

(two) Ideas for master projects

Malvina Nissim
m.nissim@rug.nl
room 1311.426

IK meeting, 28 November 2014

Project 1: Winograd Schema Challenge

Winograd Schema Challenge

Although they ran at about the same speed, Sue beat Sally because she had such a [good/bad] start.

Who had a [good/bad] start?

Answers: Sue/Sally.

I spread the cloth on the table in order to [protect/display] it.

To [protect/display] what?

Answers: the table/the cloth.

Why it is difficult, and interesting

examples are constructed such as they are:

- easily disambiguated by the human reader (ideally, so easily that the reader does not even notice that there is an ambiguity)
- not solvable by simple techniques such as selectional restrictions
- Google-proof: there is no obvious statistical test over text corpora that will reliably disambiguate these correctly

Key points

- anaphora resolution
- learning approach, but deep processing
- can we exploit paraphrases? (available)
- can we exploit scores of association norms (available)?

Data and References

- Main attempt: overall accuracy of 73% using a composite set of features in a learning setting, on this dataset:
www.hlt.utdallas.edu/~vince/data/emnlp12/train-emnlp12.txt.
Altaf Rahman and Vincent Ng (2012), “Resolving Complex Cases of Definite Pronouns: The Winograd Schema Challenge”. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 777–789, Jeju Island, Korea.
- Overview of datasets and links:
<https://www.cs.nyu.edu/davise/papers/WS.html>
- Data on association norms:
<https://sites.google.com/site/kenmcraielab/norms-data>

Project 2: Learning Ironic Patterns

(or more in general “mood patterns”)

- select one or more “moods”, possibly exploiting existing social data (e.g. newspapers who add readers’ moods to their pieces online)
- learn abstract “patterns”
Example for irony (made up!): bad before good; rarely exclamation mark; usually exaggeration via superlatives. NB: feature identification.
- learn to detect mood(s)
- per mood: continuous values rather than categorical

NB: semantic-web work on patterns also related

Initial references

- Carvalho, P., Sarmiento, L., Silva, M.J. and De Oliveira, E. (2009). Clues for Detecting Irony in User-generated Contents: Oh...? It's So Easy? ;-). In *Proceedings of the 1st International CIKM Workshop on Topic-sentiment Analysis for Mass Opinion*, pages 53–56. TSA '09. New York, NY, USA: ACM.
- Reyes, A. and Rosso, P. (2011). Mining Subjective Knowledge from Customer Reviews: A Specific Case of Irony Detection. In *Proceedings of the 2Nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, pages 118-124. WASSA'11, Stroudsburg, PA, USA: Association for Computational Linguistics.
- Vanin, A.A., Freitas, L.A., Vieira, R., and Bochernitsan, M. (2013). Some Clues on Irony Detection in Tweets. In *Proceedings of the 22Nd International Conference on World Wide Web Companion. WWW'13 Companion*, pages 635–636. WWW, Geneva, 2013.