

Segmentation of Argumentative Texts by Key Statements for Argument Mining from the Web

Motivation

- Extracting a text's fine-grained argument structure is difficult
- Exact segment boundaries are often ambiguous
- Idea: Identify main argumentative statements instead
- **Goal:** Extract key statements
 - Level-1 nodes in argument tree
 - Argumentative units that form an argument with the discussed topic
 - Can be used to infer potential key points

Key Statements Example

Thanks for the timely response. To address my opponents argument, I want to emphasize that eating meat isn't necessary for maximum physical development. All of the vitamins, minerals etc. in meat can also be found in other foods. And does the taste of meat really outweigh the costs of killing? My conclusion: Vegetarianism is a good thing because it saves animals' lives, improves one's health and helps the environment. I didn't bring religion into this debate but almost all of the major religions (even the ones that allow meat eating) agree that vegetarianism is better than eating meat.

Key Statement Segmentation

Given: Argumentative text and discussed controversial topic

Task: Find segments that contain exactly one key statement

Key statements define

- minimal text passages that have to be covered
- text passages that have to be separated

Within these limits, this allows 'flexible' segment boundaries.

Annotated Data

- 14 controversial topics from IBM KPA shared task 2021 [1]
- 50 texts from args.me corpus (1,263 sentences)
 - 147 key statements: 204 sentences, 16% of the texts
- Inter annotator agreement on a subset: 0.47 and 0.80 (Jaccard overlap of identified key statements)
- Evaluating segmentation approaches on different annotators as ground truth → F1 std. deviation 0.01 to 0.05

Matching Segments to Key Statements

- Enable matching beyond simple string matching
- Combination of three approaches:
 - 3-gram overlap thresh: 0.12
 - SequenceMatcher (difflib) thresh: 0.50
 - SBERT similarity thresh: 0.90
- Effectiveness: F1 = 0.84, Precision = 0.9, Recall = 0.79




Key Point Coverage

Assess 'severity' of missed key statements / incorrect segments

- Collect key points covered by key statements
- Calculate coverage of key points by predicted segments

PaLM _{filt}	PaLM	GPT-4 _{filt}	GPT-4	Paragr. _{filt}	Paragr.
0.70	0.74	0.87	0.87	0.57	0.70

Segmentation Results


Measure	PaLM _{filt}	PaLM	GPT-4 _{filt}	GPT-4	Paragr. _{filt}	Paragr.	Sent. _{filt}	Sent.	Ajjour _{filt}	Ajjour	Targer _{filt}	Targer
# Segments	173	285	272	470	154	347	408	1125	413	1174	465	1759
matched (Precision)	0.57	0.46	0.35	0.28	0.63	0.42	0.38	0.22	0.29	0.17	0.30	0.14
 – correct	0.43	0.31	0.24	0.18	0.36	0.22	0.21	0.08	0.18	0.07	0.18	0.05
 spurious	0.43	0.54	0.65	0.71	0.37	0.58	0.62	0.78	0.70	0.82	0.69	0.86
matched (Recall)	0.59	0.74	0.52	0.69	0.66	0.93	0.79	1.00	0.64	0.90	0.73	1.00
 – correct	0.50	0.58	0.41	0.52	0.38	0.49	0.56	0.59	0.49	0.55	0.56	0.53
 missed	0.41	0.27	0.47	0.30	0.33	0.07	0.20	0.01	0.36	0.10	0.27	0.00
F1 micro strict	0.46	0.40	0.30	0.27	0.37	0.30	0.31	0.14	0.26	0.12	0.27	0.09
F1 micro relaxed	0.58	0.56	0.42	0.41	0.65	0.58	0.52	0.36	0.41	0.30	0.44	0.25

Segmentation approaches: PaLM, GPT-4, Paragraph, Sentence, Ajjour [2], Targer – filtering non-argumentative segments (filt) by arg. classes [3]

Findings

- Coarse-grained segmentation is suitable for finding arguments in user generated content
- None of the tested approaches solves the task satisfyingly
- Filtering non-arg. segments improves effectiveness
- **Next steps:** extend dataset and test further approaches

Resources

 <https://github.com/webis-de/argmining25-argument-segmentation>

 https://webis.de/publications.html#zelch_2025a



[1] Friedman et al. 2021: Overview of the 2021 Key Point Analysis Shared Task

[2] Ajjour et al. 2017: Unit Segmentation of Argumentative Texts

[3] Reimers et al. 2019: Classification and Clustering of Arguments with Contextualized Word Embeddings