL REQUIREMENTS
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

wjb02

1 This question concerns the use of linear predictive coding (LPC) in modelling speech. The order of the LPC model is p. The n^{th} sample of the speech signal is s_n.

(a) Write an expression for the prediction error e_n in terms of the speech signal s_n and the linear prediction coefficients $a_1, \ldots, a_p$. [15%]

(b) Show how the transfer function $A(z)$ of the prediction error filter relates the speech spectrum $S(z)$ to the transform of the prediction error $E(z)$. [15%]

(c) Derive the *normal equations*

$$\sum_n s_n s_{n-j} = \sum_{k=1}^{p} a_k \sum_n s_{n-k} s_{n-j} \quad j = 1, \ldots, p$$

which are solved to obtain the optimum filter coefficients under the mean square error criterion. [20%]

(d) Explain the assumptions underlying the *autocorrelation method* and give the simplified form of the normal equations which result. [20%]

(e) Give Durbin's Algorithm for finding the optimum filter coefficients and explain how they guarantee that increasing the order of the LPC predictor reduces the mean square error. [30%]

2 (a) Draw a block diagram of the generic speech recognition architecture and give a brief description of each component. [20%]

(b) A speech recognition system is to be constructed using Hidden Markov Models (HMMs). The HMMs will have Gaussian observation distributions and will be trained as whole-word models so that each word has its own HMM. A training set of R utterances for each word in the recognition vocabulary is available, and the Baum Welch algorithm is to be used to estimate the parameters of the HMM observation distributions.

(i) Suggest an initialisation procedure for the parameters of the Gaussian observation distributions. [10%]

(ii) Give the equations defining the forward and backward probabilities used in the Baum Welch algorithm. Derive the recursions for one of the probabilities. [20%]

(iii) Give a relationship for the probability of being in state j at time t, $L_j(t)$, in terms of the forward and backward probabilities. [10%]

(iv) Give the Baum Welch reestimation formulae for the mean and variance parameters of the Gaussian observation distribution associated with each state. [20%]

(c) A large-vocabulary isolated-word system is to be trained for use with the Google Maps application. It is decided to use HMMs based on word-internal triphones.

(i) Discuss why word-internal triphones might be a better choice for this application than whole-word HMMs. [10%]

(ii) Contrast the HMM training procedure for word-internal triphone models to the training procedure for whole-word acoustic models. [10%]

3 (a) What is a weighted finite state acceptor? What is the main difference between a weighted finite state acceptor (WFSA) and a weighted finite state transducer (WFST)? [20%]

(b) Weighted finite state acceptors assign a weight to a particular string by summation ($\oplus$ operation) over the weights of all paths that generate the string, where the weight of each path is obtained as the product ($\otimes$ operation) of the path arc weights and the path initial and final weights.

(i) Define a semiring in the context of WFSAs. [10%]

(ii) Complete Table 1 to describe the attributes of three commonly used semirings. [20%]

Semiring	$\mathbb{K}$	$\oplus$	$\otimes$	$\bar{0}$	$\bar{1}$
Probability					
Log					
Tropical					

Table 1

(iii) A WFSA A can generate the sequence 'a b' via two alternative paths. What does the weight, $[\![A]\!]$('a b'), assigned to the sequence 'a b' by A represent in each of the three semirings of Table 1? [30%]

(c) Two WFSTs A and B are shown in Fig. 1, where A maps x to y and B maps y to z. Draw the transducer $A \circ B$, which maps x to z, that results from the composition of A with B in the tropical semiring. [20%]

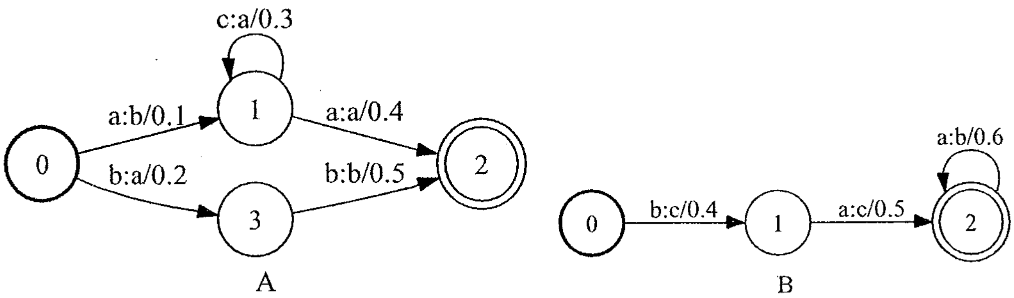

Fig. 1

4 This question concerns the use of alignment models in statistical machine translation. An English sentence of I words is denoted by the sequence $e_0^I = e_0, e_1, \ldots, e_I$, where e_0 is the additional NULL word added to the start of the sentence. A foreign sentence of J words is denoted $f_1^J = f_1, \ldots, f_J$. Alignment between the two sentences is specified by the sequence $a_1^J = a_1, \ldots, a_J$.

(a) 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%]

(b) Give the formulae of the alignment distribution under IBM models 1 and 2 and the HMM alignment model. Explain their differences. [20%]

(c) Alignment link sets B and B' are extracted from two different alignments. How is the alignment error between B and B' computed? Why is the alignment error useful for developing statistical machine translation systems? [20%]

(d) How are 'phrases' defined in the context of phrase-based statistical machine translation models? Describe the advantages of using phrases rather than words in a statistical machine translation system. [20%]

(e) Briefly describe how a word-based alignment model can be used to extract phrase pairs from a parallel corpus. [20%]

END OF PAPER