Supplemental Material for IJCAI-18 Submission 4342: Predicting Argument Persuasiveness in Student Essays

Abstract

This document contains the supplemental material for IJCAI-18 submission 4342: Predicting Argument Persuasiveness in Student Essays.

1 Introduction

This supplemental document describes our annotation procedure, including the annotation schemes and rubrics, in detail. Note that the submitted paper contains a concise summary of this procedure.

2 Annotation

2.1 Definition

As mentioned in the paper, since persuasiveness is defined on an argument, in order to annotate persuasiveness we need to define precisely what an argument is. We borrow the definition from Argumentation Theory [van Eemeren *et al.*, 2014], which defines an argument as consisting of a conclusion that is supported or attacked by a set of evidences. Given an argument tree, a non-leaf node can be interpreted as a "conclusion" that is supported or attacked by its children, which can therefore be interpreted as "evidences" for the conclusion. Hence, for the purposes of our work, an argument is composed of a non-leaf node in an argument tree and all of its children.

Each argument tree in our dataset has three to four levels. The root is a MajorClaim. Each node in the second level is a Claim that supports or attacks its parent (i.e., the MajorClaim). Each node is the third level is a Premise that supports or attacks its parent (i.e., a Claim). There is an optional fourth level consisting of nodes that correspond to Premises. Each of these Premise nodes either supports or attacks its (Premise) parent. Given the aforementioned definition of an argument and the argument trees in our dataset, an argument can be composed of (1) a MajorClaim and a set of supporting/attacking Claims; (2) a Claim and a set of supporting/attacking Premises; or (3) a Premise and a set of supporting/attacking Premises.

2.2 Annotation Scheme

Recall that the goal of our annotation is to score each argument w.r.t. its persuasiveness (see Table 1 for the rubric for scoring persuasiveness) and annotate each of its components with a set of predefined attributes that could impact the argument's persuasiveness. These attributes are motivated by theoretical work on argument persuasiveness [Petty and Cacioppo, 1984; Connor, 1990; Connor and Lauer, 1985] and Argumentation Theory [van Eemeren *et al.*, 2014]. Table 2 presents a summary of the attributes we annotate. The rest of this subsection describes these attributes and the rubrics we use to score/annotate them in detail.

Each component type (premise, claim, majorclaim) has a distinct set of attributes. All component types have two attributes in common: eloquence and specificity. *Eloquence* is how well the author uses language to convey ideas, similar to clarity and fluency. The presentation of an idea plays a part in persuasion separate from the idea itself. All other things equal, a consistently eloquent essay is more persuasive than a poorly written essay that makes the same argument using the same supporting ideas. *Specificity* refers to the narrowness of a statement's scope. Statements that are specific are more believable because they indicate an author's confidence and depth of knowledge about a subject matter. Statements that are believable are more persuasive than those which are easily doubted by a non-expert. Argument assertions (majorclaims and claims) need not be believable on their own since that is the job of the supporting evidence. For these argument components we interpret specificity to arise from the summarization of major points that will be used, and the explicit condition/extent to which the assertion holds true. The specificity of an argument assertion has a direct affect on the persuasiveness of the argument it represents. The rubrics for scoring Eloquence, Claim/MajorClaim Specificity, and Premise Specificity are shown in Table 3, 4, and 5, respectively.

MajorClaim The majorclaim acts as the representative of the entire essay. When a majorclaim is persuasive, this means that the argument as a whole is persuasive. In the argument annotated essays, sometimes there is more than one majorclaim because the author presented the thesis at the beginning and restated it at the end. In theory, these separate sections of the text should represent the same component in the argument tree, however, in some essays that have more than one majorclaim the statements are similar but not semantically equivalent. For this reason, sentences marked as majorclaims are treated individually with respect to the supporting claims.

Score	Description
6	A very strong, clear argument. It would persuade most readers and is devoid of errors that might detract from its strength or make it difficult to understand.
5	A strong, pretty clear argument. It would persuade most readers, but may contain some minor errors that detract from its strength or understandability.
4	A decent, fairly clear argument. It could persuade some readers, but contains errors that detract from its strength or understandability.
3	A poor, understandable argument. It might persuade readers who are already inclined to agree with it, but contains severe errors that detract from its strength or understandability.
2	It is unclear what the author is trying to argue or the argument is poor and just so riddled with errors as to be completely unpersuasive.
1	The author doesn't appear to make any argument (e.g. he may just describe some incident without explaining why it is important). It could not persuade any readers because there is nothing to be persuaded of. It may or may not contain detectable errors, but errors are moot since there is not an argument for them to interfere with.

Table 1: Description of the Persuasiveness scores.

Attribute	Possible Values	Applicability	Description
Relevance	1–6	C,P	The relevance of the statement to the parent statement.
PremiseType	see Section 4.2	P	The type of Premise, e.g. statistics, definition, real example
Logos/Pathos/Ethos	yes,no	M,C	Whether the argument uses the respective persuasive strategy
ClaimType	Value,Fact,Policy	C	The category of what is being claimed
Specificity	1–5	M,C,P	How detailed and specific the statement is.
Eloquence	1–5	M,C,P	How well the idea is presented. Influenced by grammar, vocabulary, and sentence structure.

Table 2: Summary of the list of attributes together with their possible values, the argument component type(s) each attribute is applicable to (**M**: MajorClaim, **C**: Claim, **P**: Premise), and a brief description.

Score	Description
5	Demonstrates mastery of English. There are no grammatical errors that distract from the meaning of the sentence. Exhibits a well thought out, flowing sentence structure that is easy to read and conveys the idea exceptionally well.
4	Demonstrates fluency in English. If there are any grammatical or syntactical errors, their affect on the meaning is negligible. Word choice suggests a broad vocabulary.
3	Demonstrates competence in English. There might be one or two errors that are noticeable but forgivable, such as an incorrect verb tense or unnecessary pluralization. Demonstrates a typical vocabulary and a simple sentence structure.
2	Demonstrates poor understanding of sentence composition and/or poor vocabulary. The choice of words or grammatical errors force the reader to reread the sentence before moving on.
1	Demonstrates minimal eloquence. The sentence contains errors so severe that the sentence must be carefully analyzed to deduce its meaning.

Table 3: Description of the Eloquence scores.

Score	Description
5	Summarizes the argument well and has a qualifier that indicates the extent to which the claim holds true. Claims that summarize the argument well must reference most or all of the supporting entities.
4	A Claim that scores a 4 must summarize the argument very well by mentioning most or all of the supporting entities, but does not have a qualifier indicating the conditions under which the claim holds true. Alternatively, the Claim may moderately summarize the argument by referencing a minority of supporting entities and contain qualifier.
3	A Claim with specificity 3 has a qualifier clause or references a minority of the supporting entities, but not both.
2	A Claim with specificity 2 does not make an attempt to summarize the argument nor does it contain a qualifier clause.
1	Simply rephrases the Majorclaim or is outside scope of the Majorclaim (entities were annotated incorrectly: Majorclaim could be used to support Claim).

Table 4: Description of the Claim and MajorClaim Specificity scores.

Since the majorclaim represents the entire argument of the essay, it is in this component that we annotate the persuasive strategies employed (i.e. *ethos, pathos* and *logos*). These attributes, ethos, pathos, logos, and evidence, are not inherent to the text identifying the majorclaim but instead summarize the child components in the argument tree. Different people are persuaded in different ways: some are persuaded by logic, some by emotion, and some by trust in a higher authority. In order to appeal to the broadest audience, usage of multiple persuasive strategies improves persuasiveness.

Score	Description
5	A score of 5 indicates an elaborate, very specific statement. The statement contains numerical data, or a historical example from the real world. There is (1) both a sufficient qualifier indicating a the extent to which the statement holds true and an explanation of why the statement is true, or (2) at least one real world example, or (3) a sufficient description of a hypothetical situation that would evoke a mental image of the situation in the minds of most readers.
4	A score of 4 indicates a more specific statement. It is characterized by either an explanation of why the statement is true, or a qualifier indicating when/to what extent the statement is true. Alternatively, it may list examples of items that do not qualify as historical events.
3	A score of 3 indicates a sufficiently specific statement. It simply states a relationship or a fact with little ambiguity.
2	A broad statement. A statement with weak qualifiers such as "maybe," "might," and "sometimes," will receive a score of 2 without other redeeming factors such as explicit examples, or elaborate reasoning. Additionally, there are few adjectives or adverbs.
1	An extremely broad statement. There is no underlying explanation, qualifiers, or real-world examples.

Table 5: Description of the Premise Specificity scores.

Score	Description
6	Anyone can see how the support relates to the parent claim. The relationship between the two entities is either explicit or extremely easy to infer. The relationship is thoroughly explained in the text because the two entities contain the same words or exhibit coreference.
5	There is an implied relationship that is obvious, but it could be improved upon to remove all doubt. If the relationship is obvious, both relating entities must have high eloquence and specificity scores.
4	The relationship is fairly clear. The relationship can be inferred from the context of the two statements. One entity must have a high eloquence and specificity scores and the other must have lower but sufficient eloquence and specificity scores for the relationship to be fairly clear.
3	Somewhat related. It takes some thinking to imagine how the entities relate. The parent entity or the child entity have low clarity scores. Two statements about the same topic but unrelated ideas within the domain of said topic would get a score of 3.
2	Mostly unrelated. It takes some major assumptions to relate the two entities. An entity may also receive this score if both entities have low clarity scores.
1	Totally unrelated. Very few people could see how the two entities relate to each other. The statement was annotated to show that it relates to the claim, but this was clearly in error.

Table 6: Description of the Relevance scores.

Type	Description
Fact	The claim states that something is true or false.
Value	The claim states that something is important, or not important, or has some other value attached to it.
Policy	The claim states that a certain law or rule should be implemented and enforced.

Table 7: Description of the Claim Types.

Claim The claim argument component possesses all of the attributes of a majorclaim in addition to a *relevance* score and a *claim_type*. In order for an argument to be persuasive, all supporting components must be relevant to the component that they support/attack. If the reader cannot tell that a statement is related to the assertion that the author intended it to support, then they cannot be persuaded. The claim_type can be *value,* (e.g. something is good or bad, important or not important, etc.) *fact,* (e.g. something is true or false) or *policy* (claiming that some action should or should not be taken). The scoring rubric for Relevance and the definitions of the claim types are shown in Tables 6 and 7, respectively.

Premise The only attribute exclusive to premises is the premise_type. This attribute takes on a discrete value from one of the following: real_example, invented_instance, analogy, testimony, statistics, definition, common_knowledge, warrant and opinion. Analogy, testimony, statistics, and definition are self-explanatory. A premise is labeled *invented_instance* when it describes a hypothetical situation,

and *definition* when it provides a definition to be used elsewhere in the argument. A premise has type *warrant* when it does not fit any other type, but serves a functional purpose to explain the relationship between two entities or clarify/quantify another statement. The *real_example* premise type indicates that the statement is a historical event that actually occurred, or something that is verfiably true about the real world. A premise has type *opinion* when the statement obviously needs further backing and could easily be the thesis of it's own persuasive essay. Loosely speaking, the most persuasive premise_types are real_example, statistics because they are associated with high specificity by their nature. The absolute weakest premise_type is the opinion due to its nature of being ungrounded, but this may be improved if the premise has its own supporting premises. The premise types are described in Table 8. Note that a premise does not have attributes that are associated with claims, including claim_type, logos, pathos, and ethos.

Table 9 summarizes the list of attributes associated with each of the three component types.

Type	Description
real_example	A historical example of something that actually happened, or a specific, non-generic, statement that is verifiably true about the real world.
invented_instance	A hypothetical situation that did not actually occur.
analogy	Stating one phenomenon or relationship is comparable to another well-known concept.
testimony	A quote from, or reference to a higher authority.
statistics	Raw numerical data or a quantitative comparison of values.
definition	An explicit definition of a term/concept.
common_knowledge	A conjecture or generalization that the author assumes most people would accept as true.
warrant	Performs one of the following functions: (1) restates the parent claim; (2) expands/upon clarifies the parent claim; or (3) explains how a sibling premise relates to the parent claim.
opinion	A controversial statement that would be unanimously accepted as an opinion. Could easily be the thesis of its own persuasive essay.

Table 8: Description of the Premise Types.

MajorClaim - Logos/Pathos/Ethos, Eloquence, Specificity
Claim - ClaimType, Relevance, Logos/Pathos/Ethos, Eloquence, Specificity
Premise - PremiseType, Relevance, Eloquence, Specificity

Table 9: Summary of the list of attributes associated with each argument component type.

2.3 An Example

To better understand our annotation scheme, we use the essay in Table 10 to illustrate how we obtain the attribute values in Table 11.

After reading the essay in its entirety and acquiring a holistic impression of the argument's strengths and weaknesses, we begin annotating the atomic argument components bottom up: starting with the leaf nodes of the argument tree. First we consider premise **P3** and score the eloquence, specificity and premise_type. This sentence has no serious grammatic or syntactic errors, has a flowing, well thought out sentence structure, and uses articulate vocabulary. As a result, **P3** gets an eloquence score of 5. Next, we consider how specific this statement is. **P3** is essentially saying that poor primary education causes illiteracy and consequently inhibits a country's development. It does not state why, or to what extent so we cannot assign a score of 4. However, it does explain a simple relationship with little ambiguity due to the lack of weak qualifiers (e.g. "maybe," "might," "sometimes") so we can assign a score of 3. The final attribute that depends on **P3** and only **P3** is the premise_type. The best premise_type for **P3** is common_knowledge because it is reasonable to assume most people would agree that poor primary education causes illiteracy, and also that illiteracy inhibits a country's development. In more formal contexts this could be classified as an opinion, but since the author is using it as an atomic unit of evidence and shows consideration for the knowledge base of the target audience, we label it as common_knowledge. Next, we reread the parent component **C1** and score the relevance of **P3**. In this case, the relationship is explicitly clear because the two components exhibit coreference; **P3** contains a reference to primary/elementary education and shows how this affects a country's inability to "take off," or transition from developing to developed. Consequently we assign a relevance score of 6 to **P3**.

Next, the annotator would score **P1** and **P2** in a similar manner, but for demonstration purposes we will now examine the scoring of **C1**. The text in **C1** shows fluency, broad vocabulary, and attention to how well the sentence structure reads so we assign an eloquence score of 5. This claim specifically says that the government should put something on their agenda, so we mark the claimtype as "Policy." The next attributes scored are derived from the combination of **C1** and its children. Specificity for a claim means how well it explains the assertion and how well it summarizes the supporting statements. **C1** contains information relevant to all the child premises: creating qualified citizens, whose role it is to provide the education, and the effect of education on a country's development. It does not contain a qualifier stating the extent to which the assertion holds true, so this claim fits the specificity score descriptor of a 4. There is no indication of an emotional appeal nor an appeal to authority of any sort, but the author attempts to use logical reasoning in order to prove their point. Hence, we mark logos as true, pathos as false, and ethos as false. The persuasiveness of a claim is mainly determined by the strength of the supporting evidence, Given that **P2**'s offers weak support and the other two premises offer decent support, this argument could persuade some reader and thus deserves a persuasiveness score of 4. Next, the parent component is factored in to the relevance score. **C1** is concerned with the endorsement of elementary education and the parent **M1** is concerned with the endorsement of both elementary and university education. Anyone can see how endorsement of elementary education relates to the endorsement of elementary education, so **C1**'s relevance is a 6.

Finally, after all of the claims have been annotated in the bottom-up method, the annotator moves on to the major claim(s). Like all components, the eloquence is among the first to be scored. **M1** shows fluency and a large vocabulary, but is terse and does not convey the idea exceptionally well so we score the eloquence as 4. The argument strategies logos, pathos, and ethos are obtained by simply taking the logical disjunction of argument strategies used in child claims. Every claim in this essay relied on logos and did not employ pathos nor ethos, so **M1** is marked with logos as true, pathos

Essay #008

Prompt: Government budget focus, young children or university?

"Education is the investment in our future", as Senator Bernie Sanders once pointed out, education, which fosters a civilized society, plays a significant role in a country's long-lasting prosperity. It is no wonder that governments throughout the world lay special emphasis on education development. As for the two integral components within the system, elementary and advanced education, there's no doubt that a government is supposed to offer sufficient financial support for both.

Concerning that elementary education is the fundamental requirement to be a qualified citizen in today's society, government should guarantee that all people have equal and convenient access to it. So a lack of well-established primary education goes hand in hand with a high rate of illiteracy, and this interplay compromises a country's future development. In other words, if countries, especially developing ones, are determined to take off, one of the key points governments should set on agenda is to educate more qualified future citizens through elementary education.
...

Table 10: An example essay. Owing to space limitations, only the first two paragraphs of the essay are shown.

		P	E	S	R	Lo	Pa	Et	cType	pType
M1	government is supposed to offer sufficient financial support for both	3	4	2		T	F	F		
C1	if countries, especially developing ones, are determined to take off, one of the key points governments should set on agenda is to educate more qualified future citizens through elementary education	4	5	4	6	T	F	F	Policy	
P1	a lack of well-established primary education goes hand in hand with a high rate of illiteracy, and this interplay compromises a country's future development	4	5	3	6					A
P2	government should guarantee that all people have equal and convenient access to it	2	3	1	6					O
P3	elementary education is the fundamental requirement to be a qualified citizen in today's society	4	5	3	6					C

Table 11: The argument components in the example in Table 13 and the scores of their associated attributes: **P**ersuasiveness, **E**loquence, **S**pecificity, **R**elevance, **Lo**gos, **Pa**thos, **Et**hos, **E**vidence, **c**laim**Type**, **p**remise**Type**

as false, and ethos as false. Next, the annotator grades the specificity of **M1** which does not have a quantifier nor does it attempt to summarize the main points of the evidence. This results in a specificity of 2. The supporting claims rely exclusively on logos, so there is no added persuasiveness from a variety of strategies (If one claim used a diversity of argument strategies, this would be reflected in the persuasiveness of that claim). The persuasiveness score is 3, the reason being that while the support provided by **C1** is reasonably strong, **M1** has two other supporting claims (not shown in the excerpt) that offer weak to poor support.

3 Data

Data samples can be accessed via the following link: `https://raw.githubusercontent.com/BCD2018/BCD2018.github.io/master/Essays.zip`. The entire dataset will be made publicly available when the paper is published.

References

[Connor and Lauer, 1985] Ulla Connor and Janice Lauer. Understanding persuasive essay writing: Linguistic/rhetorical approach. *Text-Interdisciplinary Journal for the Study of Discourse*, 5(4):309–326, 1985.

[Connor, 1990] Ulla Connor. Linguistic/rhetorical measures for international persuasive student writing. *Research in the Teaching of English*, pages 67–87, 1990.

[Petty and Cacioppo, 1984] Richard E Petty and John T Cacioppo. The effects of involvement on responses to argument quantity and quality: Central and peripheral routes to persuasion. *Journal of personality and social psychology*, 46(1):69, 1984.

[van Eemeren *et al.*, 2014] F. H. van Eemeren, B. Garssen, E. C. W. Krabbe, F. A. Snoeck Henkemans, B. Verheij, and J. H. M. Wagemans. In *Handbook of Argumentation Theory*. Springer, Dordrecht, 2014.