

# A Multi-Task Learning Framework for Multi-Target Stance Detection

Yingjie Li and Cornelia Caragea

Computer Science Department

University of Illinois at Chicago

yli300@uic.edu, cornelia@uic.edu

## Abstract

Multi-target stance detection aims to identify the stance taken toward a pair of different targets from the same text, and typically, there are multiple target pairs per dataset. Existing works generally train one model for each target pair. However, they fail to learn target-specific representations and are prone to overfitting. In this paper, we propose a new training strategy under the multi-task learning setting by training one model on all target pairs, which helps the model learn more universal representations and alleviate overfitting. Moreover, in order to extract more accurate target-specific representations, we propose a multi-task learning network which can jointly train our model with a stance (dis)agreement detection task that is designed to identify agreement and disagreement between stances in paired texts. Experimental results demonstrate that our proposed model outperforms the best-performing baseline by 12.39% in macro-averaged F1-score. Our resources are publicly available on GitHub.<sup>1</sup>

## 1 Introduction

Nowadays, people often take to social media to express their stances toward specific targets (e.g., various political figures). These stances in an aggregate can provide valuable information for obtaining insight into important events such as presidential elections. The common stance detection task is to determine from a piece of text whether the author of the text is in favor of, neutral or against to a specific target, which can be categorized as single-target stance detection (STSD) (Mohammad et al., 2016; Küçük and Can, 2020; ALDayel and Magdy, 2021). More recently, since people often comment on multiple target entities in the same text, a more challenging task, i.e., multi-target stance detection (MTSD), was designed to test whether a model can

<sup>1</sup><https://github.com/chuchun8/MTSD>

---

<b>Tweet:</b>	#Trump2016 can beat #HillaryClinton as he is easily beating #JebBush ;) People R sick and tired of career politicians.	
<b>Target 1:</b>	Donald Trump	<b>Stance 1:</b> FAVOR
<b>Target 2:</b>	Hillary Clinton	<b>Stance 2:</b> AGAINST

---

Figure 1: An example of multi-target stance detection.

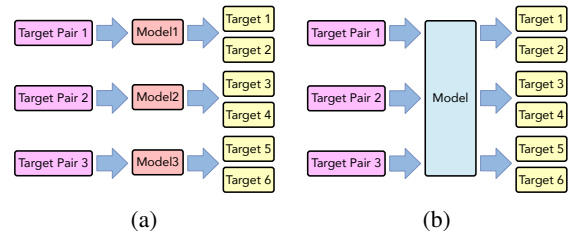


Figure 2: The left figure represents the previous “Ad-hoc” training setting in which a model is trained only on a target pair. The right figure represents the proposed “Merged” training setting in which the model is trained on all target pairs.

accurately predict the stance toward multiple targets in the same text (Sobhani et al., 2017). For example, for the tweet in Figure 1, the author aims at expressing stance toward two targets, Donald Trump and Hillary Clinton, implied by the presence of the words “beat” and “career politicians”.

**Problem statement.** Given a sentence  $x = [w_1, w_2, t_1, w_3, \dots, w_{l-1}, t_2, w_l]$ , where  $t_1$  and  $t_2$  are targets, and  $w_i$ ,  $i = 1, \dots, l$ , denotes a non-target word, the goal of MTSD is to classify the stance toward these targets into one of the three classes: {FAVOR, AGAINST, NONE}.

Previous work focused on a per-target-pair training strategy, which aims to train one model for each target pair and evaluate it on the test data corresponding to that target-pair (which we call “Ad-hoc” training). The framework is illustrated in Figure 2(a). However, the model is more likely to make predictions based on specific words without

fully considering the target information, and hence, to overfit in the “Ad-hoc” training setting. To address this, as shown in Figure 2(b), we propose a “Merged” training strategy by training one model on data from all target pairs, which helps the model learn more universal representations on the whole dataset and alleviate overfitting. Furthermore, in order to extract more accurate target-specific representations, we propose a multi-task learning network which is able to jointly train our model with a stance (dis)agreement detection task that is designed to identify agreement and disagreement between stances expressed in paired-target sentences. Results show that the proposed “Merged” training setting together with identifying whether the author expresses the same stance toward two targets are beneficial to the MTSD.

Our contributions include: 1) We propose a “Merged” training strategy for MTSD and show that models fine-tuned on the pre-trained BERTweet (Nguyen et al., 2020) perform substantially better than strong baselines. Meanwhile, the decrease in performance can be widely observed in baseline results when using the “Merged” training strategy, making it a more challenging evaluation for MTSD; 2) We propose a multi-task learning network which considers the stance (dis)agreement detection task as an auxiliary task to further improve the performance of our proposed model; 3) Our proposed model outperforms the best-performing baseline by 12.39% in macro-averaged F1-score.

## 2 Related Work

Sobhani et al. (2017) introduced the MTSD task and presented the first dataset. They also proposed an attention-based encoder-decoder (Seq2Seq) model that predicts stance labels by focusing on important parts of a tweet. Wei et al. (2018a) proposed a dynamic memory network for detecting stance. First, target-specific attention is extracted for each target. Then, a shared external memory module that maintains useful information for targets is dynamically updated. This model achieves state-of-the-art performance on the multi-target stance dataset of Sobhani et al. (2017). We used the above two works as strong baselines for our evaluation.

Sobhani et al. (2017) and Wei et al. (2018a) deal with MTSD by training one model for each target pair and the model predicts the stance toward two targets simultaneously. However, we can also solve this task by treating it as a special case of single-

target stance detection (STSD). Instead of training a model that receives two targets and a sentence as an input, we train two STSD models that receive one target and a sentence as an input, on each target pair. For the example in Figure 1, we train one STSD model for target “Donald Trump” and train another model for “Hillary Clinton” in a STSD manner.

Previous studies on STSD often employ feature engineering (Sobhani et al., 2016; Mohammad et al., 2016), Convolutional Neural Network (CNN) (Vijayaraghavan et al., 2016; Wei et al., 2016) and Recurrent Neural Network (RNN) (Zarrella and Marsh, 2016) to predict the stance for a given target. One of the major limitations is that they do not consider the target information. To address this, Augenstein et al. (2016) proposed a conditional BiLSTM encoder that learns tweet representations conditioned on the respective target. More recently, inspired by the attention mechanism (Bahdanau et al., 2015), various target-specific attention-based approaches (Du et al., 2017; Sun et al., 2018; Wei et al., 2018b; Li and Caragea, 2019, 2021) have been proposed to connect the target with the sentence representation, which is similar to aspect-based sentiment analysis (Hazarika et al., 2018; Majumder et al., 2018; Lin et al., 2019; Song et al., 2019). We compare the baseline models of STSD and MTSD with our proposed model in §4.4 using both “Ad-hoc” and “Merged” settings.

## 3 Approach

Previous work focused on an “Ad-hoc” training strategy, which fails to explore the potential of the training data and is unable to learn universal representations of targets. Moreover, we observe that STSD models that do not consider target information can still perform well on the multi-target dataset, which makes MTSD easier. Therefore, in order to learn more universal representations and better evaluate the performance of models on MTSD, we propose a “Merged” training strategy by training one model on all target pairs. More specifically, the model is trained on training data combined from all target pairs, and tested on each target pair separately to be compared with the results of the “Ad-hoc” strategy. Our proposed training strategy can be considered as a multi-task learning approach that helps the pre-trained language models to learn more generalized text representations by sharing the domain-specific information

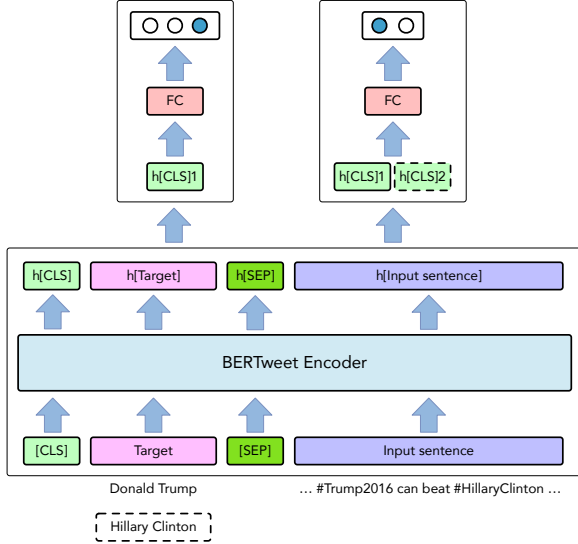


Figure 3: Model architecture.

across the related target pairs.

BERTweet (Nguyen et al., 2020) is a large-scale language model pre-trained on 850M tweets. BERTweet follows the training procedure of RoBERTa (Liu et al., 2019) and uses the same model configuration with BERT-base (Devlin et al., 2019). We fine-tune the pre-trained BERTweet on the multi-target dataset. The model architecture is shown in Figure 3. Given an input data  $x$  and a target  $t$  ( $t$  is either target 1 or target 2 in Figure 1), we formulate the input as a sequence  $s = [[CLS] t [SEP] x]$  where  $[CLS]$  encodes the sentence representation and  $[SEP]$  is used to separate the target  $t$  and sentence  $x$ . We utilize the  $[CLS]$  token  $h_{[CLS]}$  to get the prediction  $\hat{p}_1$  toward target  $t$ .

In order to learn better target-specific representations, we propose a multi-task learning network that can jointly train our model with a stance (dis)agreement detection task, which is a binary classification task where the label is 1 when the author expresses the same stance toward two targets (e.g., “FAVOR” and “FAVOR”) and 0 otherwise (e.g., “FAVOR” and “AGAINST”). More specifically, given an input data  $x$  and two targets  $t_1$  and  $t_2$ , we formulate the inputs as  $[[CLS] t_1 [SEP] x]$  and  $[[CLS] t_2 [SEP] x]$ . Then we leverage the representations of  $[CLS]$  token of two sequences to detect whether the author of the text expresses the same stance toward two targets. The (dis)agreement class probability  $\hat{p}_2$  can be computed as follows:

$$\hat{p}_2 = \text{softmax}(W_2 f(W_1 [h_{[CLS]1}; h_{[CLS]2}] + b_1) + b_2)$$

Target Pair	Total	Train	Dev	Test
Trump-Clinton	1,722	1,240	177	355
Trump-Cruz	1,317	922	132	263
Clinton-Sanders	1,366	957	137	272
Total	4,455	3,119	446	890

Table 1: Distribution of instances in our dataset.

where  $W_1 \in \mathbb{R}^{d_h * 2d_h}$ ,  $W_2 \in \mathbb{R}^{2 * d_h}$ ,  $b_1 \in \mathbb{R}^{d_h}$ ,  $b_2 \in \mathbb{R}^2$ ,  $d_h$  is the size of the hidden dimension,  $f$  is an activation function, semicolon denotes vector concatenation. Note that the main task is to identify the stance toward target  $t_1$ . The target  $t_2$  is only used in auxiliary task. Similarly, we predict the stance toward target  $t_2$  in main task where  $t_1$  is only used in auxiliary task.

Let  $D$  be a labeled training dataset and  $D_j$  be a mini-batch for the MTSD, and let  $y_1$  and  $y_2$  denote the true labels for stance detection task and (dis)agreement task, respectively. The cross-entropy loss is used to train the model. Let  $L_1$  and  $L_2$  be the loss of stance detection task and (dis)agreement task, respectively. Then the final loss is:

$$L = \sum_{i \in D_j} L_1(y_1^i, \hat{p}_1^i) + \alpha \sum_{i \in D_j} L_2(y_2^i, \hat{p}_2^i)$$

where  $i$  is the index of a data sample and  $\alpha$  is a hyper-parameter to account for the importance of the auxiliary task.  $\alpha$  is set to 0.5 in our experiments.

## 4 Experiments

### 4.1 Dataset

To test the performance of our proposed model, we use the multi-target stance dataset (Sobhani et al., 2017) of tweets annotated with stance labels with respect to two targets. This dataset contains three different target pairs: Donald Trump and Hillary Clinton, Donald Trump and Ted Cruz, Hillary Clinton and Bernie Sanders. Table 1 provides dataset statistics. Each tweet has two stance labels concerning two targets and each label has one of the values: “FAVOR”, “AGAINST” or “NONE”.

### 4.2 Evaluation Metrics

$F_{avg}^p$  and macro-average of F1-score ( $F_{macro}$ ) are adopted to evaluate the performance of our baseline models. First, the F1-score of label “Favor” and “Against” is calculated as follows:

$$F_{favor} = \frac{2P_{favor}R_{favor}}{P_{favor} + R_{favor}}$$

$$F_{against} = \frac{2P_{against}R_{against}}{P_{against} + R_{against}}$$

where P and R are precision and recall, respectively. After that, the  $F_{avg}$  is calculated as:

$$F_{avg} = \frac{F_{favor} + F_{against}}{2}$$

For each target pair, we compute the  $F_{avg}$  for each target and use the  $F_{avg}^p$ , which is calculated as the average of  $F_{avg}$  on two targets, as our evaluation metric. Moreover, we get  $F_{macro}$  by averaging  $F_{avg}^p$  on all target pairs.

### 4.3 Baseline Methods

First, we compare the proposed model with the following baselines from STSD.

**BiLSTM** (Schuster and Paliwal, 1997): A BiLSTM model that takes sentences as inputs without considering the target information.

**CNN** (Kim, 2014): The vanilla CNN that has the same input format with BiLSTM. Similarly, target information is not considered in this model.

**TAN** (Du et al., 2017): TAN is an attention-based LSTM that extracts target specific features.

**BiCE** (Augenstein et al., 2016): A BiLSTM model that uses conditional encoding for stance detection. The target information is first encoded by using a BiLSTM and the tweet is then encoded by another BiLSTM, whose state is initialised with the hidden representation of the target.

**GCAE** (Xue and Li, 2018): A model that is based on CNNs and gating mechanism, which is designed to block target-unrelated information.

**PGCNN** (Huang and Carley, 2018): Similar to GCAE, PGCNN is based on gated convolutional networks and encodes target information by generating target-sensitive filters.

The second group contains baselines from MTSD.

**Seq2Seq** (Sobhani et al., 2017): An attention-based encoder-decoder model that generates stance labels according to different parts of a tweet.

**DMAN** (Wei et al., 2018a): Using attention and memory modules to extract important information for detecting stance.

We compare the baselines of STSD and MTSD with our proposed models.

Model	Tr-CI	Tr-Cr	Cl-Sa	F <sub>macro</sub>
<b>Merged</b>				
BiLSTM	43.33	47.51	41.86	44.24
CNN	43.22	49.21	41.22	44.55
GCAE	59.07	54.28	56.13	56.49
PGCNN	59.18	54.62	50.59	54.80
TAN	43.88	50.46	45.63	46.66
BiCE	53.73	51.00	45.84	50.19
DMAN	57.43	52.62	53.87	54.64
<b>BERTweet</b>	67.38 <sup>†</sup>	70.30 <sup>†</sup>	65.64 <sup>†</sup>	67.77
<b>BERTweet-A</b>	<b>69.22<sup>†‡</sup></b>	<b>70.73<sup>†</sup></b>	<b>69.00<sup>†‡</sup></b>	<b>69.65</b>
<b>Ad-hoc</b>				
BiLSTM	58.16	52.75	52.67	54.52
CNN	59.75	55.68	56.13	57.19
GCAE	59.78	56.07	55.92	57.26
PGCNN	56.99	54.19	55.05	55.41
TAN	58.33	54.32	53.16	55.27
BiCE	58.67	53.77	51.87	54.77
Seq2Seq	56.60*	53.12*	54.72*	54.81*
DMAN	60.05	54.27	52.57	55.63
<b>BERTweet</b>	64.29 <sup>†</sup>	56.44	57.80 <sup>†</sup>	59.51
<b>BERTweet-A</b>	<b>65.55<sup>†</sup></b>	<b>57.96<sup>†</sup></b>	<b>58.17<sup>†</sup></b>	<b>60.56</b>

Table 2: Comparison with the baselines on the multi-target stance dataset (%). \*: the result is from the original paper. †: the proposed models improve the best baseline at  $p < 0.05$  with two-tailed t-test. ‡: the BERTweet-A improves the BERTweet at  $p < 0.05$  with two-tailed t-test.  $F_{macro}$  is the average of all target pairs. Bold scores are best overall.

**BERTweet** We fine-tune the BERTweet model using “Merged” and “Ad-hoc” training strategies. The pre-trained BERTweet model is fine-tuned under the PyTorch framework. When fine-tuning, the batch size is 32 and maximum sequence length is 128. We use AdamW optimizer (Loshchilov and Hutter, 2019) and the learning rate is  $2e-5$ .

**BERTweet-A** BERTweet is further improved by joint training with another stance detection task that identifies agreement and disagreement between stances in “Merged” and “Ad-hoc” training settings.

### 4.4 Results and Analysis

**Main Results** Table 2 shows the results of the comparison of our proposed models with the baselines mentioned above by using the proposed training strategy “Merged” and the “Ad-hoc” training strategy. We make the following observations.

First, the performance of baseline models that perform well in the “Ad-hoc” training setting drops heavily in our proposed “Merged” setting, especially for the BiLSTM and CNN. Specifically, the  $F_{macro}$  of BiLSTM and CNN drops by 10.28% and 12.64%, respectively. The results indicate that baseline models overfit the training data quite heavily

Model	Donald Trump
BERTweet-adhoc	46.75
<b>BERTweet-merged</b>	<b>52.51<sup>†</sup></b>

Table 3: Performance comparison of models on the target “Donald Trump” of SemEval 2016 stance dataset (%). †: the proposed BERTweet-merged improves the BERTweet-adhoc at  $p < 0.05$  with two-tailed t-test. Bold scores are best overall.

and our proposed “Merged” training strategy can serve as a better evaluation method to test whether the model learns target-specific features.

Second, different from other baselines suffering significant performance drops, BERTweet performs better in the “Merged” setting. Training all target pairs improves the  $F_{macro}$  of BERTweet from 59.51% to 67.77%, which demonstrates that BERTweet learns more universal representations with respect to targets by leveraging the data of multiple target pairs. Moreover, joint training with stance (dis)agreement detection task further improves the  $F_{macro}$  of BERTweet from 67.77% to 69.65% in the “Merged” setting. Similarly, in the “Ad-hoc” setting, the  $F_{macro}$  of BERTweet is improved from 59.51% to 60.56%, indicating that this auxiliary task is beneficial to the MTSD in both settings and helps the model put more attention on the target-related words.

Third, BERTweet-A of the “Merged” setting significantly outperforms the best-performing baseline by 12.39% in  $F_{macro}$ , showing the effectiveness of the proposed model.

**Generalization Analysis** To test the generalization ability of the BERTweet of the “Merged” setting (which we call BERTweet-merged), we train and validate the BERTweet-merged without auxiliary agreement task on the whole multi-target dataset and test it on the target “Donald Trump” of SemEval 2016 dataset (Mohammad et al., 2016) where an overall shift in the distribution of words and topics can be observed. Moreover, we train and validate the BERTweet-adhoc (BERTweet in “Ad-hoc” setting) on the target “Donald Trump” of multi-target dataset and test it on the same set of SemEval 2016 dataset to be compared with BERTweet-merged. The results are shown in Table 3. From the table, we can observe that BERTweet-merged significantly outperforms BERTweet-adhoc on the SemEval 2016 dataset, which indicates that the BERTweet model trained

in the “Merged” setting shows better generalization ability than the BERTweet model trained in the “Ad-hoc” setting.

## 5 Conclusion

In this paper, we presented a comprehensive investigation into multi-target stance detection (MTSD) and proposed a more challenging task that trains a single model on data from all target pairs instead of training a model per target pair. The new training strategy can alleviate overfitting and help the model learn more universal representations by using the data of all target pairs. Moreover, we proposed to integrate a stance (dis)agreement detection module into the proposed model as an auxiliary task to gain more accurate representations of targets. Experimental results show that the proposed model outperforms the best-performing baseline by a large margin and demonstrates its effectiveness even in the face of a more challenging evaluation. Future work includes extending the proposed training strategy and (dis)agreement task to more stance detection tasks and datasets.

## Acknowledgments

We thank the National Science Foundation for support from grants IIS-1912887 and IIS-1903963 which sponsored the research in this study. Any opinions, findings, and conclusions expressed here are those of the authors and do not necessarily reflect the views of NSF. We also thank Amazon Web Services for support for the computational resources. We are grateful to our reviewers for their insightful comments.

## References

- Abeer ALDayel and Walid Magdy. 2021. [Stance detection on social media: State of the art and trends](#). *Information Processing & Management*, 58(4):102597.
- 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.
- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. [Neural machine translation by jointly learning to align and translate](#). In *3rd International Conference on Learning Representations*.
- 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.
- 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.
- Devamanyu Hazarika, Soujanya Poria, Prateek Vij, Gangeshwar Krishnamurthy, Erik Cambria, and Roger Zimmermann. 2018. [Modeling inter-aspect dependencies for aspect-based sentiment analysis](#). In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 266–270.
- 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.
- Yoon Kim. 2014. [Convolutional neural networks for sentence classification](#). In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751.
- Dilek Küçük and Fazli Can. 2020. [Stance detection: A survey](#). *ACM Comput. Surv.*, 53(1):1–37.
- Yingjie Li and Cornelia Caragea. 2019. [Multi-task stance detection with sentiment and stance lexicons](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6298–6304.
- Yingjie Li and Cornelia Caragea. 2021. [Target-aware data augmentation for stance detection](#). In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1850–1860.
- Peiqin Lin, Meng Yang, and Jianhuang Lai. 2019. [Deep mask memory network with semantic dependency and context moment for aspect level sentiment classification](#). In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pages 5088–5094.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. [Roberta: A robustly optimized bert pretraining approach](#). *arXiv preprint arXiv:1907.11692*.
- Ilya Loshchilov and Frank Hutter. 2019. [Decoupled weight decay regularization](#). In *International Conference on Learning Representations*.
- Navonil Majumder, Soujanya Poria, Alexander Gelbukh, Md. Shad Akhtar, Erik Cambria, and Asif Ekbal. 2018. [IARM: Inter-aspect relation modeling with memory networks in aspect-based sentiment analysis](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3402–3411.
- 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.
- 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.
- Mike Schuster and Kuldip K Paliwal. 1997. [Bidirectional recurrent neural networks](#). *IEEE Transactions on Signal Processing*, 45(11):2673–2681.
- Parinaz Sobhani, Diana Inkpen, and Xiaodan Zhu. 2017. [A dataset for multi-target stance detection](#). In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 551–557.
- Parinaz Sobhani, Saif Mohammad, and Svetlana Kiritchenko. 2016. [Detecting stance in tweets and analyzing its interaction with sentiment](#). In *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics*, pages 159–169.
- Youwei Song, Jiahai Wang, Tao Jiang, Zhiyue Liu, and Yanghui Rao. 2019. [Attentional encoder network for targeted sentiment classification](#). *arXiv preprint arXiv:1902.09314*.
- 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.
- Prashanth Vijayaraghavan, Ivan Sysoev, Soroush Vosoughi, and Deb Roy. 2016. [DeepStance at SemEval-2016 task 6: Detecting stance in tweets using character and word-level CNNs](#). In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 413–419.
- Penghui Wei, Junjie Lin, and Wenji Mao. 2018a. [Multi-target stance detection via a dynamic memory-augmented network](#). In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR 2018, Ann Arbor, MI, USA, July 08-12, 2018*, pages 1229–1232.
- Penghui Wei, Wenji Mao, and Daniel Zeng. 2018b. [A target-guided neural memory model for stance detection in twitter](#). In *2018 International Joint Conference on Neural Networks, IJCNN 2018, Rio de Janeiro, Brazil, July 8-13, 2018*, pages 1–8.

Wan Wei, Xiao Zhang, Xuqin Liu, Wei Chen, and Tengjiao Wang. 2016. [pkudblab at SemEval-2016 task 6: A specific convolutional neural network system for effective stance detection](#). In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 384–388.

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.

Guido Zarrella and Amy Marsh. 2016. [MITRE at SemEval-2016 task 6: Transfer learning for stance detection](#). In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 458–463.