December 7, 2023



Jan Heinrich Reimer



Alexander Bondarenko



Maik Fröbe



Matthias Hagen

Friedrich-Schiller-Universität Jena

https://webis.de



10th Workshop on Argument Mining (ArgMining 2023)

Stance-Aware Re-Ranking for Non-factual Comparative Queries Factual comparisons



Source: https://youtube.com/watch?v=Cfkqp87ZIIc

Which tower is taller, the Eiffel Tower or the Big Ben?

Stance-Aware Re-Ranking for Non-factual Comparative Queries Factual comparisons



Source: https://youtube.com/watch?v=Cfkqp87ZIIc

Which tower is taller, the Eiffel Tower or the Big Ben?

→ comparison based on facts, no arguments required

Non-factual comparisons



Source: https://netivist.org/debate/paris-vs-london

Which city is better, Paris or London?

Non-factual comparisons



Source: https://netivist.org/debate/paris-vs-london

Which city is better, Paris or London?

choice depends on personal opinions, arguments required

Stance-Aware Re-Ranking for Non-factual Comparative Queries Argument retrieval

Existing argument search engines

- Examples: args.me [Wachsmuth et al. EMNLP'17] Or ArgumenText [Stab et al. NAACL-HLT'18]
- Goal: Relevance to query, presence/quality of arguments
- Often focus on controversial topics, e.g., 'nuclear energy'
- No search engine focused on comparative topics \rightarrow

Touché shared tasks [Bondarenko et al. CLEF'22]

- Queries compares two options (e.g., Paris vs. London)
- Goal: Arguments for/against either option
- Approaches: BM25, neural (re-)ranking, arg. guality
- → Stance not exploited by Touché participants

| 🗨 args | | | | | | | |
|--|---|--|--|--|--|--|--|
| paris | Q | | | | | | |
| All Discussions News People Pro | vs. con view - 73 arguments retrieved in 1.0 ms | | | | | | |
| PRO | CON | | | | | | |
| Resolved: on balance, the Paris attacks were justified | Even if you do want to spread your religion, you can do | | | | | | |
| ► Show full argument Rosolved: on balance, the Paris attacks were justified Don't accept if you're TheHitchslag You argue first round https://www.debate.org/debates/The-Paris- attacks-were-justified1/ | Show full argument Even if you do want to spread your religion, you can do so peacefully What their governments did was justified, as ISIS is an evil group made up of blood thirsty murderers who throw gays off of buildings, rape women and kill | | | | | | |
| Counter-Claim: Due to the fact that we have not had | https://www.debate.org/debates/The-Paris- attacks-were-justified/1/ score + | | | | | | |
| Show full argument Counter-Claim: Due to the fact that we have not had enough time to examine the effects of the Paris Climate Accord we cannot make this assumption Warrant: | You should not kill innocent people because they live in Show full argument You should not kill innocent people | | | | | | |

s://www.debate.org/debates/ls-Paris will be saving my argument for ound 2 First: L

Claim: Everyone will be happy

All Disc

its not just european nationalism thats is the effect it.

mment that does things you don ... You can not deny them of their



Stance-Aware Re-Ranking for Non-factual Comparative Queries Argument retrieval

Existing argument search engines

- Examples: args.me [Wachsmuth et al. EMNLP'17] Or ArgumenText [Stab et al. NAACL-HLT'18]
- □ Goal: Relevance to query, presence/quality of arguments
- □ Often focus on controversial topics, e.g., 'nuclear energy'
- → No search engine focused on comparative topics

Touché shared tasks [Bondarenko et al. CLEF'22]

- □ Queries compares two options (e.g., Paris vs. London)
- □ Goal: Arguments for/against either option
- □ Approaches: BM25, neural (re-)ranking, arg. quality
- → Stance not exploited by Touché participants

| 😩 args | | | | | | |
|---|--|--|--|--|--|--|
| 3 | Q | | | | | |
| ussions News People Prove | s. con view + 73 arguments retrieved in 1.0 ms | | | | | |
| | CON | | | | | |
| ed: on balance, the Paris were justified If on balance, the Paris attacks on balance, the Paris attacks the state of the state of the state were balance of the state of the state were balance of the state were balance of the state of the state of the state | Even if you do want to spread your religion, you can do > Store till argument Even if you do want to spread your religion, you can do so peacefully What their governments di was justifies at SIS's an evil group made up of blood sublidings, rape warren and kill troscivenes detain considerates the Peri- etables arene autified?! | | | | | |
| Claim: Due to the fact that we Claim: Due to the fact that we the Paris Climate Accord we nake this assumption Warrant: my opponent's claim Opponent veryone will be happy wedebate org/debates/it-Paris- action/hations/// score * | You should not kill innocent people because they live in > Show lul argument You should not kill innocent people because they live in a country with a government that does things you don't like You can not deny them of their lives and happings because of that https://www.debata.org/schatts/The Paris- tathata-weep jurited/12_score * | | | | | |

attack

that we

Claim: E

ts not just european nationalism hats is the effect it...



Re-ranking based on the stance

Stance for comparative questions (e.g. 'Which city is better, Paris or London?')



pro London I like the popular West End musicals.



pro Paris To me, Paris is the most romantic town in the world.



both equal My favorite art museums

are in London and Paris.



no stance

London has 9M inhabitants, and Paris has 10M.

Our approach

- Intuition: Helpful results should express a stance towards the options
- □ Idea: Move results without stance to the end of the ranking
- □ Evaluation: Re-rank top-5, measure retrieval effectiveness (nDCG@5)

| Rank | 1 | 2 | 3 | 4 | 5 |
|----------------|--------|---|--------|------|---|
| Original run | FACT | | FACT | | Paus Contraction of the second |
| Our re-ranking | LONDON | | Hans A | FACT | FACT |

Re-ranking based on the stance

Stance for comparative questions (e.g. 'Which city is better, Paris or London?')



pro London I like the popular West End musicals.



pro Paris To me, Paris is the most romantic town in the world.

My favorite art museums are in London and Paris.

both equal



no stance

London has 9M inhabitants, and Paris has 10M.

Our approach

- Intuition: Helpful results should express a stance towards the options
- □ Idea: Move results without stance to the end of the ranking
- □ Evaluation: Re-rank top-5, measure retrieval effectiveness (nDCG@5)

| Rank | 1 | 2 | 3 | 4 | 5 |
|----------------|--------|---|------------------|------|--|
| Original run | FACT | | FACT | | Pauls Control of the second se |
| Our re-ranking | LONDON | | Hais Contraction | FACT | FACT |

Stance-Aware Re-Ranking for Non-factual Comparative Queries Re-ranking scenarios

- 1. "Perfect" stance labels
 - □ Touché ground truth labels (artificial scenario)
 - □ Results: Significantly improved nDCG@5 for all runs, close to optimal re-rank.
- 2. Predicted stance (Touché participants)
 - Use Touché participants' stance predictions (real-world scenario)
 - □ Results: No improvements, stance detection not effective enough ($F_1 \le 0.31$)
- 3. Improved stance detection (ours)
 - **D** RoBERTa (F_1 =0.34), Flan-T5 (zero-shot, F_1 =0.39), <u>GPT-3.5</u> (4-shot, F_1 =0.49)
 - □ Results: improved nDCG@5, top-3 runs would reach rank 1 in leaderboard
- 4. Simulated stance detection (perturbed ground truth)
 - **Stopping thresholds:** $F_1 = 0.95, 0.90, \dots, 0.20$
 - □ Results: higher nDCG@5 improv. with higher F_1 , top runs "harder" to improve (higher F_1)



- □ Simple, yet effective re-ranking approach for non-factual comparative queries
- □ Significantly improves all Touché 2022 runs with "perfect" stance labels
- □ Still changes the top-3 system order with our GPT-3.5 stance detection
- □ Future work: improving stance detection, separate result lists per stance

Code and Data

- ✿ github.com/webis-de/ArgMining-23
- Contributions are welcome!

- □ Simple, yet effective re-ranking approach for non-factual comparative queries
- □ Significantly improves all Touché 2022 runs with "perfect" stance labels
- □ Still changes the top-3 system order with our GPT-3.5 stance detection
- □ Future work: improving stance detection, separate result lists per stance

Code and Data

- ✿ github.com/webis-de/ArgMining-23
- Contributions are welcome!



© Reimer, Bondarenko, Fröbe, Hagen 2023