Text Mining: Exercise week 7

By Joris van Vugt (s4279859)

1. What information from the Yahoo data do the authors use as labels for the correctness of answers?

Answers which are selected as best either by the asker or by the participants in the thread are correct. All others are wrong.

2. Explain how the authors address the problem of the vocabulary gap between question and answer without using external knowledge bases such as WordNet.

A translation model is learned, which translates question terms to answer terms.

3. Give a definition in your own words of what the authors call 'Recall@N'. Do you think that the term 'Recall@N' is a good descriptor of this definition? Please motivate your answer.

In this paper, Recall@N is defined as the number of questions which have the correct answer in the top-N retrieved answers. I think calling this metric Recall@N is misleading, because recall is very well defined in IR tasks: the number of relevant retrieved documents divided by the total number of relevant documents. Intuitively, Recall@N should indicate the fraction of relevant documents retrieved in the top-N documents. I think the metric defined in this paper is closer to what is usually called '*accuracy*' in machine learning tasks.

4. To what extent would the feature set in this paper be applicable to factoid questions? Which of the features would not be?

Answers to factoid questions are usually much shorter. Therefore, features that analyse the structure of the answer are a lot less relevant. For example features such as answer span, verb count and same sentence match would be a lot less helpful. On the other side, features like the translation from question term to answer term would still be very useful to bridge the gap between the question and aswer vocabularies.