Topic Ontologies for Arguments

Yamen Ajjour  
Hannover

Johannes Kiesel  
Weimar

Benno Stein  
Weimar

Martin Potthast  
Leipzig

Yamen Ajjour  
Leibniz Hannover University
webis.de
Motivation

- Argument mining and analysis is topic-dependent.

Example: The following argument has different stances depending on the topic.

**Argument:**

“Marijuana should be legalized since it brings economical benefits”

**Topic 1:** “Marijuana should be legalized”
**Stance:** Pro

**Topic 2:** “Does marijuana lead to social anxiety”
**Stance:** Neutral
Motivation

Supervised learning methodology for learning argument mining approaches.
Motivation

How well do argument corpora represent controversial topics?
A survey of 59 argument corpora shows that researchers take three approaches:

- Manual selection: choosing a set of topics manually
- Source-driven-greedy: a whole source is exploited
- Source-driven-sample: a source is sampled

![Bar chart showing the distribution of topic selection methods among 59 argument corpora. The chart indicates that 66% use manual selection, 22% use source-driven greedy, and 12% use source-driven sample.]
Topic Selection of Existing Argument Corpora

A survey of 59 argument corpora shows that researchers take three approaches:

- Manual selection: choosing a set of topics manually
- Source-driven-greedy: a whole source is exploited
- Source-driven-sample: a source is sampled

About a third of argument corpora cover less than 25 topics.
Ontology is an explicit specification of a conceptualization: it defines the terms of a specific domain.

A topic ontology is a directed acyclic graph, whose nodes are topics and the edges imply "is-part-of" relations.
Use Cases for Argument Topic Ontologies

Argument topic ontologies have different applications that we foresee:

- Topic Sampling
- Argument Analysis
- Argument Generation

Marijuana should be legalized since it brings profit.

Legalizing marijuana will increase tax revenue.

Marijuana should be legalized since it brings profit.
Selecting and organizing topics require domain knowledge. Hence, we resort to three authoritative sources.

<table>
<thead>
<tr>
<th>Ontology</th>
<th>Topics</th>
<th>Authors</th>
<th>Docs</th>
</tr>
</thead>
<tbody>
<tr>
<td>World Economic Forum Level-1</td>
<td>137</td>
<td>334</td>
<td>940</td>
</tr>
<tr>
<td>World Economic Forum Level-2</td>
<td>822</td>
<td>217</td>
<td>550</td>
</tr>
<tr>
<td>Wikipedia Level-1</td>
<td>14</td>
<td>78,014</td>
<td>68</td>
</tr>
<tr>
<td>Wikipedia Level-2</td>
<td>748</td>
<td>1,930</td>
<td>1</td>
</tr>
<tr>
<td>Debatepedia</td>
<td>89</td>
<td>145</td>
<td>62</td>
</tr>
</tbody>
</table>
Categorizing Arguments into Topic Ontologies

Assessing the coverage of 59 argument corpora by categorizing the arguments within a topic ontology in two ways:

1. Mapping the topic labels of an argument corpus to a topic ontology.
2. Automatic corpus units categorization within a topic ontology.
1. Mapping Topic labels to a Topic Ontology

1. Normalization

   (e.g., “Abortion should be banned” -> abortion).

2. Retrieval: using a topic label as a query to retrieve an ontology topic.

3. Annotation: is a topic label a subtopic or synonymous of the retrieved ontology topics.
Argument Corpora Topic Coverage

The proportion of ontology topics covered by each corpus using the mappings.
The distribution of corpus units over the ontology topics for the first level of World Economic Forum.
The distribution of corpus units over the ontology topics for the first level of World Economic Forum.

- The distribution of the topics is skewed.
- Only a small set of topics is covered.
2. Automatic Corpora Unit Categorization

About a third of argument corpora do not provide corpora topic labels and hence is not included in the previous analysis.

**Approach**: Semantic indexing calculates the cosine similarity between a corpus unit and the documents categorized under an ontology topic.

**Evaluation**: Pooled evaluation for 104 corpora units with a depth of five ontology topics.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Wikipedia</th>
<th></th>
<th>WEF</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Level-1</td>
<td>Level-2</td>
<td>Level-1</td>
<td>Level-2</td>
</tr>
<tr>
<td>Direct match</td>
<td>0.06</td>
<td>0.40</td>
<td>0.29</td>
<td>0.19</td>
</tr>
<tr>
<td>Semantic Indexing</td>
<td>0.43</td>
<td><strong>0.59</strong></td>
<td><strong>0.34</strong></td>
<td><strong>0.33</strong></td>
</tr>
<tr>
<td>Text2vec-SI$_{BERT}$</td>
<td><strong>0.47</strong></td>
<td>0.31</td>
<td>0.28</td>
<td>0.23</td>
</tr>
</tbody>
</table>

F1-score of the approaches
Conclusion

Contributions:

- Survey and analysis of 59 argument corpora with regard to their topics.
- Introduced topic ontologies to formally model the context of arguments.
- Effective approach to categorize arguments within a topic ontology.
- Released ontologies and topic mappings at
  https://zenodo.org/record/5180409

Findings:

- Argument corpora cover a small set of topics and the distribution is skewed.
- The generalizability of existing argument mining approaches is questionable.