UG Exam Performance Feedback
Third Year
2016/2017 Semester 2

COMP34412 Natural Language Systems

Goran Nenadic
Tingting Mu

Comments

Overall, the students demonstrated good understanding of the issues, with an average of 68.5% and standard dev. of 13%, with marks ranging between 32% and 93%. Bookwork was mostly fine as were applications.

Question 1.

Two thirds of students took this question and they answered it reasonably well overall (average mark of 68%, with std. dev. of 5%).

Part (a) was generally well answered, although not everyone used the examples provided to illustrate their views. Justifications for part (a.ii) were usually reasonable but still sometimes shallow and not mapped to the machine-translation task.

Part (b.i) contained an error in the grammar that made the sentence impossible to parse with the given set of rules. Most of answers have corrected the error and parse it, while a small number of students noted it can't be parsed. Both answers were accepted. Part (b.ii) was mainly fine, although rarely answers noted that lexical ambiguity also led to structural/syntactic ambiguity in this case.

Part (c.i) was done by almost all students. Part (c.ii) was more challenging but has been reasonably done by many. There was some confusion about left and right reduce – noting to do with how you draw your tree!

Part (d) was less well answered, with a number of answers not relating the answer to word-sense disambiguation; rather, general principles of vector space models and document similarity were discussed, but that did not bring many marks. This question aimed to go across different topics discussed in the course and could have been motivated by what has been discussed around the extended Lesk algorithm.

Part (e) was generally OK – most students identified challenges of social media wrt to NER. Discussions of rule vs. machine-learning systems was sometimes superficial and sometimes incorrect (e.g. claiming that some methods are fast and some are not reliable). The final part was a discussion point where most answers got full marks if justification was reasonable (even if not the preferred metric was chosen).

Question 2.

This was a popular question, with 85% of students selecting it. They answered it reasonably well overall (average mark of 63%, with std. dev. of 8%).

Part (a) was generally quite well answered, including the bookwork for a.i and a.ii. There was some confusion on how to calculate word likelihood. Still, most students answered correctly the practical part in a.iii, either by looking only at the local probability (ignoring the rest of the sentence) or looking at the overall sentence.

Part (b.i) was again mostly answered correctly, with very few misunderstandings of polysemy. However, b.ii was not always focused on WSD and which word weight one can use to represent the context of a target word. This was also a bit confusing for some students in b.iii – with many answers only focusing on bag of words and potentially POS tags to support WSD.

Part (c) was surprisingly quite challenging and answered vaguely. There have been quite generic and unjustified claims that rule-based methods are slow or fast; that ML approaches are (always) unreliable etc. Many answers failed to focus on syntactic analysis and why any specific solution would matter in that context.

Part (d) was again less well answered, with a number of answers misinterpreting bootstrapping and finding patterns from text as opposed to finding information we want to extract (city/state and person). The role of Wikipedia was discussed but some answers focused on layout rather than on patterns. The final part d.ii was reasonably answered with most comments focusing on variability of patterns that could be extracted from two corpora. Rarely the issues with possible noise, ambiguous patterns, data access etc. have been discussed.

Part (e) was generally OK – most students identified challenges of social media wrt to NER. Discussions of rule vs. machine-learning systems was sometimes superficial and sometimes incorrect (e.g. claiming that some methods are fast and some are not reliable). The final part was a discussion point where most answers got full marks if justification was reasonable (even if not the preferred metric was chosen).

Question 3.
A total of 26 students took this question and they answered it well overall (average mark of 30.1/40 = 75.4%, with std. dev. of 5.3/40 = 13.1%).

UG Exam Performance Feedback
Third Year
2016/2017 Semester 2

Part (a) was very well answered, with 21/26=81% of the students scored over 6 out of 8. In part i), almost all the students were able to justify the design of inverse document frequency. In part ii), some students calculated the term frequency as word counts and some as percentage normalised by the total word number. Both cases were accepted.
Part (b) was very well answered, with 22/26=85% of the students scored over 9 out of 10. In part i), almost all the students were able to explain how to use bigram model to compute sentence probability. Part ii) contained a mistake with P(british|have) missing from the table. This did not affect the students to formulate the sentence probability. Almost all the students provided the correct equation. Some students left this particular probability P(british|have) in notation, some students assumed it was equal to zero, and some assumed it was equal to one, all of which were accepted.
Part (c) was generally well answered, with 18/26=69% of the students scored over 6 out 8. Part i) was done by almost all students. Part (ii) was correctly answered by most of the students.
Part (d) was reasonably well answered, with 12/26 =46% of the students scored over 5 out of 6. In part i), a number of answers did not address the main disadvantage of being incapable of dealing with out-of-vocabulary words. In part ii), many students correctly mentioned either rule-based methods or machine learning based methods, or both.
Part (e) was less well answered, with 8/26 =31% of the students scored over 6 out of 8. Some students did not attempt this part. In part i) many students did not address the importance of defining a “best matching” metric or efficient searching. In part ii), many students were able to explain the assumption, but some students did not explain how Bayes’ rule can be used to derive the equation.