CS9180

THE UNIVERSITY OF WARWICK

MSc Examinations: Summer 2018

CS9180 Natural Language Processing

Time allowed: 2 hours.

Answer any **FOUR** questions.
Read carefully the instructions on the answer book and make sure that the particulars required are entered on **each** answer book.

Approved calculators are allowed.

---

1. (a) Write regular expressions for the following languages:
   i. the set of all vowels. Can be either in lowercase or uppercase. [1]
   ii. the set of all alphabetic strings followed by a period. [1]
   iii. the set of all alphanumeric words with one or more digits either at the beginning or at the end. It should match *2hello*, *hello33*, *2he3llo* and *hello2*, but not *he2llo* or *hello*. [2]
   iv. the set of all numbers. Numbers may contain decimal and thousand separators. The regular expression should match: *30*, *35,020.17* and *15.2*, but it should not fully match *24.120,17* (comma and period in wrong position), *12,43.5* (thousand separator in wrong place) and *15,4*. [3]

   (b) i. What are Type I and Type II errors? [3]
   ii. Let $tp = 18, fp = 2, tn = 8$ and $fn = 12$, calculate precision, recall and F1 scores. [3]

   (c) i. What is the vocabulary size ($|V|$) for the following example?
   I was very happy when I was reading the book in the library but the book was actually very expensive so I decided not to buy the book [2]
   ii. What is lemmatisation? [2]
   iii. What is the result of lemmatising the following sentence word by word: *Friends are happy living together*? [2]
   iv. What are the stems for the following words using the Porter stemmer: king, something, living? [2]

   (d) The Maximum Matching algorithm:
   i. What language is it useful for? [1]
   ii. How does it work? [3]

Continued

2. (a) i. What are n-grams? [2]

ii. Describe one way of performing extrinsic evaluation of language models, and another way of performing intrinsic evaluation. What is a disadvantage of each of these evaluation techniques? [7]

iii. What is the effect of using longer n-grams for creating language models? [3]

iv. What is the Markov assumption? [1]

v. How does the Laplace smoothing work and what is the problem of using it for language models? [2]

(b) Given a corpus with these 5 sentences:

<s> I am jumping </s>
<s> you are jumping </s>
<s> she is working </s>
<s> we are laughing </s>
<s> they are laughing </s>

i. What are the Maximum Likelihood Estimates P(are | you) and P(jumping | are)? [4]

ii. Following the Kneser-Ney smoothing method, what are the probabilities of being a novel continuation ($P_{continuation}$) for the words *are*, *jumping* and *laughing*? Do not include <s> and </s> in the calculations. [6]

Continued

3. (a) i. What is cross-validation? [2]

ii. Describe three different ways of assessing which features are good for our classifier. [3]

iii. How is semi-supervised classification different from supervised classification? [2]

iv. What are two different settings to develop multiclass classifiers? [2]

v. What is multilabel classification and how can we achieve it? [1]

vi. What is overfitting? [1]

vii. What is the difference between microaveraged and macroaveraged evaluation? [2]

viii. Describe three ways of dealing with imbalanced data in text classification. [3]

(b) i. What are the main limitations of Hidden Markov Models (HMM) and Maximum Entropy Markov Models (MEMM) that the Conditional Random Fields (CRF) overcome? [4]

ii. HMM and CRF are the sequential counterparts of what other non-sequential classifiers? [1]

iii. In HMM, what are transition probabilities and emission probabilities? [3]

iv. List two deep learning methods for sequential classification. [1]

---

4. (a) Grammars and Parsing:

   i. What are Context-Free Grammars (CFG) and what are their components? [3]

   ii. What are the differences between CFG, PCFG and lexicalised PCFG? [2]

   iii. What is the Cocke-Kasami-Younger (CKY) algorithm used for and how does it work? [2]

   iv. What is the MaltParser algorithm used for? What 4 types of operations does it perform? [3]

   (b) What are hyponyms and hypernyms, and what is the best resource we can use to identify them? [2]

   (c) Given the co-occurrence matrix in Table 1, compute the matrix of Pointwise Mutual Information (PMI) scores.

	birmingham	london	manchester	liverpool	bristol
arsenal	2	15	0	4	0
everton	0	2	2	11	1
villa	17	1	0	0	3
united	0	3	14	0	0

Table 1: Co-occurrence matrix provided to compute PMI scores.

[7]

   (d) What are the most common approaches to Named Entity Recognition (NER) in academic research and commercial applications? [2]

   (e) What is semantic role labelling? Describe the 3 steps for developing an SRL system. [4]

Continued

5. (a) Information Retrieval:

   i. What are the main advantage and disadvantage of storing positional indices over binary indices? [1]

   ii. Write an algorithm that produces the intersection of two positional indices that satisfy a proximity query. Given two lists of positional indices *p1* and *p2*, and *k* defining the distance between two query words, write a function that returns the intersection of *p1* and *p2* given distance *k*. [8]

   iii. For the following two queries:
   A: *list of universities in the united kingdom*
   B: *academics from the universities of the united states*
   Compute the Jaccard coefficient of A and B. [2]

   iv. What is TF-IDF and why is it useful for information retrieval? [3]

   v. What is the SMART notation and what is the most common notation used by search engines? [4]

   vi. Our search engine has produced the output shown in Table 2 for two queries.

Ranking	Q1	Q2
1	R	N
2	N	R
3	N	R
4	R	N
5	R	R
6	R	R
7	N	N
8	N	R
9	R	N
10	N	R

Table 2: Ranking and relevance judgements output by an information retrieval system for two queries Q1 and Q2.

Compute the Average Precision (AP) and Normalised Discounted Cumulative Gain (NDCG) for each query, as well as the Mean Average Precision (MAP) for both queries combined, up to the 10th element in all cases. Does our system perform better for Q1 or Q2? [7]

Continued

6. (a) Recommender Systems:

   i. Describe the intuition that a collaborative filtering follows to make recommendations. [2]

   ii. List two advantages and two disadvantages of collaborative filtering recommender systems. [4]

   iii. List two advantages and two disadvantages of content-based collaborative recommender systems. [4]

   iv. Define cold start and popularity bias in recommender systems. [2]

   (b) Word Embeddings:

   i. What is the process to get new vectors of reduced dimensionality by using Singular Value Decomposition (SVD)? [5]

   ii. How do we compute the similarity between two word vectors using embeddings? [2]

   iii. What are the advantages and disadvantages of SVD? [4]

   iv. Having word embeddings trained for each word, what are two commonly used approaches to get the word embeddings for a sentence as a single vector? [2]

End