Argument Classification with BERT plus Contextual, Structural and Syntactic Features as Text*

Umer Mushtaq[1] and Jérémie Cabessa[2,3]

[1] Laboratory of Mathematical Economics and Applied Microeconomics (LEMMA)
University Paris 2 – Panthéon-Assas, 75005 Paris, France
[2] Laboratory DAVID, UVSQ – Université Paris-Saclay
78000 Versailles, France
[3] Institute of Computer Science of the Czech Academy of Sciences
18207 Prague 8, Czech Republic

Abstract. In Argument Mining (AM), the integral sub-task of argument component classification refers to the classification of argument components as claims or premises. In this context, the content of the component alone does not actually suffice to accurately predict its corresponding class. In fact, additional lexical, contextual, and structural features are needed. Here, we propose a unified model for argument component classification based on BERT and inspired by the new prompting NLP paradigm. Our model incorporates the component itself together with contextual, structural and syntactic features – given *as text* – instead of the usual numerical form. This new technique enables BERT to build a customized and enriched representation of the component. We evaluate our model on three datasets that reflect a diversity of written and spoken discourses. We achieve state-of-art results on two datasets and 95% of the best results on the third. Our approach shows that BERT is capable of exploiting non-textual information given in a textual form.

Keywords: NLP · Argument Mining · Text Classification · BERT · Features as Text · Prompting

1 Introduction

Argument Mining (AM) is the automated identification and analysis of the underlying argumentational structure in natural texts [3]. Essential sub-tasks in AM include: 1) separating argument components from non-argumentative text, 2) classifying argument components to determine their role in the argumentative process, 3) given two argument components, deciding whether they are linked or not and, 4) given two linked components, deciding whether the link is supporting

* This research was supported by Labex MME-DII as well as by the Czech Science Foundation, grant AppNeCo No. GA22-02067S, institutional support RVO: 67985807.

or attacking [1,13]. AM is utilized for several popular downstream applications like Stance Recognition and Sentiment Analysis.

Argumentative discourse happens in many interesting settings. Written discourse such as essays and articles consists of structured presentation of claims and premises on a certain topic [12,16]. Organized political speeches consist of argumentative dialog between two or more candidates on several issues [11,8]. Social media platforms provide an avenue for users to debate and discuss contentious issues [15]. All three settings are inherently argumentative and are ideal for AM systems.

Text classification automatically classifies general text into pre-defined classes. In AM, it is the task of classifying argument components as either claims or premises. Claims are assertions made or positions taken for or against a particular topic and premises are evidence, justifications or warrants presented in support of claims.

For argument component classification, however, the use of different embeddings (GloVe, ELMo, FastText, etc.) alone as sentence representation do not suffice. The role of an argument component depends, among others, on its context and position in the text and thus cannot be captured by its content alone. Therefore, additional features like lexical, indicator, discourse, syntactic, contextual and structural features have been used to enrich the sentence representation of the components [17,4,7].

Transformer models have been game changers in NLP [19]. Bidirectional Encoder Representations from Transformers (BERT) are sequential models which are pre-trained on huge amounts of data in a self-supervised manner [2]. Using a transfer learning process called fine-tuning, this pre-trained BERT model is then utilized for an NLP task on a specific dataset. BERT models have been successfully used for several NLP tasks such as text classification. In fact, the BERT embedding as sentence representation outperforms earlier embeddings (GloVe, ELMo, FastText, etc.) on text classification tasks.

The 'Pre-train, Prompt, Predict' paradigm has also been a game-changer in NLP [9]. In this paradigm, task-specific supervised fine-tuning is replaced by additional self-supervised training involving textual prompts designed for specific downstream tasks. For instance, the sentiment of the sentence 'I liked the movie!' is obtained by the output of the language model on the input 'I liked the movie! The movie was [MASK].' which includes the sentence and a task specific prompt. For argument component classification, however, the straightforward prompting approach would not capture the necessary contextual, structural and syntactic information.

Based on these considerations, we propose a novel approach, inspired by prompt engineering, which incorporates – in textual form – the contextual, structural and syntactic features necessary for argument component classification. Specifically, we introduce a novel model for argument component classification which is based on the popular BERT model. Our model incorporates contextual, structural and syntactic features *as text* to build a customized and enriched BERT-based representation of the argument component. We experiment with our model on three datasets: one written essays-based, one speech-based and one written social media-based. We show that: 1) our *features as text* sentence representation model improves upon the BERT-based component only representation, 2) our structural *features as text* representation outperforms the classical approach of numerically concatenating these features with BERT embedding, and 3) our model achieves state-of-art results on two datasets and 95% of the best results on the third. Overall, we situate our work within the 'better models vs better data' question by developing task-specific and customized data as opposed to designing more complex models. We make the code available on GitHub at: https://github.com/mohammadoumar/features_as_text

This paper is structured as follows. Section 2 describes the related literature that informs our work. Section 3 presents the datasets. In Section 4, we introduce our novel *features as text* model in detail. Section 5 presents the experimental setting, results and analysis of our work. Section 5 provides concluding remarks and future directions.

2 Related works

Stab and Gurevych [17] present a features-based approach for argument component classification in the *Persuasive Essays (PE)* dataset (see Section 3). They use hand-crafted features (lexical, structural, syntactic, etc.) with Support Vector Machines (SVMs) and Conditional Random Fields (CRFs). They show that structural features, which capture the position of the component in the full text, are most useful for component classification.

Hadaddan et al. [4] use both features-based and neural network-based approaches for argument component classification in the *Yes We Can (YWC)* political debates dataset (See Section 3). In the features-based approach, they use an SVM with both Bag of Words (BoW) and a custom features set (POS, syntactic, NER, etc). In the neural network-based setting, they use both a feed-forward neural network with the custom features set and an LSTM with FastText word embedding.

Potash et al. [14] present a Joint Neural Model for simultaneous learning of argument component classification and link extraction between argument components in the *PE* and *Micro-Text Corpus (MTC)* datasets. This model consists of a Bi-LSTM encoder, a fully connected layer for component classification and an LSTM decoder for link identification. They use three methods for textual representation: Bag of Words (BoW), GloVe embedding and structural features.

Kuribayashi et al. [7] introduce an extension to the LSTM-minus-based span representation [20] where they create separate representations of the argumentative markers ('I think', 'because', etc.) and argumentative component present in the argument unit. For textual/span representation, they use GloVe and ELMo embeddings concatenated with Bag of Words (BoW) and structural features. They experiment with the *PE* and *MTC* datasets.

Mayer et al. [10] use neural network-based architectures for argument mining in a dataset of abstracts of bio-chemical healthcare trials. They combine the boundary detection and component classification tasks into one sequence tagging task. They use several static and dynamic embeddings such as BERT, GloVe, ELMo, fastText, FlairPM, etc. with various combinations of LSTMs, GRUs and CRFs as well as BERT fine-tune.

We situate ourselves within the 'better models vs better data' question. We posit that the BERT model is powerful enough to achieve improved performance if provided with task-specific enriched input data. To that end, our work is the first to investigate and implement a *features as text*, BERT-based model for argument component classification.

3 Datasets

In our work, we use three datasets for argument classification: *Persuasive Essays (PE)* [17], *Yes We Can (YWC)* [4] and *Change My View (CMV)* [6]. In this section, we present and explain the datasets.

Persuasive Essays (PE): The *PE* dataset was introduced by Stab and Gurevych [17]. It consists of 402 essays on diverse topics selected from the online portal *essayforum.com*. Each essay, which is divided into several paragraphs, consists of arguments (major claims, claims and premises) for or against a position on a controversial topic. A *MajorClaim* is a direct assertion of the author's position on the topic of the essay. A *Claim* is an assertion the author makes in support of his/her position on the topic. A *Premise* is a piece of evidence or warrant that the author presents to support his/her claim(s). For example, a snippet of the essay on the topic *'Should students be taught to compete or to cooperate?'* is given below with claim(s) in bold and premise(s) in italics:

> First of all, [**through cooperation, children can learn about interpersonal skills which are significant in the future life of all students.**]$_{claim_1}$ [*What we acquired from team work is not only how to achieve the same goal with others but more importantly, how to get along with others.*]$_{premise_1}$ [*During the process of cooperation, children can learn about how to listen to opinions of others, how to communicate with others, how to think comprehensively, and even how to compromise with other team members when conflicts occurred.*]$_{premise_2}$ [*All of these skills help them to get on well with other people and will benefit them for the whole life.*]$_{premise_3}$

Yes We Can (YWC): The *YWC* dataset was introduced by Haddadan et al. [4]. It consists of presidential and vice presidential debates in the quadrennial US presidential elections from 1960 to 2016: a total of 39 debates. The dataset consists of transcripts of these debates with claims and premises made by the candidates.

Change My Views (CMV): The *CMV* dataset is presented by Tan et al [18,6]. It is based on the "*r/changemyview*" subreddit from the social media platform *Reddit.com*. It consists of 113 threads containing argumentative conversations, made up of claims and premises, between internet users on 37 controversial topics. The statistics for all three datasets are given in Table 1.

Dataset	*Corpus Statistics*		*Component Statistics*	
Persuasive Essays (PE)	Tokens	147,271	Major Claims	751
	Sentence	7,116	Claims	1,506
	Paragraphs	1,833	Premises	3,832
	Essays	402	Total	6,089
Yes We Can! (YWC)	Speech Turns	6,601	Claims	11,964
	Sentences	34,103	Premises	10,316
	Words	676,227	Other	7,252
	Debates	39	Total	29,621
Change My View (CMV)	Words	75,078	Main Claims	116
	Paragraphs	3,869	Claims	1,589
	Topics	37	Premises	2,059
	Files	113	Total	3,764

Table 1. Corpus and component statistics for *PE*, *YWC* and *CMV* datasets. In the *CMV* dataset, Major Claims are called Main Claims.

4 Model

In this section, we introduce our novel BERT-based model for argument component classification. Our model incorporates contextual, structural and syntactic features – represented *as text* – instead of the usual numerical form. This approach enables BERT to build an enriched representation of the argument component.

4.1 BERT

BERT architecture consists of twelve encoder blocks of the Transformer model stacked together and 12 self-attention heads [2]. The self-attention heads enable BERT to incorporate bidirectional context and focus on any part of the input sequence. BERT builds a 768 dimensional representation – or embedding – of the input text sequence. In this work, as opposed to current approaches, we enrich the BERT model with textual representation of contextual, structural and syntactic features. These features are described below.

4.2 Features

Contextual features: Contextual features capture the full meaning of an argument component in its semantic and linguistic space. In our work, we use *full sentence* and *topic statement* as contextual features. The full sentence feature helps capture the presence of argumentative and/or discourse markers ('I think', 'In my opinion', etc.). These markers indicate that the component preceding or succeeding them in the sentence is more likely a claim than a premise. The topic statement feature helps discriminate between claims and premises because a claim is more likely to directly address the topic statement and, thus, be more semantically similar to it.

For both the *Persuasive Essays (PE)* and *Change My View (CMV)* datasets, the contextual features are the topic of the essay/discussion and the full sentence of the argument component. For the *Yes We Can (YWC)* dataset, in addition to the full sentence, we use candidate name and election year as topical information. We define the textual representation of contextual features as:

`contextual_features_as_text` = 'Topic: t. Sentence: s.'

where t is the topic of the essay/discussion thread or the speaker and election year of the debate speech and s is the full sentence which contains the argument component (see Example 1).

Structural features: Structural features incorporate the idea that argumentation follows a certain (perhaps fluid) pattern which can be used to discriminate between claims and premises. These features capture the location of the argument component in the whole essay and in the paragraph in which it appears. For example, claims are more likely to appear in the introductory and concluding paragraphs as well as in the beginning and towards the end of the paragraph. Premises, on the other hand, are more likely to follow a claim in the paragraph [17]. We define the textual representation of structural features as:

`structural_features_as_text` = 'Paragraph Number: n. Is in introduction: i. Is in conclusion: c. Is first in paragraph: f. Is last in paragraph: l.'

where n is the paragraph number in which the argument component is present, i is *Yes* if the argument component is in the introduction paragraph and *No* otherwise, c is *Yes* if the argument component is in the conclusion paragraph and *No* otherwise, f is *Yes* if the argument component is the first component in its paragraph and *No* otherwise, and l is *Yes* if the argument component in the last component in its paragraph and *No* otherwise (see Example 1).

Syntactic features: Part-Of-Speech (POS) involves classification of English words into categories depending on their linguistic role in a sentence. These categories include noun, verb, adjective, adverb, pronoun, preposition, conjunction, interjection, numeral, article, or determiner [5]. We define the textual representation of syntactic features as:

$$\texttt{syntactic\_features\_as\_text} = \text{`Part Of Speech tags: } t_1, t_2 ... t_n\text{'}$$

where t_i represents the POS tag of the i-th word in the argument component.

4.3 Combined features as text

We combine the textual representations of the contextual, structural and syntactic features to build an enriched BERT-based representation of the argument component. The combined representation is defined as follows:

$$\begin{aligned}\texttt{combined\_features\_as\_text} = {} & \texttt{contextual\_features\_as\_text} + \\ & \texttt{structural\_features\_as\_text} + \\ & \texttt{syntactic\_features\_as\_text}\end{aligned}$$

where '+' denotes the string concatenation operation. Note that the argument component itself is included in the full sentence.

Example 1: We consider an example from the *Persuasive Essays (PE)* dataset: argument component 398 from essay 28:

$$\begin{aligned}\texttt{argument\_component} = {} & \text{`advertising cigarettes and alcohol will definitely} \\ & \text{affect our children in negative way'}\end{aligned}$$

The contextual, structural and syntactic features of this argument component are given in Table 2.

The combined *features as text* representation of this argument component is:

> '[Topic: Society should ban all forms of advertising. Sentence: Ads will keep us well informed about new products and services, but we should also bear in mind that **advertising cigarettes and alcohol will definitely affect our children in negative way.**]$_{contextual}$ [Paragraph Number: Five. Is in introduction: No. Is in conclusion: Yes. Is first in paragraph: No. Is last in paragraph: Yes.]$_{structural}$ [Part of Speech tags: VERB, NOUN, CCONJ, NOUN, VERB, ADV, VERB, DET, NOUN, ADP, ADJ, NOUN]$_{syntactic}$'

where the argument component is in bold and the contextual, structural and syntactic features are contained in brackets. This combination of contextual features, structural features and argument component jointly form the enriched sentence representation that is input to the BERT model.

5 Results and analysis

In this section, we present and analyse our results. We use our model for two tasks: 1) *BERT fine-tune:* We fine-tune BERT on the three datasets using our novel combined features as text sentence representation. 2) *Textual vs numerical features comparison:* We fine-tune BERT and compare results in two cases: first, with our structural features as text and second, with structural features numerically concatenated with BERT sentence embedding.

Feature	*Value*
essay topic	'Society should ban all forms of advertising.'
full sentence	'Ads will keep us well informed about new products and services, but we should also bear in mind that **advertising cigarettes and alcohol will definitely affect our children in negative way.**'
para_nr	5
is_in_intro	0
is_in_conclusion	1
is_first_in_para	0
is_last_in_para	1
pos_tags	VERB, NOUN, CCONJ, NOUN, VERB, ADV, VERB, DET, NOUN, ADP, ADJ, NOUN

Table 2. Features for argument component 398 of the *PE* dataset. The component itself is in bold.

5.1 Experimental setting

For the *PE* dataset, we use the original split: 322 essays in the train set (4,709 components) and 80 essays in the test set (1,258 components). For the *YWC* dataset, we also use the original split with 10,447 components in the train set, 6,567 components in the test set and 5,226 components in the validation set. For the *CMV* dataset, we randomly set aside 90 threads for the train set (2,720 components) and 23 threads for the test set (763 components). The implementation details of the model and experiment are presented in Table 3.

5.2 Task results

The results of Task 1 and Task 2 are presented in Tables 4 and 5, respectively. State-of-the-art results are also shown in Table 4: F1 score of 0.86 for *PE* [7] and 0.67 for *YWC* [4] datasets. The results can be summarized as follows:

- Our novel *features as text* sentence representation, which incorporates contextual, structural and syntactic features as text, improves upon the BERT-based component only representation.
- Our *features as text* representation outperforms the classical approach of numerically concatenating these features with BERT embedding.
- Our model achieves state-of-art results on two datasets and 95% of the best results on the third.

5.3 Analysis

The addition of contextual, structural and syntactic features *as text* enables BERT to relate the argument component to the linguistic and argumentative flow of the whole paragraph and essay.

In Task 1, for the *PE* dataset, the contextual, structural and syntactic parts of our combined representation improve the results compared to the BERT-based component only representation. The contextual representation improves the F1 score from 0.57 to 0.68. The combined contextual, structural and syntactic representation improves the F1 score from 0.68 to 0.82 which is 95% of the state-of-the-art result (0.86) [7]. However, the state-of-the-art approach works on paragraphs which are chunked into segmented discourse units and require argumentative marker (AM) versus argumentative component (AC) distinction in sentences. In contrast, our model simply works on the sentence level and requires no AM/AC distinction to be made. Overall, the improvement achieved by structural features emphasizes the importance of the position of the argument component in written argumentative texts, like persuasive essays.

For the *CMV* dataset, our combined contextual and structural representation improves the F1 score from 0.76 to 0.79. Here, the contextual only part does not improve the results because the argument component and full sentence boundaries almost always coincide. By contrast, the structural features do improve the results, but to a lesser extent than in the *PE* dataset. This difference is explained by the fact that written text on social media platforms is less structured than written text in academic essays.

In contrast with the other datasets, for *YWC*, the combined contextual, structural and syntactic representation does not show improvement. Nevertheless, our model outperforms the state-of-the-art results in the literature (0.69 vs 0.67) [4]. These results show that the somewhat concrete linguistic and structural flow present in the written *PE* dataset and (to a lesser extent) in the *CMV* dataset is lacking in the spoken *YWC* dataset because of its extemporaneous and fluid nature.

Name	*Values*
Model	'bert-base-uncased'
Embedding dimension	768
Batch size	[16, 24, 32, 48]
Epochs	[3, 6, 8, 12]
Learning rate	[1e-5, 2e-5, 1e-3, 5e-3, 5e-5]
Warmup ratio, Weight decay, Dropout	0.1, 0.01, 0.1
Loss function	Cross Entropy Loss

Table 3. Model implementation details. We experimented with several parameter values. For each experiment, the best parameter values are available on the GitHub repository.

Sentence representation	*PE*				*YWC*			*CMV*		
	MC	*C*	*P*	*F1*	*C*	*P*	*F1*	*C*	*P*	*F1*
component only	0.49	0.41	0.81	0.57	0.71	0.68	**0.69**	0.74	0.79	0.76
sentence	0.69	0.48	0.82	0.66	0.71	0.68	**0.69**	0.70	0.77	0.74
topic + sent	0.70	0.70	0.84	0.68	0.69	0.65	0.67	0.70	0.75	0.73
sent + strct	0.85	0.68	0.91	0.81	0.70	0.68	**0.69**	0.75	0.84	**0.79**
topic + sent + strct	0.86	0.68	0.91	0.81	0.69	0.65	0.67	0.76	0.80	0.78
topic + sent + strct + synt	0.86	0.71	0.91	**0.82**	0.71	0.62	0.67	0.76	0.78	0.77
LSTM + dist [7]	0.92	0.73	0.92	**0.86**	-	-	-	-	-	-
LSTM + word emb [4]	-	-	-	-	0.70	0.68	0.67	-	-	-

Table 4. Task 1 results. Performance of our *features as text* BERT-based model on the three datasets. We report results of different combinations of features as text. *MC*, *C* and *P* represents the F1 scores for *MajorClaim*, *Claim* and *Premise*, respectively. *F1* represents the macro F1 score. The abbreviations 'strct' and 'synt' stand for structural features and syntactic features respectively. The last two rows represent the state-of-the-art results for the *PE* and *YWC* datasets.

Dataset	*Features concatenated*				*Features as text*			
	MC	*C*	*P*	*F1*	*MC*	*C*	*P*	*F1*
Persuasive Essays (PE)	0.82	0.57	0.90	0.76	0.86	0.68	0.91	**0.81**
Yes We Can! (YWC)	-	0.70	0.65	0.67	-	0.69	0.65	**0.69**
Change My View (CMV)	-	0.70	0.76	0.73	-	0.76	0.80	**0.78**

Table 5. Task 2 results. Comparison between structural features numerically concatenated to BERT embedding and our *features as text* sentence representation.

Overall, we see that our *features as text* sentence representation, which incorporates contextual, structural and syntactic features *as text*, improves upon the BERT-based component only representation. In fact, the latter representation is unable to capture two significant classification clues: the context and the structure. The context includes argumentative markers ('In my opinion', 'I think', etc.) while the structure captures the position of the argument component in argumentative text.

The results from Task 2 show that our features as text representation outperforms the classical representation where structural features are numerically concatenated with BERT embedding. For the *PE* and *CMV* datasets, the improvement in F1 scores is significant: from 0.76 to 0.82 and from 0.73 to 0.78, respectively. For the *YWC* dataset, on the other hand, the improvement is less significant: from 0.67 to 0.69. These results support our contention that the datasets for which the contextual and structural features provide the most significant information (Task 1) correspond precisely to those where the *features as text* representation performs the best (Task 2). In other words, the more significant the contextual and structural features, the better the *features as text* representation. Overall, our approach shows that BERT performs better when non-textual information is given to it as text.

6 Conclusion

In this work, we introduce a novel model for argument component classification which is based on the popular BERT model and inspired by the game-changing prompting paradigm. Our model incorporates contextual, structural and syntactic features *as text* to build an enriched BERT-based representation of the argument component.

We experiment with our model on three datasets: two written and one spoken. We obtain three main results: 1) our *features as text* sentence representation model improves upon the BERT-based component only representation, 2) our structural *features as text* representation outperforms the classical approach of numerically concatenating these features with BERT embedding and 3) our model achieves state-of-art results on two datasets and 95% of the best results on the third. To the best of our knowledge, our work is the first to investigate and implement a model based on *features as text* sentence representation.

Based on our results and analysis, we think that a systematic study to compare Argument Mining dynamics in written and spoken datasets would be of great benefit to the AM community. In terms of prospective research directions, we plan to merge our *features as text* technique with the LSTM-minus-based span representation model of Kuribayashi et al. [7]. We also intend to extend our *features as text* technique to other features such as syntactic and lexical [17].

We see our work as a first step towards a hybrid BERT-PROMPT end-to-end AM pipeline, thereby combining two dominant NLP paradigms. We think that our features as text approach opens up exciting new possibilities both for Argument Mining as well as any NLP tasks which require feature engineering. More generally, our approach can be used in other ML settings where the features can be described as text.

References

1. Cabrio, E., Villata, S.: Five years of argument mining: A data-driven analysis. In: Proceedings of the 27th International Joint Conference on Artificial Intelligence. p. 5427–5433. IJCAI'18, AAAI Press (2018)
2. Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. In: Burstein, J., Doran, C., Solorio, T. (eds.) Proceedings of NAACL-HLT 2019. pp. 4171–4186. ACL (2019)
3. Habernal, I., Wachsmuth, H., Gurevych, I., Stein, B.: The argument reasoning comprehension task: Identification and reconstruction of implicit warrants. In: Proceedings of NAACL-HLT 2018. pp. 1930–1940. ACL (2018)

4. Haddadan, S., Cabrio, E., Villata, S.: Yes, we can! mining arguments in 50 years of US presidential campaign debates. In: Proceedings of ACL 2019. pp. 4684–4690. ACL (2019)
5. Haspelmath, M.: Word Classes and Parts of Speech, pp. 16538–16545 (12 2001). https://doi.org/10.1016/B0-08-043076-7/02959-4
6. Hidey, C., Musi, E., Hwang, A., Muresan, S., McKeown, K.: Analyzing the semantic types of claims and premises in an online persuasive forum. In: Proceedings of ArgMining@EMNLP 2017. pp. 11–21. ACL (2017)
7. Kuribayashi, T., Ouchi, H., Inoue, N., Reisert, P., Miyoshi, T., Suzuki, J., Inui, K.: An empirical study of span representation in argumentation structure parsing. In: Proceedings of ACL. pp. 4691–4698. ACL (2019)
8. Lippi, M., Torroni, P.: Argument mining from speech: Detecting claims in political debates. In: Schuurmans, D., Wellman, M.P. (eds.) Proceedings of AAAI 2016. pp. 2979–2985. AAAI Press (2016)
9. Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., Neubig, G.: Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. CoRR **abs/2107.13586** (2021)
10. Mayer, T., Cabrio, E., Villata, S.: Transformer-based argument mining for healthcare applications. In: Proceedings of ECAI 2020. pp. 2108–2115. IOS Press (2020)
11. Menini, S., Cabrio, E., Tonelli, S., Villata, S.: Never retreat, never retract: Argumentation analysis for political speeches. Proceedings of AAAI 2018 **32**(1) (2018)
12. Moens, M.F., Boiy, E., Palau, R.M., Reed, C.: Automatic detection of arguments in legal texts. In: Proceedings of ICAIL 2007. p. 225–230. ACM (2007)
13. Peldszus, A., Stede, M.: From argument diagrams to argumentation mining in texts: A survey. Int. J. Cogn. Informatics Nat. Intell. **7**, 1–31 (2013)
14. Potash, P., Romanov, A., Rumshisky, A.: Here's my point: Joint pointer architecture for argument mining. In: Proceedings of EMNLP 2017. pp. 1364–1373. ACL (2017)
15. Somasundaran, S., Wiebe, J.: Recognizing stances in online debates. In: Proceedings of ACL/IJCNLP 2009. pp. 226–234. ACL (2009)
16. Song, Y., Heilman, M., Beigman Klebanov, B., Deane, P.: Applying argumentation schemes for essay scoring. In: Proceedings of ArgMining@ACL 2014. pp. 69–78. ACL (2014)
17. Stab, C., Gurevych, I.: Parsing argumentation structures in persuasive essays. Computational Linguistics **43**(3), 619–659 (2017)
18. Tan, C., Niculae, V., Danescu-Niculescu-Mizil, C., Lee, L.: Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In: Proceedings of WWW 2016. pp. 613—-624. ACM (2016)
19. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. In: Proceedings of NIPS 2017. pp. 6000—-6010. Curran Associates Inc. (2017)
20. Wang, W., Chang, B.: Graph-based dependency parsing with bidirectional LSTM. In: Proceedings of ACL 2016. pp. 2306–2315. ACL (2016)