Engineering Tripos Part IIB FOURTH YEAR

MODULE 4F11: SPEECH AND LANGUAGE PROCESSING

Examples Paper 2

All problems other than #2 and #3 are tripos level.

1. Discuss how *alignment* is used to indicate *translation equivalence*. Explain the hierarchical nature of translation equivalence.

2. The *closure* operators indicate repetition, e.g. $a^* = \{\phi, a, aa, aaa, \ldots\}$ and $a^+ = \{a, aa, aaa, \ldots\}$. Draw acceptors for the following:

   (a) $a \bigcup b$

   (b) $(a \bigcup b)^+$

   (c) $(a \bigcup b)^*$

   (d) $a^+ b^* c d^+$

3. For the vocabulary $V = \{a, b, c\}$, a per-symbol loss function is defined as following for $x, y \in V$
$$d(x,y) = \begin{cases} 0 & x == y \\ 1 & x \neq y \end{cases}$$
and $d(x, \epsilon) = d(\epsilon, x) = 1$. Draw a transducer which implements this distance and explain how this transducer can be used to compute the *edit distance* between two strings from V^*.

4. Suppose the Model-1 word translation probability distribution is to be estimated from the following two sentence pairs:
$E^{(1)} = a\, r\, p\, q\, r \quad F^{(1)} = \epsilon\, \gamma\, \rho\, \epsilon$
$E^{(2)} = o\, m\, r\, r\, a \quad F^{(2)} = \beta\, \alpha\, \epsilon\, \theta$

   (a) Explain why the maximum likelihood estimate of $p_T(\gamma|m)$ is zero.

   (b) Describe a parallel text alignment application which benefits from such harsh models.

   (c) Describe a parallel text alignment application for which models with word translation probability values of zero would be inappropriate.

   (d) Suggest a modeling strategy which avoids assigning zero probability to word translation probabilities.

5. Give formulae which describe (a) BLEU and (b) Alignment Error.

   Describe the resources needed for the calculation of both of these quantities. Explain their role in the development of statistical machine translation systems.

6. Suppose a pair of sentences f_1^J and e_1^I are known to be translations.

   (a) Write the formula for the overall translation probability under Model-1.

   (b) Describe in detail how the Model-1 (i) Viterbi likelihood and (ii) marginal likelihood can be computed for this pair of sentences using Weighted Finite State transducers.

7. A Markov process X_t takes values in $\{a, b, c\}$. The transition probability associated with the process is

X_{t-1}	X_t	$P(X_t\|X_{t-1})$
a	b	0.8
a	otherwise	0.1
b	c	0.5
b	otherwise	0.25
c	a	0.2
c	otherwise	0.4

   (a) A weighted acceptor is required which assigns weight to sequences consistent with the Markov transition probabilities. Draw this acceptor in such a way that all transition probabilities appear explicitly. Assume likelihood is to be computed with operations in the tropical semiring.

   (b) Draw an equivalent machine using *failure transitions*. Explain how this leads to a simpler machine than that of part (a).

   (c) Redraw the machine of (b) with the failure transitions replaced by epsilon transitions. Explain how this may lead to incorrect assignment of probability to some sequences. Illustrate this point by calculating the likelihood of the sequences 'a b b' and 'a c b' .

8. The following sentences are to be used as a language model training set :

   Sentence 1 : <s> a b b c </s>
   Sentence 2 : <s> a c c a b </s>
   Sentence 3 : <s> c a c c b </s>
   Sentence 4 : <s> b b c a b </s>

   (a) Tabulate the statistics needed to compute unigram, bigram, and trigram language models.

   (b) Calculate maximum likelihood unigram, bigram, and trigram language models from these statistics.

   (c) Explain *discounting* and *backing-off* and illustrate your explanation using the statistics and models from (a) and (b).

9. A pair of sentences f_1^J and e_1^I are known to be translations. Under Model-2 their alignment is described by the process a_j, $j = 1 \ldots J$, such that

$$P(f_1^J, a_1^J, J|e_1^I) = \prod_{j=1}^{J} p_{M2}(a_j|j, I, J)\, p_T(f_j|e_{a_j})$$

Note: For simplicity, disregard the sentence length distribution $p_L(J|I)$.

(a) Derive the following expression for the efficient calculation of the translation posterior

$$P(f_1^J|e_1^I) = \prod_{j=1}^{J} \sum_{i=0}^{I} p_{M2}(i|j, I, J)\, p_T(f_j|e_i)$$

Hint: $P(f_1^J|e_1^I)$ can be written $\sum_{a_1^J} P(f_1^J, a_1^J|e_1^I) = \sum_{a_1=0}^{I} \cdots \sum_{a_J=0}^{I} P(f_1^J, a_1^J|e_1^I)$.

(b) Using the result of (a), derive the following expression for the efficient calculation of the alignment link posterior probability

$$P(a_j = i|e_1^I, f_1^J) = \frac{p_{M2}(i|j, I, J)\, p_T(f_j|e_i)}{\sum_{i'=0}^{I} p_{M2}(i'|j, I, J) p_T(f_j|e_{i'})}$$

(c) Give parameter update equations for the Model-2 component distributions in terms of these posterior probabilities.

(d) Using the results of (a) and (b), give efficient 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.

(e) Suppose it is necessary to choose between $f^{(1)}$ = 'a b c d' and $f^{(2)}$ = 'a c b d' as possible translations of an English sentence e_1^I. Is Model-1 suitable for this task ? Justify your answer.

10. A pair of sentences f_1^J and e_1^I are known to be translations. Under the word-to-word alignment HMM their alignment is described by the process a_j, $j = 1 \ldots J$, such that

$$P(f_1^J, a_1^J, J|e_1^I) = \prod_{j=1}^{J} p_T(f_j|e_{a_j})\, p_{HMM}(a_j|a_{j-1}, I)$$

Note: For simplicity, disregard the sentence length distribution $p_L(J|I)$.

(a) Derive the following relationship for the efficient computation of the forward probability $\alpha_j(i) = P(a_j = i, f_1^j|e_1^I)$:

$$\alpha_j(i) = \sum_{i'} p_T(f_j|e_i)\, p_{HMM}(a_j = i|a_{j-1} = i')\, \alpha_{j-1}(i')$$

(b) Derive a similar recursion for the efficient computation of the backward probability $\beta_j(i) = P(f_{j+1}^J|a_j = i, e_1^I)$.

(c) Show that the alignment link posterior probability can be computed as

$$P(a_j = i \mid f_1^J, e_1^I) = \frac{\alpha_j(i)\beta_j(i)}{\sum_i \alpha_J(i)}$$

Hint: First show that $P(a_j = i, f_1^J | e_1^I) = \alpha_j(i)\beta_j(i)$.

(d) Show that statistics needed for estimation of the HMM transition probabilities can be found as

$$P(a_j = i, a_{j-1} = i' | f_1^J, e_1^I) = \frac{\alpha_{j-1}(i')\, p_{HMM}(i|i')\, p_T(f_j|e_i)\, \beta_j(i)}{\sum_i \alpha_J(i)}$$

(e) Give parameter update equations for the alignment HMM component distributions in terms of these posterior probabilities.

P.C.Woodland
W.J.Byrne