ENGINEERING TRIPOS PART IIB

Friday 29 April 2011 2.30 to 4

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

pcw03

1 (a) What are the basic modelling assumptions in using hidden Markov models (HMMs) to recognise speech? [15%]

(b) HMMs are to be trained using Baum-Welch re-estimation for an isolated word recognition task with each word modelled by a single HMM with a Gaussian state output distribution. 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) Hence write down maximum-likelihood re-estimation formulae for the HMM mean parameters. How are the formulae extended if multiple training sequences are used? [20%]

(iv) The HMM state output distributions are now modified to use a mixture of Gaussians. By viewing a mixture distribution as a set of parallel states, describe how the posterior probability of mixture component occupation can be computed and hence write down the re-estimation formulae for the mean vectors of the Gaussian components. [20%]

(c) If the task is altered to recognise continuous speech with whole-word HMMs, describe how Baum-Welch training can be used with a database of continuous speech and how the HMM parameters should be initialised. [15%]

2 A large vocabulary continuous-speech recognition system based on hidden Markov models (HMMs) is being designed. Initially the system is to use monophone HMMs with an M-component mixture of diagonal covariance Gaussians as the state output distribution. A d-dimensional observation vector is used based on filter-bank log energy values. There are 45 phones to be modelled by the system. A unigram language model is used.

The following changes are being proposed (in order) for the system. For each proposed change, briefly describe the proposed change and the impact on the number of parameters to be estimated, computational cost and expected recognition performance.

(a) Mel frequency cepstral coefficients (MFCCs) are used in place of log filter-bank energies as the observation vector. [20%]

(b) The observation vector is extended to include first and second differential coefficients. [20%]

(c) Each HMM diagonal covariance matrix is replaced by a full covariance matrix. [15%]

(d) Cross-word triphones are used with decision-tree state tying as the HMM modelling units. [25%]

(e) A bigram language model replaces the unigram language model. [20%]

3 (a) Explain why morphologically rich languages and the differences in word order between languages make machine translation difficult. [10%]

(b) The quality of two automatic translations *mt1* and *mt2* are to be evaluated against a reference translation *ref*. These are shown in the following table:

mt1	he got home much later than he had planned
mt2	he came home at a later time than expected
ref	he came home later than he expected

(i) Describe the BLEU score used to measure translation quality. [20%]

(ii) Considering a maximum order of N=3, which of the two translation hypotheses has a higher BLEU score? Justify your answer based on N-gram precisions. [10%]

(c) A pair of sentences $e_1^I = e_1 \ldots e_I$ and $f_1^J = f_1 \ldots f_J$ are known to be translations of each other. Their word-to-word alignment is described by the alignment process $a_1^J = a_1 \ldots a_J$. Derive the alignment likelihood $P(f_1^J, a_1^J, J|e_1^I)$ under Model 1 and Model 2, and explain their differences. [25%]

(d) Consider an aligned parallel corpus of two sentence pairs of lengths $I^{(1)} = I^{(2)} = 5$, $J^{(1)} = J^{(2)} = 4$, and with the following estimated alignments:

$$a^{(1)} : a_1^{(1)} = 2\,, a_2^{(1)} = 1\,, a_3^{(1)} = 3\,, a_4^{(1)} = 5$$
$$a^{(2)} : a_1^{(2)} = 3\,, a_2^{(2)} = 0\,, a_3^{(2)} = 2\,, a_4^{(2)} = 4$$

Describe how Model 2 alignment probabilities can be estimated from counts. Find two example Model 2 alignment probabilities from this corpus. [20%]

(e) Contrast the estimation procedure in part (d) with the use of the EM algorithm, discussing any possible advantages. [15%]

4 (a) Give the general formula which describes the statistical machine translation (SMT) process using the source-channel formulation. Describe the two models involved and the kind of data used to estimate them. [20%]

(b) In phrase-based SMT the translation and alignment model operates at the phrase level, so that:

$$P(f_1^J, v_1^K, a_1^K, u_1^K | e_1^I) = P(f_1^J | v_{a_1} \dots v_{a_K})\, P(a_1^K | v_1^K, u_1^K)\, P(v_1^K | u_1^K)\, P(u_1^K | e_1^I)$$

(i) What is the meaning of v, u and a? What is the sequence of steps needed to go from the words of one language to the words in the other language? Associate each of the above probability distributions to the corresponding steps in the translation process. [30%]

(ii) Draw an example weighted finite state transducer (WFST) which implements $P(v_1^K | u_1^K)$. [10%]

(iii) Draw an example un-weighted phrase segmentation transducer. [10%]

(iv) Write the phrase-based SMT process as a composition of WFSTs. [15%]

(v) Draw an example bigram language model acceptor with an exact implementation of the back-off. [15%]

END OF PAPER