Detecting Implicit Aspects in Consumer Reviews for Sentiment Analysis

Kim Schouten, Flavius Frasincar, and Franciska de Jong

Word-based co-occurrence approach

**Annotated sentences**

- Part-of-Speech filter:
  - Only include words that have a specific POS.

**Unseen sentence**

- \( \hat{T} \times \mathbf{C} \times \frac{1}{n} = \text{Aspect scores} \)

Semantics-based co-occurrence approach

**Annotated sentences**

- Use Word Sense Disambiguation (Lesk, 1986) to link words to WordNet concepts when available

**Unseen sentence**

- \( \mathbf{W} \times \mathbf{C} \times \frac{1}{n} = \text{Aspect scores} \)

Results

- **Evaluation**
  - Moving from a traditional word-based to a semantics-based approach is beneficial. Without including the concept relations, the semantic approach performed roughly the same as the word-based approach. This could be due to errors in the word sense disambiguation. However, when including the concept relations, performance increases significantly.
  - It is especially the case for the product reviews, a data set which has more different aspects but less training data. We hypothesize that for small data sets, the information added by the concept relations is valuable in detecting implicit aspects. Even for larger data sets, there is still a significant effect.

**Future Work**

- Some of the possibilities for future work are:
  1. Adjust the method such that it can detect multiple implicit aspects per sentence. This is especially useful for the restaurant data, where a large number of sentences have more than one implicit aspect.
  2. Combine the two training steps: the genetic optimization for the concept relation weights and the linear search for the threshold value, into one optimization algorithm. A proper training for the POS filter could also be included.
  3. Incorporate domain-specific ontology concepts and relations.

Acknowledgements

- The **word-based** method is described in:

- The **semantics-based** method is created together with students Nienke de Boer, Tjian Lam, Marijtje van Leeuwen, and Ruud van Leeuwen, and is currently submitted for publication.

- **Data**
  - Restaurant reviews (SemEval-2014 training data)
  - Electronic product reviews (30k b. i.e. 2006)

- **Evaluation & Conclusion**
  - These results are obtained using 10-fold cross-validation. For each method, the best performing POS filter is used.