Theory-based Argument Quality for Advanced Argument Retrieval
Opportunities and Challenges
Anne Lauscher @Touché 2021
Argument Retrieval
Argument Retrieval
Argument Retrieval

Why should we allow gay marriage?
Show me arguments!
Argument Retrieval

Why should we allow gay marriage?
Show me good arguments!
Argument Retrieval

Argument retrieval should be guided by the quality of arguments!

Why should we allow gay marriage? Show me good arguments!
"Our human assessors will label the retrieved documents manually, both for their general topical relevance and for the rhetorical quality, i.e., "well-writtenness" of the document:

1. whether a document contains arguments (…), and whether the text has a good style of speech (…);
2. whether the text has a proper sentence structure and is easy to read and follow, whether it can be well understood;
3. whether it includes profanity, has typos, and makes use of other detrimental style choices."
“Our human assessors will label the retrieved documents manually, both for their general topical relevance and for the rhetorical quality, i.e., "well-writtenness" of the document:

(1) whether a document contains arguments (… ) and whether the text has a good style of speech (… ),

(2) whether the text has a proper sentence structure and is easy to read and follow, whether it can be well understood,

(3) whether it includes profanity, has typos, and makes use of other detrimental style choices.”
“Our human assessors will label the retrieved documents manually, both for their general topical relevance and for the rhetorical quality, i.e., "well-writtenness" of the document:
(1) whether a document contains arguments (…) and whether the text has a good style of speech (…),
(2) whether the text has a proper sentence structure and is easy to read and follow, whether it can be well understood,
(3) whether it includes profanity, has typos, and makes use of other detrimental style choices.”
“Our human assessors will label the retrieved documents manually, both for their general topical relevance and for the rhetorical quality, i.e., "well-writtenness" of the document:

1. Whether a document contains arguments (…) and whether the text has a good style of speech (…),
2. Whether the text has a proper sentence structure and is easy to read and follow, whether it can be well understood,
3. Whether it includes profanity, has typos, and makes use of other detrimental style choices.”
“Our human assessors will label the retrieved documents manually, both for their general topical relevance and for the rhetorical quality, i.e., "well-writtenness" of the document:
(1) whether a document contains arguments (…) and whether the text has a good style of speech (…),
(2) whether the text has a proper sentence structure and is easy to read and follow, whether it can be well understood,
(3) whether it includes profanity, has typos, and makes use of other detrimental style choices.”
What is this? And what can we gain?
Theory-based Argument Quality
(Wachsmuth et al., 2017a)
Theory-based Argument Quality
(Wachsmuth et al., 2017a)
Theory-based Argument Quality
(Wachsmuth et al., 2017a)
Theory-based Argument Quality
(Wachsmuth et al., 2017a)
Theory-based Argument Quality
(Wachsmuth et al., 2017a)
Theory-based Argument Quality

Should we allow gay marriage?
Theory-based Argument Quality

Should we allow gay marriage?

No, *$%§"$%&&&!*****
Theory-based Argument Quality

Should we allow gay marriage?

No, *$%^$%&!!!
Theory-based Argument Quality

Should we allow gay marriage?

Yes, <claim>, because

<premise>
<premise>
<premise>
Theory-based Argument Quality

Should we allow gay marriage?

Yes, <claim>, because

<premise>

<premise>
Opportunities for Argument Retrieval
Opportunities for Argument Retrieval

Holistic assessment of "good" arguments based on a theoretically-grounded notion
Opportunities for Argument Retrieval

- Holistic assessment of “good” arguments based on a theoretically-grounded notion
- More interpretable
Opportunities for Argument Retrieval

- Holistic assessment of “good” arguments based on a theoretically-grounded notion
- More interpretable
- Potential for more focused and targeted, e.g., user-specific, retrieval
Where do we stand?
Taxonomy of theory-based AQ
(Wachsmuth et al., 2017a,b)

- assessment is challenging
- crowdsourcing is possible
- guidelines and task need to be simplified
Taxonomy of theory-based AQ
(Wachsmuth et al., 2017a,b)

- assessment is challenging
- crowdsourcing is possible
- guidelines and task need to be simplified

But, until last year,

- No large corpus
- No computational model
Grammarly Argument Quality Corpus (GAQCorpus)

First multi-domain corpus and largest English corpus annotated with theory-based Argument Quality scores


GAQCorpus
Simplifications

● Reduction of the taxonomy
  ○ Keep overall Argument Quality and three higher-level dimensions
  ○ Translate lower-level aspects to guiding questions

● Instruction Modifications

● 5-point scale
GAQCorpus
Annotation Process

- Guideline development with four expert annotators
  - Fluent or native in English
  - Advanced degree in Linguistics
- Pilot studies with crowd and experts
- Web Interface
GAQCorpus
Validation of our Simplifications

- 200 randomly sampled arguments and **gold annotations** from Dagstuhl-ArgQuality-Corpus-V2 (Wachsmuth et al., 2017a)
- Crowd-sourced annotations from (Wachsmuth et al., 2017b, TvsP)
GAQCorpus

Validation of our Simplifications

- 200 randomly sampled arguments and gold annotations from Dagstuhl-ArgQuality-Corpus-V2 (Wachsmuth et al., 2017a)
- Crowd-sourced annotations from (Wachsmuth et al., 2017b, TvsP)
- Crowd-sourced annotations with our simplified guidelines (Ours)
GAQCorpus
Validation of our Simplifications

- 200 randomly sampled arguments and gold annotations from Dagstuhl-ArgQuality-Corpus-V2 (Wachsmuth et al., 2017a)
- Crowd-sourced annotations from (Wachsmuth et al., 2017b, TvsP)
- Crowd-sourced annotations with our simplified guidelines (Ours)

<table>
<thead>
<tr>
<th></th>
<th>Cogency</th>
<th>Effectiveness</th>
<th>Reasonableness</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ours</td>
<td>0.46</td>
<td>0.48</td>
<td>0.48</td>
<td>0.55</td>
</tr>
<tr>
<td>TvsP</td>
<td>0.27</td>
<td>0.38</td>
<td>0.13</td>
<td>0.43</td>
</tr>
</tbody>
</table>

Krippendorff’s alpha between expert and crowd annotations
GAQCorpus Data

- Debate forums (Debates)
  - Convince Me
  - Change My View
- Community Q&A forums (CQA)
  - Yahoo Answers: Law & Ethics
- Review forums (Reviews)
  - Yelp restaurant reviews
GAQCorpus
Results

- Total arguments: 5,285
- Three domains
- Portions:
  - Crowd annotations only (10 votes)
  - Expert annotations (1-3 votes)
  - Overlapping portions with expert and crowd annotations
GAQCorpus Results

- Total arguments: 5,285
- Three domains
- Portions:
  - Crowd annotations only (10 votes)
  - Expert annotations (1-3 votes)
  - Overlapping portions with expert and crowd annotations

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Train</th>
<th>Dev</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>CQA</td>
<td>2,085</td>
<td>1,109</td>
<td>476</td>
<td>500</td>
</tr>
<tr>
<td>Debates</td>
<td>2,100</td>
<td>1,093</td>
<td>469</td>
<td>538</td>
</tr>
<tr>
<td>Reviews</td>
<td>1,100</td>
<td>700</td>
<td>300</td>
<td>100</td>
</tr>
<tr>
<td>All</td>
<td>5,285</td>
<td>2,902</td>
<td>1,245</td>
<td>1,138</td>
</tr>
</tbody>
</table>
Experiments

Models

Argument Length
SVR with lexical features
SVR with semantic features
Feature-rich SVR (Wachsmuth et al., 2016)
Single Task Learning w. BERT (BERT ST)
Experiments
Models

Argument Length
SVR with lexical features
SVR with semantic features
Feature-rich SVR (Wachsmuth et al., 2016)
Single Task Learning w. BERT (BERT ST)
Experiments

Results

Pearson correlations of our model predictions with the annotation scores for the Community Q&A domain on our newly created GAQcorpus.
Experiments

Results

Pearson correlations of our model predictions with the annotation scores for the Community Q&A domain on our newly created GAQcorpus.
Experiments

Models

Argument Length
SVR with lexical features
SVR with semantic features
Feature-rich SVR
Single Task Learning w. BERT (BERT ST)
Experiments
Models

Argument Length
SVR with lexical features
SVR with semantic features
Feature-rich SVR
Single Task Learning w. BERT (BERT ST)

Flat MT Learning (BERT MT Flat)
Experiments

Models

Argument Length
SVR with lexical features
SVR with semantic features
Feature-rich SVR
Single Task Learning w. BERT (BERT ST)

Flat MT Learning (BERT MT Flat)

Hierarchical MT Learning (BERT MT Hier)
Concatenation of the hidden document representation with the predictions for the lower-level dimensions for predicting overall AQ
Experiments

Results

Pearson correlations of our model predictions with the annotation scores for the Community Q&A domain on our newly created GAQcorpus.

Overall, the multi-task models outperform the single task model in 9 out of 12 cases.
What challenges are we facing?
Challenge 1: Resources

A first larger scale corpus is available
Challenge 1: Resources

A first larger scale corpus is available

- only covers the four higher-level dimensions (overall AQ, cogency, effectiveness, reasonableness)
- only covers three domains
- only covers English
Challenge 1: Resources

A first larger scale corpus is available

- only covers the four higher-level dimensions
  (overall AQ, cogency, effectiveness, reasonableness)
- only covers three domains
- only covers English

We need to consider more domains and languages. And what about an even finer-grained assessment?
Challenge 2: Advanced Knowledge

Intrinsically evaluating the quality makes sense, but there is much more to it ...
Challenge 2: Advanced Knowledge

Intrinsically evaluating the quality makes sense, but there is much more to it ...

- Knowledge about the cultural background
- Knowledge about the audience
- Knowledge about the speaker
- Commonsense knowledge & world knowledge
The Role of Knowledge in Argumentation

The Role of Knowledge in Argumentation

The Role of Knowledge in Argumentation

The Role of Knowledge in Argumentation

The Role of Knowledge in Argumentation

The Role of Knowledge in Argumentation

The Role of Knowledge in Argumentation

The Role of Knowledge in Argumentation


Argumentation knowledge pyramid

useful for ...

- many NLU task, e.g., common sense knowledge, world knowledge
- almost all NLP models, e.g., traditional linguistic features, word embeddings

Linguistic knowledge

Argumentation-specific knowledge

General knowledge

Task-specific knowledge
The Role of Knowledge in Argumentation

The Role of Knowledge in Argumentation

The Role of Knowledge in Argumentation

The Role of Knowledge in Argumentation

Challenge 2: Advanced Knowledge

Intrinsically evaluating the Quality makes sense, but there is much more to it ...

- Knowledge about the cultural background
- Knowledge about the audience
- Knowledge about the speaker
- Commonsense knowledge & world knowledge

Where to obtain the information needed?
How to model it?
Challenge 3: Ethics

If argument retrieval influences people’s opinions ...

... and argument retrieval is guided by argument quality

... and we also have evidence that argumentative corpora are biased

(Spliethover and Wachsmuth, 2020)
Challenge 3: Ethics

If argument retrieval influences people’s opinions …

… and argument retrieval is guided by argument quality

… and we also have evidence that argumentative corpora are biased (Spliethover and Wachsmuth, 2020)

How can we ensure that the models’ assessments are not unfairly biased?
Theory-based AQ for Advanced Argument Retrieval

Our Opportunities

Our Position

Our Challenges
Theory-based AQ for Advanced Argument Retrieval

Our Opportunities

- Holistic perspective
- Theoretically grounded
- Improved interpretability
- More focused/targeted retrieval

Our Position


Our Challenges


Theory-based AQ for Advanced Argument Retrieval

**Our Opportunities**
- Holistic perspective
- Theoretically-grounded
- Improved interpretability
- More focused/targeted retrieval

**Our Position**
- Data and computational models for three domains
- Intrinsic assessment

**Our Challenges**
Theory-based AQ for Advanced Argument Retrieval

Our Opportunities
- Holistic perspective
- Theoretically-grounded
- Improved interpretability
- More focused/targeted retrieval

Our Position
- Data and computational models for three domains
- Intrinsic assessment

Our Challenges
- Resources for more domains and languages, finer-grained assessment?
- Advanced knowledge?
- Ethical aspects, bias?
Theory-based AQ for Advanced Argument Retrieval

Our Opportunities
- Holistic perspective
- Theoretically-grounded
- Improved interpretability
- More focused/targeted retrieval

Our Position
- Data and computational models for three domains
- Intrinsic assessment

Our Challenges
- Resources for more domains and languages, finer-grained assessment?
- Advanced knowledge?
- Ethical aspects, bias?

Thank you very much for your interest!
References


