

“Why my essay received a 4?”: A Natural Language Processing Based Argumentative Essay Structure Analysis Appendix

Appendix

Table 1. Labels and explanations ^a

Labels (Ours)	Labels (Original)	Explanations
(not included)	Lead	An introduction that begins with a statistic, a quotation, a description, or some other device to grab the reader’s attention and point toward the thesis.
Position	Position	An opinion or conclusion on the main question.
Claim	Claim	A claim that supports the position.
	Counterclaim	A claim that refutes another claim or gives an opposing reason to the position.
	Rebuttal	A claim that refutes a counterclaim.
Evidence	Evidence	Ideas or examples that support claims, counterclaims, or rebuttals.
(not included)	Concluding Statements	A concluding statement that restates the claims.

^a The explanations were taken directly from the competition webpage: <https://www.kaggle.com/competitions/feedback-prize-2021/data>

ALGORITHM 1: Voting Scheme Algorithm

Input: A vector of models’ prediction results on the sentence prediction_results

Output: Label of the sentence

position, evidence, claim = 0

predict_label is consisted of the prediction results from five models.

for each predict_label in prediction_results

if predict_label == ‘position’

 position += 1

else if predict_label == ‘evidence’

 evidence += 1

else

 claim += 1

end if

end for

if max (position, evidence, claim) == position

```

    return 'position'
else if max (position, evidence, claim) == evidence
    return 'evidence'
else
    return 'claim'
end if

```

Table 2. Essays data analysis with different scores

Scores	Number of Essays	Average Sentences	Average Positions	Average Claims	Average Evidence
2	215	4.237	0.237	2.73	1.27
3	221	6.466	0.443	4.33	1.692
4	805	10.375	0.758	6.896	2.722
5	1018	12.83	0.836	9.1	2.898
6	2929	17.822	1.092	12.955	3.774
7	2094	21.538	1.267	16.127	4.144
8	4062	26.2	1.591	19.839	4.772
9	1228	30.629	1.8	23.368	5.46
10	1032	32.99	2.059	25.095	5.836
11	302	36.662	2.172	27.76	6.728
12	84	40.976	2.988	30.44	7.548

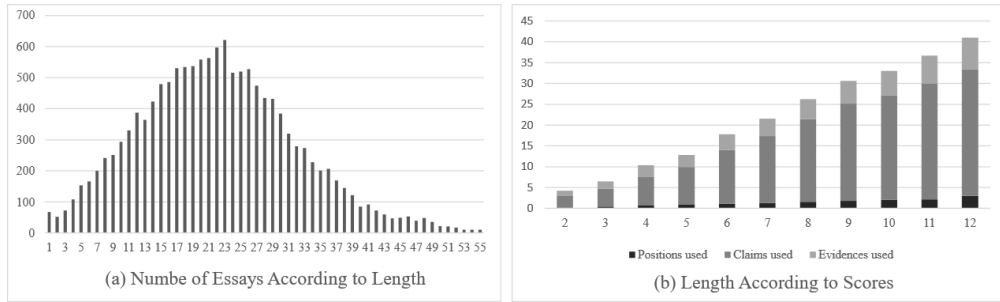


Fig. 1. Length analysis results