

Argument Component Identification im Stile von Stab und Gurevych

Hugo Meinhof, 815220

March 1, 2024

Abstract

some abstract stuff goes here [Stab und Gurevych](#)

trainingsdaten

402 von Stab und Gurevych annotierte Schüleressays bilden die Grundlage für die Trainingsdaten dieses Projekts. Für jeden Essay enthält der Datensatz die Information, wo sich die Kernbehauptung des Textes - der MajorClaim, sonstige Behauptungen zum Stützen der zentralen Aussage (/Grundaussage)- die Claims - und die Prämissen befinden, die mittels Statistiken und anderen Belegen die These untermauern. Die verwendeten Texte stammen von [essayforum.com](#) und wurden zufällig ausgewählt. über die Autoren ist nichts weiter bekannt, allerdings ist es denkbar, dass Englisch nicht ihre Muttersprache ist und sie diese im schulischen Rahmen lernen. Stab und Gurevych versahen die Essays anschließend mit Tokens und speicherten, wo sich Spannen befinden.

Datenherkunft: Schüleressays/ learner Essays von nichtmuttersprachlern (englisch) Mehrere Essays zu selben Themen; von Stab und gurevych annotiert Von [essayforum.com](#); randomly selected

quelle: <https://aclanthology.org/J17-3005.pdf>

Claim: semantische klasse, die eine behauptung aufstellt; stützt majorclaim MajorClaim: kernbehauptung(en) des textes prämissen: unterstützung/ untermauerung der Claims (zb statistiken)

über baselines schreiben!

“For finding the best-performing models, we conduct model selection on our training data using 5-fold cross-validation”

die essays wurden tokenisiert und gespeichert wo sich spannen befinden, welche rolle sie haben, und welche tokens dazugehören. diese daten werden von renes script erstellt was war das. damit modeelle damit lernen können, muss ein dataset erstellt werden, welches die daten aufbereitet und dem modell verständlich sortiert. dies ist die größte aufgabe beim training. da alle trainierten modelle auf dem selben datensatz basieren, gibt es für alle zusammen ein gemeinsames dataset. viele extraktions, aufbereitungs, und matching schritte bleiben für alle modelle gleich. unterschiede gibt es im grunde nur im letzten aufbereitungsschritt. das wird von den verschiedenen configs des datasets gehandhabt. es gibt für jedes modell eine eigene config, die den selben namen trägt, wie das modell. da sich die modelle nur darin unterscheiden wie die trainingsdaten aufbereitet sind, bedeutet das auch, dass ein trainingsscript für alle modelle verwendet werden kann, in dem nur die config angepasst werden muss. ich habe zudem darauf

geachtet, dass die configs die selben namen tragen wie die modelle, damit alles reibungslos abläuft
 bessere erklärung des trainings scripts. das war jedoch nicht schon immer so. angefangen habe
 ich mit je einem trainings script pro modell. das ist zwar auf der einen seite nicht so anpassbar wie
 ein sript für alle, welches über command line arguments angepasst werden kann, hat jedoch auf
 der anderen seite den klaren vorteil, dass so ein einzelnes modell erstmal trainiert und ausgetestet
 werden kann, ohne dabei andere im hinterkopf behalten zu müssen.

1 evaluation

Model Epochs	Makro-f1		
full_labels		spans	
4	0.579	4	0.898
5	0.631	5	0.905
6	0.716	6	0.908
7	0.740	7	0.910
8	0.752	8	0.911
9	0.757	9	0.912
10	0.759	10	0.912
		simple	
		4	0.769
		5	0.782
		6	0.790
		7	0.796
		8	0.800
		9	0.801
		10	0.801
		sep_tok_full_labels	
		4	0.749
		5	0.801
		6	0.816
		7	0.824
		8	0.832
		9	0.837
		10	0.838
		sep_tok	
		4	0.843
		5	0.849
		6	0.858
		7	0.864
		8	0.864
		9	0.865
		10	0.867

Table 1: 5-fold cross-validation of the macro-f1

2 training

3 Stab und Gurevych zusammenfassung

Argument Component Identification	Makro-f1
Stab und Gurevytch	
Human upper bound	0.886
Baseline majority	0.259
Baseline heuristic	0.642
CRF all features	0.867
Meinhof	
spans	0.912

Table 2: Argument Component Identification (5-fold cross-validation)

Argument Component Classification	Makro-f1
Stab und Gurevytch	
Baseline majority	0.257
Baseline heuristic	0.724
SVM only structural	0.746
SVM all without prob and emb	0.771
SVM without genre-dependent	0.742
SVM all features	0.773
<i>SVM and CRF (all features)</i>	no-eval
Meinhof	
full_labels	0.759
simple	0.801
sep_tok_full_labels	0.838
sep_tok	0.867
<i>full_pipe</i>	<i>TO-DO</i>

Table 3: Argument Component Classification (5-fold cross-validation)