

Debate Outcome Prediction using Automatic Persuasiveness Evaluation and Counterargument Relations

Daiki Shirafuji^{1*}, Rafal Rzepka^{1,2} and Kenji Araki¹

¹Graduate School of Information Science and Technology, Hokkaido University, Japan

²RIKEN Center for Advanced Intelligence Project (AIP)

{d.shirafuji, rzepka, araki}@ist.hokudai.ac.jp

Abstract

Debates play an important educational role and proper argumentation has a power to change people's stance on a given topic. Existing NLP research on persuasiveness of argumentation calculates it for single arguments or ranks arguments by their level of conviction. Our work extends this research by considering counterarguments. They can weaken or strengthen persuasiveness of a given argument, hence we propose novel methods to calculate persuasiveness with opinions existing in opposite stances. We create a corpus from a site containing debates where users evaluate discussions and choose winning side of a debate, allowing the calculation of opinion change. We experimentally confirmed 60.12% accuracy in the proposed debate outcome prediction task proving that additional counterargument-related information is capable to improve baseline methods.

1 Introduction

In recent years, the area of argumentation mining has become popular [Green *et al.*, 2014; Al-Khatib *et al.*, 2016; Stab *et al.*, 2018]. This topic spreads from extracting argument components (e.g. claims, premises) to predicting persuasiveness of an argument [Van Eemeren *et al.*, 2014; Cabrio and Villata, 2018]. Recent works on persuasiveness address student essays [Farra *et al.*, 2015] and debate's argument [Persing and Ng, 2017].

During a debate, various real life problems are discussed, such as whether to support death penalty or to ban guns. One of the purposes of a debate is to determine which side has won and to provoke forming richer opinions on societal issues. Human beings judge the winning side of a debate generally by assessing argument persuasiveness and counterarguments. However, researchers have concentrated only on counterargument retrieval [Wachsmuth *et al.*, 2018], predicting the persuasiveness of argumentation [Persing and Ng, 2017] or comparing persuasiveness of a pair of arguments [Habernal and Gurevych, 2016] in debate processing. To the best of authors' knowledge, works on counterarguments did

not consider the persuasiveness of both argument and counterargument, and research on persuasiveness has not yet considered how persuasiveness is affected by a counterargument, hence researchers have not dealt with several arguments within a particular debate. Therefore, existing methods are not capable of automatic debate winner selection. This problem is crucial because the debate is a competition of comparing persuasiveness between attacking and defending sides. Moreover, existing methods for estimating persuasiveness cannot address real life problems because these methods do not consider whether an argument is rebutted or not.

Researchers have not dealt with above-mentioned problems, and this paper proposes and studies a task of predicting debate outcome for evaluating persuasiveness of several arguments. Predicting debate outcome means automatically selecting the winner from Pro side (*For*¹) and Con side (*Against*) of a debate on a given topic with counterargument relations. Usually one debate consists of several arguments, and there is a need to compare not only pairs of arguments within one side as in previous works, but all arguments separated into two camps (*For* and *Against*) to be capable to predict the winning side. Some researchers worked on predicting debate outcome task before [Potash and Rumshisky, 2017], but their research did not consider persuasiveness of each argument. Moreover, they only focused on the final audience poll, which can become the bias. Therefore, we propose a method for debate outcome prediction considering arguments persuasiveness and the audience bias to debate themes.

For this task, we provide a new corpus of 321 debates with third party evaluation, retrieved from idebate.org site presenting debates on various topics where anyone can vote for more, in their opinion, convincing argumentation.

In short, main contributions of this article are:

- (1) A corpus for judging the winner side;
- (2) Appropriate task setting for debate outcome prediction;
- (3) Methods for the proposed task.

2 Related Work

In an essay, a debate or a discussion, the most important issue is whether the argumentation can change people's stance or not [Tan *et al.*, 2016]. Therefore, argumentation mining has become focused on persuasion of an

*Contact Author

¹We expressed the words we defined originally in *italic letter*.

argumentation [Carlile *et al.*, 2018; Persing and Ng, 2017; Durmus and Cardie, 2018]. Persuasion research concentrates on debates and discussions, and makes effort to measure the absolute value of persuasiveness [Wei *et al.*, 2016] and to compare persuasiveness of an argument pair [Habernal and Gurevych, 2016; Hidey and McKeown, 2018] or rank arguments persuasiveness [Tan *et al.*, 2016; Cano-Basave and He, 2016].

However, argumentation itself is not enough to determine the persuasiveness because arguments are usually being rebutted. In order to achieve better winning side prediction, we need to consider counterarguments [Habernal *et al.*, 2018]. Counterargument retrieval is the major task in argumentation mining that identifies *attack* or *support* relations of arguments [Cocarascu and Toni, 2017]. In those studies, it was usual to use prior topic knowledge, but [Wachsmuth *et al.*, 2018] proposed methods independent of knowledge. We have to consider counterarguments prediction in the debate outcome prediction task.

There are several studies for predicting outcomes of a debate. However, most of them focused on only one or two themes [Strapparava *et al.*, 2010]. Those research cannot be applied to general debate in terms of persuasiveness. To the authors’ best knowledge, Potash and Rumshisky are the first to study general debate outcome prediction [Potash and Rumshisky, 2017]. They achieved the best accuracy (71%) using a Recurrent Neural Network with Attention architecture. The dataset they used is debates only with the final favorability of audience. It is visible which stance, *For* or *Against*, is more supported with the final favorability, but audience supports is prone to bias or influenced by preconceptions for a given debate theme. Their research did not consider the audience bias, which has a large effect on debate results. It is preferable to compare the final audience favorability with the audience favorability before a debate because audience usually has an opinion which changes during the actual debate. In addition, Potash and Rumshisky did not consider persuasiveness of each argument, which plays a significant role in debates. To tackle these problems, we decided to deal with debate outcome prediction with argument persuasiveness and audience favorability before and after a debate.

3 Corpus

This section introduces two corpora which we use for our proposed task. First one is used in [Persing and Ng, 2017] for predicting persuasiveness level of an argument. Second one is for debate outcome prediction task. Our proposed method for this task is novel, hence no corpus with annotated winner’s side (before/after debating) had existed before. Therefore, we constructed a new corpus, automatically retrieved from idebate.org. This corpus contains a set of debates with the debate outcomes. Besides, we define “winner” in debate as the side which persuaded more people than the other side.

3.1 Persing and Ng’s corpus

Persing and Ng’s corpus contains a subset of 165 debates extracted from idebate.org. Each debate includes *Motion* which expresses the debate theme, and has 7.3 arguments on average (1,208 arguments in total). Argument is an opinion on

the *Motion*, and every argument belongs to a stance (*For* or *Against*). Arguments are divided into *Assertions* and *Justifications*. *Assertion* is the debater’s main opinion written about the reason why this person agrees or disagrees with *Motion*. *Justification* explains the *Assertion* in detail usually with references and logical explanation. Table 1 shows an example of argument divided into components. In their corpus, Persing and Ng annotated arguments with Argument Persuasiveness (AP).

AP is the persuasiveness score of arguments on a 6 point rating scale, where 6 indicates that the argument is very persuasive and clear, while 1 means that it is an unclear argument. For example, AP of the argument shown in Table 1 is 6.

3.2 Our Corpus

We retrieved the debate data from idebate.org, acquiring 321 debate themes (*For*: 148 and *Against*: 173), almost twice more than in the case of corpus of [Persing and Ng, 2017] and not overlapped with their corpus. Each debate theme has 7.55 arguments on average. On idebate.org, the third party evaluation is performed by the site visitors for almost all of the debates, and includes opinion rating before and after reading a debate and these results are open to the public. There are five evaluation categories: Strongly *For* (SF), Mildly *For* (MF), Don’t Know (DK), Mildly *Against* (MA) and Strongly *Against* (SA). The evaluation from the visitors is shown as a percentage, and the evaluation of a debate on the topic “This House believes Tennessee is correct to protect teachers who wish to explore the merits of creationism” is shown in Table 2 as an example.

From the data we obtain the debate outcome with the following equation.

$$2 * SF + MF - (2 * SA + MA) \quad (1)$$

To estimate whether arguments in a debate have changed people’s stance or not, we subtract “before” from “after” values (Equation (1)), and if the result is larger than zero, the *For* side is assigned as the debate’s winner. For example, if we substitute Equation (1) with values from Table 2, *For* side wins because the result will be larger than zero, even though the percentage of *Against* side is bigger than *For*.

We also show arguments used in this example in Table 3.

4 Proposed Methods

Our method for the task of debate outcome prediction is divided into the following three steps: AP estimation, argument similarity and discourse parsing. In these steps, we treat persuasiveness estimation of a debate as a main source for the result prediction via SVM [Cortes and Vapnik, 1995].

AP estimation We calculate AP in the range of 1 to 6 for arguments in our corpus using the Persing and Ng’s corpus as train data and by employing Support Vector Machine (SVM) with ten features, which produced better results than the method of [Persing and Ng, 2017]. The ten features are as follows: number of grammar errors, subjectivity indicators, number of first plural pronouns, number of citations, number of content lemmas only in *Justification*, *Assertion*

<i>Motion</i>	This house believes Quebec should secede from Canada
<i>Assertion</i>	International Law Mandates Quebec be allowed Independence
<i>Justification</i>	International law recognizes Quebec’s right to self-determination and denying them self-determination is therefore a violation of international law. International law recognizes the right of all peoples to self-determination. The international community has decided that it is oppressive to individuals to live under a government that is systematically incapable or unwilling to protect them and their interests.[1] The Quebecois have been systematically denied adequate representation in the federal government of Canada. Quebecois legislation protection their basic rights to retain their language and culture have been met with contempt[2] and legal action by the federal Canadian government and courts.[3] This is but one example of the very clear denial of basic representation and self-governance that afflicts the Quebecois in Canada. Therefore, Quebec has the legal right to self-determination and independence in international law. [1] “Reference re Secession of Quebec”, Supreme Court of Canada, 1998, 2 S.C.R. 217, < http://scc.lexum.org/en/1998/1998scr2-217/1998scr2-217.html > [2] “Maxime Bernier on Quebec law: ‘We don’t need Bill 101’”, The Canadian Press, 4 February 2011, < http://www.ctv.ca/CTVNews/Canada/20110204/bernier-law-110204/ > [3] Hudon, R., „Bill 101”, The Canadian Encyclopedia, < http://www.thecanadianencyclopedia.com/index.cfm?PgNm=TC&Params=A1ARTA0000744 >

Table 1: An argument components (*Motion*, *Assertion* and *Justification*)

	SF	MF	DK	MA	SA
Before	20%	11%	3%	9%	57%
After*	25%	13%	0%	0%	63%

* The percentage of After is not equal to 100 due to rounding off numbers

Table 2: The percentage of *For* and *Against* votes

length, number of content lemmas only in *Assertion*, number of words in *Justification*, number of subject matches in discourse relation (between two sentences in an argument), and number of transitional phrases in *Justification*.

To confirm the error rate of AP estimation, we performed 10-fold cross-validation with Persing and Ng’s corpus, and the AP estimation results are evaluated with two scoring metrics: E, which is the error rate, and ME, which measures the mean distance between a prediction and the correct value of AP. Persuasiveness calculated with data from Persing and Ng’s corpus resulted with E=0.64 and ME=1.18 when only AP estimation was used. This result may be capable for predicting debate outcome because the mean distance between a prediction and the correct value of AP is approximately 1 (the range of ME is 0 to 5).

Argument similarity For retrieving counterarguments against an argument, we use cosine similarity over word2vec [Mikolov *et al.*, 2013] trained with Google News dataset between the *Justification* in an argument without stop words and arguments in the opposing side within the same debate. In order to obtain the best threshold of cosine similarity, we tested all of our proposed methods while changing the threshold for the debate outcome prediction task using our corpus. We changed the threshold within 0.325 and 0.775 in increments of 0.025. As a result, 0.55 performed the best, therefore we set the threshold to this value.

Discourse parsing We assumed that counterarguments can be partially discovered with the additional help of transitional phrases such as “but” or “because” appearing in *Justification*. Therefore, we extract a sentences proceeding and follow-

ing those transitional phrases using PDTB-styled discourse parser [Lin *et al.*, 2014].

4.1 Basic SVM-based Approach

In this method, we utilize AP in the debate outcome prediction with SVM (default parameters).

The input to SVM is a vector of AP. This vector is derived from For APs and Against APs. We firstly make one vector with eight elements: sorted For APs, and sorted Against APs. In the next step, two vectors are concatenated. Finally, we input the vector, which length is 16, and the result is calculated. If the number of *For/Against* arguments is smaller than eight, the vector was padded with zeroes.

Results are computed with 4-cross validation, and this evaluation procedure is identical in other experiments described below.

4.2 Similarity-based Approach

In this approach, we only use AP estimation and argument similarity calculation. At first, the method estimates AP for all arguments in a debate, and when an argument is analyzed and its AP is lower than the AP of other arguments in the opposing side and argument similarity is over the threshold, we multiply AP of the argument with a value, automatically decided as follows. A certain value changes by d which is the distance between AP of the input argument and remaining arguments. To obtain the highest accuracy of debate outcome task, we tested this approach with following value $V[d]$ which is calculated by Equation (2).

$$V[d] = SV - i * (d - 1) \quad (2)$$

where SV is the standard value within 0.8 and 1.0 in increments of 0.01, and i is the intervals: 0.0025, 0.005, 0.01, 0.0125, 0.015, 0.0175 and 0.02. The best results were achieved when $SV=1$ and $i=0.0025$, so that $V[1]=1.0$ and $V[5]=0.96$. For example, in the case where the argument’s AP equals 2, the similarity with the other argument is over the threshold and AP is 5, $V[5-2]=0.98$ is multiplied by the input argument’s AP: 2.

<i>Motion</i>	This House believes Tennessee is correct to protect teachers who wish to explore the merits of creationism
<i>For Assertion 1</i>	Freedom of speech should apply to teachers as much as anyone else
<i>For Assertion 2</i>	Teaching creationism as well as evolution gives students freedom to choose
<i>For Assertion 3</i>	The bill does not exclude evolution just allows room for other theories
<i>Against Assertion 1</i>	Teachers should not have freedom to teach whatever they wish as fact
<i>Against Assertion 2</i>	Children should have the freedom not to be misled
<i>Against Assertion 3</i>	Tennessee is not seeking to protect freedom of speech
<i>Against Assertion 4</i>	As it is not science creationism should not even be covered by the Tennessee law

Table 3: An example of *For* and *Against* assertions for one theme

In addition, we tested the case of $i=0$, which means $V[d]$ does not depend on d , for comparison. In this case, the highest result was achieved when SV was equal 0.99.

4.3 PDTB-styled Approach 1 (PDTB1)

In this approach, we add discourse parsing to the Similarity Approach. In Similarity Approach, we get the cosine similarity between an argument’s *Justification* and other argument’s *Justification*, but in this approach, we extract sentences which are in discourse relations with discourse parsing, and calculate the similarity between an argument’s *Justification* and other’s sentences which are in discourse relations. The best case of $V[d]$ is when $SV=0.92$ and $i=0.0025$. In addition, we tested the case of $i=0$ for comparison. The best accuracy was achieved when SV was equal 0.92.

4.4 PDTB-styled Approach 2 (PDTB2)

In this approach, in addition to discourse parsing, we calculate similarity of the first sentence in the *Justification* of the other side argument because we assume that the first sentence of *Justification* may be mentioned in the counterargument. The best prediction results for $V[d]$ are achieved when $SV=0.95/0.96$ and $i=0.0025$. Additionally, we performed an experiment for the case of $i=0$ and $SV=0.93$ obtained the most accurate estimation.

5 Evaluation Results

Researchers have not dealt with the task of debate outcome prediction yet, hence three prediction methods using statistics are proposed as baselines. First one is the median baseline (a). This method predicts the debate winner by comparing median of AP in *For* and *Against* stances. Two remaining baselines use average (b) and summation (c) instead of median.

Table 4 shows the accuracy of each system described earlier. Variables (SV and i) are selected to achieve the best accuracy. The rows indicating $i=0$ mean that the value which is multiplied with AP is not altered by the distance of AP. The best result is 60.1% achieved by PDTB2 method when $SV=0.93$ and $i=0$. All of our proposed methods are improved by approximately 0.07. Moreover, PDTB1 showed accuracy superior to the Similarity Approach and PDTB2 accuracy is better than in the case of PDTB1. Therefore, it can be said that the discourse relations are beneficial for comparison of persuasiveness of an argument and the counterarguments. However, the accuracy of all methods is higher when the value of i is 0 than in the case where $i < 0$. Moreover, the value of i which provides the best results are 0.0025, hence

Method	Parameters	Accuracy
Median	None	0.4953
Average	None	0.5109
Summation	None	0.5140
Basic	None	0.5855
Similarity	$SV=0.99, i=0$	0.5825
	$SV=1, i=0.0025$	0.5855
PDTB1	$SV=0.92, i=0$	0.5918
	$SV=0.92, i=0.0025$	0.5887
PDTB2	$SV=0.93, i=0$	0.6012
	$SV=0.95/0.96, i=0.0025$	0.5950

Table 4: Accuracy of each method

it might be better to set i to a smaller value. This also suggests that the prediction accuracy does not necessarily depend on the distance between persuasiveness of an argument and counterargument.

6 Conclusion and Future Work

In this paper, we proposed a new appropriate task of debate outcome prediction and four methods with cosine similarity and PDTB-styled discourse parser were introduced. The highest accuracy (60.1%) was achieved by PDTB2 method showing that similarity between arguments (on both *For* or *Against* sides) when combined with discourse parsing-based information, are capable to improve the accuracy. However, the results also suggest that debate outcome (decision on which side has won the debate) may be independent from the distance between persuasiveness of an argument and its counterargument.

In our experiments, we did not consider all discourse information from the discourse parser, therefore we plan to perform series of experiments to investigate the categories in PDTB (e.g. “CONTINGENCY” or “COMPARISON”) with our methods in order to improve the accuracy of our proposed task further.

In the corpora we use, an argument is separated into *Motion*, *Assertion* and *Justification* components, but it is rather artificial division. Therefore, there is a need for an and automatic way to discover other elements of a debate, e.g. Claim, Premise, Anecdote and Assumption [Ajjour *et al.*, 2017; Stab and Gurevych, 2014].

We did not consider values from Equation (1) when calculating the change before and after reading all arguments of a

debate. This change may differ in size depending on a theme, therefore, in the next step we plan to examine the role of this value when predicting debate outcome.

In addition, both corpora include a clear stance (*For* or *Against*), but in general arguments are not categorized in such manner, therefore, we have to add stance prediction algorithm similar to the one proposed by [Chen *et al.*, 2018] in order to be able to predict outcomes of debates from other resources.

References

- [Ajjour *et al.*, 2017] Yamen Ajjour, Wei-Fan Chen, Johannes Kiesel, Henning Wachsmuth, and Benno Stein. Unit segmentation of argumentative texts. In *Proceedings of the 4th Workshop on Argument Mining*, pages 118–128, 2017.
- [Al-Khatib *et al.*, 2016] Khalid Al-Khatib, Henning Wachsmuth, Matthias Hagen, Jonas Köhler, and Benno Stein. Cross-domain mining of argumentative text through distant supervision. In *Proceedings of the 2016 conference of the north american chapter of the association for computational linguistics: human language technologies*, pages 1395–1404, 2016.
- [Cabrio and Villata, 2018] Elena Cabrio and Serena Villata. Five years of argument mining: a data-driven analysis. In *IJCAI*, pages 5427–5433, 2018.
- [Cano-Basave and He, 2016] Amparo Elizabeth Cano-Basave and Yulan He. A study of the impact of persuasive argumentation in political debates. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1405–1413, 2016.
- [Carlile *et al.*, 2018] Winston Carlile, Nishant Gurrupadi, Zixuan Ke, and Vincent Ng. Give me more feedback: Annotating argument persuasiveness and related attributes in student essays. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 621–631, 2018.
- [Chen *et al.*, 2018] Di Chen, Jiachen Du, Lidong Bing, and Ruifeng Xu. Hybrid neural attention for agreement/disagreement inference in online debates. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 665–670, 2018.
- [Cocarascu and Toni, 2017] Oana Cocarascu and Francesca Toni. Identifying attack and support argumentative relations using deep learning. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1374–1379, 2017.
- [Cortes and Vapnik, 1995] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, Sep 1995.
- [Durmus and Cardie, 2018] Esin Durmus and Claire Cardie. Exploring the role of prior beliefs for argument persuasion. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, volume 1, pages 1035–1045, 2018.
- [Farra *et al.*, 2015] Noura Farra, Swapna Somasundaran, and Jill Burstein. Scoring persuasive essays using opinions and their targets. In *Proceedings of the Tenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 64–74, 2015.
- [Green *et al.*, 2014] Nancy Green, Kevin Ashley, Diane Litman, Chris Reed, and Vern Walker. Proceedings of the first workshop on argumentation mining. In *Proceedings of the First Workshop on Argumentation Mining*. Association for Computational Linguistics, 2014.
- [Habernal and Gurevych, 2016] Ivan Habernal and Iryna Gurevych. Which argument is more convincing? analyzing and predicting convincings of web arguments using bidirectional lstm. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 1589–1599, 2016.
- [Habernal *et al.*, 2018] Ivan Habernal, Henning Wachsmuth, Iryna Gurevych, and Benno Stein. Before name-calling: Dynamics and triggers of ad hominem fallacies in web argumentation. *arXiv preprint arXiv:1802.06613*, 2018.
- [Hidey and McKeown, 2018] Christopher Thomas Hidey and Kathleen McKeown. Persuasive influence detection: The role of argument sequencing. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
- [Lin *et al.*, 2014] Ziheng Lin, Hwee Tou Ng, and Min-Yen Kan. A pdtb-styled end-to-end discourse parser. *Natural Language Engineering*, 20(2):151–184, 2014.
- [Mikolov *et al.*, 2013] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositional-ity. In *Advances in neural information processing systems*, pages 3111–3119, 2013.
- [Persing and Ng, 2017] Isaac Persing and Vincent Ng. Why can’t you convince me? modeling weaknesses in unper-
suasive arguments. In *IJCAI*, pages 4082–4088, 2017.
- [Potash and Rumshisky, 2017] Peter Potash and Anna Rumshisky. Towards debate automation: a recurrent model for predicting debate winners. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2465–2475, 2017.
- [Stab and Gurevych, 2014] Christian Stab and Iryna Gurevych. Annotating argument components and relations in persuasive essays. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 1501–1510, 2014.
- [Stab *et al.*, 2018] Christian Stab, Tristan Miller, Benjamin Schiller, Pranav Rai, and Iryna Gurevych. Cross-topic argument mining from heterogeneous sources. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3664–3674, 2018.
- [Strapparava *et al.*, 2010] Carlo Strapparava, Marco Guerini, and Oliviero Stock. Predicting persuasiveness in political discourses. In *LREC*, 2010.

- [Tan *et al.*, 2016] Chenhao Tan, Vlad Niculae, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In *Proceedings of the 25th international conference on world wide web*, pages 613–624. International World Wide Web Conferences Steering Committee, 2016.
- [Van Eemeren *et al.*, 2014] Frans H Van Eemeren, Bart Garssen, Erik CW Krabbe, A Francisca Snoeck Henkemans, Bart Verheij, and Jean HM Wagemans. *Handbook of argumentation theory*. 2014.
- [Wachsmuth *et al.*, 2018] Henning Wachsmuth, Shahbaz Syed, and Benno Stein. Retrieval of the best counterargument without prior topic knowledge. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 241–251, 2018.
- [Wei *et al.*, 2016] Zhongyu Wei, Yang Liu, and Yi Li. Is this post persuasive? ranking argumentative comments in online forum. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pages 195–200, 2016.