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

TECHNISCHE UNIVERSITÄT DARMSTADT

UKP

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

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

Fork me on GitHub

Summary

Argumentation mining
- Analysing arguments in text according to a particular argumenation 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

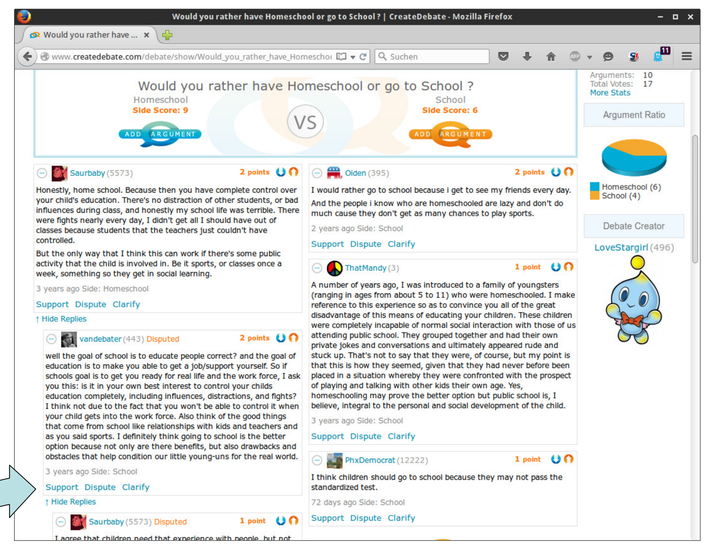
Algorithms

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, cluster count, …
- 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

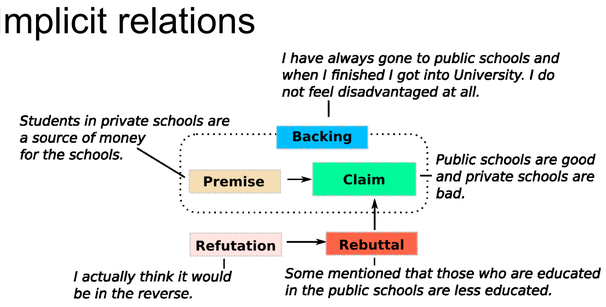
Doc #2823 (article comment, public-private-schools)

Claim *I agree - Kids can do great in the public school system and parents DO need to be involved.* *The more people leave, the worse its going to become.* *The public school system lets them deal with real life too, unfortunate that it may be but that is what's out there in college and the work force too.* *There are still great*
Premise *teachers in the public schools - lets stand behind them.*

Doc #2224 (forumpost, single-sex-education)

Backing *I went to an all boys school* - *Can't say I particularly liked it, I would of much*
Claim *preferred gone to a co-ed.* *It is closer to the 'real world' that way. Kids should*
Premise *grow up in the company of both sexes... They will be more at ease around the opposite sex when they are older and it just makes sense.* *If it is purely education you are concerned about (and not so much behaviour), our year (at a private school) went shockingly bad in OP scores. We were the worst in 12 years and were beaten by LOTS of co-ed and public schools... So you can never tell. In saying that my sister really enjoyed going to an all girls school. Her year went really well too. Ask your daughters what they would prefer...*
Backing *Btw, I work at a co-ed school at the moment and the kids there get on just fine.*

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



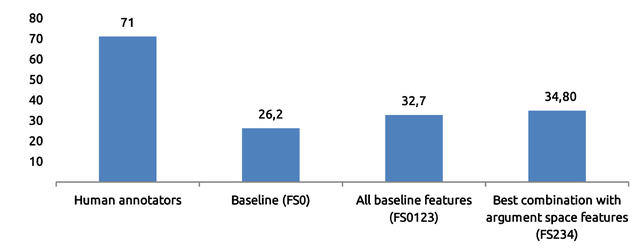
Data representation
- BIO tagging, 11 classes

Doc #2224 (forumpost, single-sex-education)

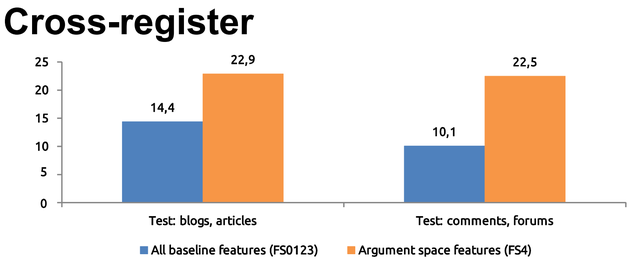
Backing-B *I*	Claim-I *it*
Backing-I *went*	Claim-I *,*
Backing-I *to*	Claim-I *I*
Backing-I *an*	Claim-I *would*
Backing-I *all*	Claim-I *of*
Backing-I *boys*	Claim-I *much*
Backing-I *school*	Claim-I *preferred*
O -	Claim-I *gone*
Claim-B *Can't*	Claim-I *to*
Claim-I *say*	Claim-I *a*
Claim-I *I*	Claim-I *co-ed*
Claim-I *particularly*	Claim-I *.*
Claim-I *liked*	Premise-B *It*
Claim-I *it*	Premise-I *is*
Claim-I *,*	*...*

Results

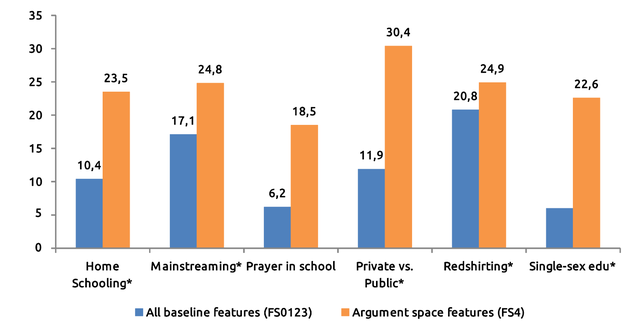
Cross-validation



Cross-register



Cross-domain



Avg. F1 scores for Claim-B, Claim-I, Premise-B, Premise-I, Backing-B, Backing-I, and **O**

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).

VolkswagenStiftung