Exploiting Debate Portals for Argumentation Mining in User-Generated Web Discourse

Ivan Habernal and Iryna Gurevych
Ubiquitous Knowledge Processing Lab
https://www.ukp.tu-darmstadt.de/

This work has been supported by the Volkswagen Foundation as part of the Lichtenberg Professorship Program under grant I/82806 and by the German Institute for Educational Research (DIPF).

Get the code and data: github.com/habernal/emnlp2015

Summary

Argumentation mining
- Analysing arguments in text according to a particular argumentation model
- Supervised methods
- Expensive annotation costs
- Scarce labeled data

Research question
- Can we gain useful information from noisy, unlabeled, yet argumentative debate portals „for free“?

Findings
- By clustering unlabeled arguments and sentences using their representation in embeddings space and using projection into this space by supervised algorithm, we get improvement
- About 10% in cross-validation
- Up-to 90% in cross-domain and cross-register evaluation

Algorithms

Baseline feature sets
- N-grams baseline (FS0)
- Structure and syntax (FS1)
  - first and last three words, relative position, number of POS n-grams
  - maximum dependency tree depth
  - constituency tree production rules, number of sub-clauses
- Sentiment and topic (FS2)
  - Sentiment distribution (5 feat.)
  - LDA topic model distribution
- Semantic and discourse (FS3)
  - PropBank Semantic frames
  - Coreference chain
  - PTDB-style discourse parser

Classification
- DKPro TC framework for experimental setup and feature engineering
- DKPro Core for linguistic annotations
- SVM-HMM for sequence labeling

Argument space features

Semi-supervised extension (FS4)
- Cluster sentences and arguments from debate portals using embeddings (word2vec)
- Parameters: TF-IDF weighting, addition, averaging, ...
- Use cluster centroids as a space for projecting argument components, feed into supervised classification

Gold-standard Data

Source
- 340 English documents (90k tokens)
- 6 topics: homeschooling, mainstreaming, prayer in schools, public vs. private school, redshirting, single-sex schools
- 4 registers: comments to articles, forum posts, blogs, newswire opinionated articles

Annotated examples

Dec #3053 (article comment, public-private schools)

- The more people leave, the worse it's going to become. The public schools are there, and there will be more private schools. There are still great teachers in the public schools - just stuck behind them.

Dec #3224 (forum post, single-sex education)

- My daughter is 14. She is in the 7th grade. My school is a very competitive academic school. They only accept girls of certain intellectual capability. If it is purely education you are concerned about and not just social status, then you are better off in a private school. If it is purely social status, and you want your daughter to be around other girls, then you would be better off in a public school.

Data representation
- BIO tagging, 11 classes

Annotation model
- Extended Toulmin’s model of argument
  - 5 argument component types
  - Implicit relations

Classification
- DKPro TC framework for experimental setup and feature engineering
- DKPro Core for linguistic annotations
- SVM-HMM for sequence labeling

Results

Cross-validation


Cross-register


Cross-domain