Argumentation Technology for Artificial Intelligence
Part 3: Argument Retrieval

Philipp Cimiano, Benno Stein, and Henning Wachsmuth
September 24th, 2019
<table>
<thead>
<tr>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Argument Retrieval Problems</td>
</tr>
<tr>
<td>3.2 Argument Ranking</td>
</tr>
<tr>
<td>3.3 Resources</td>
</tr>
<tr>
<td>3.4 Argument Search Engines</td>
</tr>
<tr>
<td>3.5 Shared Tasks</td>
</tr>
</tbody>
</table>
Basic Argument Model

**Conclusion**  
*Mankind will be able to travel to other galaxies.*

**Premise 1**  
*Photon drives can take you up to relativistic velocities.*

**Premise 2**  
*In August 2019 Lightsail2 demonstrated its functioning.*

**Premise 3**  
*NASA announces progress on torpor (human hibernation).*

**Argument:**
- A conclusion (claim) supported by premises (reasons).  [Walton et al. 2008]
  Conclusion and premises are considered as propositions.

- Conveys a stance on a controversial topic.  [Freeley and Steinberg, 2009]
  Assignment of truth values to the propositions:
  \( I(“Mankind will be able to travel to other galaxies.”) = 1, \)
  \( I(“Photon \ldots”) = 1, \ldots \)

- The mechanism ("calculus", "argumentation type") to obtain ("derive") the conclusion from the premises is let implicit and is usually informal.
Basic Argument Model

**Thesis / Major claim** \( t \)  Human beings will colonize other planets.

\[
A_{pro} = \left\{ \begin{array}{l}
    c_1 \quad \text{Mankind will be able to travel to other galaxies.} \\
    p_1 \quad \text{Photon drives can take you up to relativistic velocities.} \\
    p_2 \quad \text{In August 2019 Lightsail2 demonstrated its functioning.} \\
    p_3 \quad \text{NASA announces progress on torpor (human hibernation).}
\end{array} \right. 
\]

**Note:** \( c_1 \succ t \)

- “\( c_1 \) supports \( t \)” (entailment in a cogent, nonobligatory sense)
- “\( t \) is compatible with \( c_1 \)” (but the real argumentation focus)
Basic Argument Model

Thesis / Major claim $t$  
*Human beings will colonize other planets.*

$A_{pro}$  
$c_1$  
*Mankind will be able to travel to other galaxies.*  
$p_1$  
*Photon drives can take you up to relativistic velocities.*  
$p_2$  
*In August 2019 Lightsail2 demonstrated its functioning.*  
$p_3$  
*NASA announces progress on torpor (human hibernation).*

$A_{con}$  
$c_2$  
*Mankind will never explore other galaxies.*  
$p_4$  
*Matter cannot pass through wormholes.*  
$p_5$  
*Hawkins explained why time travel is impossible.*

- The standard interpretation $\mathcal{I}$ of all propositions, $t$, $c_i$, $p_j$, is 1 (true).

Note:  
- $c_1 \approx \neg c_2$  
  "$\neg c_2$ is a paraphrase of $c_1$"
  
  $\Rightarrow c_2$ can be expressed as $c_1$ with opposite truth assignment, $\mathcal{I}(c_1) = 0$, $\mathcal{I}(c_2) = 1$
Retrieval Problems \( \Pi_{rel} \)

**Query**  Will human beings colonize other planets?

Given in \( \Pi_{rel} \):

- information need, expressed as query, \( q \in Q \)
- set of arguments, \( A = \{(c_1, P_1), (c_2, P_2), \ldots, (c_n, P_n)\} \)
  - (possibly hidden) human selection of the relevant arguments, \( A_q^*, q \in Q \)

Sought in \( \Pi_{rel} \):

- a relevance function \( \rho : Q \times A \to \{0, 1\} \), such that . . .
- the macro-averaged \( F \)-measure (precision, recall) regarding \( A_q^*, q \in Q \), is maximum

\[ c_1 \quad \text{Mankind will be able to travel to other galaxies.} \]

\[ p_1 \quad \text{Photon drives can take you up to relativistic velocities.} \]

\[ p_2 \quad \text{In August 2019 Lightsail2 demonstrated its functioning.} \]

\[ p_3 \quad \text{NASA announces progress on torpor (human hibernation).} \]
**Query**  
*Will human beings colonize other planets?*

Given in $\Pi_{\text{rank}}$:
- information need, expressed as query, $q \in Q$
- set of relevant arguments, $A_q = \{(c_1, P_1), (c_2, P_2), \ldots, (c_m, P_m)\}$
- *(possibly hidden)* human ranking of the relevant arguments, $\pi^*_{A_q}$, $q \in Q$

Sought in $\Pi_{\text{rank}}$:
- a ranking function $\sigma : Q \times \mathcal{P}(A) \to \Pi$, such that...
  - the mean rank correlation $\bar{\tau}$ regarding $\pi^*_{A_q}$, $q \in Q$, is maximum
Retrieval Problems (3) – (6)

3. $\Pi_{\text{counter}}$  Retrieve the “best” counterargument
   Given: query $q$, argument set $A$, argument $A$

4. $\Pi_{\text{sameside}}$  Retrieve (all) arguments with the same stance
   Given: argument set $A$, argument $A$

5. $\Pi_{\text{argdoc}}$  Is the document argumentative?
   Given: document $d$

6. $\Pi_{\text{argquery}}$  Is the query argumentative?
   Given: query $q$

Notes:

- $\Pi_{\text{counter}}$ can be cast as $\Pi_{\text{rank}}$ if the query is negated.
- $\Pi_{\text{docquery}}$ and $\Pi_{\text{argquery}}$ are decision problems.
- $\Pi_{\text{counter}}$ and $\Pi_{\text{sameside}}$ can be cast as decision problems as well.
- Challenge: development of domain-independent or “topic-agnostic” approaches.
Take Home Messages

3.1 Argument Retrieval Problems

- basic argument model: \((c, \{p_1, p_2, \ldots, p_{|c|}\})\)

- relevant retrieval problems: \(\Pi_{rel}, \Pi_{rank}, \Pi_{counter}\)
3.1 Argument Retrieval Problems

3.2 Argument Ranking

3.3 Resources

3.4 Argument Search Engines

3.5 Shared Tasks
A Topic-Agnostic Solution for $\Pi_{\text{rank}}$

+ Query

Retrieval

“Argumentative” documents

Argument mining

Arguments

Relation detection

Paraphrase analysis

Graph analysis

Argument graph

“PageRank”

Ranking
Query: Reintroduce death penalty?

Premises:
1. The death penalty doesn't deter people from committing crimes.
2. Death penalty should be abolished.
3. A survey of the UN on the relation between the death penalty and homicide rates gave no support to the deterrent hypothesis.

Conclusion:
- The death penalty doesn't deter people from committing serious violent crimes.
- A survey of the UN on the relation between the death penalty and homicide rates gave no support to the deterrent hypothesis.

Death penalty should be abolished.
It does not prevent people from committing crimes.
Original PageRank \[\text{[Page et al. 1999]}\]

1. ground relevance + recursive relevance
2. \(d_j\) links to \(d_i\) \(\leadsto\) increase \(\text{PageRank}(d_i)\)
3. reward exclusive links
4. uniform ground relevances (sum to 1)

\[ p(d_i) = (1 - \alpha) \cdot \frac{1}{|D|} + \alpha \cdot \sum_j \frac{p(d_j)}{|D_j|} \]

\[ \hat{p}(c_i) = (1 - \alpha) \cdot \frac{p(d_i) \cdot |D|}{|A|} + \alpha \cdot \sum_j \frac{\hat{p}(c_j)}{|P_j|} \]

ArgRank \[\text{[Wachmuth/Stein, 2017]}\]

1. ground strength + recursive relevance
2. \(c_i\) premise for \(c_j\) \(\leadsto\) increase \(\text{ArgRank}(c_i)\)
3. reward exclusive premises
4. ground strength \(\leadsto\) PageRank

"Reversal of Evidence"

PageRank: Author cannot enforce links to her web page.
ArgRank: Author cannot enforce use of her argument.
From Premise Scores to Argument Ranks
From Premise Scores to Argument Ranks

Thesis $t$

Premises

Conclusion

$\approx$

$\approx$

$\approx$

$\approx$

$\approx$

$\approx$

  (rank)

Conclusion

$p_1: 0.01$

$p_2: 0.12$

$p_3: 0.03$

How to weigh the premise scores of the matching arguments?
(maximum, average, etc.)
Ranking with Argument Graphs

Case Study: Graph Construction

Construction of a raw graph using 57 corpora from the Argument Web:

28,875 Argument units, used in
17,877 Arguments

Processing steps towards an argument graph:

3,113 Conclusions with \( \geq 1 \) argument, where
498 have multiple premises, from which
70 have a relevant claim, from which
32 are used in 110 intelligible arguments.
Ranking with Argument Graphs
Case Study: Graph Construction

Construction of a raw graph using 57 corpora from the Argument Web:

- 28,875 Argument units, used in 17,877 Arguments

Processing steps towards an argument graph:

- 3,113 Conclusions with ≥ 1 argument, where
  - 498 have multiple premises, from which
  - 70 have a relevant claim, from which
  - 32 are used in 110 intelligible arguments.

Acquisition of a ranking ground truth:

- 7 experts from NLP and IR ranked all arguments (110) for each conclusion (32)
- \( \tau = 0.59 \) as highest agreement between two experts (mean: \( \tau = 0.36 \))
### Ranking with Argument Graphs

#### Case Study: Results

<table>
<thead>
<tr>
<th>Ranking approach</th>
<th>Premise score computation</th>
<th></th>
<th></th>
<th></th>
<th>Sum</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Minimum</td>
<td>Average</td>
<td>Maximum</td>
<td></td>
<td></td>
<td>Best</td>
</tr>
<tr>
<td>1. PageRank</td>
<td>0.01</td>
<td>0.02</td>
<td>0.11</td>
<td>0.28</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>2. Frequency</td>
<td>-0.10</td>
<td>-0.03</td>
<td>-0.01</td>
<td>0.10</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>3. Similarity</td>
<td>-0.13</td>
<td>-0.05</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>4. Sentiment</td>
<td>0.01</td>
<td>0.11</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>5. Most premises</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>6. Random</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>

Approach 1: An argument’s relevance corresponds to the ArgRank of its premises.
(2) A Topic-Agnostic Solution for $\Pi_{\text{counter}}$ [idebate]
(2) A Topic-Agnostic Solution for $\Pi_{\text{counter}}$

**Idea:** Given an argument $A$, the best counterargument $\overline{A}^*$ employs premises that are similar wrt. topic, but takes the opposite stance.

→ Consider both similarities to the premises and conclusion [Walton 2009] :

$A$

\[ \begin{array}{c|c|c} \text{Premises} & \phi_c & \text{Counter-argument} \\ \hline A & \overline{A} \end{array} \]

How to compute these similarities?

How to combine these similarities?

(= What is a sensible hypothesis space of promising model functions?)
(2) A Topic-Agnostic Solution for $\Pi_{\text{counter}}$

**Idea:** Given an argument $A$, the best counterargument $\overline{A}^*$ employs premises that are similar wrt. topic, but takes the opposite stance.

→ Consider both similarities to the premises and conclusion [Walton 2009] :

![Diagram showing the relationship between argument $A$, premises, conclusion, counter-argument $\overline{A}$, and similarity measures $\varphi_P$ and $\varphi_C$.]

Proposed model function to rank counterarguments [Wachsmuth et al., 2018] :

$$R(A, \overline{A}) = \alpha \cdot (\varphi_{\text{conclusion}} \circ \varphi_{\text{Premises}}) - (1 - \alpha) \cdot (\varphi_{\text{conclusion}} \circ \varphi_{\text{Premises}})$$

where

- $\varphi$ combines both word and embedding similarities
- $\circ \in \{\text{min, max, +, }\ast\}$
- $\alpha \in [0; 1]$
(2) A Topic-Agnostic Solution for $\Pi_{\text{counter}}$

Corpus and Analysis

<table>
<thead>
<tr>
<th>Theme</th>
<th>Debates</th>
<th>Points</th>
<th>Counters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Culture</td>
<td>46</td>
<td>278</td>
<td>278</td>
</tr>
<tr>
<td>Digital freedoms</td>
<td>48</td>
<td>341</td>
<td>341</td>
</tr>
<tr>
<td>Economy</td>
<td>95</td>
<td>590</td>
<td>588</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sport</td>
<td>23</td>
<td>130</td>
<td>130</td>
</tr>
<tr>
<td>$\sum$</td>
<td>1069</td>
<td>6779</td>
<td>6753</td>
</tr>
</tbody>
</table>

Corpus:
- based on the iDebate.org portal
- Download: ArguAna Counterargs

Retrieval experiments (selected results):

<table>
<thead>
<tr>
<th>Find the best counterargument within . . .</th>
<th>True-to-false ratio</th>
<th>Accuracy*</th>
</tr>
</thead>
<tbody>
<tr>
<td>all counters of the same debate</td>
<td>1 : 3</td>
<td>0.75</td>
</tr>
<tr>
<td>all counters of the same theme</td>
<td>1 : 136</td>
<td>0.54</td>
</tr>
<tr>
<td>all arguments of the entire portal</td>
<td>1 : 2 800</td>
<td>0.32</td>
</tr>
</tbody>
</table>

* The parameters for $R(A, \overline{A})$ were determined by a systematic ranking analysis.
3.2 Argument Ranking

- topic agnostic solution for $\Pi_{\text{rank}}$, based on Page rank
- topic agnostic solution for $\Pi_{\text{counter}}$, realized via grid search
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Argument Retrieval Problems</td>
</tr>
<tr>
<td>3.2</td>
<td>Argument Ranking</td>
</tr>
<tr>
<td>3.3</td>
<td>Resources</td>
</tr>
<tr>
<td>3.4</td>
<td>Argument Search Engines</td>
</tr>
<tr>
<td>3.5</td>
<td>Shared Tasks</td>
</tr>
</tbody>
</table>
## Argumentation-Related Resources

<table>
<thead>
<tr>
<th>Leverage effort</th>
<th>Resource type</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>low</td>
<td><strong>Technology</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Visual inspection</td>
<td>Argument Web</td>
</tr>
<tr>
<td></td>
<td>Acquisition, Tagging</td>
<td>Truthmapping</td>
</tr>
<tr>
<td></td>
<td><strong>Corpora</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Argumentative structure analysis</td>
<td>AIFdb data</td>
</tr>
<tr>
<td></td>
<td>Argumentation quality analysis</td>
<td>IBM Debater data</td>
</tr>
<tr>
<td></td>
<td>Stance detection</td>
<td>UKP data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Webis data</td>
</tr>
<tr>
<td>medium</td>
<td><strong>Debate portals</strong></td>
<td>Kialo</td>
</tr>
<tr>
<td></td>
<td>English</td>
<td>idebate</td>
</tr>
<tr>
<td></td>
<td>German</td>
<td>Debatepedia</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Argumentia</td>
</tr>
<tr>
<td>high</td>
<td><strong>Discussion pages</strong></td>
<td>changemyview</td>
</tr>
<tr>
<td></td>
<td>Focus on persuasion</td>
<td>reddit</td>
</tr>
<tr>
<td></td>
<td>Controversial issues</td>
<td>WikiTalk</td>
</tr>
<tr>
<td></td>
<td>Focus on deliberation</td>
<td></td>
</tr>
<tr>
<td>very high</td>
<td><strong>Articles</strong></td>
<td>New York Times</td>
</tr>
<tr>
<td></td>
<td>Editorials, Essays</td>
<td>ACL anthology</td>
</tr>
<tr>
<td></td>
<td>Legal</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Scientific publications</td>
<td></td>
</tr>
</tbody>
</table>
The Argument Web

AIFdb Corpora
Structured argument data in uniform format

AIFdb Search
Search interface for argument resources

ARG-tech API
Several argument web services

Argublogging
Widget for argument annotation in blogs

OVA
Online visualization and analysis of arguments

Arvina
Dialogue platform based on AIFdb
Take Home Messages

3.3 Resources

- plethora of resources, which are mainly of academic nature
- tools, corpora, debates, discussions, argumentative texts
3.1 Argument Retrieval Problems

3.2 Argument Ranking

3.3 Resources

3.4 Argument Search Engines

3.5 Shared Tasks
Vision of Argument Search*

Arguments in future web search:

- support forming opinions
- make it easy to find relevant arguments
- deliberation: learn about other views
- education: learn to debate

Search results should . . .

- rank the best arguments highest
- cover diverse aspects
- cover reliable and heterogeneous sources
- be up-to-the-minute
- be traceable and evaluable

Argument Search Engines
Basic Elements and Process

Crawling  Mining  Cleansing  Indexing  Filtering  Ranking  Presentation

Sources  Candidate documents  Candidate arguments  Model-conform arguments

Acquisition

Index  Relevant arguments

Retrieval

Ranked arguments  Argument map

Result
Argument Search Engines

Basic Elements and Process

Acquisition paradigm [Ajjour et al., KI 2019]:

- distribution of processing steps regarding offline time and online time
- tradeoff between precision, recall, and topicality
Acquisition Paradigms

(a) args.me [Demo]

- Research focus: argument ranking
- Supervision level: medium (distantly supervised)

- Effectiveness profile: high precision, low recall
- Stance balance: guaranteed
- Efficiency: high
### Acquisition Paradigms

#### (b) IBM Debater

- **Research focus:** debating technology
- **Supervision level:** medium (recognized source)

- **Effectiveness profile:** high precision, high recall on topic
- **Stance balance:** guaranteed
- **Efficiency:** high

#### Diagram:

- **Topic-specific retrieval**
- **Mining**
- **Cleansing**
- **Indexing**
- **Filtering**
- **Ranking**

  - Wikipedia documents
  - Candidate arguments
  - Model-conform arguments
  - Index
  - Relevant arguments
  - Ranked arguments

- **Online**
- **Offline**
Acquisition Paradigms

Research focus: argument mining
Supervision level: low

Effectiveness profile: low precision, high recall
Stance balance: cannot be guaranteed
Efficiency: low
Designing a ranking algorithm:

- Analyze conclusions, premises, or both?
- Use fulltext or elite terms only?
- Exploit metadata and sentiment?
- Analyze relations between arguments?
### Ranking Paradigms in IR

<table>
<thead>
<tr>
<th>Year</th>
<th>Empirical Models</th>
<th>Probabilistic Models</th>
<th>Language Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960</td>
<td>Boolean, VSM</td>
<td>ProbabilityIndex, 2-Poisson</td>
<td>LanguageModel, LDA, Doc2Vec</td>
</tr>
<tr>
<td>1970</td>
<td>FuzzySet, GVSM</td>
<td>BIM, BII</td>
<td>MixtureUnigram, pLSI</td>
</tr>
<tr>
<td>1980</td>
<td>LSI</td>
<td>Inquery, BeliefNet</td>
<td>LanguageModel, LDA</td>
</tr>
<tr>
<td>1990</td>
<td>Genre</td>
<td>BestMatch</td>
<td>pLSI</td>
</tr>
<tr>
<td>2000</td>
<td>SuffixTree</td>
<td>DivRand, WebGenre</td>
<td>LDA</td>
</tr>
<tr>
<td>2010</td>
<td>DivRand</td>
<td>CL-ESA, ESA</td>
<td>MixtureUnigram</td>
</tr>
<tr>
<td>2015</td>
<td>CL-ESA</td>
<td>LearningToRank</td>
<td>pLSI</td>
</tr>
</tbody>
</table>

#### BM25

\[
\rho(q, A) \equiv BM25(q, A) = \sum_{i=1}^{\|q\|} \frac{IDF(t_i) \cdot TF(t_i, A) \cdot (k_1 + 1)}{TF(t_i, A) + k_1 \cdot (1 - b + b \cdot \frac{LEN(A)}{avdl})}
\]
New research indicates that *Divergence from Randomness* and *Learning to Rank* are significantly more effective retrieval models to address $\Pi_{rank}$.

[Pottast et al., SIGIR 2019]
More on args.me

Argument sources:

<table>
<thead>
<tr>
<th>#</th>
<th>Debate Portal</th>
<th>Argument Units</th>
<th>Arguments</th>
<th>Debates</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>idebate.org</td>
<td>16 084</td>
<td>15 384</td>
<td>698</td>
</tr>
<tr>
<td>2</td>
<td>debatepedia.org</td>
<td>34 536</td>
<td>33 684</td>
<td>751</td>
</tr>
<tr>
<td>3</td>
<td>debatewise.org</td>
<td>39 576</td>
<td>33 950</td>
<td>2 252</td>
</tr>
<tr>
<td>4</td>
<td>debate.org</td>
<td>210 340</td>
<td>182 198</td>
<td>28 045</td>
</tr>
<tr>
<td>5</td>
<td>forandagainst.com</td>
<td>29 255</td>
<td>26 224</td>
<td>3 038</td>
</tr>
<tr>
<td>∑</td>
<td></td>
<td>329 791</td>
<td>291 440</td>
<td>34 784</td>
</tr>
</tbody>
</table>

Design decisions:

- **Argument model**: conclusion + 1 premise with stance information
- **Query**: free text phrase, interpreted as AND query
- **Retrieval**: exact matching against conclusion
- **Ranking**: BM25F based on conclusion (1.0), premise (0.5), and debate (0.2)
More on args.me  [args.me]

Top queries  (Sep.’17 – Apr.’19):

<table>
<thead>
<tr>
<th>Query</th>
<th>Absolute</th>
<th>Relative</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 climate change</td>
<td>251</td>
<td>3.5%</td>
</tr>
<tr>
<td>2 feminism</td>
<td>193</td>
<td>2.7%</td>
</tr>
<tr>
<td>3 abortion</td>
<td>158</td>
<td>2.2%</td>
</tr>
<tr>
<td>4 trump</td>
<td>146</td>
<td>2.0%</td>
</tr>
<tr>
<td>5 brexit</td>
<td>128</td>
<td>1.8%</td>
</tr>
<tr>
<td>6 death penalty</td>
<td>73</td>
<td>1.0%</td>
</tr>
<tr>
<td>7 google</td>
<td>58</td>
<td>0.8%</td>
</tr>
<tr>
<td>8 vegan</td>
<td>57</td>
<td>0.8%</td>
</tr>
<tr>
<td>9 nuclear energy</td>
<td>56</td>
<td>0.8%</td>
</tr>
<tr>
<td>10 donald trump</td>
<td>47</td>
<td>0.7%</td>
</tr>
</tbody>
</table>

Coverage of 1082  Wikipedia controversial issues:

- 78% match with $\geq 1$ argument
- 42% match with $\geq 1$ conclusion
More on args.me [args.me]

Presentation and Analytics

PRO
Abortion is the ending of pregnancy by the removal or destruction of a human fetus or embryo. It is the legal and medical process of terminating a pregnancy. Abortion is a medical procedure that can be performed at various stages of pregnancy. Abortion is a human right and a fundamental aspect of reproductive freedom. Abortion is legal in many countries, but it is illegal in some countries.

CON
Abortion is the ending of pregnancy by the removal or destruction of a human fetus or embryo. It is the legal and medical process of terminating a pregnancy. Abortion is a medical procedure that can be performed at various stages of pregnancy. Abortion is a human right and a fundamental aspect of reproductive freedom. Abortion is legal in many countries, but it is illegal in some countries.

There are many good and bad sides to abortion. But just...
More on args.me

Presentation and Analytics
More on args.me  [args.me]

Presentation and Analytics
3.4 Argument Search Engines

- from “what” search engines to “why” search engines
- acquisition paradigm trades precision, recall, and topicality
<table>
<thead>
<tr>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Argument Retrieval Problems</td>
</tr>
<tr>
<td>3.2 Argument Ranking</td>
</tr>
<tr>
<td>3.3 Resources</td>
</tr>
<tr>
<td>3.4 Argument Search Engines</td>
</tr>
<tr>
<td>3.5 Shared Tasks</td>
</tr>
</tbody>
</table>
Same Side Stance Classification

Task: Given two arguments regarding a certain topic, decide whether or not the two arguments have the same stance.

Topic: “Gay marriage should be legalized.”

**Argument 1**
Marriage is a commitment to love and care for your spouse till death. This is what is heard in all wedding vows. Gays can clearly qualify for marriage according to these vows, and any definition of marriage deduced from these vows.

**Argument 2**
Marriage is the institution that forms and upholds for society, its values and symbols are related to procreation. To change the definition of marriage to include same-sex couples would destroy its function.

○≠○
different side

**Argument 1**
Marriage is a commitment to love and care for your spouse till death. This is what is heard in all wedding vows. Gays can clearly qualify for marriage according to these vows, and any definition of marriage deduced from these vows.

**Argument 2**
Gay marriage should be legalized since denying some people the option to marry is discriminatory and creates a second class of citizens.

○=○
same side
Same Side Stance Classification

Task Rationale

Same side classification needs not to distinguish topic-specific pro- / con-vocabulary.

→ "Only" argument similarity within a stance needs to be assessed.

→ Same side classification may be solved in a topic-agnostic fashion.

Applications:

- measure the bias strength within argumentation
- structure a discussion
- find out who or what is challenging me in a discussion
- filter wrongly labeled stances in a large argument corpus
- ...
Same Side Stance Classification

Tasks Details

Two topics (domains):

1. Should gay marriage be legalized?
2. Should abortion be legalized?

Within domain setting:

Training. Instances from both domains.
Test. Instances from both domains.

Cross domain setting:

Training. Instances from abortion.
Test. Instances from gay marriage.

Form of an instance:

1. Name of the topic (domain) $d$.
2. Argument 1 from $A_d$.
3. Argument 2 from $A_d$.
4. One of $\{\circ=\circ, \circ\neq\circ\}$.

Timeline:

8.6. 2019: Training data online.
## Same Side Stance Classification

### Results “Within Domain”

<table>
<thead>
<tr>
<th>Team</th>
<th>Gay marriage</th>
<th>Abortion</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre</td>
<td>Rec</td>
<td>Acc</td>
</tr>
<tr>
<td>Trier University</td>
<td>0.90</td>
<td>0.73</td>
<td>0.83</td>
</tr>
<tr>
<td>Leipzig University</td>
<td>0.80</td>
<td>0.78</td>
<td>0.79</td>
</tr>
<tr>
<td>IBM Research</td>
<td>0.73</td>
<td>0.63</td>
<td>0.70</td>
</tr>
<tr>
<td>TU Darmstadt</td>
<td>0.74</td>
<td>0.56</td>
<td>0.68</td>
</tr>
<tr>
<td>Düsseldorf University</td>
<td>0.76</td>
<td>0.35</td>
<td>0.62</td>
</tr>
<tr>
<td>LMU</td>
<td>0.53</td>
<td>1.00</td>
<td>0.55</td>
</tr>
</tbody>
</table>

...
Same Side Stance Classification
Results “Cross Domain”: Abortion → Gay marriage

<table>
<thead>
<tr>
<th>Team</th>
<th>Gay marriage (large)</th>
<th>Gay marriage (small)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre</td>
<td>Rec</td>
</tr>
<tr>
<td>LMU</td>
<td>0.67</td>
<td>0.53</td>
</tr>
<tr>
<td>TU Darmstadt</td>
<td>0.64</td>
<td>0.59</td>
</tr>
<tr>
<td>IBM Research</td>
<td>0.62</td>
<td>0.49</td>
</tr>
<tr>
<td>Paderborn University</td>
<td>0.60</td>
<td>0.38</td>
</tr>
<tr>
<td>Trier University</td>
<td>0.69</td>
<td>0.16</td>
</tr>
<tr>
<td>Düsseldorf University</td>
<td>0.72</td>
<td>0.53</td>
</tr>
</tbody>
</table>

Most of the submitted classifiers are robust regarding:

- imbalances between domain proportions in training and test
- imbalances between domain proportions within test
- imbalances between same side / different side proportions
Argument Retrieval Task @ CLEF 2020

**Task 1: Supporting argumentative conversations**
- **Scenario:** Users search for arguments on controversial topics
- **Task:** Retrieve “strong” pro/con arguments on the topic
- **Data:** 300,000 “arguments” (short text passages)

**Task 2: Answering comparative questions with arguments**
- **Scenario:** Users face personal decisions from everyday life
- **Task:** Retrieve arguments for “Is X better than Y for Z?”
- **Data:** ClueWeb12 or ChatNoir

- Run submissions similar to “classical” TREC tracks
- Software submissions via TIRA
Take Home Messages

3.5 Shared Tasks

- build community for relevant problems
- acquire knowledge and best practices, improve benchmarks
## 3.1 Argument Retrieval Problems

basic argument model, relevant retrieval problems

## 3.2 Argument Ranking

topic-agnostic solution for $\Pi_{\text{rank}}$ and $\Pi_{\text{counter}}$

## 3.3 Resources

resources: tools, corpora, debates, discussions, argumentative texts

## 3.4 Argument Search Engines

acquisition paradigm trades between precision, recall, and topicality

## 3.5 Shared Tasks

build community, acquire knowledge, improve benchmarks