Which argument is more convincing?

Analyzing and predicting convincingness of Web arguments using bidirectional LSTM

ACL 2016, Berlin

TECHNISCHE UNIVERSITÄT DARMSTADT

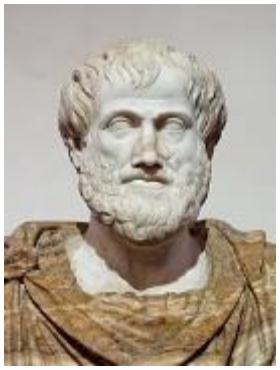
Ivan Habernal

Iryna Gurevych

What makes a “good” argument?

TECHNISCHE UNIVERSITÄT DARMSTADT

- Perceiving argument as “good” one depends on logos, ethos, pathos (Aristotle, 367-322 BC)



- Validity of arguments in Informal Logic (Walton, 1989)

- Different audiences differently (Mercier and Sperber, 2011)

- Relevance, Acceptability, Sufficiency (Johnson and Blair, 2006)

- Aristotle and George Kennedy (translator). 1991. *On Rhetoric: A Theory of Civil Discourse*. Oxford University Press, USA.
- Hugo Mercier and Dan Sperber. 2011. Why do humans reason? Arguments for an argumentative theory. *The Behavioral and Brain Sciences*, 34(2):57–74.
- Douglas N. Walton. 1989. *Informal Logic: A Handbook for Critical Argument*. Cambridge University Press.
- Ralph H. Johnson and Anthony J. Blair. 2006. *Logical Self-Defense*. International Debate Education Association.

What makes a “good” argument?

TECHNISCHE UNIVERSITÄT DARMSTADT

- How to apply for everyday argumentation on the Web?
  - ***Prompt**: Should physical education be mandatory in schools?*
    ***Stance**: Yes!*

> physical education should be mandatory cuhz 112,000 people have died in the year 2011 so far and it’s because of the lack of physical activity and people are becoming obese!!!!

Applying these theories on everyday arguments falls short empirically… (Boudry et al., 2014)

Maarten Boudry, Fabio Paglieri, and Massimo Pigliucci. 2015. The Fake, the Flimsy, and the Fallacious: Demarcating Arguments in Real Life. *Argumentation*, 29(4):431–456.

Purpose of argumentation – to convince

- Take a holistic approach
- How **convincing** is this argument?
  - **Prompt**: Should physical education be mandatory in schools? **Stance**: Yes!

physical education should be mandatory cuhz 112,000 people have died in the year 2011 so far and it's because of the lack of physical activity and people are becoming obese!!!!

YES, because some children don't understand anything excect physical education especially rich children of rich parents.

Let's reformulate the problem…

TECHNISCHE UNIVERSITÄT DARMSTADT

- Which argument is **more convincing**?
  - Prompt: Should physical education be mandatory in schools? Stance: Yes!

physical education should be mandatory cuhz 112,000 people have died in the year 2011 so far and it's because of the lack of physical activity and people are becoming obese!!!!

A1

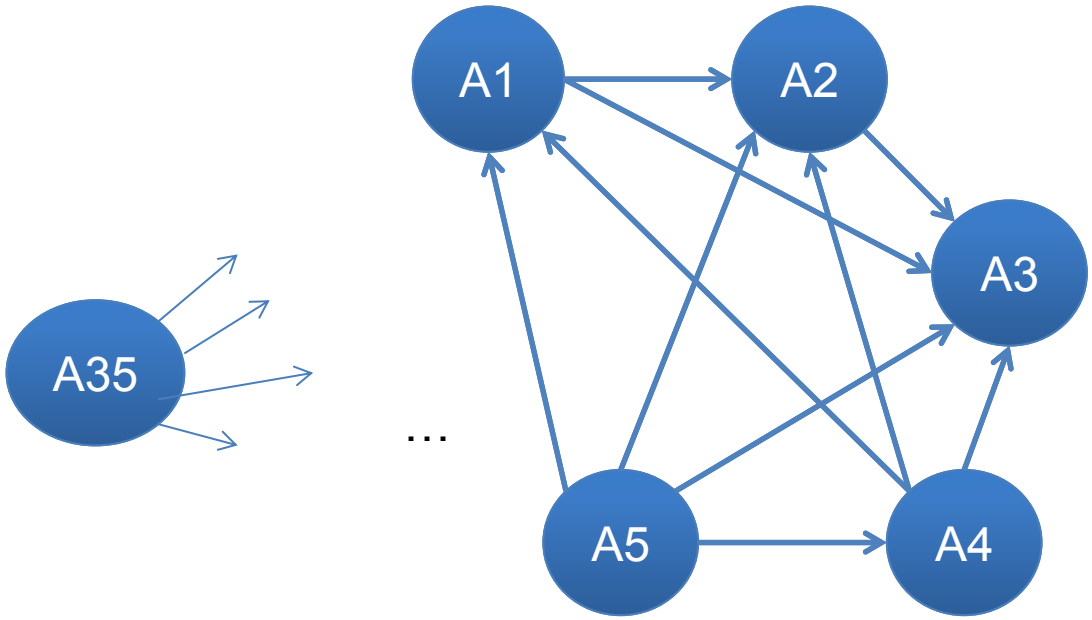
YES, because some children don't understand anything excect physical education especially rich children of rich parents.

A2

- And why?
  - A1>A2, *because* A1 uses statistics, and doesn't make assumptions.
  - A1>A2, *because* A1 talks about the importance of health.
  - A1>A2, *because* A1 provides a health-related argument.
  - A1>A2, *because* A2 is very harsh and attacks
  - A1=A2, *because* Neither A1 or A2 cite evidence to support their claims.

…and scale up…

- Which argument is more convincing?
  - Prompt: Should physical education be mandatory in schools? Stance: Yes

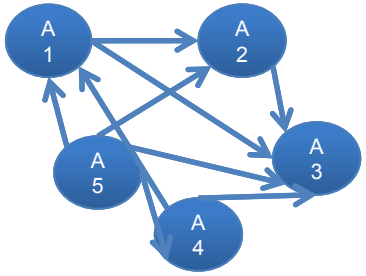


  - … up to ~35 nodes per topic
  - Sources: `createdebate.com` and `procon.org`

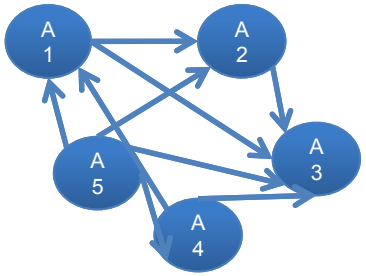
…and scale up

TECHNISCHE UNIVERSITÄT DARMSTADT

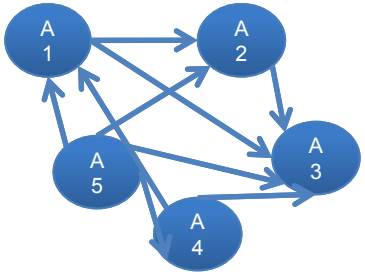
- Topic 1: Ban Plastic Water Bottles? Yes



Topic 2: Should parents use spanking? No



- Topic 3: Is the school uniform a good or bad idea?



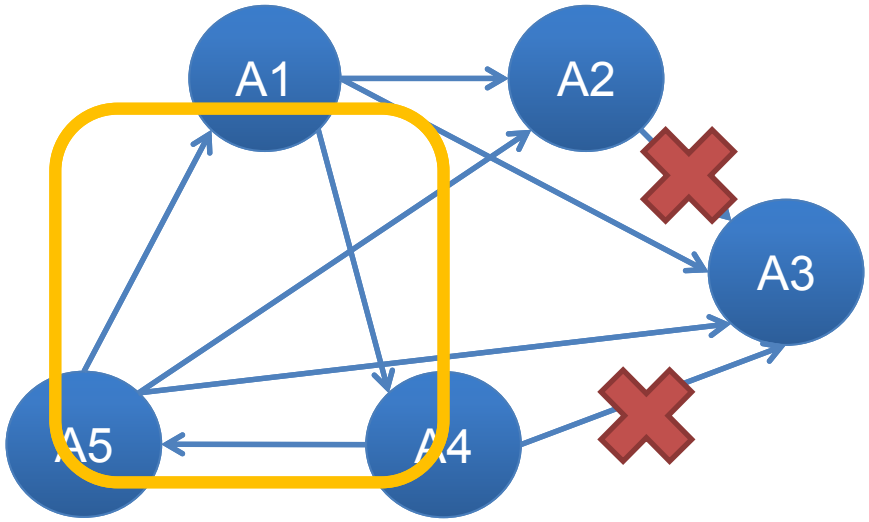
Topic 4:……

- 32 Topics, Total 16,927 pairs, each 5 workers (85k assignments)
- Best MTurk comments:
  - *“Interesting. I hope we're building some kind of cool NLP data-set here?”*
  - *“I am no one to judge but I really hope these [authors] aren't college student. :D”*
  - *“I NEED SOME MONEY PLEASE GOD BLESS”*

Data cleaning

TECHNISCHE UNIVERSITÄT DARMSTADT

- Remove the “worst” 5% using MACE



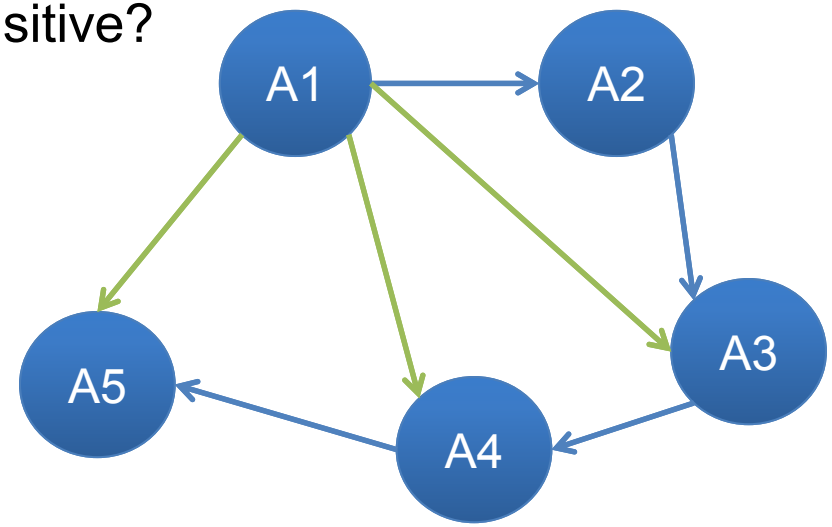
- What about cycles?

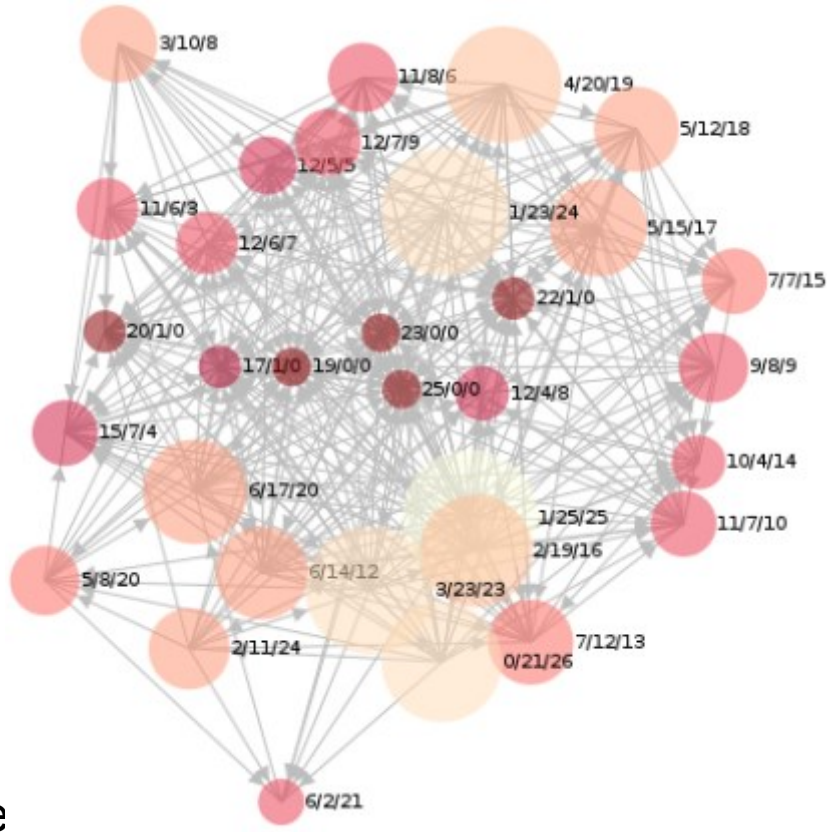
Hovy, D., Berg-Kirkpatrick, T., Vaswani, A., & Hovy, E. (2013). Learning Whom to Trust with MACE. In Proceedings of NAACL-HLT 2013 (pp. 1120–1130). Atlanta, Georgia: Association for Computational Linguistics. http://www.aclweb.org/anthology/N13-1132

Are cycles inevitable?

TECHNISCHE UNIVERSITÄT DARMSTADT

- Is the relation “is more convincing than” transitive?





- Not all edges are equal, if we consider score based on MACE
  - When building the graph taking better edge first, we get much better “transitivity score” (10,8 nodes on average)

Clean data

TECHNISCHE UNIVERSITÄT DARMSTADT

- UKPConvArg1 corpus

Dataset	Size	Instance type	Size per topic	Reasons
UKPConvArgAll	16,081	argument pair	502.5	56,446
UKPConvArgStrict	11,650	argument pair	364.1	44,121
UKPConvArgRank	1,052	argument	32.9	

- Ranking: run PageRank over an argument graph

- Reliability performed on pilot study
  - The best-ranked worker for each argument pair
  - The average global rank of this hypothetical worker was 11 ± 6.6. This rank can be interpreted as a decently performing worker
  - Obtained score reached 0.935 accuracy

Experiment 1 – binary relation prediction

TECHNISCHE UNIVERSITÄT DARMSTADT

- "Which one of these two arguments is more convincing?"
  - UKPConvArg1 data, 11k argument pairs
- Method 1: SVM with RBF kernel
  - Plenty of linguistically motivated features
- Method 2: Bidirectional Long Short-Term Memory Network
  - GloVe embeddings, 64 output neurons from each LSTM, dropout
- Cross-topic evaluation (31 topics training, remaining one test)

Method	Average accuracy
SVM	0.78
BiDi-LSTM	0.76

  - Significantly better p = 0.0414 using two-tailed Wilcoxon signed-rank test

Experiment 2 – argument ranking

TECHNISCHE UNIVERSITÄT DARMSTADT

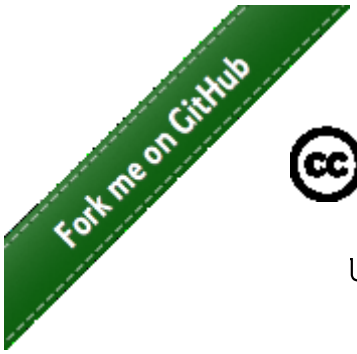
- "Rank the ~35 arguments for each topic with respect to their convincingness"
  - UKPConvArgRank, 1052 arguments
- Method 1: SVM with RBF kernel for regression
- Method 2: Bidirectional Long Short-Term Memory Network
  - GloVe embeddings, 64 output neurons from each LSTM, dropout
  - Output layer: linear activation
- Cross-topic evaluation (31 topics training, remaining one test)

Method	Pearson's r	Spearman's rho
SVM	0.351	0.402
BiDi-LSTM	0.270	0.354

Discussion of results

TECHNISCHE UNIVERSITÄT DARMSTADT

- "Traditional" SVM with rich linguistic features outperforms BLSTM in both tasks, but
  - Employed features require heavy language-specific preprocessing machinery (lemmatizer, POS tagger, parser, NER, sentiment analyzer)
  - By contrast, BLSTM only requires pre-trained embedding vectors, while delivering comparable results
- We only experimented with vanilla LSTMs
  - Recent developments of deep neural networks (especially attention mechanisms or grid-LSTMs) open up many future possibilities to gain performance



Download the data and beat our results!

ub.com/UKPLab/acl2016-convincing-arguments

But what really makes a convincing argument?

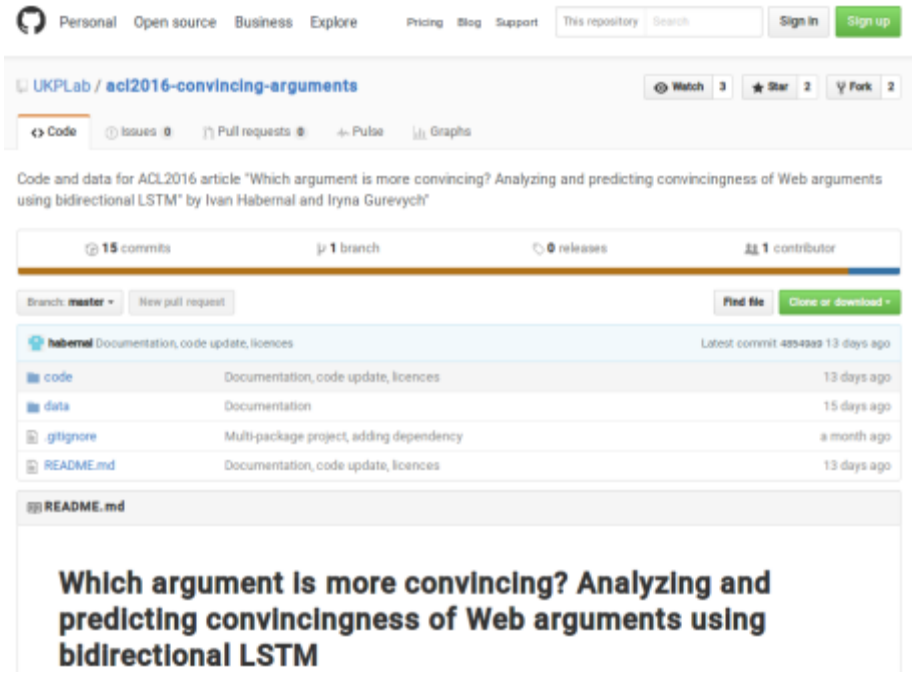
TECHNISCHE UNIVERSITÄT DARMSTADT

- Remember, we have 44,121 free-text explanations written by MTurkers
  - “giving examples or actual reasons”
  - “explaining the reasoning or logical coherence”
  - “confirmation bias”
  - “non-sense”
  - “emotional aspects and rhetorical moves”
  - …
- Follow up – drop by at EMNLP 2016!

Ivan Habernal and Iryna Gurevych (2016). **What makes a convincing argument? Empirical analysis and detecting attributes of convincingness in Web argumentation**. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP), to appear.

Conclusions

- New problem – argument convincingness
- New data – large scale
- New tasks – argument pair evaluation, argument ranking
- State of the art methods – BLSTM, feature-rich SVM



Data formats

XML

Both `UKPConvArg1-full-XML` and `UKPConvArg1Strict-XML` have the same XML format. Here is a commented excerpt from the `evolution-vs-creation_evolution.xml` file.

```
<?xml version="1.0"?>
<list>                                                    <!-- the root element -->
  <annotatedArgumentPair>                                 <!-- is a list of annotated argument pairs -->
    <id>803_794</id>                                      <!-- id of the first and the second argument -->
    <arg1>                                                <!-- first argument with original metadata -->
      <author>http://www.convinceme.net/profile/318/seanohagan.html</author>
      <voteUpCount>0</voteUpCount>
      <voteDownCount>0</voteDownCount>
      <stance>Evolution</stance>
      <text>I have to contradict phro and say that the peppered moths do show evidence of evolution.
       The data may have been insufficient, but evolution did occur.
       When different alleles are expressed due to external factors, this is evolution.</text>
      <id>803</id>
      <!-- some files also have "originalHTML" element with the text in escaped HTML -->
    </arg1>
    <arg2>                                                <!-- second argument with original metadata -->
      <author>http://www.convinceme.net/profile/633/yole136.html</author>
      <voteUpCount>0</voteUpCount>
      <voteDownCount>0</voteDownCount>
      <stance>Evolution</stance>
      <text>You can actually see evolution happen. Fruit Flies are quite useful for this experiment
        since the breed, live, and die so quickly. You have to understand evolution happens because
        of mutations and they survive because those mutations have made it easier for the creature
        to survive then the others. ie. natural selection. Speciation is also an example of
        evolution... different species come about because they have adapted to a slightly different
        enviroment. If you look you can see evolution</text>
      <id>794</id>
```

Thank you!

TECHNISCHE UNIVERSITÄT DARMSTADT

Which argument is more convincing? Analyzing and predicting convincingness of Web arguments using bidirectional LSTM

Ivan Habernal

Iryna Gurevych

Fork me on GitHub

creative commons

thub.com/UKPLab/acl2016-convincing-arguments