”PageRank“ for Argument Relevance

Henning Wachsmuth, Benno Stein, Yamen Ajjour
henning.wachsmuth@uni-weimar.de
The future of search

**Pro death penalty**

**#1 Retribution**
[http://www.bbc.co.uk](http://www.bbc.co.uk) (36 other sources...)
Real justice requires people to suffer for their wrongdoing in a way adequate for the crime.

**#2 Deterrence**
[http://www.debate.org](http://www.debate.org) (15 other sources...)
By executing convicted murderers, would-be murderers are deterred from killing people.

**#3 Prevention of re-offending**
[http://www.bbc.co.uk](http://www.bbc.co.uk) (25 other sources...)
Those executed cannot commit further crimes. Imprisonment does not protect sufficiently.

**Con death penalty**

**#1 Execution of the innocent**
[http://www.bbc.co.uk](http://www.bbc.co.uk) (81 other sources...)
As long as human justice remains fallible, the risk of executing the innocent can never be eliminated.

**#2 Right to live**
[http://www.amnesty.org](http://www.amnesty.org) (102 other sources...)
Everyone has an inalienable human right to live, even those who commit murder.

**#3 Failure to deter**
[http://www.procon.org](http://www.procon.org) (24 other sources...)
There is no scientific proof that executions have a greater deterrent effect than life imprisonment.
Mining of relevant arguments

- **Argument mining**
  - Identifies arguments in natural language text
  - Does not assess relevance

- **Argument relevance**
  - Contribution to conclusion on an issue (Walton, 2006)
  - Often perceived subjectively

"The death penalty legitimizes an irreversible act of violence. As long as human justice remains fallible, the risk of executing the innocent can never be eliminated."

"The death penalty doesn’t deter people from committing serious violent crimes. The thing that deters is the likelihood of being caught and punished."

- **Research question**
  - Can we develop an "objective" relevance measure?
Argument relevance at web scale

- **Exploit web scale for objective relevance**
  - Ignore content and inference of argument (for now)
  - Decide relevance structurally

- **Key hypothesis**
  - Relevance of a conclusion depends on what other arguments across the web use it as a premise
  - Author cannot control who "cites" a conclusion in this way

- **Assume perfect argument mining technology**
  - Build argument graph for the web
  - Adapt PageRank algorithm to arguments

"PageRank, a method for rating web pages objectively and mechanically, effectively measuring human interest."

"Why not adapt PageRank to arguments?"
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.

It does not deter people from committing serious violent crimes.

The death penalty should be abolished.

Even if it did, is it acceptable to pay for predicted future crimes of others?

"PageRank" for Argument Relevance – Henning Wachsmuth, Benno Stein, Yamen Ajjour
PageRank for argument relevance

- **Original PageRank score** of a web page $d$ (Page et al., 1999)
  
  
  $p(d) = (1 - \alpha) \cdot \frac{1}{|D|} + \alpha \cdot \sum_i \frac{p(d_i)}{|D_i|}$
  
  same score for each page
  
  ground relevance
  
  recursive relevance
  
  page $d_i$ links to $d$
  
  # pages $d_i$ links to

- **Adapted PageRank score** of an argument unit $c$

  $\hat{p}(c) = (1 - \alpha) \cdot \frac{p(d)}{|D|} + \alpha \cdot \sum_i \frac{\hat{p}(c_i)}{|P_i|}$

  PageRank of page $d$ containing $c$
  
  ground relevance
  
  recursive relevance
  
  conclusion $c_i$ uses $c$ as premise
  
  # premises of $c_i$

- **Argument relevance** is aggregation of premise scores
  
  - Minimum, average, maximum, or sum
A large ground-truth argument graph

- **No use of argument mining here**
  - Evaluation of PageRank without noise

- **Construction of a ground-truth argument graph**
  - 57 argument corpora at [www.aifdb.org](http://www.aifdb.org)
  - Merged all arguments except for duplicates
  - Units assumed to match if they span the same text
  - Computed PageRank for each unit

- **17,877 arguments with 31,080 different units**

```
   0  1  2  3  4  5–9  10–122
   0  1  0  2  5  3  2
372 595 846 63 88 288 50
```

- 0
- 1
- 2
- 3
- 4
- 5–9
- 10–122

usage as conclusion

```
   0  1  2  3  4  5  6–8
   0  1  2  3  4  5  6
95 172 123 172 170 170 11
```

usage as premise
Benchmark argument relevance rankings

- **No objective relevance judgments available**
  - Use average judgments as a proxy

- **Filtering of general claims from the graph**
  - 3113 conclusions with >1 argument, 498 with premises used multiply
  - 70 classified as claims of general interest by 2 annotators (Cohen’s $\kappa = .69$)
  - 32 have 2–6 ”real“ arguments (Cohen’s $\kappa = .63$)

- **Creation of relevance rankings for the 32 claims**
  - 110 arguments ranked by 7 annotators (mean Kendall’s $\tau = .36$, highest $\tau = .59$)

"Strawberries are the best choice for your breakfast meal."

"Berries are superfoods because they’re so high in antioxidants without being high in calories, says Giovinazzo MS, RD, a nutritionist at Clay health club and spa, in New York City."

Thanks to our group!

PageRank for Argument Relevance – Henning Wachsmuth, Benno Stein, Yamen Ajjour
Impact of PageRank

- **Evaluation of unsupervised ranking approaches**
  - PageRank of premises
  - Frequency of premises
  - Similarity of units
  - Sentiment of premises
  - Number of premises
  - Random ranking

\[
\hat{p} \sum c \sim P
\]

- Each for minimum, average, maximum, and sum aggregation

- **Experiment on ground-truth graph**
  - Rank arguments with each approach
  - Correlate with benchmark rankings

- **Results**
  - PageRank with sum aggregation best
  - Consistently outperforms frequency
  - Notable correlation despite ignorance of content and inference

---

<table>
<thead>
<tr>
<th>#</th>
<th>Dimension</th>
<th>$\tau$</th>
<th>best</th>
<th>worst</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PageRank</td>
<td>0.28</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>Number</td>
<td>0.19</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>Sentiment</td>
<td>0.12</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>Frequency</td>
<td>0.10</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>Similarity</td>
<td>0.02</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>Random</td>
<td>0.00</td>
<td>8</td>
<td>7</td>
</tr>
</tbody>
</table>

"PageRank" for Argument Relevance – Henning Wachsmuth, Benno Stein, Yamen Ajjour
Towards argument search engines

**Contributions**
- Approach to assess argument relevance structurally
- Dataset with argument rankings
- First empirical evidence that relevance depends on the reuse of conclusions

**Major open challenges...**
- Arguments must be mined robustly from the web
- Identification of reuse is hard
- PageRank only for relevant candidates

**... but web scale helps**
- Prefer precision over recall
- Start with reliable sources and limited domains
- Refine argument graph step by step