

# Experiments on Student-Essay and Debatepedia Corpora

# Student Essays

Only “majorclaim” as conclusion. “claim-for”, “claim-against” and “premise” as premise.

Training Set:

- 598 Conclusion (12.3 %)
- 4280 Premise (87.7 %)

Test Set:

- 153 Conclusion (12.6 %)
- 1058 Premise (87.4 %)

Experiment Setup:

- Training Set was oversampled
- Random Forest Classifier with 300 Trees and 35 Random Features

Features					Number of Features	Baseline Accuracy	Results			
Token-N-Grams	POS-N-Grams	Lemma-N-Grams	Length	Position			Accuracy	F-Measure Conclusion	F-Measure Premise	F-Measure Combined
x					280	87.4 %	86.54 %	0.276	0.926	0.844
	x				563	87.4 %	86.4575 %	0.128	0.927	0.826
		x			316	87.4 %	87.0355 %	0.308	0.928	0.850
			x		2	87.4 %	60.4459 %	0.195	0.738	0.669
				x	4	87.4 %	85.4666 %	0.591	0.912	0.871
x	x				843	87.4 %	87.0355 %	0.234	0.929	0.841
	x	x			879	87.4 %	87.1181 %	0.235	0.930	0.842
	x	x		x	883	87.4 %	91.7424 %	0.667	0.953	0.917
	x	x	x	x	885	87.4 %	91.7424 %	0.664	0.953	0.916

# Student Essays

Only “premise” as premise. “majorclaim”, “claim-for” and “claim-against” as conclusion.

Training Set:

- 1797 Conclusion (36.8 %)
- 3081 Premise (63.2 %)

Test Set:

- 460 Conclusion (37.99 %)
- 751 Premise (62.01 %)

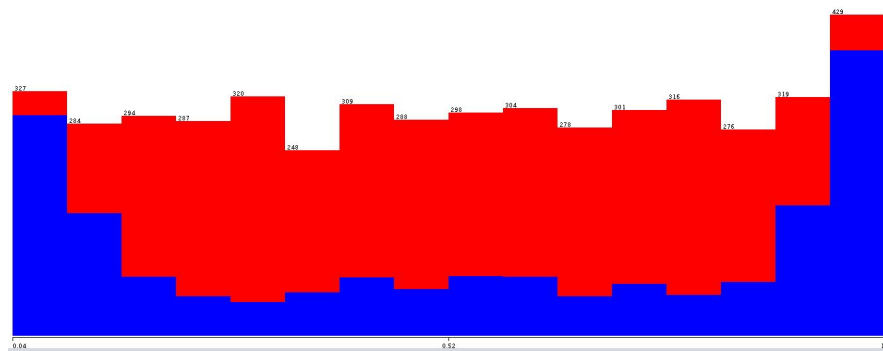
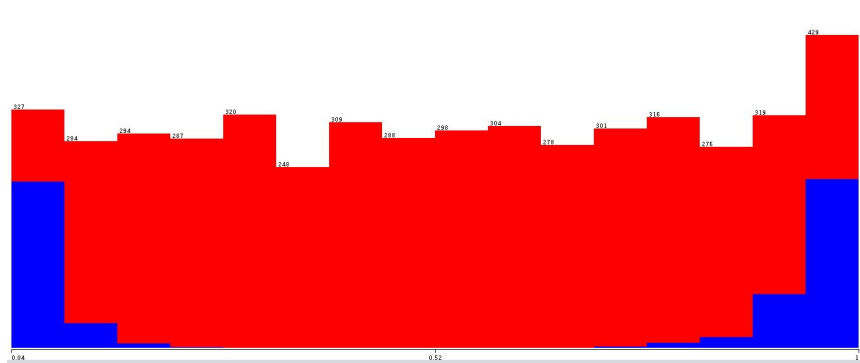
Experiment Setup:

- Training Set was oversampled
- Random Forest Classifier with 300 Trees and 35 Random Features

Features					Number of Features	Baseline Accuracy	Results			
Token-N-Grams	POS-N-Grams	Lemma-N-Grams	Length	Position			Accuracy	F-Measure Conclusion	F-Measure Premise	F-Measure Combined
x					280	62.01 %	68.9513 %	0.572	0.756	0.686
	x				563	62.01 %	63.8315 %	0.487	0.721	0.632
		x			316	62.01 %	68.621 %	0.562	0.755	0.682
			x		2	62.01 %	55.161 %	0.481	0.605	0.558
				x	4	62.01 %	74.5665 %	0.638	0.804	0.741
x	x				843	62.01 %	67.3823 %	0.517	0.754	0.664
	x	x			879	62.01 %	67.4649 %	0.512	0.756	0.663
	x	x		x	883	62.01 %	<b>76.218 %</b>	<b>0.656</b>	<b>0.818</b>	<b>0.757</b>
	x	x	x	x	885	62.01 %	76.1352 %	0.653	<b>0.818</b>	0.755

# Student Essays

The position of an argumentative discourse unit in an essay is highly indicative whether the unit is a conclusion or not.



Position of argumentative discourse units in the essays. Blue (conclusion), red (premise).

Left diagram shows the distribution of the positions of units when only “majorclaim” was labeled as conclusion.

Right diagram shows the distribution of the positions of units when all claims were labeled as conclusions.

# Debatepedia

## Training Set:

- 9244 Conclusion (24.36 %)
- 28697 Premise (75.64 %)

## Test Set:

- 2312 Conclusion (24.37 %)
- 7174 Premise (75.63 %)

## Experiment Setup:

- Training Set was oversampled
- Random Forest Classifier with 250 Trees and 35 Random Features

Features				Number of Features	Baseline Accuracy	Results			
Token-N-Grams	POS-N-Grams	Lemma-N-Grams	Length			Accuracy	F-Measure Conclusion	F-Measure Premise	F-Measure Combined
x				263	75.63 %	88.5199 %	0.769	0.924	0.886
	x			694	75.63 %	93.1689 %	0.858	0.955	0.931
		x		313	75.63 %	88.6253 %	0.770	0.924	0.887
			x	2	75.63 %	82.7641 %	0.712	0.877	0.837
x	x			957	75.63 %	93.4641 %	0.865	0.957	0.935
	x	x		1007	75.63 %	93.1848 %	0.862	0.956	0.933
x	x		x	959	75.63 %	93.5062 %	0.867	0.957	0.935
	x	x	x	1009	75.63 %	93.4219 %	0.866	0.956	0.934
x	x	x	x	1272	75.63 %	93.1794 %	0.861	0.955	0.932



# Cross-Domain-Experiments

- Random Forest Classifier with 250 Trees and 35 Random Features
- Train on Debatepedia and evaluate on Student-Essays
- Train on Student-Essays and evaluate on Debatepedia

Training Set	Test Set	Labels	Features			Number of Features	Baseline Accuracy	Results			
			Token-N-Grams	POS-N-Grams	Length			Accuracy	F-Measure Conclusion	F-Measure Premise	F-Measure Combined
Debatepedia (Training)	Student-Essays (Full)	“majorclaim” as conclusion	x	x		957	62.93 %	19.8062 %	0.229	0.165	0.173
Debatepedia (Training)	Student-Essays (Full)	all claims as conclusion	x	x		957	62.93 %	42.3715 %	0.552	0.191	0.325
Debatepedia (Training)	Student-Essays (Full)	“majorclaim” as conclusion	x	x	x	959	62.93 %	28.7896 %	0.234	0.334	0.322
Debatepedia (Training)	Student-Essays (Full)	all claims as conclusion	x	x	x	959	62.93 %	47.3148 %	0.553	0.359	0.431
Student-Essays (Full)	Debatepedia (Training)	“majorclaim” as conclusion	x	x		845	75.64 %	75.2748 %	0.060	0.858	0.663
Student-Essays (Full)	Debatepedia (Training)	all claims as conclusion	x	x		845	75.64 %	78.63 %	0.481	0.865	0.772
Student-Essays (Full)	Debatepedia (Training)	“majorclaim” as conclusion	x	x	x	847	75.64 %	75.4435 %	0.058	0.859	0.664
Student-Essays (Full)	Debatepedia (Training)	all claims as conclusion	x	x	x	847	75.64 %	79.5235 %	0.510	0.871	0.783