Evaluation of Scientific Elements for Text Similarity in Biomedical Publications

Mariana Neves, Daniel Butzke, Barbara Grune
German Centre for the Protection of Laboratory Animals (Bf3R), German Federal Institute for Risk Assessment (BfR), Max-Dohrn-Str. 8-10, 10589, Berlin, Germany

Rhetorical elements from scientific publications provide a more structured view of the document and allow algorithms to focus on particular parts of the text.

We surveyed the literature for previously proposed schemes for rhetorical elements and present an overview of its current state of the art.

We also searched for available tools using these schemes and applied four tools for our particular task of ranking biomedical abstracts based on text similarity.

**Summary of previous work based on selected features supported by the schemes:**
- Abstracts and Full-text
- Entity and Relation
- Biomedical
- Ontology or levels
- Corpus available
- Tool available

**Evaluation of the available tools on a biomedical use case for text similarity:**

We evaluated the tools for the task of text similarity: given an input document that describes an animal experiment, we would like to mine similar candidate documents that may also be potential alternatives to animal testing.

Our definition of similarity requires that:
- both input and candidate documents should have similar research goal and comparable outcomes,
- however, the methods in the input document should be substantial different from those in the candidate documents.

We calculated the similarity between the input and candidate documents, either based on the whole text or on selected rhetorical elements as provided by the tools. We used the TextFlow tool for text similarity.

**Data available at:**
https://github.com/mariananeves/scientific-elements-text-similarity

**Conclusions:**
- A considerable improvement can be obtained when using ArgumIndi wt. the original ranking returned by PubMed and to the Text Flow baseline.
- However, there is still much room for improvement: the scores are still far below the possible maximum values.

Data available at: https://github.com/mariananeves/scientific-elements-text-similarity

<table>
<thead>
<tr>
<th>Tools</th>
<th>Categories</th>
<th>Corpus</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>AZ</td>
<td>ABB, TEXTUAL, ORN. BACKG., CONTRAST, BASIC, OTHER</td>
<td>80 (Golds and Morris, 2002) and 20 (Minta et al., 2006)</td>
<td>CL, no</td>
</tr>
<tr>
<td>CoreC</td>
<td>[Level 1] Hypothesis, Motivation, Background, Goal, Object, Method, Experiment, Model, Observation, Result, Conclusion</td>
<td>225 (Shah et al., 2016)</td>
<td>open</td>
</tr>
<tr>
<td>Dr. Inventor</td>
<td>Approach, Challenge, Background, Outcome, Focus Work</td>
<td>40 (Romans and Sep- gian, 2016)</td>
<td>CO, open</td>
</tr>
<tr>
<td>MAZEA</td>
<td>background, gap, purpose, method, result, conclusion</td>
<td>608 abstracts (Belay et al., 2019)</td>
<td>LS, open</td>
</tr>
<tr>
<td>PIBOSO</td>
<td>Populations, Intervention, Background, Outcome, Study Design, Other</td>
<td>1,890 abstracts (Kan et al., 2013)</td>
<td>No, open</td>
</tr>
<tr>
<td>PubMedRCT</td>
<td>background, objective, method, result, conclusion</td>
<td>20,000 and 200,000 abstracts (Kan et al., 2013)</td>
<td>No, open</td>
</tr>
<tr>
<td>Willie</td>
<td>FOCUS, POLARITY, CERTAINTY, EVIDENCE, DIRECTIONALITY</td>
<td>10,000 sentences (Shah et al., 2006)</td>
<td>No, open</td>
</tr>
<tr>
<td>ScilTB</td>
<td>USAGE, RESULT, MODEL, PART,WHOLE, TOPIC, COMPARISON</td>
<td>576 abstracts (Kan et al., 2013)</td>
<td>CS, open</td>
</tr>
<tr>
<td>[Course level]</td>
<td>Attribution, Background, Context, Effect, Comparison, Condition, Context, Evaluation, Enactment, Explanation, Explanation, Joint, Main-mean, Progression, Same-end, Summary, Temporal</td>
<td>786 abstracts (Yang and Li, 2016)</td>
<td>CL, open</td>
</tr>
</tbody>
</table>

**Identification of the schemes for which tools are readily available for use:**
- Achakulvisut et al. (Achakulvisut et al., 2018) (PubMedRCT schema)
- ArgumIndi (Lauscher et al., 2018a) (Dr. Inventor schema extended)
- MAZEA tool and schema (Dayrell et al., 2012) (MAZEA schema)
- Prasad and Kan (Prasad and Kan, 2017) (ScienceIE schema)

**Summary of the results from the two baselines (two first rows) and when using the selected tools. The maximum scores represent the maximum value of P@10, R@10 and F@10 that could have been obtained by any of the approaches.**

**Performance of the single labels in the re-ranking task.**