Give me More Feedback: Annotating Argument Persuasiveness and Related Attributes in Student Essays

Winston Carlile Nishant Gurrapadi Zixuan Ke Vincent Ng
Human Language Technology Research Institute
University of Texas at Dallas

ACL 2018

Automated Essay Scoring

- Current work's focus: **holistic** scoring, summarizing quality with one number
  - provides limited feedback to students
- A few attempts to address this problem by scoring a particular dimension of essay quality, such as coherence, technical errors, relevance to prompt, etc.
- Little work on scoring **argument persuasiveness** despite its being one of the most important dimensions of persuasive essay quality
  - Exception: Persing & Ng (2015)
  - Problems with P&N's persuasiveness-scored essay corpus
    - Only the "overall' argument was scored
    - The resulting score does not explain why the argument is (un)persuasive
      - Provides limited feedback to students on how to improve arguments

Goal

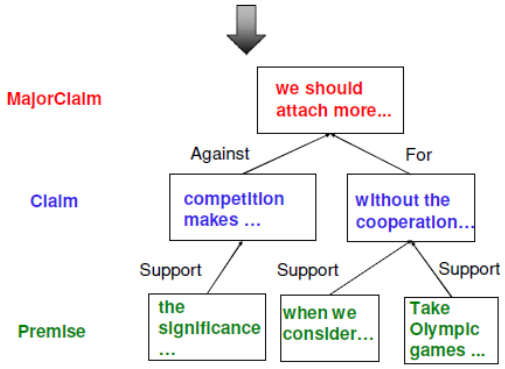
- Annotate a corpus of persuasive student essays that addresses the problems of P&N's corpus via designing annotation schemes and scoring rubrics
  - Score **each** argument's persuasiveness
  - Annotate the **attributes** of an argument that can impact its persuasiveness

Corpus

- 102 essays randomly chosen from the Argument Annotated Essays corpus
  - Each essay was annotated by Stab & Gurevych with an argument tree

Prompt: Should students be taught to compete or to cooperate?

… **we should attach more importance to cooperation during primary education.** First of all, …On the other hand, **the significance of competition is that**… Hence… **competition makes the society more effective**. However, **when we consider about the question that how to win the game**… **Take Olympic games for instance**… **Therefore without the cooperation there would be no victory of competition.**



Annotation

- **Definition:** for the purposes of our work, an argument is composed of a node in an argument tree and all of its children, if any
  - a non-leaf node can be interpreted as a conclusion supported/attacked by its children, which can be interpreted as evidences for the conclusion
  - a leaf node can be interpreted as an unsupported conclusion
- **Goal:** annotate each argument with its persuasiveness and a set of predefined attributes that could impact an argument's persuasiveness

Attribute	Possible Values	Applicability	Description
Persuasiveness	1-6	MC,C,P	How persuasive the argument is
Specificity	1-5	MC,C,P	How specific the statement is
Eloquence	1-5	MC,C,P	How well the idea is presented
Evidence	1-6	MC,C,P	How well the supporting statements support their parent
Logos/Ethos/Pathos	Yes,No	MC,C	Whether the argument uses the respective persusive strategy
Relevance	1-6	C,P	Relevance to the parent statement
ClaimType	Value,Fact ,Policy	C	The category of what is claimed
PremiseType		P	The type of premise
Strength	1-6	P	How well a single statement contributes to persuasiveness

Annotation Procedure

- Two human annotators who were both native speakers of English were first familiarized with the rubrics and definitions and then trained on five essays
- 30 essays were doubly annotated for computing inter-annotator agreement
- Each of the remaining essays was annotated by one of the annotators
- Score/Class distributions by component type:

Specificity	1	2	3	4	5
MC	0	73	72	32	8
C	80	259	155	59	14
P	64	134	238	173	98

Evidence	1	2	3	4	5	6
MC	3	62	57	33	16	14
C	246	115	85	80	35	6
P	614	28	12	26	15	12

Eloquence	1	2	3	4	5
MC	3	19	116	42	5
C	23	106	320	102	16
P	24	97	383	154	49

Persuasiveness	1	2	3	4	5	6
MC	3	62	60	28	17	15
C	82	278	84	74	39	10
P	8	112	145	249	123	70

ClaimType	Fact	Value	Policy
C	368	145	54

Relevance	1	2	3	4	5	6
C	1	33	58	132	97	246
P	5	45	59	145	147	306

PremiseType	Real example	Invented instance	Analogy	Testimony	Statistics	Definition	Common knowledge	warrant
P	93	53	2	4	15	3	493	44

Logos	Yes	No
MC	181	4
C	304	263

Pathos	Yes	No
MC	67	118
C	59	508

Ethos	Yes	No
MC	16	169
C	9	558

- Inter-annotator agreement (Krippendorff's alpha):

Attribute	MC	C	P
Persuasiveness	.739	.701	.552
Specificity	.560	.530	.690
Eloquence	.590	.580	.557
Evidence	.755	.878	.928
Relevance		.678	.555
Strength			.549
Logos	1	.842	
Pathos	.654	.637	
Ethos	1	1	
ClaimType		.589	
PremiseType			.553

- Persuasiveness agreement exhibits a downward trend as the component type narrows
- Evidence agreement exhibits an upward trends as the component type narrows
- Eloquence has one of the lowest agreement
- Specificity has low agreement in claims and major claims
- Relevance agreement for premises is one of the lowest

Analysis of Annotations

- To understand whether the attributes are useful for predicting persuasiveness, we compute the Pearson's Correlation Coefficient (PC) between Persuasiveness and each attribute along with the corresponding p-value
- Among the correlations that are significant at the $p < .05$ level, Persuasiveness is positively correlated with Specificity, Evidence, Eloquence, and Strength.
- Support in the form of statistics and examples is positively correlated with Persuasiveness
- Logos and invented_instance have significant correlations with Persuasiveness, but the correlation is weak

Attribute	PC	p-value
Specificity	.5680	0
Relevance	-.0435	.163
Eloquence	.4723	0
Evidence	.2658	0
Strength	.9456	0
Logos	-.1618	0
Ethos	-.0616	.1666
Pathos	-.0835	.0605
CType:Fact	.0901	.1072
CType:Value	-.0858	.1251
CType:Policy	-.0212	.7046
PType:real_example	.2414	0
PType:invented_instance	.0829	.0276
PType:analogy	.0300	.4261
PType:testimony	.0269	.4746
PType:statistics	.1515	0
PType:definition	.0278	.4608
PType:common_knowledge	-.2948	0
PType:warrant	.0198	.6009

- **Oracle experiment**: to understand how well these attributes, when used together, can explain persuasiveness, we train 3 linear SVM regressors, one for each component type, to score an arguments persuasiveness using gold attribute's as features
- Five-fold cross validation results (in terms of PC and ME (mean absolute error) show that they largely can

	MC	C	P	Avg
PC	.9688	.9400	.9494	.9495
ME	.0710	.1486	.0954	.1061