The University of Amsterdam at TAC 2008 Question Answering track

Valentin Jijkoun   Maarten de Rijke

ISLA, University of Amsterdam
`http://ilps.science.uva.nl/`

Abstract: We describe the participation of the University of Amsterdam's ILPS group in the Question Answering track of the TAC 2008. We used a simple system based on lexicon-based identification of opinionated sentences and the answer extraction module of our factual question answering system. The results indicate that filtering out sentences that are unlikely to contain opinions does improve the end-to-end performance of the system.

1   System description

Systems participating in the TAC 2008 Question Answering track were required to find precise answers to English questions in a large corpus of blogs (TREC Blog06 corpus). The test questions were opinionated: they asked either for lists of entities that are sources or targets of specific opinions (rigid list questions, such as *Name US senators who support tax reform.*) or for reasons and other details of specific opinions (squishy list questions, such as *Why do countries want to have nuclear power plants?*).
   For our participation, we used a simple system consisting of five modules:

1. *Blog post retrieval*: for every questions, we retrieved top 500 blog posts from the Blog06 corpus, using a language modeling-based retrieval engine [Balog et al., 2008];

2. *Sentence filtering*: in the content of the retrieved blogs, we only retained sentences containing at least one subjective word from the subjectivity lexicon described in [Wilson et al., 2005];

3. *Question classification*: from every question, we extracted the expected answer type using the question classification module of our factual question answering system [Jijkoun et al., 2008]. Questions for which the classifier was able to determine the answer type (a named entity type) were considered to be rigid list questions; other questions were treated as squishy list;

4. *Answer extraction*: For rigid questions, we extracted all named entities of the relevant types from the filtered sentences; for squishy questions, we simply extracted all sentences along with their similarity scores (similarity with the question);

5. *Answer selection*: For every question, we returned a list of 20 top ranked answers, ranking either by frequency (rigid) or by similarity (squishy).

We produced two runs: with and without sentence filtering by opinionatedness.

2 Results

We show the evaluation results (the F-score of precision and recall) for rigid and squishy list questions in the table below.

Run	F (rigid)	F (squishy)
Filtering	0.070	0.088
No filtering	0.063	0.081

As we expected, filtering out sentences that are unlikely to contain opinions does improve the end-to-end performace of the system on both types of questions. The improvement for rigid questions is statistically significant (with sign test at p=0.02).

3 Acknowledgments

References

[Balog et al., 2008] Balog, K., Meij, E., Weerkamp, W., He, J., and de Rijke, M. (2008). The university of amsterdam at trec 2008: Blog, enterprise, and relevance feedback. In *TREC 2008 Working Notes*.

[Jijkoun et al., 2008] Jijkoun, V., Hofmann, K., Ahn, D., Khalid, M., van Rantwijk, J., de Rijke, M., and Tjong Kim Sang, E. (2008). The university of amsterdam's question answering system at qa@clef 2007. In *Advances in Multilingual and Multimodal Information Retrieval*, volume 5152, pages 344–351.

[Wilson et al., 2005] Wilson, T., Wiebe, J., and Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP 2005)*, pages 347–354.