The University of Amsterdam at the 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 at 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.

<table>
<thead>
<tr>
<th>Run</th>
<th>F (rigid)</th>
<th>F (squishy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filtering</td>
<td>0.070</td>
<td>0.088</td>
</tr>
<tr>
<td>No filtering</td>
<td>0.063</td>
<td>0.081</td>
</tr>
</tbody>
</table>

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

This research was supported by the E.U. IST program of the 6th FP for RTD under project MultiMATCH contract IST-033104, by the DuOMAn project carried out within the STEVIN programme which is funded by the Dutch and Flemish Governments (http://www.stevin-tst.org) under project number STE-09-12, and by the Netherlands Organisation for Scientific Research (NWO) under project numbers 220-80-001, 017.001.190, 640.001.501, 640.002.501, 612.066.512, 612.061.814, 612.061.-815, and by the Virtual Laboratory for e-Science project (http://www.vl-e.nl), which is supported by a BSIK grant from the Dutch Ministry of Education, Culture and Science and is part of the ICT innovation program of the Ministry of Economic Affairs.

## References

