

Sentence-Level Polarity Classification: Sentiment Analysis of PDT

Kateřina Veselovská
Hejnice, September 2010

Outline

- An Introduction to Polarity Classification
- Sentence-Level Classification
 - Task & Motivation
 - Feature Design
- Application of SA to PDT
 - Phases
 - Experiments
 - Further goals

Outline

- An Introduction to Polarity Classification
- Sentence-Level Classification
 - Task & Motivation
 - Feature Design
- Application of SA to PDT
 - Phases
 - Experiments
 - Further goals

Polarity Classification

- subfield of Opinion Mining
- two different types of text classification in Opinion Mining
 - **Subjectivity Detection**
 - Does a text represent an **opinion or a fact?**
*The Czech President resides in the Prague Castle vs.
The current Czech President is very popular in Europe.*
 - **Polarity Classification**
 - Given an opinionated text, is the opinion expressed in the text **positive or negative?**
I like functors vs. I hate subfunctors

Polarity Classification

- Polarity is mostly indicated by **polar expressions** (e.g. *nice*, *awful* etc.)
- Polar expressions are *not frequent* content words
- Polar expressions are subject to context influences:
 - Example *negation*:
 - I [*don't like*]- *subfunctors*.

Outline

- An Introduction to Polarity Classification
- Sentence-Level Classification
 - Task & Motivation
 - Feature Design
- Application of SA to PDT
 - Phases
 - Experiments
 - Further goals

The Task

- Decide whether a given sentence is either an overall positive or negative opinion
- All sentences to be classified are assumed to be subjective and carrying either positive or negative overall polarity

Polarity Classification at Sentence Level

- More fine-grained polarity classification than document-level classification is needed for NLP tasks such as
 - Question Answering
 - Text Summarization
- Subjectivity/polarity is usually not uniformly distributed across a document

Difference to Document-Level Classification

- At document level text classification relies very much on redundancy
- The correct interpretation of every word is NOT necessary
- There are so many other cues suggesting positive polarity

Difference to Document-Level Classification

- At sentence level the correct interpretation of every word is more important
 - E.g. one needs to determine whether *like* is a verb and hence a positive polar expression or just a preposition.

Why PDT?

- it is expected that some syntactic (and hypersyntactic) relations are useful for identification of sentence polarity
 - negation
 - sentential modality marking
 - discourse relations
 - intersentential coreferential relations
 - depth of the polarity item in the tree

Contribution of this work

- Inspection of various linguistic features derived from sentence structure
- Various combinations of linguistic features, polarity features and bag-of-words features in supervised machine learning

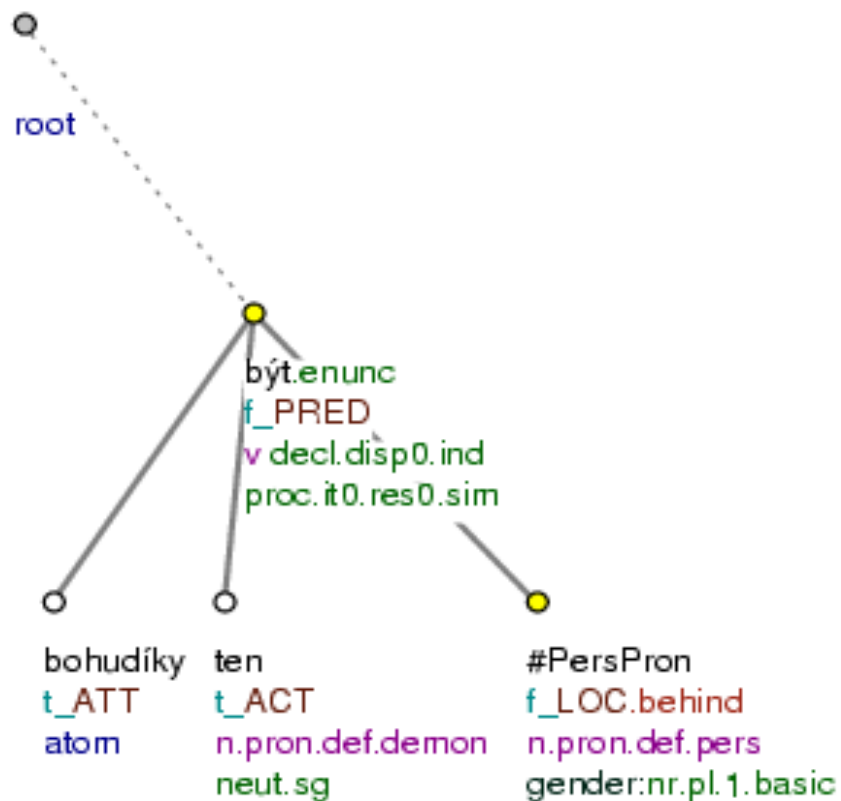
Outline

- An Introduction to Polarity Classification
- Sentence-Level Classification
 - Task & Motivation
 - Feature Design
- Application of SA to PDT
 - Phases
 - Experiments
 - Further goals

Linguistic Word-Level Features

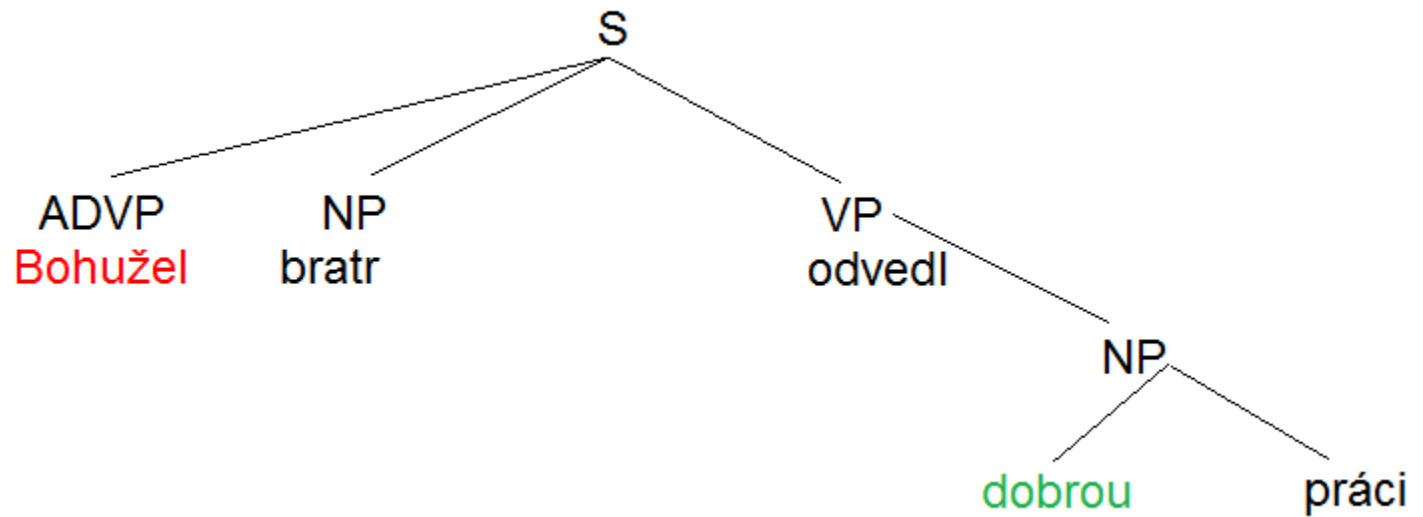
- Linguistic features: linguistic properties of words
 - (Generalized) parts of speech:
tags: *noun, verb, adjective, adverb etc.*
 - Position (of leaf node representing word in the dependency tree)
 - Main predicate
 - Clause type

Position Feature



Je to bohudíky za námi.

Position Feature



- Main predicate

- more predictive towards polarity than other words

[The Pakistani government *supports*[pred] president Bush and his *war on terror*]+. (Wiegand 2009)

- Clause type

- main clause vs. subordinate clause

Outline

- An Introduction to Polarity Classification
- Sentence-Level Classification
 - Task & Motivation
 - Feature Design
- **Application of SA to PDT**
 - Phases
 - Experiments
 - Further goals

Application of SA to PDT - Phases

- Building a subjectivity lexicon
 - based on Frequency Dictionary of Czech (Čermák et al., 2009)
- Detection of polarity items in PDT
 - tags: positive, negative, neutral, undecidable
 - fully automatic annotation
 - series of manual controls
- Analysis of the annotation using statistical methods

Application of SA to PDT - Experiments

- Psycholinguistic field
 - DmDX measurement (F. Smolík)
 - detection of polarity items in given sentences
 - correct vs. incorrect
 - reaction time
 - focused on ambiguous sentences
 - Electrodermal responses
(J. Lukavský, P. Šlechta)

Institute of Psychology, Academy of Sciences

Application of SA to PDT – Further Goals

- submitted GAČR proposal
(we won't give up!)
- TAČR proposal
 - www.ataxo.com – Czech division of the Internet marketing agency
(research and development manager J. Šlerka)



THANK YOU



THANK YOU
(FOR BRAINSTORMING)

References

- Akkaya, C., Wiebe, J., and Mihalcea, R.: Subjectivity Word Sense Disambiguation. In *Proceedings of EMNLP 2009*.
- Banea, C., Mihalcea, R., Wiebe, J.: A Bootstrapping Method for Building Subjectivity Lexicons for Languages with Scarce Resources. In *The Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008)*, 2008.
- Banea, C., Mihalcea, R., Wiebe, J., Hassan, S.: Multilingual Subjectivity Analysis Using Machine Translation. *Conference on Empirical Methods in Natural Language Processing (EMNLP 2008)*, 2008.
- Liu, B.: *Sentiment Analysis and Subjectivity*. Invited Chapter for the *Handbook of Natural Language Processing*, Second Edition. Marcel Dekker, Inc: New York, 2009.
- Narayanan, R., Liu, B., and Choudhary, A.: Sentiment Analysis of Conditional Sentences. In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-09)*. August 6-7, 2009. Singapore.
- Ruppenhofer, J., Somasundaran, S., and Wiebe, J.: Finding the Sources and Targets of Subjective Expressions. In *The Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008)*, 2008.
- Somasundaran, S., Namata, G., Wiebe, J., and Getoor, L.: Supervised and Unsupervised Methods in Employing Discourse Relations for Improving Opinion Polarity Classification. In *Proceedings of EMNLP 2009*.
- Somasundaran, S., Wiebe, J., and Ruppenhofer, J. Discourse Level Opinion Interpretation. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING-2008)*, 2008.
- Wiebe, J., Wilson, T., Bruce, R., Bell, M., and Martin, M. Learning subjective language. *Computational Linguistics* 30 (3), 2004.
- Wilson, T. *Fine-Grained Subjectivity and Sentiment Analysis*. PhD Dissertation, Intelligent Systems Program, University of Pittsburgh, 2008.