

Multi-Task Stance Detection with Sentiment and Stance Lexicons

Yingjie Li

Computer Science
University of Illinois at Chicago
Chicago, IL, USA
yli300@uic.edu

Cornelia Caragea

Computer Science
University of Illinois at Chicago
Chicago, IL, USA
cornelia@uic.edu

Abstract

Stance detection aims to detect whether the opinion holder is in support of or against a given target. Recent works show improvements in stance detection by using either the attention mechanism or sentiment information. In this paper, we propose a multi-task framework that incorporates target-specific attention mechanism and at the same time takes sentiment classification as an auxiliary task. Moreover, we used a sentiment lexicon and constructed a stance lexicon to provide guidance for the attention layer. Experimental results show that the proposed model significantly outperforms state-of-the-art deep learning methods on the SemEval-2016 dataset.

1 Introduction

With the rapid growth of social media, user opinions towards various targets, e.g., politicians and religion, are abundant. These opinions can help optimize management systems and can gain insight into important events, e.g., presidential elections. The stance detection task aims to determine whether people are in favor of, against, or neutral towards a specific target. This task is similar to the three-way aspect-level sentiment analysis that determines sentiment polarity towards aspect terms. However, different from aspect-level sentiment analysis, the target in stance detection might not be explicitly mentioned in a given sentence. Consider the following example tweet: *@realDonaldTrump is the only honest voice of the @GOP and that should scare the shit out of everyone! #SemST*. Target: *Hillary Clinton*; Stance: *Against*; Sentiment: *Positive*. Observe that even though target Hillary Clinton does not appear in this tweet, we can still infer that opinion holder is less likely to be in favor of Hillary Clinton. Therefore, identifying target related information is of vital importance to stance detection.

Previous studies to stance detection used feature engineering (Mohammad et al., 2016b), Convolutional Neural Networks (CNNs) (Vijayaraghavan et al., 2016; Wei et al., 2016) and Recurrent Neural Networks (RNNs) (Zarrella and Marsh, 2016). However, they failed to take target information into considerations. In order to address this issue, several target-specific attention mechanisms (Du et al., 2017; Zhou et al., 2017; Sun et al., 2018) have been proposed to embed target information into sentence representations.

Sentiment information has also been proven to be beneficial to stance detection. For example, Sobhani et al. (2016) found that sentiment lexicon features are useful for stance detection when combined with other features. Instead of directly integrating sentiment features into vector representations, Sun et al. (2018) proposed a hierarchical attention network that learns the importance of sentiment information automatically. Later, Sun et al. (2019) proposed a joint model that determines stance and sentiment simultaneously. However, the attention mechanism is not considered in this model.

Motivated by recent advances in multi-task learning (Liu et al., 2016, 2017; Balikas et al., 2017; Yu et al., 2018; Cohan et al., 2019) and the effectiveness of sentiment information in stance detection (Sobhani et al., 2016; Sun et al., 2018, 2019), we propose an attention-based multi-task framework that takes sentiment classification as an auxiliary task for stance detection. Specifically, we first encode words of input sentences in fixed-length dense vectors (Mikolov et al., 2013; Bojanowski et al., 2017) and then feed them as input to Bidirectional Long Short-Term Memory networks (BiLSTM) (Hochreiter and Schmidhuber, 1997; Schuster and Paliwal, 1997). After that, the attention mechanism (Bahdanau et al., 2015) is used to extract important words for sentiment

representation. Besides, a target-specific attention layer is designed to identify important words related to a given target. The main task incorporates sentiment information by concatenating the outputs of the two attention layers. Moreover, in order to provide guidance to the attention mechanism, sentiment and stance lexicon features are integrated into the final loss. We show that the proposed model, called AT-JSS-Lex, achieves competitive results on the SemEval-2016 dataset (Mohammad et al., 2016a,b, 2017).

Our contributions are summarized as follows: First, we propose a joint sentiment and stance model (AT-JSS-Lex) based on multi-task learning that improves stance detection with the help of sentiment information and integrates both sentiment attention and target-specific attention. Second, we propose a novel formulation of the loss function that uses an existing sentiment lexicon and a stance lexicon, that we specifically constructed for this task, to guide the attention mechanism. As part of our contributions, we will make the stance lexicon publicly available. Finally, we show that the proposed AT-JSS-Lex model achieves remarkable improvements in performance over strong baselines and prior works on the SemEval-2016 stance detection dataset.

2 Model

2.1 Multi-Task Learning Architecture

The overall architecture of the proposed model is shown in Figure 1. For sentiment classification (the auxiliary task), the input sentence ($S = s_1, s_2, \dots, s_n$) is first sent to an embedding layer and each word is represented by a dense vector ($X = x_1^{d_1}, x_2^{d_1}, \dots, x_n^{d_1}$), where n is the sentence length and d_1 is the dimension of the word embedding. Then, a BiLSTM is used for feature extraction. At time step i , the hidden vector of forward LSTM is calculated based on previous hidden vector h_{i-1} and current input vector x_i , $\vec{h}_i = LSTM(x_i, \vec{h}_{i-1})$. The hidden vector \overleftarrow{h}_i of backward LSTM is defined in a similar way. After concatenating the two hidden vectors, we obtain $h_i = [\vec{h}_i; \overleftarrow{h}_i]$.

We adopt the attention mechanism (Bahdanau et al., 2015) to improve the three-class sentiment classification. Attention weight α_i^1 is computed as:

$$\alpha_i^1 = \frac{\exp(e_i)}{\sum_{j=1}^n \exp(e_j)} \quad (1)$$

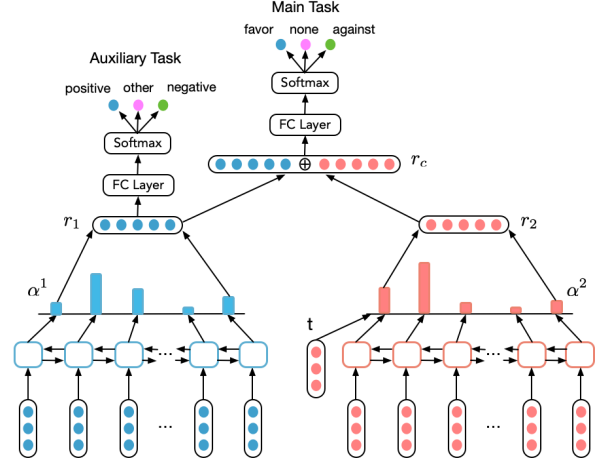


Figure 1: Our multi-task model for stance detection.

where e_i is calculated based on sentence summary vector s , the final hidden vector of BiLSTM, and hidden vector h_i :

$$e_i = v_s^T \tanh(W_s s + W_i h_i + b_s) \quad (2)$$

where $W_s, W_i \in \mathbb{R}^{2d_2 \times 2d_2}$ and $v_s, b_s \in \mathbb{R}^{2d_2}$. d_2 is the dimension of hidden units of LSTM. The final vector representation is the weighted sum of hidden vectors:

$$r_1 = \sum_{i=1}^n \alpha_i^1 h_i \quad (3)$$

At last, a fully-connected layer and a softmax layer are applied to get label distribution.

Different from the attention layer of auxiliary task, the attention layer of main task integrates target embedding t , similar to (Zhou et al., 2017). The target embedding t is the word embedding of the target word. For example, for target ‘‘Abortion,’’ t is the word vector of the word ‘‘Abortion.’’ For multi-word targets (e.g., Hillary Clinton), we use the average of the constituent word vectors (e.g., the average of the vectors corresponding to ‘‘Hillary’’ and ‘‘Clinton’’). Then e'_i can be written as:

$$e'_i = v_t^T \tanh(W_t t + W'_i h_i + b_t) \quad (4)$$

where $W_t \in \mathbb{R}^{2d_2 \times d_1}$, $W'_i \in \mathbb{R}^{2d_2 \times 2d_2}$, $t \in \mathbb{R}^{d_1}$, v_t and $b_t \in \mathbb{R}^{2d_2}$. Then, α_i^2 and r_2 are computed in the same way as Eq. (1) and Eq. (3), respectively. The final vector representation of main task is the concatenation of r_1 and r_2 .

In the training stage, cross-entropy loss is used to train the model and the loss function L is defined as follows:

$$L = \lambda L_{main} + (1 - \lambda) L_{aux} \quad (5)$$

Target	#Train	%Favor	%Against	%None	#Test	%Favor	%Against	%None
Atheism	513	17.9	59.3	22.8	220	14.5	72.7	12.7
Climate	395	53.7	3.8	42.5	169	72.8	6.5	20.7
Feminism	664	31.6	49.4	19.0	285	20.4	64.2	15.4
Hillary	689	17.1	57.0	25.8	295	15.3	58.3	26.4
Abortion	653	18.5	54.4	27.1	280	16.4	67.5	16.1
Total	2914	25.8	47.9	26.3	1249	23.1	51.8	25.1

Table 1: Data distribution of SemEval-2016 dataset.

where λ is a hyper-parameter to be tuned, determining the weight of stance detection task. L_{main} and L_{aux} are loss functions of main task and auxiliary task, respectively.

2.2 Lexicon Loss

Previous works showed that the attention mechanism is beneficial for stance detection. However, the attention mechanism does not always work well due to the size of training data and inability to identify target information. In order to address these issues, we propose a reformulation of the loss in Eq. (5) that employs both sentiment and stance lexicons to improve stance detection. Our full model is called AT-JSS-Lex.

We use the sentiment lexicon¹ by Hu and Liu (2004) and construct a stance lexicon² (of almost 2,000 words) with words related to the five targets in the SemEval-2016 dataset. Specifically, we construct a stance lexicon for each target from the training data available from the SemEval 2016 dataset and from an extra 1,000 tweets for each target that we collected from Twitter using specific hashtags. For example, for the target “Hillary Clinton,” hashtags such as “#Crooked-Hillary,” “#hillaryforprison,” and “#Hillary2016” are used to collect more tweets. After data collection, we manually extracted the directly related words (e.g., Killary, Shillary, and Hilly) and indirectly related words (e.g., Trump, Benghazi, and abortion) from each tweet. We ended up with around 400 lexicon words for each target.

The intuition of using sentiment and stance lexicons is to provide guidance to the attention layer. Specifically, given an input sentence, we mark sentiment and stance lexicon words as 1 and mark as 0 the remaining words that are not present in any of the two lexicons in order to obtain a lexicon vector lex . For example, the lexicon vector for “Celebrity atheism is beginning to irk me #init-

forthemoney #SemST” is [0 1 0 0 0 1 0 0 0]. Note that the words “atheism” and “irk” are from the stance and sentiment lexicons, respectively. The final loss from Eq. (5) is then defined as:

$$L = \lambda L_{main} + (1 - \lambda) L_{aux} + \beta (\alpha_{norm} - lex)^2 \quad (6)$$

where β is a hyperparameter that determines the importance of lexicon loss and α_{norm} is the normalization of α_{sum} . α_{sum} is the summation of attention weights α^1 and α^2 . We normalize the α_{sum} by dividing with the maximum value. Through the minimization of the loss function, ideally, α_{norm} gets closer to lex , i.e., the vector components of α_{norm} get closer to 1 when the corresponding components in lex are 1 (and closer to 0 otherwise), which enforces the model to learn higher attention weights for important words.

3 Experiments and Analysis

3.1 Datasets and Experimental Settings

We use SemEval-2016 Task 6.A to test the performance of our proposed model. This dataset contains five different targets: “Atheism,” “Climate Change is a Real Concern” (“Climate”), “Feminist Movement” (“Feminism”), “Hillary Clinton” (“Hillary”) and “Legalization of Abortion” (“Abortion”). Table 1 shows the distribution of these targets in the dataset. Each tweet has a stance label (“Favor,” “Against” or “None”) and a sentiment label (“Positive,” “Negative” or “Other”).

We sampled about 15% of the training data as validation data to tune the parameters. Word vectors are initialized using fastText word embeddings³ (Bojanowski et al., 2017) with dimension 300. The size of hidden units in LSTM is set to 200. The dropout rate is 0.5 for dense layer and 0.2 for LSTM layer. The learning rate of Adam optimizer (Kingma and Ba, 2015) is 0.001. The size of the dense layer for the auxiliary task and

¹<https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon>

²<https://github.com/chuchun8/EMNLP19-Stance>

³<https://github.com/facebookresearch/fastText/blob/master/docs/crawl-vectors.md>

Model	Atheism	Climate	Feminism	Hillary	Abortion	$MacF_{avg}$	$MicF_{avg}$
	F_{avg}	F_{avg}	F_{avg}	F_{avg}	F_{avg}		
AT-JSS-Lex	69.22	59.18	61.49	68.33	68.41	65.33	72.33
AT-JSS	70.81	57.71	60.14	67.06	67.16	64.58	72.06
JSS	66.87	54.28	58.35	64.74	62.48	61.34	68.96
AT-BiLSTM	66.31	52.56	59.50	64.51	66.10	61.80	71.86
BiLSTM	64.88	43.68	57.93	58.81	60.86	57.23	68.44
SVM-ngram	65.19	42.35	57.46	58.63	66.42	58.01	68.98
JOINT	66.78	50.60	59.35	62.47	61.58	60.16	69.22
TAN	59.33	53.59	55.77	65.38	63.72	59.56	68.79
AS-biGRU-CNN	66.76	43.40	58.83	57.12	65.45	58.31	69.42
HAN	70.53	49.56	57.50	61.23	66.16	61.00	69.79
TGMN-CR	64.60	43.02	59.35	66.21	66.21	59.88	71.04

Table 2: Performance comparison of stance detection on the SemEval-2016 dataset. The numbers are percentages.

main task is 50 and 100, respectively. λ is 0.7 and β is 0.025 for all targets. L2 regularization is applied to the loss function and the regularization parameter is set to 0.01.

3.2 Evaluation Metrics

F_{avg} , macro-average of F1 score ($MacF_{avg}$) and micro-average of F1 score ($MicF_{avg}$) are adopted to evaluate the performance of proposed model. Firstly, the F1 score of label ‘‘Favor’’ and ‘‘Against’’ is calculated as follows:

$$F_{favor} = \frac{2P_{favor}R_{favor}}{P_{favor} + R_{favor}} \quad (7)$$

$$F_{against} = \frac{2P_{against}R_{against}}{P_{against} + R_{against}} \quad (8)$$

where P and R are precision and recall respectively. Then the F1 average is calculated as:

$$F_{avg} = \frac{F_{favor} + F_{against}}{2} \quad (9)$$

Note that the label ‘‘None’’ is not discarded during training. However, the label ‘‘None’’ is not considered in the evaluation because we are only interested in labels ‘‘Favor’’ and ‘‘Against’’ in this task.

We average the F_{avg} on each target to get $MacF_{avg}$. Moreover, we get $MicF_{avg}$ by calculating F_{favor} and $F_{against}$ across all targets.

3.3 Results

First, an ablation experiment is used to determine the importance of each component of our proposed model for stance detection:

- AT-JSS-Lex is a lexicon integrated multi-task model with attention mechanisms.
- AT-JSS is a model that shares the same architecture with AT-JSS-Lex, but has no lexicon loss.

- JSS is a joint sentiment and stance model similar to AT-JSS, but without the attention mechanisms.
- AT-BiLSTM is a BiLSTM model with stance attention.
- BiLSTM is a single task model that only exploits BiLSTM to detect stance.

Table 2 (top) shows the results of this ablation study. As we can see from the table, AT-JSS-Lex performs best on ‘‘Climate,’’ ‘‘Feminism,’’ ‘‘Hillary,’’ and ‘‘Abortion.’’ Moreover, AT-JSS-Lex has the best $MacF_{avg}$ and $MicF_{avg}$ scores when compared with the other models. Experimental results show that $MacF_{avg}$ and $MicF_{avg}$ drop by 0.75% and 0.27% when we remove the lexicon component, indicating that lexicon information contributes to stance detection (except on ‘‘Atheism’’). Note that AT-JSS-Lex also outperforms AT-BiLSTM model, which shows that the proposed multi-task framework has better performance than single task with attention mechanism.

Second, we compare the proposed model with the following baseline methods (all experimental results of baseline methods are retrieved from original papers):

- SVM-ngram (Mohammad et al., 2016b) is trained by using word n-grams and character n-grams features, surpassing the best model in SemEval-2016 competition.
- JOINT (Sun et al., 2019) is a joint model that exploits sentiment information to improve stance detection task without attention mechanism.
- TAN (Du et al., 2017) is an attention-based LSTM model that extracts important part of given text.

Target	Sentences and Attention Weights														
Climate	I can literally sense the global warming NOW														
	0.00	0.03	0.04	0.05	0.02	0.34	0.52	0.00							
Atheism	Holy Mary Mother of God, pray for us sinners now ... Amen. #rosary #teamjesus #God #SemST														
	0.07	0.19	0.11	0.01	0.21	0.05	0.00	0.00	0.23	0.00	...	0.05	0.00	0.00	0.07

Figure 2: Visualization of learned attention weights.

- AS-biGRU-CNN (Zhou et al., 2017) is another attention-based model that adds CNN layer after attention-based LSTM model to extract target specific features.
- HAN (Sun et al., 2018) is a hierarchical attention model leveraging various linguistic features.
- TGMN-CR (Wei et al., 2018) uses attention and memory modules to extract important information for detecting stance.

Table 2 (bottom) shows the results of this comparison as well. We can observe that AT-JSS-Lex model outperforms all baseline models except on “Atheism.” Specifically, the proposed model outperforms JOINT model by 5.17% and 3.11% in $MacF_{avg}$ and $MicF_{avg}$, demonstrating the effectiveness of attention mechanism and lexicon information. In addition, our AT-JSS-Lex model also performs better than attention-based models (TAN, HAN, AS-biGRU-CNN and TGMN-CR), showing that multi-task learning can benefit the stance detection task.

3.4 Attention Visualizations

In Figure 2, we list two input sentences and visualize the attention weights learned during the training process. The color indicates the importance of a word in given sentences, the darker the more important. We can observe that the words “global” and “warming” that are closely related to the target “Climate,” have high attention weights in the first sentence. Likewise, in the second sentence, the words “Mary,” “Mother,” “God,” and “sinners,” are highlighted and show that the proposed model can pay attention to target-related words.

3.5 Error Analysis

Mislabeled data and compound hashtags are two challenging factors that increase the difficulty of

further improving the stance classification. For example, consider the following tweet: *Watch @DameLillard bring the Blazers to the playoff #Beast #SemST*. Target: *Atheism*; Stance: *Against*; Sentiment: *Positive*. Even though the stance label of this tweet is “Against,” however, we can observe that the content of this sentence is nothing related to the target.

For SemEval 2016 dataset, sometimes it is hard to predict the stance without considering the compound hashtags. Here is an example: *@Reince This is very credible! Good work! America is desperately in need of good leadership. #VoteGOP #NoHillary #SemST*. Target: *Hillary Clinton*; Stance: *Against*; Sentiment: *Positive*. It would be very difficult to infer the correct stance label if we do not consider the hashtags “#NoHillary” and “#VoteGOP.” Therefore, inability to separate compound hashtags results in the loss of important target information.

4 Conclusion and Future Work

In this paper, we propose an attention-based multi-task learning framework and integrate lexicon information to achieve better performance. Experimental results show that our model outperforms state-of-the-art deep learning methods for this task. Moreover, visualization results indicate the capability of our model to capture essential information. Future work includes exploiting unsupervised learning to generate target-related lexicon and incorporating more labels (e.g., emotion classification) for multi-task learning.

Acknowledgments

We thank our reviewers for their valuable comments and feedback that helped improve the quality of our paper. This research is supported by NSF awards 1541155, 1741353 and 1526542 to Cornelia Caragea. The computation for this project was performed on Amazon Web Services.

References

- 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*.
- Georgios Balikas, Simon Moura, and Massih-Reza Amini. 2017. Multitask learning for fine-grained twitter sentiment analysis. In *Proceedings of the 40th international ACM SIGIR conference on research and development in information retrieval*, pages 1005–1008.
- Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.
- Arman Cohan, Waleed Ammar, Madeleine van Zuylen, and Field Cady. 2019. Structural scaffolds for citation intent classification in scientific publications. *arXiv preprint arXiv:1904.01608*.
- 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.
- Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8):1735–1780.
- 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*, pages 168–177.
- Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations*.
- Pengfei Liu, Xipeng Qiu, and Xuanjing Huang. 2016. Recurrent neural network for text classification with multi-task learning. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, pages 2873–2879.
- Pengfei Liu, Xipeng Qiu, and Xuanjing Huang. 2017. Adversarial multi-task learning for text classification. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1–10.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.
- Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiao-Dan Zhu, and Colin Cherry. 2016a. A dataset for detecting stance in tweets. In *LREC*.
- Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016b. Semeval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 31–41.
- Saif M Mohammad, Parinaz Sobhani, and Svetlana Kiritchenko. 2017. Stance and sentiment in tweets. *ACM Transactions on Internet Technology (TOIT)*, 17(3):26.
- Mike Schuster and Kuldip K Paliwal. 1997. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11):2673–2681.
- 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.
- Qingying Sun, Zhongqing Wang, Shoushan Li, Qiaoming Zhu, and Guodong Zhou. 2019. Stance detection via sentiment information and neural network model. *Frontiers of Computer Science*, 13(1):127–138.
- 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, Wenji Mao, and Daniel Zeng. 2018. A target-guided neural memory model for stance detection in twitter. In *2018 International Joint Conference on Neural Networks (IJCNN)*, 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.
- Jianfei Yu, Luis Marujo, Jing Jiang, Pradeep Karuri, and William Brendel. 2018. Improving multi-label emotion classification via sentiment classification with dual attention transfer network. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1097–1102.
- 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.

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.