Evaluation of Scientific Elements for Text Similarity in Biomedical Publications

Bundesinstitut für Risikobewertung

DEUTSCHES ZENTRUM ZUM SCHUTZ VON VERSUCHSTIEREN

Mariana Neves, Daniel Butzke, Barbara Grune

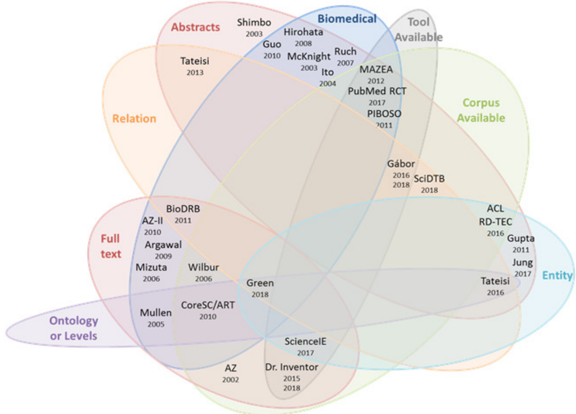
German Centre for the Protection of Laboratory Animals (Bf3R), German Federal Institute for Risk Assessment (BfR), Max-Dohrn-Str. 8-10, 10589, Berlin, Germany

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.

Short survey on existing schemes for rhetorical elements in scientific publications:



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

Identification of the schemes for which available corpora are available:

	Tools	Categories	Corpora	Topic
Sentence/Phrase	AZ	AIM, TEXTUAL, OWN, BACKGROUND, CONTRAST, BASIC, OTHER	80 (Teufel and Moens, 2002) and 20 (Mizuta et al., 2006)	CL, bio
	CoreSC	[Level 1] Hypothesis, Motivation, Background, Goal, Object, Method, Experiment, Model, Observation, Result, Conclusion	225 (Liakata et al., 2010)	chem
	Dr. Inventor	Approach, Challenge, Background, Outcomes, Future Work	40 (Ronzano and Saggion, 2015)	CG
	MAZEA	background, gap, purpose, method, result, conclusion	645 abstracts (Dayrell et al., 2012)	phy, eng, LS
	PIBOSO	Population, Intervention, Background, Outcome, Study Design, Other	1,000 abstracts (Kim et al., 2011)	bio
	PubMedRCT	background, objective, method, result, conclusion	20,000 and 200,000 abstracts (Dernoncourt and Lee, 2017)	bio
	Wilbur	FOCUS, POLARITY, CERTAINTY, EVIDENCE, DIRECTIONALITY	10,000 sentences (Shatkay et al., 2008)	bio
Ent.	ScienceIE	Task, Process, Material	500 (Augenstein et al., 2017)	CS
Relation	Gábor	USAGE, RESULT, MODEL, PART_WHOLE, TOPIC, COMPARISON	500 abstracts (Gábor et al., 2018)	CL
	SciDTB	[Coarse level] Attribution, Background, Cause-effect, Comparison, Condition, Contrast, Elaboration, Enablement, Evaluation, Explain, Joint, Manner-means, Progression, Same-unit, Summary, Temporal	798 abstracts (Yang and Li, 2018)	CL
Hybrid	Green	[Levels 1-3] 1. Causation, 1.1 One Group, 1.1.1 Agreement Arguments, 1.1.2 Eliminate Candidates, 1.1.3 Explanation-Based, 1.2 Two Group, 1.2.1 Difference, 1.2.2 Analogy (Causal), 1.2.3 Explanation-Based, 2. Other, 2.1 Classification, 2.2 Confirmation	one (Green, 2018)	bio

Identification of the schemes for which tools are readily available for use:

- Achakulvisut et al. (Achakulvisut et al., 2018) (PubMedRCT schema)
- ArguminSci (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)

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.

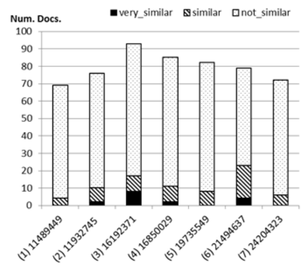
Our definition of similarity requires that:

- 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.

(a) Data:

- Seven input documents from Medline (identifiers in figure on the right).
- For each input document, we collected the top 200 documents (titles and abstracts) retrieved from PubMed's "similar articles" functionality.
- A biomedical researcher manually validated at least the top 100 documents with regards to three degrees of similarity: **very similar, similar and not similar.**



(b) Evaluation

Comparison to two baselines:

- the original order of the candidate documents as returned by PubMed's "similar articles" functionality;
- string similarity based on the whole text (title and abstract) without any preprocessing on the text.

Tools	P@10	R@10	F@10
PubMed	0.30	0.33	0,31
Title+Abstract	0.43	0.51	0.45
Achakulvisut et al	0.44	0.52	0.47
ArguminSci	0.47	0.56	0.50
MAZEA	0.4	0.47	0.42
Prasad and Kan	0.44	0.54	0.47
Min score	0.14	0.16	0.15
Max score	0.83	1.0	0.90

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.

Tools	Labels	P@10	R@10	F@10
Achakulvisut	Background	0.28	0.32	0.30
	Objective	0.33	0.41	0.35
	Methods	0.31	0.40	0.34
	Results	0.20	0.25	0.22
	Conclusions	0.23	0.26	0.24
ArguminSci	Background	0.23	0.25	0.24
	Challenge	0.23	0.26	0.24
	Approach	0.26	0.32	0.28
	Outcome	0.41	0.50	0.44
	Future Work	0.33	0.41	0.35
MAZEA	Background	0.24	0.28	0.25
	Purpose	0.24	0.25	0.25
	Method	0.30	0.37	0.32
	Result	0.28	0.32	0.30
	Conclusion	0.23	0.30	0.25
Prasad	Process	0.37	0.48	0.40
	Material	0.31	0.35	0.33
	Task	0.28	0.36	0.31

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

Conclusions:

- A considerable improvement can be obtained when using ArguminSci wrt. 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**

GERMAN FEDERAL INSTITUTE FOR RISK ASSESSMENT