

PGT Exam Performance Feedback

2018/2019 Semester 2

COMP61332 Text Mining

Sophia Ananiadou
Riza Batista-Navarro

Comments Question 1.

- (i) This was a bookwork question and many answered it correctly.
- (ii) The question involved some critical thinking, they had to assign the correct semantic role to some sentences: most students did well in this question. Most students missed the fact they had to give a definition of semantic role and most assign wrongly the semantic role of sentence 4.
- (iii) This was a bookwork question and most students answered it correctly.
- (iv) This involved a critical understanding of the process of distant supervision. The students did not answer the question satisfactorily, missing the stages of distant supervision (matching criterion and label transfer). Most answered half the question.
- (v) C-value question. Although this question was covered in the lecture notes, the vast majority of students did not answer the question. None answered what is involved in the calculation of the statistical part of C-value.

Question 2.

- (i) This was a bookwork question, nevertheless one third of the students did not answer it correctly. The majority of students gave a simple list of NLP applications but did not provide the brief justification how terms can be used in these applications.
- (ii) This question involved critical thinking, comparing index terms to technical terms. Some students answered the question by repeating what has been asked. Overall the response was below average.
- (iii) Normalisation. This was a critical question mixed with some bookwork. The majority of students answered it satisfactorily.

Question 3. (i) Syntactic ambiguity. This question involves analysis of the sentence to determine where the syntactic ambiguity is coming from. Majority of the students answered it correctly; those who did not were mostly confused thinking that "down" could possibly be attached to "dog" (whereas the possible heads are "chased" and "cat"). (ii) Dependency parse trees. This question was answered correctly by majority of the students. (iii) Phrase structure trees. This question was answered correctly by majority of the students. (iv) Noun phrase annotation using BIO. This question was answered correctly by majority of the students. Most of those who answered it incorrectly made the mistake of annotating only the head nouns (i.e., not including the determiners which are part of noun phrases). (v) BIO vs inline notation. This question was answered correctly by almost all students.

Question 4. (i) Dictionary-based NER. This question was answered correctly by almost all students. (ii) Rule-based NER. Most students answered this question correctly. Some students however provided unsatisfactory answers, unable to provide examples of function words (e.g., "cancer", "carcinoma") that can be used to form rules. (iii) NER performance metrics. While only few students came up with the correct number of true positives (6), false positives (1) and false negatives (6), almost all students demonstrated knowledge of how precision, recall and F-score are correctly calculated. (iv) Event extraction. Unfortunately most students provided unsatisfactory answers to this question. Only few were able to distinguish between simple and complex events, and between positive-polarity and negative-polarity events.
