

Lexicon Guided Attentive Neural Network Model for Argument Mining

Jian-Fu Lin,¹ Kuo Yu Huang,¹ Hen-Hsen Huang,^{2,3} Hsin-Hsi Chen^{1,3}

¹ Department of Computer Science and Information Engineering,
National Taiwan University, Taiwan

² Department of Computer Science, National Chengchi University, Taiwan

³ MOST Joint Research Center for AI Technology and All Vista Healthcare, Taiwan
{cflin, kyhuang}@nlg.csie.ntu.edu.tw,
hhhuang@nccu.edu.tw, hhchen@ntu.edu.tw

Abstract

Identification of argumentative components is an important stage of argument mining. Lexicon information is reported as one of the most frequently used features in the argument mining research. In this paper, we propose a methodology to integrate lexicon information into a neural network model by attention mechanism. We conduct experiments on the UKP dataset, which is collected from heterogeneous sources and contains several text types, e.g., microblog, Wikipedia, and news. We explore lexicons from various application scenarios such as sentiment analysis and emotion detection. We also compare the experimental results of leveraging different lexicons.

1 Introduction

Argument Mining (AM) is an emerging research area that has drawn more and more attention since around 2010. Recently, Project Debater from IBM has shown such an AI machine supported by argument mining techniques can do well at arguing. The task of AM can be divided into a few stages: (1) Extracting argumentative components from large texts, i.e., boundary detection or segmentation; (2) Classifying the extracted components into classes. In general, an argumentative component can be categorized into "Claim", which usually contains conclusions and stance toward the given topic, or "Premise", which contains reasoning or evidence used to support or attack a claim; (3) Predicting the relations between the identified argumentative components, i.e., supporting and attacking (Cabrio and Villata, 2018). Some works also consider more complex relations such as recursively support/attack the relations themselves rather than merely build relations between components (Peldszus and Stede, 2013).

Argument detection and classification can improve legal reasoning (Moens et al., 2007), policy

formulation (Florou et al., 2013), and persuasive writing (Stab and Gurevych, 2014). In this paper, we focus on mining argumentative components from a large collection of documents and further classifying them into roles of support/opposition. Our model is based on the recurrent neural network (RNN), which has been widely used in natural language processing tasks (Cho et al., 2014). With the help of the attention mechanism (Bahdanau et al., 2015), RNN can further attend on the key information.

We propose a novel attention mechanism that is guided by argumentative lexicon information. Lexicon information is reported as one kind of the most frequently used features in argument mining (Cabrio and Villata, 2018). Previous works on AM have tried to integrate lexical features into the learning models (Levy et al., 2017; Nguyen and Litman, 2015; Rinott et al., 2015). These lexicons are mostly composed by human beings or derived by hand-crafted rules, and result in domain-specificity. That is, it may fail to be used for other domains. In the contrast of scarcity of general lexicon for AM, lexical resources are abundant in other fields like sentiment analysis, opinion mining, and emotion detection (Hu and Liu, 2004; Mohammad and Turney, 2013; Kiritchenko and Mohammad, 2016). As a more general domain, AM may get the benefits of not only in-domain lexicon, but also out-domain lexicons.

The contribution of this work is two-fold: (1) We propose an attention mechanism to leverage lexicon information. (2) In the face of the scarcity of argument lexicon, we explore several different types of lexicons to verify whether outside resources are useful for AM tasks.

The rest of this paper is organized as follows. Section 2 summarizes related works about AM. The dataset and linguistic resources used for experiments are shown in Section 3. We introduce

our model in Section 4 and show the experimental results in Section 5. We also look into the errors made by our best model in Section 6. Section 7 makes a discussion on experimental results and concludes this work.

2 Related Works

Neural networks have been used in varieties of AM tasks. To improve the vanilla LSTM model, Stab et al. (2018a) use attention mechanism to fuse topic and sentence information together. In the work of Laha and Raykar (2016), they present several bi-sequence classification models on different datasets. However, rather than using some sophisticated architecture such as attention, it considers only different concatenation or condition method on the output of LSTM. Eger et al. (2017) propose an end-to-end training model to mining argument structure, identifying argument components.

Besides syntactic and positional information, lexical information is also reported as one of the most used features in argument mining task (Cabrio and Villata, 2018). In some similar research fields such as sentiment analysis and emotion mining, a number of works have been proposed to combine lexical information with the NN models. Teng et al. (2016) use lexical scores as the weights and do the weighted sum over the outputs of the LSTM model, in order to derive the sentence scores. Zou et al. (2018) determines attention weights using lexicon labels, which lead the model to focus on the lexicon words. Bar-Haim et al. (2017) proposes an idea of expanding lexicons to improve stance classifying task.

However, in AM, seldom works directly combine lexicon with models. By using discourse feature, Levy et al. (2018) generates weak labels and use weak supervision. Shnarch et al. (2018) also present a methodology to blend such weak labeled data with high quality but scarce labeled data for AM. Al-Khatib et al. (2016) consider the distant supervision method. Most of these works use the end-to-end training paradigm with the outside resources only for generating the weak label, which may not fully leverage the information of the lexicons.

3 Resources

In this section, we introduce the dataset used to evaluate the performance of our proposed model. Besides, we describe each lexicon in brief and

show how to perform the data preprocessing.

3.1 Data

We conduct the experiments on the dataset released by Stab et al. (2018b).¹ The dataset includes 25,492 sentences over eight topics that are randomly selected from an online list of controversial topics.² The selected topics, which are considered as queries, are used to retrieve documents from heterogeneous sources via the Google search engine. Among these sentences, 4,944 of them are supporting arguments, 6,195 are opposing arguments, and 14,353 are non-argument sentences. This dataset is commonly used for sentential argument identification task. Levy et al. (2018) collect a dataset with around 1.5 million sentences over 150 topics from Wikipedia. However, only 2,500 of them are labeled. It may not be sufficient for training a model, especially for neural network models.

The definition of argumentative components differs from dataset to dataset. In the dataset used in this work, an argumentative component is a span of text with reasoning or evidence, which is able to either support or oppose a topic (Stab et al., 2018b).

3.2 Lexicon resource

To improve the baseline model, we consider several existing lexicons across different domains. We first explore the claim lexicon that is built for argument mining task (Levy et al., 2017). We also include the lexicon resources often used in sentiment analysis (Hu and Liu, 2004) and emotion detection (Mohammad and Turney, 2013). We postulate that the resources for these application scenarios may have the potential for argument mining. We further develop a model based on the general purpose lexicon, WordNet (Miller, 1995).

These resources are applied in different ways. We use the claim lexicon (Levy et al., 2017), the sentiment lexicon (Hu and Liu, 2004), and the emotion lexicon (Mohammad and Turney, 2013) to extract critical words from the i -th input sentence C_i , forming a sentence A_i . In contrast, we consult WordNet (Miller, 1995) to expand the short topic T_i into the corresponding A_i .

¹https://www.informatik.tu-darmstadt.de/ukp/research_6/data

²<https://www.procon.org/>, <https://www.questia.com/library/controversial-topics>

Claim Lexicon. The claim lexicon is a lexicon containing words with argumentative characteristics. Levy et al. (2017) use the appearance of the term "that" as a weak signal of sentences containing argumentative components. After collecting nearly 1.86M sentences, they compute the prior probability of the term "that" $P(\text{that})$ occurring in a sentence, and the probabilities $P(\text{that}|w_i)$, where $P(\text{that}|w_i)$ denotes the probability of a sentence having the term "that" and the word w_i is in the suffix after the main concept (i.e. the target entity in a controversial topic), given the sentence containing w_i . Those words with a probability $P(\text{that}|w_i) > P(\text{that})$ are included in their proposed claim lexicon, resulting a lexicon with around 600 claim words. The lexicon was then used for designing sentence pattern rules called claim sentence query (CSQ). They believe the claim lexicon can help detect sentences containing argument.

Sentiment Lexicon. Hu and Liu (2004) built a sentiment lexicon that contains around 6,800 words. Each word is labeled as negative and/or positive. We construct an additional sentence \mathcal{A} by extracting the words that are in both sentiment lexicon and the input sentence \mathcal{C} , regardless whether they are positive or negative.

Emotion Lexicon. The emotion lexicon built by Mohammad and Turney (2013) contains around 14,200 words. Each word in the lexicon is given eight emotion labels. An emotion in the lexicon could be one of eight emotions, including anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. The labels are defined as follows:

$$\text{label}(w_i, e_j) = \begin{cases} 1, & \text{if } w_i \text{ associated with } e_j \\ 0, & \text{otherwise} \end{cases}$$

Where w_i is a word, and e_j is one of the eight emotions. In the experiment, we select the words that have at least one emotion labeled as 1, resulting a list of 6,468 words. We then use this list to create an additional sentence \mathcal{A} from the input sentence \mathcal{C} .

WordNet. To expand a topic \mathcal{T} composed of words $w_1^{\mathcal{T}}, w_2^{\mathcal{T}}, \dots, w_K^{\mathcal{T}}$, we expand each of the words in it. For each word $w_i^{\mathcal{T}}$, we use WordNet (Miller, 1995) to find its corresponding synonyms. We then put the found synonyms together, forming an additional sentence \mathcal{A} , an expanded version of topic \mathcal{T} .

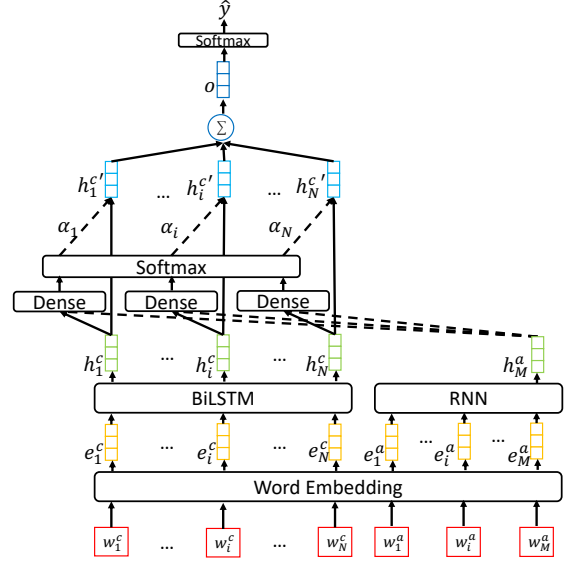


Figure 1: The architecture of Lexicon Guided Attentive Neural Network Model

4 Model

This section describes the development of the baseline model and the proposed model. To identify sentence-level argumentative components, the model is given a sentence \mathcal{C} , which contains a sequence of words $w_1^c, w_2^c, \dots, w_N^c$ and a topic \mathcal{T} with words $w_1^{\mathcal{T}}, w_2^{\mathcal{T}}, \dots, w_K^{\mathcal{T}}$. The input word sequence is then encoded as a sequence of word embeddings via the GloVe word vectors. The pre-trained word vectors with the dimension of 100 released by Pennington et al. (2014) are used. Based on the given input, the model makes a prediction \hat{y} for the given sentence, i.e., classifying it as *supporting argument*, *opposing argument*, or *non-argument*. For comparison, we implement a baseline model with the vanilla bidirectional LSTM (BiLSTM).

In order to exploit the linguistic knowledge, Lei et al. (2018) highlight the sentiment words of the input sentence, computing attention weight for each word with them. By integrating the sentiment lexicon into the neural network model, the work successfully improves the performance in sentiment analysis. This work proposes a model that integrates an outside lexicon resource into attention mechanism (Vaswani et al., 2017). For each input sentence, we compose an additional sentence \mathcal{A} , which contains words $w_1^a, w_2^a, \dots, w_M^a$ based on the given lexicon. The additional sentence \mathcal{A} is then forwarded to the embedding layer together with input sentence \mathcal{C} . The output of

model	F_1	P_{arg+}	P_{arg-}	R_{arg+}	R_{arg-}
BiLSTM	.5337 ± .0123	.4521 ± .0391	.4832 ± .0393	.2911 ± .1139	.4816 ± .1276
ClaimLex*	.5684 ± .0222	.4736 ± .0322	.5075 ± .0450	.3756 ± .1072	.5011 ± .0854
SentimentLex*	.5718 ± .0165	.4937 ± .0365	.5125 ± .0414	.3590 ± .1043	.5240 ± .0889
EmotionLex*	.5695 ± .0129	.4920 ± .0369	.5036 ± .0356	.3524 ± .0861	.5264 ± .1006
WordNet*	.5788 ± .0142	.4846 ± .0292	.5191 ± .0376	.3724 ± .0818	.5235 ± .0772

Table 1: The results of the baseline model and the proposed model with different lexicon resources. The highest score of each column is highlighted in bold font.

embedding layer is the sequences $e_1^c, e_2^c, \dots, e_N^c$ and $e_1^a, e_2^a, \dots, e_M^a$, representing the embedded sentences \mathcal{C} and \mathcal{A} , respectively. Then, e_i^c is fed into BiLSTM and results in h_i^c at the corresponding time step. As for \mathcal{A} , we add an RNN to collect its information and take the output h_M^a at the last time step as its representation. Though Lei et al. (2018) use an LSTM to encode the sentimental sentences, we do not follow their approach. In our work, the simple RNN outperforms the LSTM in the preliminary experiments.

The attention weight of the i -th word (i.e. α_i) is determined by the concatenation of the output of the BiLSTM h_i^c and the output of the RNN (i.e. h_M^a), which is given the additional sentence \mathcal{A} as the input:

$$\alpha_i = \frac{\exp(\sigma([h_i^c; h_M^a]))}{\sum_{i=1}^N \exp(\sigma([h_i^c; h_M^a]))} \quad (1)$$

where α_i indicates the attention weight of i -th word of the input sentence, and $[h_i^c; h_M^a]$ indicates the concatenation of i -th hidden state and the RNN output state. The scoring function $\sigma(\cdot)$ is designed as:

$$\sigma([h_i^c; h_M^a]) = \tanh(W_c[h_i^c; h_M^a]) \quad (2)$$

where W_c indicates trainable parameters.

All the weighted hidden states are then summed up, and connected to a fully connected layer for the final prediction:

$$o = \sum_{i=1}^N h_i^{c'} \quad (3)$$

$$h_i^{c'} = \alpha_i h_i^c \quad (4)$$

Figure 1 illustrates the architecture of our model.

5 Experiments

Because most of the lengths of input sentences are less than 60 and most of the lengths of additional

sentences \mathcal{A} are less than 20, we truncate them into lengths of 60 and 20 respectively. The dataset has 25,492 sentences in total. We conduct 5-fold cross validation for evaluating our model.

To evaluate our approaches, we report the average macro F_1 as ternary setting, precision and recall of predicting supporting arguments (P_{arg+}, R_{arg+}), and precision and recall of predicting opposing arguments (P_{arg-}, R_{arg-}). We run paired t-test for each proposed model in comparison with the baseline model, and mark the models having statistical significance (i.e. p -value < 0.05) with a wildcard. As the result shown in Table 1, we can observe that the proposed models benefit from the information from the adopted lexicons, improving the performance of argumentative components identification. The best model, which uses WordNet to expand topic \mathcal{T} , outperforms the baseline model by 4.5 percentage in F_1 . The proposed model with the lowest F_1 score (i.e. ClaimLex) still outperforms the baseline by 3.4 percentage. Furthermore, the best performance reported by Stab et al. (2018b) on the same dataset is 0.4285 in macro F_1 , which is the result of only incorporating topic information into their models. This shows the impact of the lexicon information.

However, we can also observe that the result of integrating claim lexicon (Levy et al., 2017) is out of our expectation though it is a resource for argument mining. Possible reasons are figured out as follows. Firstly, the lexicon is built based on a strong assumption, i.e., the present of the term "that" indicates a high probability of the occurrence of argumentative components. Secondly, the lexicon has only 586 words, indicating a very small coverage with the whole vocabulary. Thirdly, the lexicon built from the sentences across 100 different topics contains a number of domain-specific words such as "LGBTQ" and "militarily". Highlighting of these domain-specific words may cause noise when the topic is unrelated to them.

	Topic	Sentence	Prediction	Annotation
S1	death penalty	Advocates of death penalty cite examples on how imposing the death sentence or abolishing it have affected crime rate.	attack	support
S2	death penalty	Pain of Death: Executing a person can be quick and painless, or executing a person can be slow and painful.	support	non-arg
S3	gun control	If you can get to the phone to call 911, if you are strong enough to hold off your attacker until the police arrive, and if you can wait 15 minutes in the city or 45 minutes in the country for law enforcement to arrive while you struggle with the intruder, then you might make it.	support	attack
S4	cloning	Our of respect for human clones (human beings in every respect), a ban on human cloning should be opposed.	support	attack
S5	gun control	We should have more of it!	non-arg	support

Table 2: The examples that our best model fails to correctly predict. The sentences predicted/annotated as non-argumentative ones are abbreviated to *non-arg*.

6 Error Analysis

To know better what kind of sentence would mislead our model to make wrong predictions, we randomly sample the sentences with error from our best model, i.e., lexicon-guided attentive neural network model with WordNet. After looking into these errors, we find that the causes of a wrong prediction can be briefly categorized into the following cases. Some illustrations of the errors are listed in Table 2: (1) The sentences that have ambiguous words or state an open question can easily lead our model to predict the sentences’ labels from non-argumentative to argumentative, or predict the labels from one stance, i.e., supporting or attacking, to the other. For example, both “*imposing*” and “*abolishing*” are shown in S1, which may cause the model to fails on correctly detecting the stances. S2 states an open question on the influence of death penalty, but the model mistakes it for an argumentative sentence; (2) When arguing over an issue, people may use irony to attack the opposite stance. Such statement may mislead the model, as S3 has shown; (3) We also find that our model may predict wrongly with the appearance of double negation. The part of the sentence S4, “*a ban on human cloning should be opposed*”, conveys the supporting stance with a double negative statement. With a limited amount of training data, the model may not be able to comprehend rela-

tively complicated syntax.

On the other hand, some examples in the dataset may have been wrongly annotated. According to [Stab et al. \(2018b\)](#), arguments are defined as a span of text having reasoning or evidence that can be used to support or oppose a topic. S5 does explicitly declare its supporting stance, but nevertheless has no reasoning or evidence.

7 Conclusion

In this work, we propose a novel approach to leverage the lexicon from both in-domain and out-of-domain sources for the task of argumentative component mining. We explore several sources from different application scenarios, from claim lexicon ([Levy et al., 2017](#)) to other domain resources such as sentiment analysis ([Hu and Liu, 2004](#)), emotion detection ([Mohammad and Turney, 2013](#)), and the general domain lexicon resource ([Miller, 1995](#)). Experimental results confirm the effectiveness of the integration of lexicon information. The scarcity of the resources in argument mining is also highlighted in the discussion.

Acknowledgments

This research was partially supported by Ministry of Science and Technology, Taiwan, under grants MOST-106-2923-E-002-012-MY3, MOST-107-2634-F-002-011-, MOST-108-2634-F-002-008-.

References

- Khalid Al-Khatib, Henning Wachsmuth, Matthias Hagen, Jonas Köhler, and Benno Stein. 2016. [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, San Diego, California. Association for Computational Linguistics.
- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. [Neural machine translation by jointly learning to align and translate](#). *CoRR*, abs/1409.0473.
- Roy Bar-Haim, Lilach Edelstein, Charles Jochim, and Noam Slonim. 2017. [Improving claim stance classification with lexical knowledge expansion and context utilization](#). In *Proceedings of the 4th Workshop on Argument Mining*, pages 32–38, Copenhagen, Denmark. Association for Computational Linguistics.
- Elena Cabrio and Serena Villata. 2018. [Five years of argument mining: a data-driven analysis](#). In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, pages 5427–5433. International Joint Conferences on Artificial Intelligence Organization.
- Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. [Learning phrase representations using rnn encoder–decoder for statistical machine translation](#). In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734, Doha, Qatar. Association for Computational Linguistics.
- Steffen Eger, Johannes Daxenberger, and Iryna Gurevych. 2017. [Neural end-to-end learning for computational argumentation mining](#). In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017)*, volume Volume 1: Long Papers, pages (11–22). Association for Computational Linguistics.
- Eirini Florou, Stasinou Konstantopoulos, Antonis Koukourikos, and Pythagoras Karampiperis. 2013. [Argument extraction for supporting public policy formulation](#). In *Proceedings of the 7th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pages 49–54.
- Minqing Hu and Bing Liu. 2004. [Mining and summarizing customer reviews](#). In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '04*, pages 168–177, New York, NY, USA. ACM.
- Svetlana Kiritchenko and Saif Mohammad. 2016. [The effect of negators, modals, and degree adverbs on sentiment composition](#). In *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 43–52, San Diego, California. Association for Computational Linguistics.
- Anirban Laha and Vikas Raykar. 2016. [An empirical evaluation of various deep learning architectures for bi-sequence classification tasks](#). In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 2762–2773, Osaka, Japan. The COLING 2016 Organizing Committee.
- Zeyang Lei, Yujiu Yang, and Min Yang. 2018. [Sentiment lexicon enhanced attention-based lstm for sentiment classification](#).
- Ran Levy, Ben Bogin, Shai Gretz, Ranit Aharonov, and Noam Slonim. 2018. [Towards an argumentative content search engine using weak supervision](#). In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2066–2081, Santa Fe, New Mexico, USA. Association for Computational Linguistics.
- Ran Levy, Shai Gretz, Benjamin Sznajder, Shay Hummel, Ranit Aharonov, and Noam Slonim. 2017. [Unsupervised corpus-wide claim detection](#). In *Proceedings of the 4th Workshop on Argument Mining*, pages 79–84, Copenhagen, Denmark. Association for Computational Linguistics.
- George A. Miller. 1995. [Wordnet: A lexical database for english](#). *Commun. ACM*, 38(11):39–41.
- Marie-Francine Moens, Erik Boiy, Raquel Mochales Palau, and Chris Reed. 2007. [Automatic detection of arguments in legal texts](#). In *Proceedings of the 11th International Conference on Artificial Intelligence and Law, ICAIL '07*, pages 225–230, New York, NY, USA. ACM.
- Saif M. Mohammad and Peter D. Turney. 2013. [Crowdsourcing a word-emotion association lexicon](#). 29(3):436–465.
- Huy Nguyen and Diane Litman. 2015. [Extracting argument and domain words for identifying argument components in texts](#). In *Proceedings of the 2nd Workshop on Argumentation Mining*, pages 22–28, Denver, CO. Association for Computational Linguistics.
- Andreas Peldszus and Manfred Stede. 2013. [From argument diagrams to argumentation mining in texts: A survey](#). *Int. J. Cogn. Inform. Nat. Intell.*, 7(1):1–31.
- Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. [Glove: Global vectors for word representation](#). In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.

- Ruty Rinott, Lena Dankin, Carlos Alzate Perez, Mitesh M. Khapra, Ehud Aharoni, and Noam Slonim. 2015. [Show me your evidence - an automatic method for context dependent evidence detection](#). In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 440–450, Lisbon, Portugal. Association for Computational Linguistics.
- Eyal Shnarch, Carlos Alzate, Lena Dankin, Martin Gleize, Yufang Hou, Leshem Choshen, Ranit Aharonov, and Noam Slonim. 2018. [Will it blend? blending weak and strong labeled data in a neural network for argumentation mining](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 599–605, Melbourne, Australia. Association for Computational Linguistics.
- Christian Stab and Iryna Gurevych. 2014. 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.
- Christian Stab, Tristan Miller, and Iryna Gurevych. 2018a. Cross-topic argument mining from heterogeneous sources using attention-based neural networks. *arXiv preprint arXiv:1802.05758*.
- Christian Stab, Tristan Miller, Benjamin Schiller, Pranav Rai, and Iryna Gurevych. 2018b. [Cross-topic argument mining from heterogeneous sources](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3664–3674, Brussels, Belgium. Association for Computational Linguistics.
- Zhiyang Teng, Duy Tin Vo, and Yue Zhang. 2016. [Context-sensitive lexicon features for neural sentiment analysis](#). In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1629–1638, Austin, Texas. Association for Computational Linguistics.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. [Attention is all you need](#). In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 5998–6008. Curran Associates, Inc.
- Yicheng Zou, Tao Gui, Qi Zhang, and Xuanjing Huang. 2018. [A lexicon-based supervised attention model for neural sentiment analysis](#). In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 868–877, Santa Fe, New Mexico, USA. Association for Computational Linguistics.