Siamese Neural Network for Same Side Stance Classification

Milad Alshomary
Computational Social Science Group
Department of Computer Science
Paderborn University
milad.alshomary@upb.de

Henning Wachsmuth
Computational Social Science Group
Department of Computer Science
Paderborn University
henningw@upb.de

Abstract

Classifying the stance of an argument towards its target is an important step in many applications of computational argumentation. A simpler variant of stance classification was proposed as a shared task recently, called *same-side stance classification*: Given two arguments on the same topic, decided whether they have the same stance. In this paper, we present our approach to the shared task, exploring the potential of modeling same-side stance as a similarity learning task. For this purpose, we train a siamese neural network on pairs of arguments represented in an embedding space. In the two scenarios of the shared task, *within topics* and *cross topics*, our approach achieved an accuracy of 0.53 and 0.56 respectively.

1 Introduction

In computational argumentation, stance classification is the task of identifying the position of a claim or a whole argument (usually either *pro* or *con*) towards some target, such as a controversial topic or another claim. Identifying the stance of a natural language argument is a major step in argument search engines (Wachsmuth et al., 2017), debating technologies (Bar-Haim et al., 2017), and many other downstream applications.

The *same-side stance classification* task,[1] a simplified variant of stance classification, was proposed as a shared task in the context of the RATIO research program on argumentation,[2] and its results were presented at the 6th Workshop on Argument Mining.[3] The task is defined as:

> *Given two arguments on the same topic, classify whether the arguments have the same stance towards the topic or not.*

As suggested by the organizers, solving this task does not require knowledge about the topic of the argument, but focuses more on modeling features of the argument pairs actually capturing stance, thus making the task potentially easier. Still, knowing whether two arguments are "on the same side" helps in many downstream tasks, e.g., for structuring discussions, for measuring the bias in a debate, and for propagating a (known) stance of an argument to other arguments.

To approach same-side stance classification, a dataset was provided in the shared task, where each instance consists of two textual natural language arguments from a debate portals, along with a text covering the topic they address. Two experimental set-ups were introduced: (1) *within topics* and (2) *cross topics*. In the former, the training set and the test set contain the same topics. In the latter, the test topis are disjunct from the training topics.

In this paper, we investigate the hypothesis that arguments with the same stance are more lexically and/or semantically similar than those with different stance. In particular, we explore the potential of similarity-learning approaches in addressing the given task. To this end, we first represent each argument in an embedding space derived from the words they span. Then, we learn to map the arguments to a new space where similar arguments (having the same stance) are closer to each other, and other arguments are further away. Concretely, we represent each argument by document embedding computed by the Flair library (Akbik et al., 2018), which is the average of the contextual string embedding of each word in the argument. To learn same-stance similarity, we then employ a Siamese Neural network (Bromley et al., 1994) that is trained to minimize the distance between positive pairs and to maximize it for the negative ones.

For both experimental set-ups, we evaluate our approach by first tuning its parameters on the val-

[1] Same-side task, https://sameside.webis.de
[2] RATIO, http://www.spp-ratio.de
[3] ArgMining, https://argmining19.webis.de

idation set and then evaluating it on the test set. *Within topics*, our approach achieved an accuracy of 53% in the shared task, whereas it classified 56% of the *cross topics* test cases correctly. While these values are rather in the middle of the task leaderboard, our analysis provides insights into the adequacy of siamese neural networks for the task. It seems likely that using the provided topic information and/or experimenting with different embedding techniques would boost their effectiveness.

2 Related Work

The detection of stance as *pro* or *con* (and possibly *none*, *neutral*, or similar) is a crucial step in many technologies related to computational argumentation. Much research has been dedicated to this task (Stede and Schneider, 2018). Among this, Bar-Haim et al. (2017) tackled the classification of the stance of a claim towards a topic, and Persing and Ng (2016) constructed a dataset to study stance detection on student essays. Also, Krejzl and Steinberger (2016) proposed a SemEval task where, given a tweet and a target phrase, the goal was to identify the stance of the tweet towards this target. Unlike these works, the paper at hand focuses on the *same-side* stance classification task where the actual stance does not matter, but only whether two texts have the same stance.

Basically, we seeks to learn a similarity function that reflects the likelihood of two arguments having the same stance. For this we first represent these arguments in a semantic embedding space. A big body of research has investigated different ways for learning word embeddings, including (Bojanowski et al., 2017; Peters et al., 2018; Akbik et al., 2018). Representing sentences and larger units of text in an embedding space is more complicated task. Although there has been many approaches proposed for this task, such as (Kiros et al., 2015; Arora et al., 2017), simply taking the average embedding of the sentence's words has been proven to be a strong baseline (Conneau et al., 2018).

Given argument embeddings, we transfer them to a new embedding space where arguments with same stance are similar. To this end, we employ a siamese neural network, which have been first introduced by Bromley et al. (1994) to approach the task of signature verification. Later, its architecture has been utilized for metric learning in tasks such as face verification (Chopra et al., 2005), visual pattern recognition (Hu et al., 2014), and many others. In natural language processing, siamese neural networks have been used, e.g., for learning sentence similarity (Mueller and Thyagarajan, 2016) and text categorization (Shih et al., 2017).

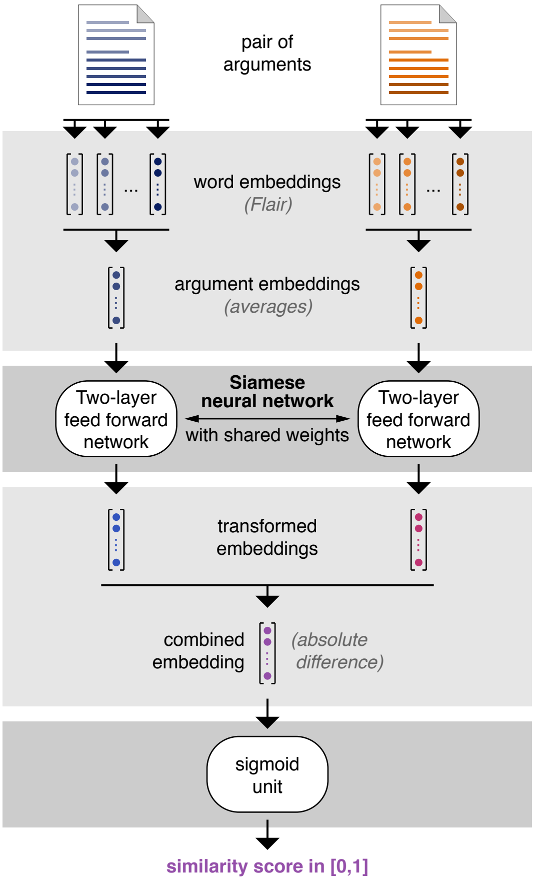

Figure 1: Our approach to classify same-side stance: First, we represent the two given arguments as embeddings. A siamese neural network then transforms these embeddings. Finally, we infer a similarity score from the difference between the transformed embeddings.

3 Approach

We hypothesized that arguments having the same stance towards a given topic are usually more similar semantically than arguments with opposite stance. To model this similarity, we represent each argument in an embedding space and then learn a similarity function that reflects the likelihood of having the same stance. Figure 1 gives an overview of our approach, detailed in the following.

Concretely, we map each argument to an embedding using the contextual string embedding model proposed by Akbik et al. (2018). The model utilizes a character-level LSTM (Graves, 2013) which is trained to predict the next character given a sequence of previous characters. The LSTM thus generates for each character x_t in a given string a predictive distribution $P(x_t|x_0, \ldots, x_{t-1})$, encoded in the hidden state h_t of the LSTM. Building on this, Akbik et al. (2018) trained a bi-directional LSTM model, which consists of two LSTMs that process the string in forward (left-to-right) and in a backward (right-to-left) manner. Thereby, each character gets two hidden state representations, h_t^f and h_t^b. Then, the embedding of a word thats spans over $x_2 \ldots x_k$ is constructed by concatenating the forward hidden state h_{k+1}^f after the last character x_k and the backward hidden state h_1^b before the first, x_2. The embedding of the whole argument text is obtained by averaging the embeddings of all its words.

Afterwards, we utilize a siamese neural network to learn a similarity function over the encoded arguments. The input of the neural network is pairs of arguments encoded as vectors in the embedding space and a label y indicating whether the two arguments have same stance or not. The encoded arguments are then passed through two feed-forward neural networks that share their weights. An absolute difference is computed from the two output representations that is finally passed through one layer with a single output and a sigmoid activation. As a loss function L, we use binary cross entropy to minimize the difference between predicted scores ($\hat{y}$) and the true labels (y):

$$L \;=\; -y \cdot \log(\hat{y}) \;+\; (1-y) \cdot \log(1-\hat{y})$$

The idea behind is to make arguments with the same stance as similar as possible and those with opposite stance as dissimilar as possible.

4 Experiments

This section describes experiments with our approach within the shared task as well as their results.

Implementation As mentioned above, we use the contextual string embeddings of Akbik et al. (2018). Specifically, we resort to the pretrained model provided in the Flair library, which is trained over news articles.[4] We represent each argument as a vector of 4096 dimensions. The siamese neural network we employed is implemented as a two feed-forward neural networks of two layers with ReLU as an activation function. The two layers share weights, resulting in an output vector of 128 dimensions. For the shared task, both models were trained on batches of size 16 using Adam optimizer (Kingma and Ba, 2014).

	Within Topics			Cross Topics		
System	**Prec.**	**Rec.**	**Acc.**	**Prec.**	**Rec.**	**Acc.**
ReCAP (v1)	**0.85**	0.66	**0.77**	**0.72**	**0.72**	**0.73**
Leipzig University	0.79	0.73	**0.77**	**0.72**	**0.72**	0.72
HHU SSSC	0.70	0.33	0.60	**0.72**	0.53	0.66
DBS	0.53	**1.00**	0.55	0.67	0.53	0.63
UKP	0.68	0.52	0.64	0.64	0.59	0.63
IBM Research	0.69	0.59	0.66	0.62	0.49	0.60
Our approach	0.59	0.19	0.53	0.60	0.38	0.56
ReCAP (v2)	0.65	0.24	0.56	0.70	0.11	0.53
sam	0.51	0.58	0.51	0.51	0.52	0.51
ACQuA (v1)	0.53	0.57	0.54	0.50	0.57	0.50
ACQuA (v2)	0.50	0.50	0.50	0.50	0.57	0.50

Table 1: Precision, recall, and accuracy of our approach and all other systems in both scenarios of the shared task (sorted by cross-topic accuracy). Systems with *v1* and *v2* were submitted with different configurations.

Training 63,903 training argument pairs on two topics (“abortion” and “gay marriage”) were provided by the task organizers. For the *within topics* scenario, we randomly split the provided data into a training set (44732 instances) and a validation set (19171 instances). Then, we chose the model with the best accuracy on the validation set. In the *cross topics* scenario, we randomly sampled 1000 pairs of arguments on “gay marriage” for validation. We trained our model on the provided training set and chose the configuration that performed best on the validation set. In particular, this configuration achieved an accuracy of 0.72 in the *within topics* scenario and 0.54 in the *cross topics* scenario.

Results Table 1 shows the final results of our approach on the held-out test set in the shared task, in comparison to all other participating systems. We achieved an accuracy of 0.53 *within topics*, and 0.56 in the *cross topics* scenario. All top approaches on the leaderboard fine-tuned some variant of BERT (Devlin et al., 2019) on the task. A similar approach to ours is *HHU SSSC*, which also used a siamese neural network but with embeddings generated by BERT. The high effectiveness obtained just by using BERT suggests that also our approach might benefit from integrating it.

Looking at the accuracy drop within topics from validation (0.79) to test (0.53), our approach seems to have overfitted to the specific content of the train-

[4] Flair, github.com/zalandoresearch/flair

ing arguments. Interestingly, its effectiveness on across topics remains stable, indicating that the siamese neural network does learn something general to the task of same-side stance classification.

5 Conclusion

Same-side stance classification is a simplified version of stance classification where the goal is to classify whether two arguments on the same topic have the same stance or not. In this paper, we have presented the approach that we participated with in the first same-side shared task. Our approach was meant to explore the potential of modeling the task as similarity learning using a siamese neural network. The resulting model achieved 0.53 accuracy in the *within topics* test set and 0.56 on the *cross topics* test set, putting it roughly into the middle of the leaderboard. Unlike us, the best systems all utilized BERT embeddings.

A follow-up work could study the integration of siamese neural networks with embeddings such as those from BERT. Besides, so far we refrained from integrating the given topic into our approach for simplicity. Making use of topic information to solve the task may also be worth attempting.

References

Alan Akbik, Duncan Blythe, and Roland Vollgraf. 2018. Contextual string embeddings for sequence labeling. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1638–1649, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Sanjeev Arora, Yingyu Liang, and Tengyu Ma. 2017. A simple but tough-to-beat baseline for sentence embeddings. 5th International Conference on Learning Representations, ICLR 2017 ; Conference date: 24-04-2017 Through 26-04-2017.

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.

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.

Jane Bromley, Isabelle Guyon, Yann LeCun, Eduard Säckinger, and Roopak Shah. 1994. Signature verification using a “siamese” time delay neural network. In *Advances in neural information processing systems*, pages 737–744.

Sumit Chopra, Raia Hadsell, Yann LeCun, et al. 2005. Learning a similarity metric discriminatively, with application to face verification. In *CVPR (1)*, pages 539–546.

Alexis Conneau, Germán Kruszewski, Guillaume Lample, Loïc Barrault, and Marco Baroni. 2018. What you can cram into a single vector: Probing sentence embeddings for linguistic properties. In *ACL*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT*.

Alex Graves. 2013. Generating sequences with recurrent neural networks. *ArXiv*, abs/1308.0850.

Junlin Hu, Jiwen Lu, and Yap-Peng Tan. 2014. Discriminative deep metric learning for face verification in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1875–1882.

Diederik P. Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980.

Ryan Kiros, Yukun Zhu, Russ R Salakhutdinov, Richard Zemel, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. 2015. Skip-thought vectors. In *Advances in neural information processing systems*, pages 3294–3302.

Peter Krejzl and Josef Steinberger. 2016. Uwb at semeval-2016 task 6: stance detection. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 408–412.

Jonas Mueller and Aditya Thyagarajan. 2016. Siamese recurrent architectures for learning sentence similarity. In *Thirtieth AAAI Conference on Artificial Intelligence*.

Isaac Persing and Vincent Ng. 2016. Modeling stance in student essays. In *ACL*.

Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In *Proc. of NAACL*.

Chin-Hong Shih, Bi-Cheng Yan, Shih-Hung Liu, and Berlin Chen. 2017. Investigating siamese lstm networks for text categorization. In *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 641–646. IEEE.

Manfred Stede and Jodi Schneider. 2018. *Argumentation Mining*. Number 40 in Synthesis Lectures on Human Language Technologies. Morgan & Claypool.

Henning Wachsmuth, Martin Potthast, Khalid Al-Khatib, Yamen Ajjour, Jana Puschmann, Jiani Qu, Jonas Dorsch, Viorel Morari, Janek Bevendorff, and Benno Stein. 2017. Building an argument search engine for the web. In *Proceedings of the 4th Workshop on Argument Mining*, pages 49–59, Copenhagen, Denmark. Association for Computational Linguistics.