

Stance Classification in Argument Search

A presentation about my master's thesis

Philipp Heinisch

phhei@campus.uni-paderborn.de

<https://www.philippeheinisch.de>

Paderborn University – supervisor Henning Wachsmuth

August 12, 2019

Stance Classification in Argument Search

... what an author argues to be true ...

Stance Classification in Argument Search



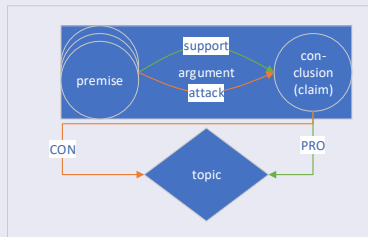
PRO ... *what an author argues to be true* ... **CON**

Stance Classification in Argument Search



PRO ... *what an author argues to be true* ... **CON**

Stance Classification: not trivial task of natural language processing



given a topic + *relevant* claim: determine whether supports (*PRO*) or contests (*CON*) the topic

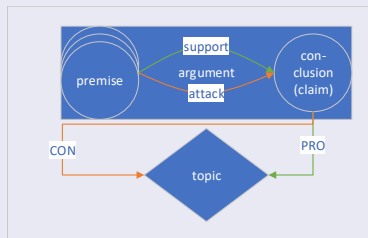
Text → analysis → framework (APACHE UIMA)

disambiguation, negation, terms of language Hasan and Ng (2013), ... // Machine Learning (statistical inference by (big) data) Abu-Mostafa et al. (2017)

Stance Classification in Argument Search

PRO ... *what an author argues to be true* ... **CON**

Stance Classification: not trivial task of natural language processing



given a topic + *relevant* claim: determine whether supports (*PRO*) or contests (*CON*) the topic

Text → analysis → framework (APACHE UIMA)

disambiguation, negation, terms of language Hasan and Ng (2013), ... // Machine Learning (statistical inference by (big) data) Abu-Mostafa et al. (2017)

Example

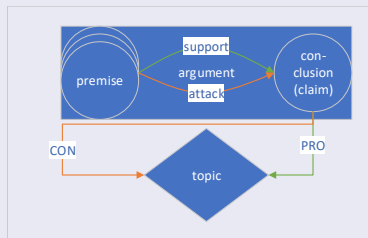
The assumption of God's existence is reasonable with the cosmological argument.

Atheism is wrong.

Stance Classification in Argument Search

PRO ... *what an author argues to be true* ... **CON**

Stance Classification: not trivial task of natural language processing



given a topic + *relevant* claim: determine whether supports (*PRO*) or contests (*CON*) the topic

Text → analysis → framework (APACHE UIMA)

disambiguation, negation, terms of language Hasan and Ng (2013), ... // Machine Learning (statistical inference by (big) data) Abu-Mostafa et al. (2017)


Example

The assumption of God's existence is reasonable with the cosmological argument.

PRO Atheism is wrong.

Stance Classification in Argument Search

Automatic stance classification in Argument Search: Conclusion → Query (Wachsmuth et al. (2017))

 args

→

Page 3 of 1126 arguments, 583 pro, 543 con (retrieved in 2048.5ms) Pro vs. Con View | Topic Space View

Pro

Paddy where is your proof and evidence for u...

<http://www.debate.org/debates/Atheism-has-no-proof-and-Evidence-that-A...>

Paddy where is your proof and evidence for u disproving God is u Paddy say no u got no evidence that u cannot disprove God then u cannot disprove God thus u cannot be a Atheist because u cannot disprove God that is why on one can be a Atheist a Theist is a Theist define what Theist means Paddy to the voters out there Paddy has failed to provide proof and evidence that **atheism** is accurate and correct paddy is not a atheist why because paddy dose not have any proof and evidence that **atheism** is accurate and correct Paddy has shown that on one can be a atheist just by saying I cannot prove that **atheism** is 'accurate and correct' thus **Atheism** is madness **Atheism** being defeated, Checkmate Paddy Checkmate Paddy Checkmate Paddy Checkmate Paddy Checkmate Paddy Checkmate Paddy Checkmate Paddy Checkmate Paddy Checkmate Paddy :)

Supported conclusion: Atheism has no proof and Evidence that Atheism is accurate and correct ▲ score

All bible science is wrong. I see no use in trying...

http://www.forandagainst.com/Creationism_Is_Wrong

All bible science is **wrong**. I see no use in trying to disprove idiotic theories like the earth ... matter for long anyways. **Atheism** is the fastest growing group on earth. We'll have ... ▼ score

Con

When atheists criticise religions for the largely...

<http://www.debatepedia.org/en/index.php/Debate/Atheism>

When atheists criticise religions for the largely negative impact **atheism** has had on society, they are overlooking their own negative impact.

Attacked conclusion: Atheism ▲ score

"Paddy what proof and Evidence do you have that...

<http://www.debate.org/debates/Atheism-has-no-proof-and-Evidence-that-A...>

"Paddy what proof and Evidence do you have that say's Someone can be an atheist, even if **atheism** were to be conclusively disproven." **Atheism** is a position. You can be an atheist whether ... ▼ score

Hello and thank you for accepting! I personally...

<http://www.debate.org/debates/Atheism-is-a-Religion/12/>

Hello and thank you for accepting! I personally find the first definition of **atheism** ("archaic ... atheists would define **atheism** this way. I understand that my opponent simply copied the ... ▼ score

Stance Classification in Argument Search

Automatic stance classification in Argument Search: Conclusion → Query (Wachsmuth et al. (2017))

The screenshot shows the 'args' web interface. At the top, there's a search bar with the query 'Atheism is wrong'. Below the search bar, it indicates 'Page 3 of 1126 arguments, 583 pro, 543 con (retrieved in 2048.5ms)'. The interface is split into two columns: 'Pro' and 'Con'.

Under the 'Pro' column, the first argument is titled 'Paddy where is your proof and evidence for u...' with a URL. The text of the argument is partially visible. Below this argument, there is a blue box containing the text: 'Supported conclusion: Atheism has no proof and Evidence that Atheism is accurate and correct ▲ score'.

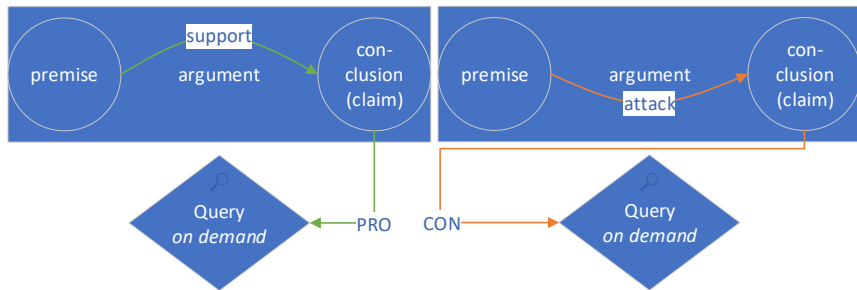
Under the 'Con' column, the first argument is titled 'When atheists criticise religions for the largely...' with a URL. The text of the argument is partially visible. Below this argument, there is a blue box containing the text: 'Attacked conclusion: Atheism ▲ score'.

At the bottom of the 'Pro' column, there is another argument titled 'All bible science is wrong. I see no use in trying...' with a URL. The text of the argument is partially visible.

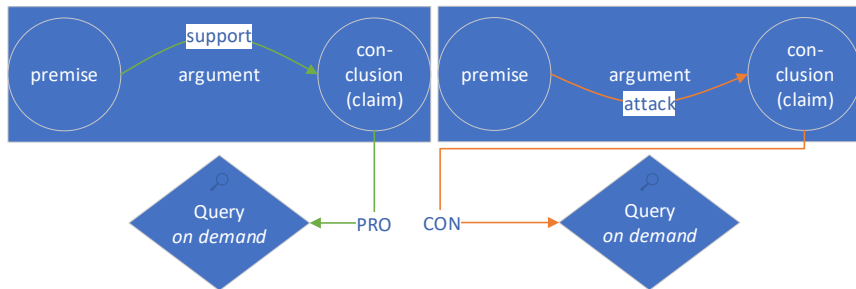
Arguments – evaluate, debate, convince, decision making

“Computational argumentation is expected to play a critical role in the future of web search.” Wachsmuth et al. (2017)

Task

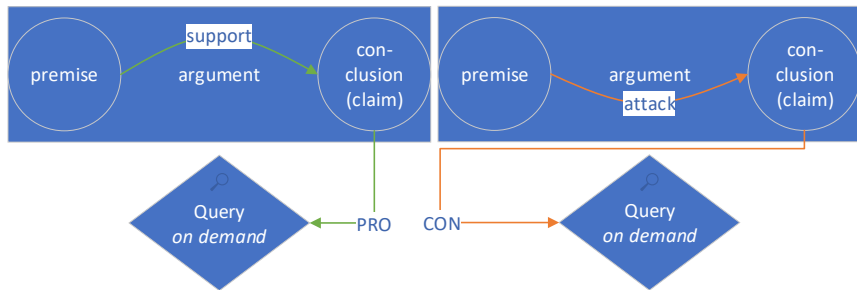


Task



Can we do it better with NLP?

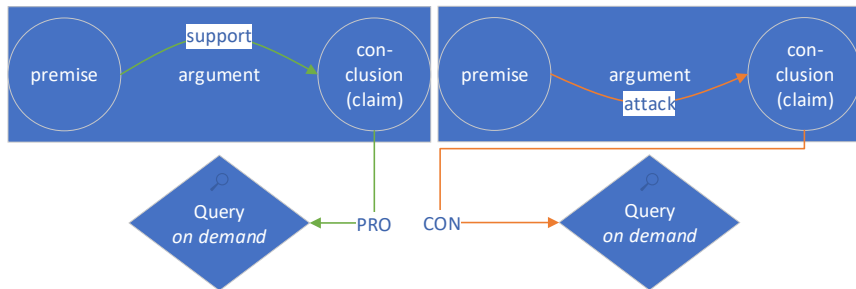
Task



Can we do it better with NLP?

- Investigate a state-of-the-art-approach for stance classification

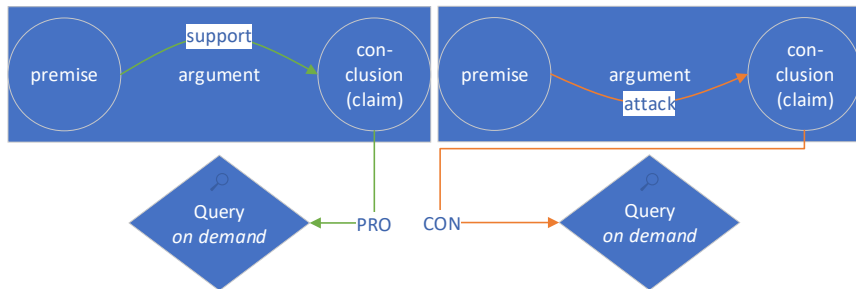
Task



Can we do it better with NLP?

- Investigate a state-of-the-art-approach for stance classification
- Determine the differences to the argument search case

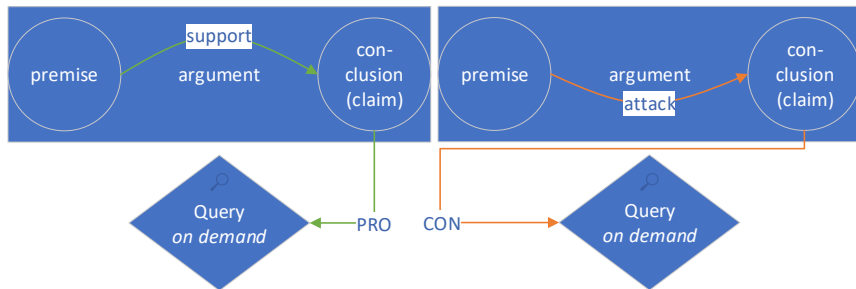
Task



Can we do it better with NLP?

- Investigate a state-of-the-art-approach for stance classification
- Determine the differences to the argument search case
- Adapt the approach, implement

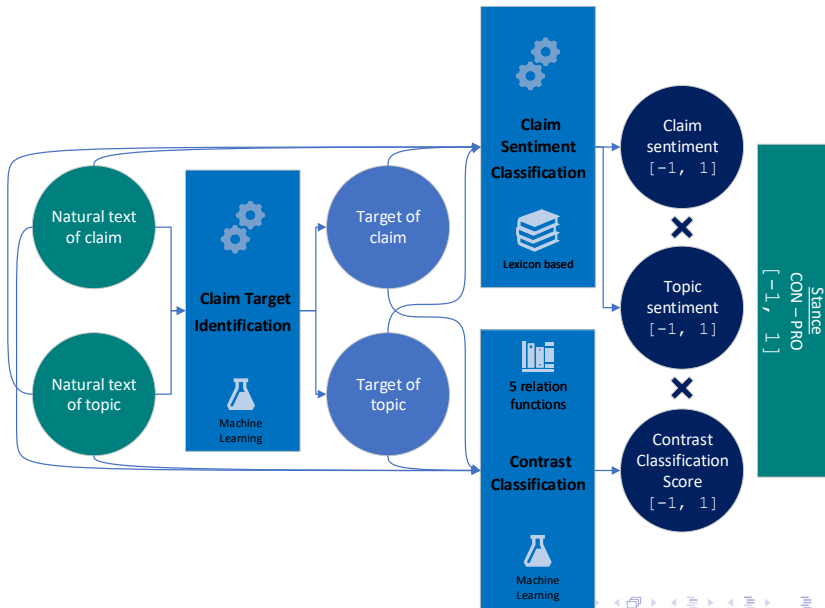
Task



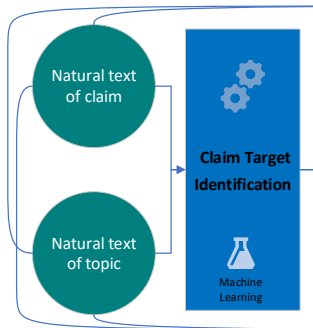
Can we do it better with NLP?

- Investigate a state-of-the-art-approach for stance classification
- Determine the differences to the argument search case
- Adapt the approach, implement
- evaluate

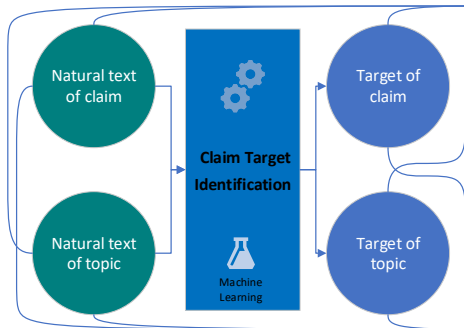
The modular approach of IBM (Bar-Haim et al. (2017a) + Bar-Haim et al. (2017b))



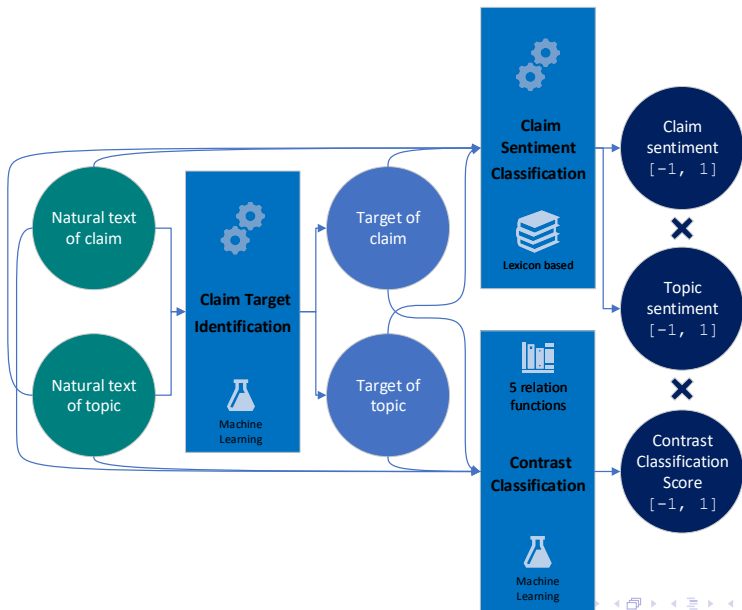
The modular approach of IBM (Bar-Haim et al. (2017a) + Bar-Haim et al. (2017b))



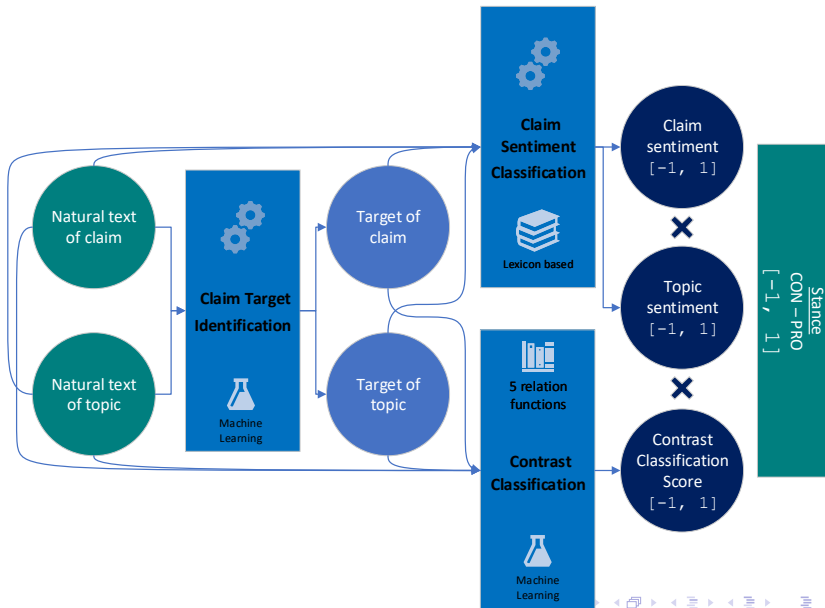
The modular approach of IBM (Bar-Haim et al. (2017a) + Bar-Haim et al. (2017b))



The modular approach of IBM (Bar-Haim et al. (2017a) + Bar-Haim et al. (2017b))



The modular approach of IBM (Bar-Haim et al. (2017a) + Bar-Haim et al. (2017b))



The approach of IBM: example & evaluation

Current step: BOT::input

0 ⇌ PRO

*The assumption of God's existence is
reasonable with the cosmological
argument.*

Atheism is wrong.

The approach of IBM: example & evaluation

Current step: Claim Target Identification

0 ⇌ PRO

*The assumption of **God's existence**
is reasonable with the cosmological
argument.*

Atheism is wrong.

The approach of IBM: example & evaluation

Current step: Claim Sentiment Classification

$$1 \cdot -1 = -1 \not\Rightarrow PRO$$

*The assumption of **God's existence**
is reasonable with the cosmological
argument.*

***Atheism** is wrong.*

The approach of IBM: example & evaluation

Current step: Contrast Classification

$$1 \cdot -1 \cdot -1 = 1 \Leftrightarrow PRO$$

*The assumption of **God's existence**
is reasonable with the cosmological
argument. ⇔*

⇔ **Atheism** is wrong.

The approach of IBM: example & evaluation

Current step: output

$$1 \cdot -1 \cdot -1 \Leftrightarrow PRO$$

*The assumption of **God's existence**
is reasonable with the cosmological
argument. ⇔*

⇔ **Atheism** is wrong.

The approach of IBM: example & evaluation

Current step: output

$$1 \cdot -1 \cdot -1 \Leftrightarrow PRO$$

*The assumption of **God's existence**
is reasonable with the cosmological
argument. ⇔*

⇔ **Atheism** is wrong.

Training example of IBM

Government may not prohibit the expression of an idea simply because society finds the idea offensive or disagreeable.

CON

This house believes that it is sometimes right for the government to restrict freedom of speech.

Algorithm of IBM not freely available

details of algorithms of subtasks (feature types), some used components (e.g. ESG Parser)

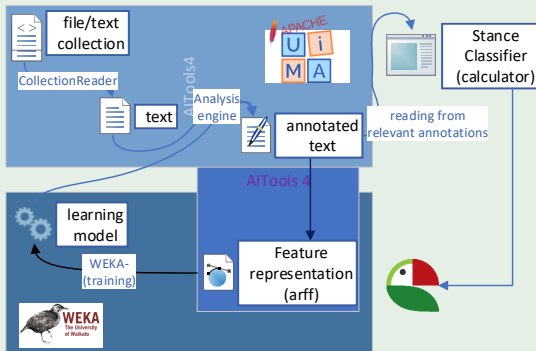
IBM \leftrightarrow Argument Search: implementation (contribution)

Algorithm of IBM not freely available

details of algorithms of subtasks (feature types), some used components (e.g. ESG Parser)

Reimplement and reinvent Ferrucci and Lally (2003) + Frank et al. (2016)

invent efficient
(recursive) algorithm
(e.g. path in WordNet
feature) +
uses freely available
approaches (e.g.
Stanford Parser)



IBM handles a claim2claim-scenario

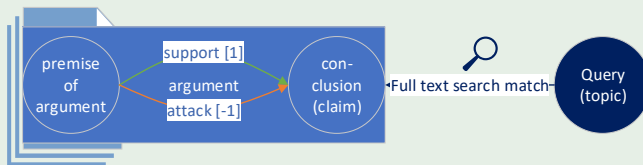
approach for elevated claims with context (Wikipedia)

IBM \leftrightarrow Argument Search: different scenarios (contribution)

IBM handles a claim2claim-scenario

approach for elevated claims with context (Wikipedia)

Argument search is a argument2query-scenario

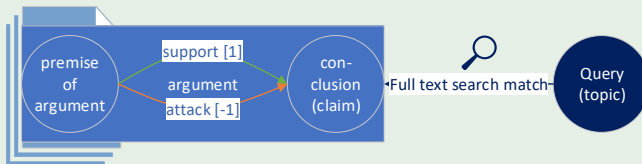


IBM \leftrightarrow Argument Search: different scenarios (contribution)

IBM handles a claim2claim-scenario

approach for elevated claims with context (Wikipedia)

Argument search is a argument2query-scenario



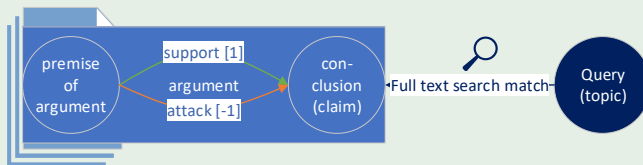
- Query preprocessing: mostly only single words / single phrases (default handler for target identification and sentiment classification)

IBM \leftrightarrow Argument Search: different scenarios (contribution)

IBM handles a claim2claim-scenario

approach for elevated claims with context (Wikipedia)

Argument search is a argument2query-scenario



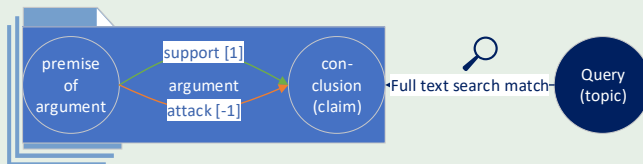
- Query preprocessing: mostly only single words / single phrases (default handler for target identification and sentiment classification)
- Handling spelling and punctuation mistakes

IBM \leftrightarrow Argument Search: different scenarios (contribution)

IBM handles a claim2claim-scenario

approach for elevated claims with context (Wikipedia)

Argument search is a argument2query-scenario



- Query preprocessing: mostly only single words / single phrases (default handler for target identification and sentiment classification)
- Handling spelling and punctuation mistakes
- finally: adapting Stance-formula for argument's premises a :
$$Stance(a, \text{topic}) = S(a, c) \cdot Stance(c, \text{topic})$$

Algorithm of IBM not on-demand?

efficiency decrease because of morphemes extraction, Contrast Classification with many pairs for calculation, ...

Algorithm of IBM not on-demand?

efficiency decrease because of morphemes extraction, Contrast Classification with many pairs for calculation, ...

Efficiency research



research of (pipeline) parts of algorithms:

- effectiveness – efficiency-trade-offs
- suggestions of alternatives (e.g. use of syllables)
- ...

IBM \leftrightarrow Argument Search: creation of sample of args.me (contribution)

- selected 5 top-query-topics like *feminism* Ajjour et al. (2019) + 5 query-topics

IBM \leftrightarrow Argument Search: creation of sample of args.me (contribution)

A Query	B Conclusion	C Supported conclusion (PRO) or attacked conclusion (CON)	D PRO or CON? (manuel rat- ing)
feminism			
1	Feminism is bad	CON	?
2

- selected 5 top-query-topics like *feminism* Ajjour et al. (2019) + 5 query-topics
- for each topic: search for 10 distinct supported conclusions + 10 attacked ($10 \cdot (10 + 10) = 200$ claims)

IBM ↔ Argument Search: creation of sample of args.me

(contribution)

A Query	B Conclusion	C Supported conclusion (PRO) or attacked conclusion (CON)	D PRO or CON? (manuel rating)
feminism			
1	Feminism is bad	CON	<i>PRO</i>
2

- selected 5 top-query-topics like *feminism* Ajjour et al. (2019) + 5 query-topics
- for each topic: search for 10 distinct supported conclusions + 10 attacked ($10 \cdot (10 + 10) = 200$ claims)
- 3 annotators (thanks to Anna Xenia Stephan, Shahbaz Syed and Milad Alshomary)
 - 31 out of 200: disagreement (e.g. stance of “If there is a God, then it is most definitely the Christian God.” towards “God”)
 - Fleiss’ kappa score: 0.79 (substantial agreement)

IBM ↔ Argument Search: creation of sample of args.me

(contribution)

A Query	B Conclusion	C Supported conclusion (PRO) or attacked conclusion (CON)	D PRO or CON? (manuel rating)
feminism			
1	Feminism is bad	CON	<i>PRO</i>
2

- selected 5 top-query-topics like *feminism* Ajjour et al. (2019) + 5 query-topics
- for each topic: search for 10 distinct supported conclusions + 10 attacked ($10 \cdot (10 + 10) = 200$ claims)
- 3 annotators (thanks to Anna Xenia Stephan, Shahbaz Syed and Milad Alshomary)
 - 31 out of 200: disagreement (e.g. stance of “If there is a God, then it is most definitely the Christian God.” towards “God”)
 - Fleiss’ kappa score: 0.79 (substantial agreement)
- → ++annotated corpus

Two corpora



IBM Debater® – Claim Stance Dataset

- train: 25 topics (1,039 claims)
 - test: 30 topics (1,355 claims)
- of Wikipedia



Sample of args.me

- train: –
 - test: 20 topics (200 claims)
- of args.me (387,606 arguments)

Experimental setup

Two corpora



IBM Debater® – Claim Stance Dataset

- train: 25 topics (1,039 claims)
 - test: 30 topics (1,355 claims)
- of Wikipedia



Sample of args.me

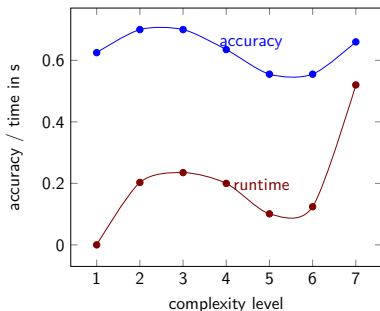
- train: –
 - test: 20 topics (200 claims)
- of args.me (387,606 arguments)

Final setup

- Training + hyperparameter-tuning with IBM corpus (for each module)
- implementing an evaluation class including parameters for activating/deactivating the modules and datasets
- tested various configurations with the two corpora

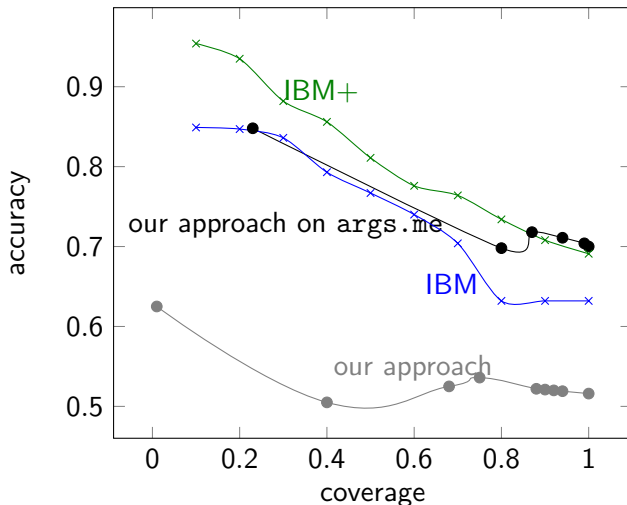
Experiment results: do not to use all available features and subcomponents.

effectiveness (accuracy) + efficiency
(average time per stance prediction):



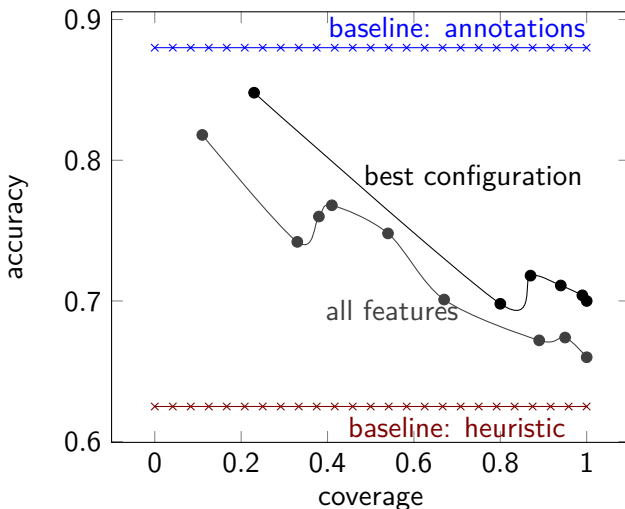
IBM	args.me	time
heuristic		
52%	63%	1ms
best (reduced target identification (e.g. no Wikipedia), $\mathcal{R}(x_c, x_t) = 1$) [only 2% are contrastive in args.me]		
52%	70%	200-250ms
all but TT4J-PhraseChunker		
49%	56%	100-160ms
all		
50%	66%	500-650ms

Experiment results: best configuration vs. IBM



Evaluation with continuous model: $|Stance(claim, topic)| \geq \Theta$

Experiment results: outperform the current heuristic in args.me



Conclusion

Our approach increases significantly accuracy in `args.me` $\alpha \approx 0.025$



modular NLP-approach (target identification, sentiment classification, contrast classification) fits for Argument Search, too (adapted open domain): + $\approx 8\%$ accuracy

Conclusion

Our approach increases significantly accuracy in `args.me` $\alpha \approx 0.025$



modular NLP-approach (target identification, sentiment classification, contrast classification) fits for Argument Search, too (adapted open domain): + $\approx 8\%$ accuracy

Effectiveness-Efficiency-problem: not more than 1s per 20 stance classifications? (Brutlag et al. (2008))



efficient algorithms, cache-solutions + pre-basic-annotations for the conclusions // parallelism \Rightarrow runtime our approach is not critical

Conclusion

Our approach increases significantly accuracy in `args.me` $\alpha \approx 0.025$



modular NLP-approach (target identification, sentiment classification, contrast classification) fits for Argument Search, too (adapted open domain): + $\approx 8\%$ accuracy

Effectiveness-Efficiency-problem: not more than 1s per 20 stance classifications? (Brutlag et al. (2008))



efficient algorithms, cache-solutions + pre-basic-annotations for the conclusions // parallelism \Rightarrow runtime our approach is not critical

Stance Classification – improvements with further work!



stance classification as known problem – further work can be done: including contextual features like author constraints, expanding (sentiment) lexicons, ...

References I

- Yaser S. Abu-Mostafa, Malik Magdon-Ismael, and Hsuan-Tien Lin. 2017. *Learning from data*. AMLBook.
- Yamen Ajjour, Henning Wachsmuth, Johannes Kiesel, Martin Potthast, Matthias Hagen, and Benno Stein. 2019. Data Acquisition for Argument Search: The args.me Corpus. In *KI 2019*, Kassel, Germany.
- Roy Bar-Haim, Indrajit Bhattacharya, Francesco Dinuzzo, Amrita Saha, and Noam Slonim. 2017a. Stance Classification of Context-Dependent Claims. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 251–261, Valencia, Spain. IBM, Association for Computational Linguistics.
- Roy Bar-Haim, Lilach Edelstein, Charles Jochim, and Noam Slonim. 2017b. Improving Claim Stance Classification with Lexical Knowledge Expansion and Context Utilization . In *Proceedings of the 4th Workshop on Argument Mining*. Association for Computational Linguistics.

References II

- Jake D. Brutlag, Hilary Hutchinson, and Maria Stone. 2008. User Preference and Search Engine Latency. resreport, Google, Inc.
- David Ferrucci and Adam Lally. 2003. Accelerating corporate research in the development, application and deployment of human language technologies. In *Proceedings of the HLT-NAACL 2003 workshop on Software engineering and architecture of language technology systems*. Association for Computational Linguistics.
- Eibe Frank, Mark A. Hall, and Ian H. Witten. 2016. *Data Mining: Practical Machine Learning Tools and Techniques*, fourth edition, chapter The WEKA Workbench. Morgan Kaufmann. Online Appendix.
- Kazi Saidul Hasan and Vincent Ng. 2013. Stance Classification of Ideological Debates: Data, Models, Features, and Constraints. In *International Joint Conference on Natural Language Processing*, pages 1348–1356, Richardson, TX 75083-0688. Human Language Technology Research Institute, University of Texas at Dallas.

Henning Wachsmuth, Martin Potthast, Khalid Al Khatib, Yamen Ajjour, Jana Puschmann, Jiani Qu, Jonas Dorsch, Viorel Morari, Janek Bevendorff, and Benno Stein. 2017. Building an Argument Search Engine for the Web. In *Proceedings of the 4th Workshop on Argument Mining*. Association for Computational Linguistics.