

# Infusing Knowledge from Wikipedia to Enhance Stance Detection

Zihao He<sup>1,2</sup> Negar Mokhberian<sup>1,2</sup> Kristina Lerman<sup>1</sup>

<sup>1</sup>Information Sciences Institute, University of Southern California

<sup>2</sup>Department of Computer Science, University of Southern California

{zihaohe, nmokhber}@usc.edu lerman@isi.edu

## Abstract

Stance detection infers a text author’s attitude towards a target. This is challenging when the model lacks background knowledge about the target. Here, we show how background knowledge from Wikipedia can help enhance the performance on stance detection. We introduce **Wikipedia Stance Detection BERT (WS-BERT)** that infuses the knowledge into stance encoding. Extensive results on three benchmark datasets covering social media discussions and online debates indicate that our model significantly outperforms the state-of-the-art methods on target-specific stance detection, cross-target stance detection, and zero/few-shot stance detection.<sup>1</sup>

## 1 Introduction

Stance detection aims to automatically identify author’s attitude or standpoint (favor, neutral, against) towards a specific target or topic using text as evidence (Mohammad et al., 2016; Augenstein et al., 2016; Jang and Allan, 2018; Somasundaran and Wiebe, 2010; Stefanov et al., 2020). To precisely capture the stance towards a target, background knowledge about the target is often necessary, especially in cases where the text does not explicitly mention the target, as shown in Figure 1. People have wide-ranging background knowledge regarding various targets and use it to infer the implicit stance in a statement. However, machines by default do not have such knowledge and previous works on stance detection (Allaway and McKeown, 2020; Allaway et al., 2021; Liang et al., 2021; Augenstein et al., 2016; Siddiqua et al., 2019; Sun et al., 2018; Li et al., 2021b; Hardalov et al., 2021) fail to incorporate such knowledge in modeling stances.

In this paper, we propose to utilize background knowledge from Wikipedia about the target as a

<sup>1</sup>Code and data are publicly available at <https://github.com/zihaohe123/wiki-enhanced-stance-detection>.

<b>Target:</b> Donald Trump	<b>Stance:</b> Favor
<b>Document:</b> All Republican presidents are better!	
<b>Background knowledge required:</b> Donald Trump was a Republican president.	
<b>Target:</b> LeBron	<b>Stance:</b> Favor
<b>Document:</b> James is a successful basketball player.	
<b>Background knowledge required:</b> “LeBron” and “James” refer to the same person.	

Figure 1: Two examples of stance detection where background knowledge is required.

bridge to enable the model’s deeper understanding of the target, thus improving its performance on stance detection. We crawl the Wikipedia pages for the targets and use them as external textual information. To infuse this information into stance detection, we propose **Wikipedia Stance Detection BERT (WS-BERT)**, which integrates the representation of Wikipedia knowledge into that of documents and targets. Depending on the textual style of the documents, we introduce two variants of WS-BERT. We conduct a comprehensive set of experiments on three recently published benchmark datasets for stance detection that include social media discussions and online debates, covering three sub-tasks of stance detection: target-specific stance detection, cross-target stance detection, and zero/few-shot stance detection. Significant improvements over the state-of-the-art methods on all datasets and sub-tasks demonstrate the superiority of our model in terms of effectiveness and broad applicability.

**Related Work.** Baly et al. (2018, 2020) use Wikipedia pages of a news medium as an additional source of information to predict the factuality and bias of the medium. However, they use static pretrained BERT (Devlin et al., 2019) embeddings of the Wikipedia pages without finetuning, failing to align the pretrained embeddings to the domain of the target task. Hanawa et al. (2019) first propose to make use of the external knowledge from Wikipedia for stance detection; however,

the authors only consider the promote/suppress relations between the texts and Wikipedia, which require a large amount of manual annotations to extract; in addition, a substantial amount of knowledge that is not captured by such relations is ignored; in contrast, WS-BERT utilizes the original Wikipedia textual knowledge and does not proactively exclude any information. Zhang et al. (2020) propose SEKT to extract external word-level semantic and emotion knowledge, which fails to capture the global relationship between the document and the target; moreover, such a model is designed for cross-target stance detection and is hardly applicable to target-specific and zero/few-shot stance detection. Liu et al. (2021) utilize commonsense knowledge from a knowledge graph by extracting the two-hop paths between entities in the targets and in the documents; however, the existence of such paths do not always hold true and we found that a well-finetuned BERT without external knowledge can achieve performance comparable with it, as shown in Section 3.6.

## 2 Methodology

### 2.1 Problem Definition

Let  $D = \{(x_i = (d_i, t_i, w_i), y_i)\}_{i=1}^N$  denote  $N$  examples, with input  $x_i$  consisting of a document  $d_i$ , target  $t_i$ , and Wikipedia text  $w_i$  about the target, and a stance label  $y_i \in \{\text{favor, against, neutral}\}$  as output. The goal is to infer  $y_i$  given  $x_i$ .

### 2.2 Encoding Wikipedia Knowledge

For the background knowledge, we use the raw text of Wikipedia pages instead of a Wikipedia knowledge graph because 1) a knowledge graph is more structured but inevitably suffers information loss when being constructed; Liu et al. (2021) uses a commonsense knowledge graph to enhance stance detection, which is outperformed by our method that simply uses raw texts, as shown in Section 3.6; 2) in addition, raw text is much more readily accessible and needs less preprocessing, especially for newly emerging targets.

To incorporate background knowledge about targets from Wikipedia, we propose Wikipedia Stance Detection BERT (WS-BERT). Depending on the textual style (formal vs. informal) of the documents, we introduce two variants of WS-BERT, namely WS-BERT-Single, for dealing with formal documents, and WS-BERT-Dual, for dealing with informal documents. Below we elaborate on the

architectures of these models.

**Infusing Wikipedia knowledge with formal documents.** When documents are written in a formal style as Wikipedia articles, we use BERT that is also pretrained on Wikipedia articles to collectively encode the document  $d$ , the target  $t$ , and the Wikipedia knowledge  $w$ . Previous works (Allaway et al., 2021; Liang et al., 2021; Li et al., 2021a; Glandt et al., 2021; Allaway and McKeown, 2020; Liu et al., 2021) treat the document  $d$  and the target  $t$  as a sequence pair and use BERT to encode it, with the input format as “[CLS]  $d$  [SEP]  $t$  [SEP]”. Since BERT was originally designed to deal with at most two sequences<sup>2</sup>, to encode the Wikipedia knowledge in addition to the document-target pair, we merge the document and the target into a single sequence and redesign the input format as “[CLS] Text:  $d$  Target:  $t$  [SEP]  $w$  [SEP]” as shown in Figure 2(a). Such an input format enables  $d$ ,  $t$ , and  $w$  to attend to each other during the encoding process. The pooled output of the final layer [CLS] embedding is used as the final representation of the input  $x$ . Since one BERT is used, we call the model WS-BERT-Single.

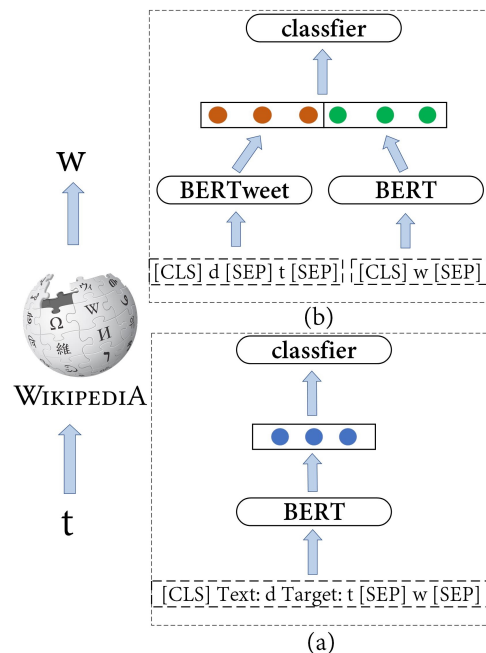


Figure 2: Architecture of (a) WS-BERT-Single and (b) WS-BERT-Dual.

**Infusing Wikipedia knowledge with informal documents.** Social media has become a popular platform for people to express their views on

<sup>2</sup>There do exist some works that have tried to make it encode three sequences simultaneously by using three [SEP] tokens (Xu et al., 2021).

public figures or political events. The opinions of online users are documented by noisy and casual user-generated texts. Such texts have a different distribution than the Wikipedia corpus that BERT is pretrained on. In this case, we use BERTweet (Nguyen et al., 2020) or COVID-Twitter-BERT (Müller et al., 2020) that is pretrained on social media texts to encode the document-target pair, and use the vanilla BERT to encode Wikipedia knowledge, as shown in Figure 2(b). We encode the document-target pair and the Wikipedia knowledge separately with two language models so as to minimize domain shift between the training examples used in this paper and the original pretraining corpora of the language models. We concatenate the two pooled outputs of the final layer [CLS] embeddings from two language models as the final representation of the input  $x$ .<sup>3</sup> We call this model WS-BERT-Dual since we use two BERT-based language models.

### 2.3 Stance Prediction

The final representation from WS-BERT is fed into a single fully-connected layer and softmax layer to predict the stance label  $\hat{y} \in \{\text{favor, against, neutral}\}$ , which is optimized by a cross-entropy loss.

## 3 Experiments

### 3.1 Datasets

We evaluate the proposed WS-BERT model on three newly published datasets since 2020. For the targets in the three datasets, we use summaries of the fetched Wikipedia pages as the textual background knowledge.

P-Stance (Li et al., 2021a) is for target-specific and cross-target stance detection and it consists of tweets related to three politicians “Biden”, “Sanders” and “Trump”. We manually fetched the individual Wikipedia pages of the three politicians.

COVID-19-Stance (Glandt et al., 2021) is a dataset of pandemic-related tweets for target-specific stance detection and contains four targets: “Anthony Fauci”, “stay-at-home orders”, “wear a face mask”, and “keeping school closed”. The titles of the Wikipedia pages used are “Anthony Fauci”, “COVID-19 lockdowns”, “Face masks during the

COVID-19 pandemic in the United States”, and “Impact of the COVID-19 pandemic on education”. Locating these Wikipedia pages is also a manual process.

Varied Stance Topics (VAST) (Allaway and McKeown, 2020) is for zero/few-shot stance detection and comprises comments from *The New York Times* “Room for Debate” section on a large range of topics covering broad themes. It has ~6000 targets. We use an API<sup>4</sup> to crawl the Wikipedia pages of them. For the targets that have multiple related Wikipedia pages, we choose the first one recommended by the API. For the targets that do not have any Wikipedia pages (~200, e.g., “salt preference” and “tennis fans”), we use the targets themselves as background knowledge, with no additional information introduced.

### 3.2 Evaluation Metric

Following previous works (Mohammad et al., 2016, 2017), we adopt macro-average of F1-score as the evaluation metric. For P-Stance where the examples only have two stance labels,  $F_{\text{avg}} = (F_{\text{favor}} + F_{\text{against}})/2$ . For COVID-19-Stance and VAST that have three stance labels,  $F_{\text{avg}} = (F_{\text{favor}} + F_{\text{against}} + F_{\text{neutral}})/3$ .

### 3.3 Experimental Setup

We use WS-BERT-Dual in experiments on P-Stance and COVID-19-Stance, both of which consist of tweets. Following the setup in their original papers (Li et al., 2021a; Glandt et al., 2021), for P-Stance, we use BERTweet as the document-target encoder, and for COVID-19 Stance, we use COVID-Twitter-BERT as the the document-target encoder; for both datasets, BERT-base is used to encode Wikipedia knowledge. On VAST that comprises online debates, we use BERT-base to jointly encode the document-target-knowledge tuple.

All models are implemented using PyTorch. The Wikipedia summaries are truncated to a maximum of 512 tokens. We train the models using Adam optimizer with a batch size of 32 for a maximum of 100 epochs with patience of 10 epochs. The weight decay is set to  $5e - 5$ . To speed up the training process we only finetune the top layers of the Wikipedia encoder in WS-BERT-Dual. We search the learning rate in  $\{1e - 5, 2e - 5\}$  and the number of Wikipedia encoder layers to finetune in  $\{1, 2\}$ .

<sup>3</sup>Admittedly, concatenation of the two vectors seems naive, but it achieves satisfactory performance as shown in Section 3.4 and 3.5; more sophisticated ways to fuse them like cross-attention count towards our future work.

<sup>4</sup><https://pypi.org/project/wikipedia/>

On target-specific and zero/few-shot stance detection, we follow the standard train/validation/test splits of the three datasets. On cross-target stance detection, the model is trained on the train set of the source target, evaluated on the validation set of the source target, and tested on the combination of train, validation, and test set of the destination target, following the setup in P-Stance. The results are reported from the model with the best performance on the validation set.

### 3.4 Target-specific Stance Detection

For target-specific stance detection on P-Stance and COVID-19-Stance, we train a model separately for each target and test it on the same target.

**Baselines.** On P-Stance we compare to the baselines TAN (Du et al., 2017), BiCE (Augenstein et al., 2016), PGNN (Huang and Carley, 2018), BERT, and BERTweet. On COVID-19-Stance we compare to TAN, ATGRU (Zhou et al., 2017), GCAE (Xue and Li, 2018), COVID-Twitter-BERT, COVID-Twitter-BERT-NS (Xie et al., 2020), and COVID-Twitter-BERT-DAN (Xu et al., 2020).

Method	Trump	Biden	Sanders	Avg.
TAN	77.1	77.6	71.6	75.1
BiCE	77.2	77.7	71.2	75.4
PGCNN	76.9	76.6	72.1	75.2
GCAE	79.0	78.0	71.8	76.3
BERT	78.3	78.7	72.5	76.5
BERTweet	82.5	81.0	78.1	80.5
BERTweet†	85.2	82.5	78.5	82.1
WS-BERT-Dual	<b>85.8</b>	<b>83.5</b>	<b>79.0</b>	<b>82.8</b>

Table 1: Macro-average F1 scores of target-specific stance detection on P-Stance. BERTweet is implemented in (Li et al., 2021a) and BERTweet† is implemented in this paper.

Method	Fauci	Home	Mask	School	Avg.
TAN	54.7	53.6	54.6	53.4	54.1
ATRGRU	61.2	52.1	59.9	52.7	56.5
GCAE	64.0	64.5	63.3	49.0	60.2
CT-BERT	81.8	80.0	80.3	75.5	79.4
CT-BERT-NS	82.1	78.4	83.3	75.3	79.8
CT-BERT-DAN	83.2	78.7	82.5	71.7	79.0
CT-BERT†	83.0	83.6	83.8	81.7	83.0
WS-BERT-Dual	<b>83.6</b>	<b>85.0</b>	<b>86.6</b>	<b>82.2</b>	<b>84.4</b>

Table 2: Macro-average F1 scores of target-specific stance detection on COVID-19-Stance. CT-BERT (short for COVID-Twitter-BERT) represents COVID-Twitter-BERT implemented in (Glandt et al., 2021) and CT-BERT† represents the model implemented in this paper.

**Results and Analysis.** Results for P-Stance and COVID-19-Stance are shown in Table 1 and Table

2. On P-Stance, BERTweet† outperforms the baselines on all targets, and WS-BERT-Dual further improves the performance and achieves the new state-of-the-art. On COVID-19-Stance, COVID-Twitter-BERT† outperforms all the baselines on targets except “Fauci”, including the self-training baseline COVID-Twitter-BERT-NS and the domain adaptation baseline COVID-Twitter-BERT-DAN, both of which are trained using some additional external data. However, WS-BERT-Dual augmented with background knowledge outperforms state-of-the-art on all targets. Therefore, even on target-specific stance detection, where the models are fed sufficient data to learn the target, background knowledge about the target still helps improve performance.

### 3.5 Cross-target Stance Detection

We use P-Stance for cross-target stance detection, where the model is trained on one target, e.g., “Trump”, and tested on another, e.g., “Biden.”

**Baselines.** We use BERTweet as a strong baseline, which is the most performant method reported in (Li et al., 2021a).

Target	BERTw	BERTw†	WS-BERT-D
Trump→Biden	58.9	52.2	<b>68.3</b>
Trump→Sanders	56.5	53.0	<b>64.4</b>
Biden→Trump	63.6	66.8	<b>67.7</b>
Biden→Sanders	67.0	68.5	<b>69.0</b>
Sanders→Trump	58.7	60.0	<b>63.6</b>
Sanders→Biden	73.0	74.6	<b>76.8</b>
Avg.	63.0	62.5	<b>68.3</b>

Table 3: Macro-average F1 scores of cross-target stance detection on P-Stance. Trump→Biden indicates that the model is trained on “Donald Trump” and tested on “Joe Biden”. BERTweet is implemented in (Li et al., 2021a) and BERTweet† is implemented in this paper.

**Results and Analysis.** Results are shown in Table 3. We see that our implementation of BERTweet† outperforms BERTweet when the model is trained on “Biden” and “Sanders”. After infusing Wikipedia knowledge, WS-BERT-Dual enhances the performance on all six target pairs compared to BERTweet† and achieves the new state-of-the-art. Notably, the performance gains on “Trump”→“Biden” and “Trump”→“Sanders” are the biggest, which we argue is because the tweets about “Trump” mention the other two targets less, so that the model trained on “Trump” learns little knowledge transferable to the other two targets. In this case, background knowledge about “Biden” or “Sanders” brings huge information gains, leading to

substantial performance improvement. In addition, compared to performance gain on target-specific stance detection, the gains in performance are more noticeable on this cross-target task, which signifies that background knowledge from Wikipedia is more important when the test target is outside of the training set.

### 3.6 Zero-shot and Few-shot Stance Detection

Finally, we evaluate our model on zero-shot and few-shot stance detection using VAST, where the model is trained on thousands of targets and evaluated on targets that are not seen in the training data (zero-shot learning) and are seen just a few times in the training data (few-shot learning).

**Baselines.** We compare our model to BERT, TGA-Net (Allaway and McKeown, 2020), BERT-GCN (Lin et al., 2021), and CKE-Net (Liu et al., 2021).

Method	Zero-shot	Few-shot	Overall
TGA-Net	66.6	66.3	66.5
BERT	68.5	68.4	68.4
BERT-GCN	68.6	69.7	69.2
CKE-Net	70.2	70.1	70.1
BERT <sup>†</sup>	70.1	70.0	70.0
WS-BERT-Single	<b>75.3</b>	<b>73.6</b>	<b>74.5</b>

Table 4: Macro-average F1 scores of zero-shot and few-shot stance detection on VAST. BERT is implemented in (Liu et al., 2021) and BERT<sup>†</sup> is implemented in this paper.

**Results and Analysis.** Results are shown in Table 4. CKE-Net extracts the links between entities in targets and documents from a knowledge graph so as to make use of the commonsense knowledge. However, a well-finetuned BERT<sup>†</sup> implemented in this paper achieves performance on par with it, putting the effectiveness of CKE-Net into question. WS-BERT-Single significantly improves the performance on both zero-shot and few-shot learning by a huge margin, thus creating new state-of-the-art. We argue that such nontrivial performance gain is due to the presence of many targets in VAST that are difficult for the model to understand without background knowledge, such as “b-12” (a vitamin) and “2big2fail”.

As mentioned in Section 3.1, the Wikipedia pages of the thousands of targets in VAST are retrieved by an API. Admittedly, such an automated process might incur noisy information because the retrieved pages are not guaranteed to be the most relevant ones, and the summaries might miss useful

content. However, even with the noise, our method manages to outperform the state-of-the-art baselines significantly, with an improvement in F1 of 4.5%. Such a huge improvement demonstrates the robustness of our method in handling the noisy external knowledge: when the model is trained with noisy Wikipedia summaries, it learns to deal with such perturbations; as a result, during inference, with noisy external knowledge, it is still able to infer the correct stance.

Moreover, the improvement on zero-shot learning is more observable compared to that on few-shot learning, because in few-shot learning the model is able to attend to some examples in the training data to understand the targets, while in zero-shot learning the model is not exposed to the targets at all, in which case background knowledge is of more importance.

## 4 Conclusion

In this paper we propose to utilize background knowledge about targets from Wikipedia to enhance stance detection. We propose WS-BERT with two variants to encode such knowledge. Such a simple yet effective method achieves state-of-the-art performance on three benchmark datasets and on three sub-tasks: in-target stance detection, cross-target stance detection, and zero/few-shot stance detection. The comprehensive and growing list of topics covered by Wikipedia ensures that our method will adapt to newly emerging targets.

In the future, we plan to investigate incorporating knowledge about entities in the input documents, in addition to knowledge about the targets. Since Wikipedia pages may contain subjective opinions towards the targets, how to prevent the model from being negatively impacted by such bias when modeling the knowledge remains a promising research direction. Moreover, background knowledge from relevant news articles might also be helpful for inferring stances.

## Acknowledgements

We sincerely thank the reviewers for their insightful and constructive comments and suggestions that helped improve the paper. This research was supported in part by DARPA under contract HR001121C0168.

## References

- Emily Allaway and Kathleen McKeown. 2020. [Zero-Shot Stance Detection: A Dataset and Model using Generalized Topic Representations](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8913–8931, Online. Association for Computational Linguistics.
- Emily Allaway, Malavika Srikanth, and Kathleen McKeown. 2021. [Adversarial learning for zero-shot stance detection on social media](#). In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4756–4767, Online. Association for Computational Linguistics.
- Isabelle Augenstein, Tim Rocktäschel, Andreas Vlachos, and Kalina Bontcheva. 2016. [Stance detection with bidirectional conditional encoding](#). In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 876–885, Austin, Texas. Association for Computational Linguistics.
- Ramy Baly, Georgi Karadzhov, Dimitar Alexandrov, James Glass, and Preslav Nakov. 2018. [Predicting factuality of reporting and bias of news media sources](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3528–3539, Brussels, Belgium. Association for Computational Linguistics.
- Ramy Baly, Georgi Karadzhov, Jisun An, Haewoon Kwak, Yoan Dinkov, Ahmed Ali, James Glass, and Preslav Nakov. 2020. [What was written vs. who read it: News media profiling using text analysis and social media context](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3364–3374, Online. Association for Computational Linguistics.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of deep bidirectional transformers for language understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Jiachen Du, Ruifeng Xu, Yulan He, and Lin Gui. 2017. [Stance classification with target-specific neural attention networks](#). In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pages 3988–3994.
- Kyle Glandt, Sarthak Khanal, Yingjie Li, Doina Caragea, and Cornelia Caragea. 2021. [Stance detection in COVID-19 tweets](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1596–1611, Online. Association for Computational Linguistics.
- Kazuaki Hanawa, Akira Sasaki, Naoaki Okazaki, and Kentaro Inui. 2019. [Stance detection attending external knowledge from wikipedia](#). *Journal of Information Processing*, 27:499–506.
- Momchil Hardalov, Arnav Arora, Preslav Nakov, and Isabelle Augenstein. 2021. [Cross-domain label-adaptive stance detection](#). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9011–9028, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Binxuan Huang and Kathleen Carley. 2018. [Parameterized convolutional neural networks for aspect level sentiment classification](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1091–1096, Brussels, Belgium. Association for Computational Linguistics.
- Myungha Jang and James Allan. 2018. [Explaining controversy on social media via stance summarization](#). In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 1221–1224.
- Yingjie Li, Tiberiu Sosea, Aditya Sawant, Ajith Jayaraman Nair, Diana Inkpen, and Cornelia Caragea. 2021a. [P-stance: A large dataset for stance detection in political domain](#). In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 2355–2365, Online. Association for Computational Linguistics.
- Yingjie Li, Chenye Zhao, and Cornelia Caragea. 2021b. [Improving stance detection with multi-dataset learning and knowledge distillation](#). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6332–6345, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Bin Liang, Yonghao Fu, Lin Gui, Min Yang, Jiachen Du, Yulan He, and Ruifeng Xu. 2021. [Target-adaptive graph for cross-target stance detection](#). In *Proceedings of the Web Conference 2021*, pages 3453–3464.
- Yuxiao Lin, Yuxian Meng, Xiaofei Sun, Qinghong Han, Kun Kuang, Jiwei Li, and Fei Wu. 2021. [BertGCN: Transductive text classification by combining GNN and BERT](#). In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1456–1462, Online. Association for Computational Linguistics.
- Rui Liu, Zheng Lin, Yutong Tan, and Weiping Wang. 2021. [Enhancing zero-shot and few-shot stance detection with commonsense knowledge graph](#). In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 3152–3157.

- Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. [SemEval-2016 task 6: Detecting stance in tweets](#). In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 31–41, San Diego, California. Association for Computational Linguistics.
- Saif M Mohammad, Parinaz Sobhani, and Svetlana Kiritchenko. 2017. Stance and sentiment in tweets. *ACM Transactions on Internet Technology (TOIT)*, 17(3):1–23.
- Martin Müller, Marcel Salathé, and Per E Kummervold. 2020. Covid-twitter-bert: A natural language processing model to analyse covid-19 content on twitter. *arXiv preprint arXiv:2005.07503*.
- Dat Quoc Nguyen, Thanh Vu, and Anh Tuan Nguyen. 2020. [BERTweet: A pre-trained language model for English tweets](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 9–14, Online. Association for Computational Linguistics.
- Umme Aymun Siddiqua, Abu Nowshed Chy, and Masaki Aono. 2019. [Tweet stance detection using an attention based neural ensemble model](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1868–1873, Minneapolis, Minnesota. Association for Computational Linguistics.
- Swapna Somasundaran and Janyce Wiebe. 2010. [Recognizing stances in ideological on-line debates](#). In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 116–124, Los Angeles, CA. Association for Computational Linguistics.
- Peter Stefanov, Kareem Darwish, Atanas Atanasov, and Preslav Nakov. 2020. [Predicting the topical stance and political leaning of media using tweets](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 527–537, Online. Association for Computational Linguistics.
- Qingying Sun, Zhongqing Wang, Qiaoming Zhu, and Guodong Zhou. 2018. [Stance detection with hierarchical attention network](#). In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2399–2409, Santa Fe, New Mexico, USA. Association for Computational Linguistics.
- Qizhe Xie, Minh-Thang Luong, Eduard Hovy, and Quoc V Le. 2020. Self-training with noisy student improves imagenet classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10687–10698.
- Chang Xu, Cécile Paris, Surya Nepal, Ross Sparks, Chong Long, and Yafang Wang. 2020. Dan: Dual-view representation learning for adapting stance classifiers to new domains. In *ECAI 2020*, pages 2260–2267. IOS Press.
- Yichong Xu, Chenguang Zhu, Ruochen Xu, Yang Liu, Michael Zeng, and Xuedong Huang. 2021. Fusing context into knowledge graph for commonsense question answering. In *Workshop on Commonsense Reasoning and Knowledge Bases*.
- Wei Xue and Tao Li. 2018. [Aspect based sentiment analysis with gated convolutional networks](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2514–2523, Melbourne, Australia. Association for Computational Linguistics.
- Bowen Zhang, Min Yang, Xutao Li, Yunming Ye, Xiaofei Xu, and Kuai Dai. 2020. [Enhancing cross-target stance detection with transferable semantic-emotion knowledge](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3188–3197, Online. Association for Computational Linguistics.
- Yiwei Zhou, Alexandra I Cristea, and Lei Shi. 2017. Connecting targets to tweets: Semantic attention-based model for target-specific stance detection. In *International Conference on Web Information Systems Engineering*, pages 18–32. Springer.