Two hours

Question ONE is COMPULSORY

UNIVERSITY OF MANCHESTER
SCHOOL OF COMPUTER SCIENCE

Natural Language Systems

Date: Monday 19th January 2015
Time: 09:45 - 11:45

Please answer Question ONE in Section A and TWO Questions from Section B.

This is a CLOSED book examination

The use of electronic calculators is permitted provided they are not programmable and do not store text
You should answer question 1: each part of this question carries 6 marks

1. a) Transformation-based learning (TBL) uses templates that match the tags and forms of items in the context of a given word to decide whether its tag should be changed. Suppose that some initial tagger had assigned the tags \([\text{pron}, \text{verb}, \text{pron}, \text{pron}, \text{noun}, \text{verb}, \text{adj}]\) to the text ‘I saw that that man was wrong’ where the correct tags are \([\text{pron}, \text{verb}, \text{comp}, \text{det}, \text{noun}, \text{verb}, \text{adj}]\). What instantiations would be assigned to the template \\
\[ t1(A, B, C) \# \text{tag}: A > B \leftarrow \text{tag}: C @ [-1]. \]

on the first pass of the TBL learning algorithm? [6 marks]

b) i) Define precision, recall and F-measure [3 marks].

ii) Suppose you had two taggers \(T_1\) and \(T_2\), where \(T_1\) assigned the tag det to the words ‘a’ and ‘the’ and left all others untagged and \(T_2\) assigned the tag noun to every word. Which of these would have higher precision and which would have higher recall? Should you use \(T_1\) first and then back-off to \(T_2\) or vice-versa? [3 marks]

c) Consider the following texts:

(1) the cat chased the mouse.
(2) a cat chased a mouse.
(3) the dog chased the cat.

i) Construct vector space representations of each of these sentences. [3 marks]

ii) What are the TF-IDF scores of the words ‘the’, ‘a’ and ‘cat’ in the three sentences? [3 marks]
d)  
   i) What are formants? What kinds of sounds can be distinguished by looking at
      the formants in a section of a speech signal?  
   [3 marks]
   
   ii) One of the signals shown below is someone saying ‘potato’, the other is someone
       saying ‘banana’. Say which is which, and explain your reasoning.  
   [3 marks]

   ![Sound waves](image)

   e) Define lexical and structural ambiguity, and give examples of each.  
   [3 marks]
   
   Say which of them causes more problems for machine translation systems and
   explain why.  
   [3 marks]

   TOTAL  
   [30 marks]
Section B

You should answer two questions from this section: each question carries 35 marks

2. a) Describe the MALT dependency parsing algorithm. What are the key data structures and the three major operations used in this algorithm? You should illustrate your answer by using it to obtain the tree defined by

\[
['\text{know'}]>\text{'I'}: \text{subject, 'know'}>\text{'saw'}: \text{comp, 'saw'}>\text{'she'}: \text{subject, 'saw'}>\text{'him'}: \text{object}
\]

for the sentence ‘I know she saw him’.

[10 marks]

b) Show that the time complexity of the standard MALT algorithm is linear in the length of the input text. Nivre has argued that parsing algorithms have to find a trade-off between speed, accuracy and robustness. Which of these three is MALT weakest on?

[8 marks]

c) Suppose that the target analysis of the sentence ‘I am eating it’ is ['eating'] > ‘I’: subject, ‘eating’ > ‘it’: object, ‘am’ > ‘eating’: mv]. Show that it is not possible to obtain this analysis using the three standard operations, and suggest one way of extending this set of actions that will lead to this analysis.

[10 marks]

d) Explain how a treebank can be used to provide training data in order to enable MALT to choose which operation to apply in a given situation.

[7 marks]

TOTAL [35 marks]
3. a) Describe the basic page rank (PR) algorithm for finding the most important nodes in a network. You should illustrate your answer with reference to the network shown below. [10 marks]

![Network Diagram]

b) What would happen if you removed the link from B to C and ran the basic algorithm until it converged? How can you overcome this problem? What is the role of the ‘damping factor’ in this variant of the basic algorithm? [7 marks]

c) Describe the personalised (PPR) version of the algorithm and explain how it can be used for laying extra emphasis on specific nodes. [5 marks]

d) i) Explain how you could use the glosses provided in WordNet to construct a network to which PPR can be applied, and explain how this can be used for word sense disambiguation. [10 marks]

   ii) Why does the fact that the nodes in this network are synsets but the items in the glosses are words cause a problem? [3 marks]

TOTAL [35 marks]
4. a) What are Markov networks and how do they differ from ‘hidden’ Markov networks (HMMs)? What are the roles of the emission probabilities and transition probabilities in these networks? You should illustrate your answer with a concrete example. [8 marks]

b) The HTK (Cambridge Hidden Markov Model Toolkit) trains a collection of individual HMMs and then uses a grammar to connect these together to construct a recognition system.

   i) What do the individual HMMs correspond to and what is the role of the phonetic transcription in training them? What happens during the ‘forced alignment’ phase in training the HTK? [7 marks]

   ii) How does using HMMs with three internal states allow the HTK to capture information about phoneme lengths? [6 marks]

   iii) How is the grammar used for constructing a single complex network out of the individual HMMs? [6 marks]

c) Explain how a speech synthesiser might be used to generate training data for a speech recogniser. What problems might arise when you try to do this, and how might you overcome these? [8 marks]

TOTAL [35 marks]
5. a) Describe the dynamic time warping algorithm. You should illustrate your discussion by showing the first few steps in the conversion of ‘bat’ into ‘coat’. [10 marks]

b) Consider the following two passages:

- Explain how a speech synthesiser might be used to generate training data for a speech recogniser. What problems might arise when you try to do this, and how might you overcome these?
- Expliquez comment un synthétiseur de parole pourrait être utilisé pour générer les données d’apprentissage pour un son discours. Quels sont les problèmes pouvant survenir lorsque vous essayez de le faire, et comment vous pourriez surmonter ?

i) How might you use this algorithm to find potential translation pairs in the following two passages? [6 marks]

ii) What other information might you exploit in these two passages when looking for translation pairs? [4 marks]

c) How might you use this algorithm in combination with the WordNet hypernym links to determine whether ‘I saw an old man’ entailed ‘I have seen a human’? Would you expect this approach to entailment to have high precision or high recall? [8 marks]

d) What would happen if you applied this algorithm to the pairs {‘I know she loves me’ ⊢ ‘she likes me’}, {‘I know she likes me’ ⊢ ‘she loves me’}, {‘I doubt she loves me’ ⊢ ‘she loves me’}, {‘I doubt she likes me’ ⊢ ‘she dislikes me’}? How might using labelled dependency trees help with these examples? [7 marks]

TOTAL [35 marks]