Target Really Matters: Target-aware Contrastive Learning and Consistency Regularization for Few-shot Stance Detection

Rui Liu[1,2], Zheng Lin[1,2][*] Huishan Ji[1,2], Jiangnan Li[1,2], Peng Fu[1] and Weiping Wang[1]

[1]Institute of Information Engineering, Chinese Academy of Sciences
[2]School of Cyber Security, University of Chinese Academy of Sciences
{liurui1995,linzheng,jihuishan,lijiangnan}@iie.ac.cn
{fupeng,wangweiping}@iie.ac.cn

Abstract

Stance detection aims to identify the attitude from an opinion towards a certain target. Despite the significant progress on this task, it is extremely time-consuming and budget-unfriendly to collect sufficient high-quality labeled data for every new target under fully-supervised learning, whereas unlabeled data can be collected easier. Therefore, this paper is devoted to few-shot stance detection and investigating how to achieve satisfactory results in semi-supervised settings. As a target-oriented task, the core idea of semi-supervised few-shot stance detection is to make better use of target-relevant information from labeled and unlabeled data. Therefore, we develop a novel target-aware semi-supervised framework. Specifically, we propose a target-aware contrastive learning objective to learn more distinguishable representations for different targets. Such an objective can be easily applied with or without unlabeled data. Furthermore, to thoroughly exploit the unlabeled data and facilitate the model to learn target-relevant stance features in the opinion content, we explore a simple but effective target-aware consistency regularization combined with a self-training strategy. Experimental results demonstrate that our approach can achieve state-of-the-art performance on multiple benchmark datasets in the few-shot setting.

1 Introduction

Stance detection is intended to identify the attitude of opinions towards certain targets, where labels can be favor, against, and neutral. For example, the opinion *"True equality allows all to be born."* is against the target *"Legalization of Abortion"*. The settings of stance detection can be generally divided into in-target and cross-target ones. Specifically, in-target stance detection aims to train an exclusive classifier for prediction on the same set of targets (Mohammad et al., 2016b; Augenstein et al., 2016; Li et al., 2021). Obviously, in-target stance detection is the most ideal setting when there is sufficient labeled data. However, in practice, a severe challenge is the scarcity of annotations for new targets. Cross-target stance detection (Zhang et al., 2020a; Xu et al., 2018; Wei and Mao, 2019) is to train on the labeled data of one source target and test on the destination target, which is based on an assumption that there is a strong correlation between the two targets. No doubt the harsh demand from the assumption above limits the extension of cross-target stance detection, in which it still requires a large amount of annotated data for the source target. In this paper, we focus on the in-target few-shot stance detection with unlabeled data and limited labeled data, which is to alleviate the demand for human supervision.

Different from common classification tasks (e.g., sentiment classification), the identification of stance is heavily dependent on the specific target (Siddiqua et al., 2019). As a target-oriented task, the key problem of few-shot stance detection is how to thoroughly exploit the target-relevant information from the limited labeled data and the unlabeled data. Existing methods like supervised contrastive learning (Gunel et al., 2021) and semi-supervised learning (Sohn et al., 2020) prove significant effectiveness in the few-shot setting. When only a few labeled samples are provided, supervised contrastive learning (SCL) hopes to improve the representation ability of the model to a certain extent by leveraging class label information. However, such an objective ignores target label information, the crucial clue, which plays a vital role in stance detection. When the unlabeled data is available, semi-supervised learning (SSL) like self-training and consistency regularization, is a way of bringing unlabeled data into full play, which helps to overcome the scarcity of sufficient annotated data. However, self-training algorithm (Glandt

---

[*]Zheng Lin is the corresponding author.

et al., 2021; Li et al., 2021) only encourages the student network to mimic the teacher network's label predictions simply. Moreover, consistency regularization (Xie et al., 2020a; Sohn et al., 2020) constrains the model to make consistent predictions of the same example under some task-agnostic data augmentation strategies. Neither of them digs deeply into target-relevant stance features in the opinion content.

To further tackle the challenges above, we propose a novel **S**emi-supervised framework with **T**arget-aware **C**ontrastive learning and **C**onsistency regularization (**STCC**). First, we introduce a target-aware contrastive learning objective to consider both target and stance label information, which promotes the distinction and isolation of samples from different targets, as well as different classes. Since the target-aware contrastive learning objective learns more distinguishable representations, the model trained here can be used in semi-supervised learning as a better teacher model. Second, we combine two approaches to SSL: consistency regularization and self-training. Specifically, we design a simple but effective target-aware data augmentation strategy for consistency regularization, i.e., masking the corresponding target for every unlabeled example, so that the model can perceive the target-related information contained in the content as much as possible. Moreover, our proposed target-aware contrastive learning objective can be naturally extended to semi-supervised learning with additional unlabeled data. Specifically, the pseudo-labels generated by the teacher model are used for the contrastive learning objective. Finally, we conduct comprehensive experiments on two large datasets of SemEval-2016 and COVID-19-Stance to verify the effectiveness of our framework for few-shot stance detection. The experimental results demonstrate that our approach achieves the state-of-the-art performance[1].

2 Methodology

2.1 Problem Definition

The semi-supervised few-shot stance detection is to train a classifier by leveraging labeled and unlabeled data, which identifies the users' stance from the context and the corresponding target. Formally, given a collection of limited labeled data $\mathcal{X} = \{(x_i, t_i, y_i)\}_{i=1}^{N_l}$ and a collection of unlabeled

data $\mathcal{U} = \{x_i, t_i\}_{j=1}^{N_u}$, where x_i is the opinion content, t_i is the corresponding target, y_i is the stance label, N_l is the number of the labeled data and N_u is the number of the unlabeled data.

2.2 BERT Model

First of all, we select the pretrained model BERT as the encoder. For the labeled data $\mathcal{X}$, we concatenate the content x_i and the target t_i of each sample in the following format: [CLS] $\boldsymbol{t}_i$ [SEP] $\boldsymbol{x}_i$ [SEP] and utilize BERT to process it. We then feed the representation $\boldsymbol{h}_i^{[CLS]}$ of [CLS] from the last layer of BERT into the final classification layer. Finally, we compute the probability distribution with the softmax function: $p(\hat{y}_i|x_i, t_i) = softmax(W_h h_i^{[CLS]})$, where W_h is a trainable matrix. We fine-tune the model by minimizing the cross-entropy loss:

$$\mathcal{L}_{ce} = -\frac{1}{|\mathcal{X}|} \sum_i^{|\mathcal{X}|} CE(p(\hat{y}_i|x_i, t_i), y_i), \quad (1)$$

where CE denotes the cross entropy loss function.

2.3 Target-aware Contrastive Learning

The core idea of contrastive learning is to bring the representations of positive sample pairs closer, and push the negative sample pairs farther, to learn more distinguishable representations. In supervised contrastive learning, samples under the same label match each other as positive pairs and samples of different labels match as negative pairs. Given the index $I = \{1, \cdots, B\}$ in a batch, the supervised contrastive learning loss is formulated as:

$$\mathcal{L}_{sup}^i = -\sum_{i=1}^{B} \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} \log \frac{e^{sim(\boldsymbol{h}_i, \boldsymbol{h}_j / \tau)}}{\sum_{k \in C(i)} e^{sim(\boldsymbol{h}_i \cdot \boldsymbol{h}_k / \tau)}},$$

(2)

where B is the batch size, $\mathcal{N}_i = \{\boldsymbol{h}_j | i \neq j, y_i = y_j\}$ is the positive examples of $\boldsymbol{h}_i$, $|\mathcal{N}_i|$ is the number of examples labeled as y_i in the same batch, $C(i) \equiv I \backslash \{i\}$, and τ is temperature parameter.

It is our goal to integrate target label information into contrastive training, enabling the model to learn the target-specific information adequately under the setting of few labeled data. Specifically, as shown in Figure 1, target-aware contrastive training (TCL) tries to make intra-target representations being more compact in the feature space and inter-target ones more distinguishable. The target-aware
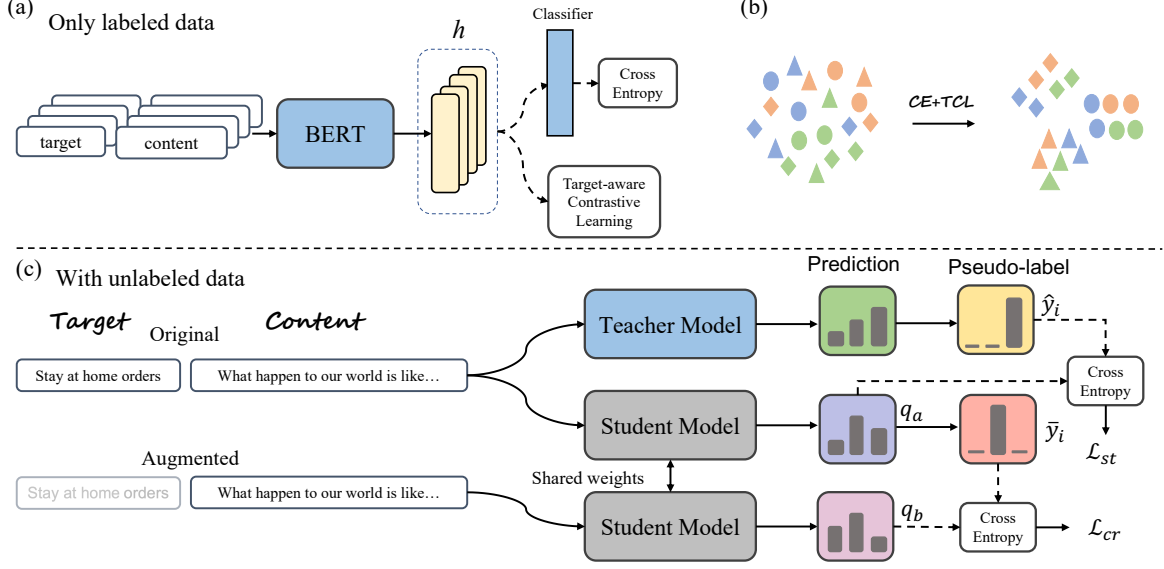
Figure 1: Structure overview of our target-aware semi-supervised framework. There are mainly three parts in the figure: **(a)** The teacher model, aiming at generating pseudo-labels, is trained on labeled data with a combination of cross-entropy (CE) loss and target-aware contrastive learning (TCL) objective. **(b)** an illustration for TCL, in which different shapes stand for different targets, and different colors represent different stance labels. **(c)** is the semi-supervised learning procedure. The teacher model comes from part (a), fed with original data, and its prediction is then softmaxed and regarded as pseudo-labels $\hat{y}_i$. Simultaneously, the original data are also fed into the student model, generating a prediction q_a and pseudo-labels $\bar{y}_i$. Then, pseudo-labels $\hat{y}_i$ from the teacher model and q_a are used to compute the self-training loss $\mathcal{L}_{st}$. Afterwards, in another branch, the student model is fed with augmented data to generate a prediction q_b. Consistency regularization term $\mathcal{L}_{cr}$ is then computed by q_b and $\bar{y}_i$.

contrastive learning loss can be written as:

$$\mathcal{L}_{tcl} = -\sum_{i=1}^{B} \mathcal{L}_{tcl}^{i} \qquad (3)$$

$$\mathcal{L}_{tcl}^{i} = \frac{1}{|\mathcal{T}_i|} \sum_{j \in \mathcal{T}_i} \log \frac{e^{sim(\boldsymbol{h}_i, \boldsymbol{h}_j / \tau)}}{\sum_{k \in C(i)} e^{sim(\boldsymbol{h}_i \cdot \boldsymbol{h}_k / \tau)}}, \quad (4)$$

where $\mathcal{T}_i = \{\boldsymbol{h}_j | i \neq j, y_i = y_j, t_i = t_j\}$ is the positive examples of $\boldsymbol{h}_i$, $|\mathcal{T}_i|$ is the number of examples with the target t_i labeled as y_i in the same batch, and τ is temperature parameter.

The above trained model can be utilized alone or as a teacher model to participate in semi-supervised training. Note that TCL can also be applied to semi-supervised learning. For the labeled data, the annotated labels are used for picking positive and negative pairs. For the unlabeled data, the teacher model generates pseudo-labels as a complement.

2.4 Target-Aware Consistency Regularization in Semi-supervised Learning

Consistency regularization utilizes data augmentation to add perturbations for the unlabeled data. An ideal model is ought to make consistent predictions of samples before and after adding perturbation. Therefore, such property makes it possible for

us to tailor target-aware data augmentation strategy, which can facilitate model mining the target-relevant stance features in the content. Practically, given the prediction distribution of the original data, $q_a = p_s(\bar{y}_i | x_i, t_i)$, the $\bar{y}_i = argmax(q_a)$ is used as pseudo-labels in the later process. There is a simple but effective way of acquiring the augmented version of the unlabeled data for stance detection here. Specifically, the prediction of an augmented sample $q_b = p_s(\tilde{y}_i | x_i)$ from the model can be generated by masking the corresponding target t_i of the content x_i.

$$\mathcal{L}_{cr} = -\frac{1}{|\mathcal{U}|} \sum_{i}^{|\mathcal{U}|} CE(q_b, \bar{y}_i), \qquad (5)$$

Note that conventional consistency regularization performs a supervised training from the labeled data as well, of which the loss is computed by Eq (1). Such a process utilizes the pseudo-labels generated by the model under training. In consistency training, the model trained from scratch has a low accuracy and high entropy, which prevents the model from achieving good accuracy (Xie et al., 2020b). Therefore, we incorporate the self-training strategy into consistency training to

Algorithm 1 Target-aware Semi-supervised Learning for Few-shot Stance Detection

---

Require: Labeled data $\mathcal{X}$ and unlabeled data $\mathcal{U}$.
Require: $\Delta = 20$ ▷ The increment every step
  1: $K = 100$ ▷ The initial threshold
  2: $t = 1$ ▷ The time step
  3: Train the teacher model θ_t using labeled data via Eq (1) and Eq (4).
  4: **repeat**
  5:     $K = K - \Delta$
  6:     Generate pseudo labels $\hat{y}_i$ using the teacher model θ_t for unlabeled data.
  7:     Select unlabeled data with top $K\%$ high-confidence pseudo labels.
  8:     Update student model θ_s using the combination of labeled data and unlabeled data via Eq (7).
  9:     Using the student model θ_s as the teacher model θ_t for next iteration.
 10:     $t = t + 1$
 11: **until** $K \neq 0$
 12: **return** θ_s

---

stabilize the entire training process. Instead of using the model trained from scratch, we train a teacher model ahead to generate high-quality pseudo-labels. Pseudo-labels generated by the teacher model can stabilize the training of a student model that uses these generated pseudo-labels, following the loss function:

$$\mathcal{L}_{st} = -\frac{1}{|\mathcal{U}|} \sum_{i}^{|\mathcal{U}|} CE(q_a, \hat{y}_i), \qquad (6)$$

where, $\hat{y}_i$ is the pseudo labels generated by the teacher model. Inspired by a self-training method of curriculum labeling (Cascante-Bonilla et al., 2021), which applies self-paced curriculum learning principles in each self-training cycle, we select samples of the top $K\%$ highest confidence from the entire unlabeled dataset in each iteration with an increment Δ.

The total loss for the semi-supervised learning is as follows:

$$\mathcal{L} = \mathcal{L}_{ce} + \mathcal{L}_{st} + \lambda_a \mathcal{L}_{cr} + \lambda_b \mathcal{L}_{tcl}, \qquad (7)$$

where λ_a and λ_b are scalar weighting hyper-parameters.

The procedure of our semi-supervised framework is summarized in Algorithm 1.

3 Experiments

3.1 Datasets and evaluation

The macro-averaged F1 is used as the evaluation metric for all datasets. We conducted experiments using two well-known standard benchmarks, SemEval-2016 and COVID-2019-Stance.

SemEval-2016 (Mohammad et al., 2016a) is the earliest dataset to detect users' stance from tweets, which contains 6 targets, specifically, "Atheism", "Climate Change is a Real Concern", "Feminist Movement", "Hillary Clinton", and "Legalization of Abortion" and "Donald Trump". The data with "Donald Trump" in the original dataset is not split into the training and test sets. We split this target's data for training and testing with a ratio of 5:2. The processed dataset has 3414 tweets for training and 1456 for testing. Additionally, we split the original training set in a ratio of 9:1 into training and validation subsets.

COVID-19-Stance (Glandt et al., 2021) consists of 6,133 tweets for stance towards four targets relevant to COVID-19 health mandates, specifically "Anthony S. Fauci, M.D.", "Keeping Schools Closed", "Stay at Home Orders", and "Wearing a Face Mask". This dataset has 4533 samples for training, 800 samples for validation, and 800 samples for testing.

3.2 Model comparisons

We compare our method with several strong baseline methods:

(1) *Some general models trained only using labeled data*:

CrossNet (Xu et al., 2018): It is a BiLSTM model for cross-target stance with an aspect-specific attention layer.

BERT (Devlin et al., 2019): It is a transformer-based language model pre-trained by two self-supervised tasks.

ProtoNets (Snell et al., 2017): It aims to learn prototypes embedding for each class, in which the model makes predictions by computing distances to prototype representations of each class.

(2) *Fine-tuning objectives:*

SCL (Gunel et al., 2021): A contrastive learning objective combining the label information with the self-supervised contrastive learning.

PT-HCL (Liang et al., 2022): This method uses a pre-text task to distinguish the types of the stance

data, then integrates the type information into the supervised contrastive learning.

(3) *Semi-supervised learning methods using the unlabeled data and few labeled data*:

Prompt (Schick and Schütze, 2021): The prompt-based learning is a fine-tuning strategy leveraging language prompts as contexts to stimulate knowledge from pre-trained Language Models, which can has the flexibility to use or not use unlabeled data.

UDA (Xie et al., 2020a): A consistency training algorithm enforcing the model predictions to be consistent between an unlabeled example and its augmented version.

ST (Glandt et al., 2021): A vanilla self-training method transferring the knowledge from the teacher model to the student model iteratively.

UPS (Rizve et al., 2021): An uncertainty-aware pseudo-label selection framework that leverages the prediction uncertainty to guide the pseudo-label selection procedure.

CL (Cascante-Bonilla et al., 2021): A pseudo-label selection framework with a hand-crafted curriculum choice strategy, which selects unlabeled samples progressively from high confidence to low confidence.

Finally, we add the model **BERT w/ full data** trained by the full labeled data to present the upper bound of the performance for all few-shot methods.

3.3 Implementation Details

In the few-shot setting, we randomly select 5, 10, and 20 samples for each target for training under different settings. For example, the size of the labeled data for SemEval-2016 under the 5-shot setting is 30. In order to obtain a relatively uniform distribution, samples of different labels, which are favor, against, and none, are picked by a ratio of 2:2:1. Under the setting of semi-supervised learning, in each of the datasets above, the rest of the data in the training set is used as the unlabeled data. We implement our model using PyTorch[2] and BERT-base from huggingface Transformers[3] is used as the backbone. The models are optimized by AdamW and the batch size is set as 32. An iteration for the self-training procedure is set as 20 epochs. We report the average results of the models using a fixed set of 5 random seeds. We set λ_a to 1.0. As few-shot learning is extremely

[2]https://pytorch.org/
[3]https://huggingface.co/docs/transformers/index

sensitive to hyper-parameters, we conduct a grid hyper-parameters search based on the performance on the validation data for learning rate, temperature τ and scalar weighting λ_b on two different datasets across different shot sizes. The details are in Appendix A.

3.4 Main Results

Table 1 shows the results from baselines on SemEval2016 and COVID-19, including the performance with and without unlabeled data. We select BERT-base as the base encoder of these baselines, except for CrossNet. Table 1 is divided into two parts by whether unlabeled data is available. Models in the upper part are only trained by N-shot labeled data (minimal few-shot setting), while results in the lower part come from semi-supervised learning (semi-supervised few-shot setting).

First of all, generally, our method shows substantial improvement compared to baseline models under minimal few-shot setting. Then, specifically, the comparison of results from TCL and SCL under different datasets validates that the application of target label information in contrastive learning further improves the performance of models. Considering the significant improvement of our model over PT-HCL, it is clear that the distinguishing of whether a data is sensitive to the target or not under the minimal few-shot setting does not make the most of target-specific information.

In addition, we also would like to acknowledge that, SCL performs better than TCL under 5-shot learning on SemEval2016, but no such phenomenon is observed on COVID-19. A possible explanation is that the short length of texts of SemEval2016 impedes the model from learning general representations because TCL has to distinguish the belonged target and stance at the same time, making the number of data in each potential cluster too small to learn. Note that there is no such phenomenon in models with unlabeled data, indicating the application of unlabeled data alleviates the scarcity of data and enables STCC to develop its full potential. Moreover, the model CrossNet based on a traditional BiLSTM performs poorly under the setting of few-shot stance detection here.

Next, from the lower part of Table 1, where the results under semi-supervised learning are shown, we find that STCC outperforms all other semi-supervised methods by a great margin. Considering different settings of numbers of labeled data,

Model	SemEval2016			COVID19		
	5	10	20	5	10	20
without unlabeled data						
CrossNet (Xu et al., 2018)	29.82	33.85	35.37	31.20	34.93	44.32
BERT (Devlin et al., 2019)	41.12	44.45	49.72	32.45	36.85	50.83
ProtoNets(Snell et al., 2017)	41.50	44.13	48.72	33.90	40.50	48.44
SCL (Gunel et al., 2021)	**48.02**	49.40	52.22	37.40	42.23	52.83
PT-HCL (Liang et al., 2022)	34.72	39.56	45.22	-	-	-
Prompt (Schick and Schütze, 2021)	37.88	41.80	43.74	34.96	37.46	47.52
TCL (Ours)	47.32	**51.41**	**53.47**	**40.27**	**46.80**	**53.52**
with unlabeled data						
Prompt (Schick and Schütze, 2021)	37.93	42.41	43.80	34.96	49.42	47.11
UDA (Xie et al., 2020a)	46.86	46.77	50.87	40.27	47.02	53.52
ST (Glandt et al., 2021)	48.35	51.12	55.01	42.08	47.66	55.67
UPS (Rizve et al., 2021)	43.45	48.11	52.73	41.45	44.37	53.87
CL (Cascante-Bonilla et al., 2021)	48.92	51.34	55.42	40.96	50.22	56.86
STCC (Ours)	**52.84**	**55.00**	**57.11**	**44.38**	**52.26**	**58.06**
BERT w/ full data		68.34			73.12	

Table 1: Summary of test results for few-shot stance detection using the shot size of 5, 10, 20 for training. The best results are in bold.

STCC exceeds the best baselines by an average of 3.09% on SemEval 2016 and 1.84% on COVID-19. It is also verified that BERT performs better with the help of unlabeled data from the same target. At last, the increase in numbers of labeled data guarantees a steady growth for all semi-supervised learning methods, especially outstanding for ours. Besides, the performance of prompt-tuning methods whose backbone is $BERT_{base}$ is not ideal. A possible reason is that such methods depend on the generalization ability of large pre-trained models and hand-crafted prompt designs, which are not the focus of our work.

3.5 Ablation Study

As shown in Table 2, we conduct an ablation study to inspect the importance of the components in **STCC** on SemEval2016, including the target-aware contrastive learning (TCL), the target-aware consistency regularization (TCR), and the self-training procedure (ST). It is clear that the removal of either one of our three independent modules causes the drop in performance, especially for the self-training procedure. A possible explanation is that the training of a model from the very beginning introduces consistency training under a low accuracy, forcing the model to stay in a condition of high entropy.

Moreover, compared with "-TCR&TCL" (i.e.,

Model	5-shot	10-shot	20-shot
STCC	52.84	55.00	57.11
-TCR	51.32	53.53	56.48
-TCL	49.91	52.32	56.18
-ST	43.45	47.70	52.65
-ST&TCR	47.32	51.41	53.47
-ST&TCL	45.19	46.29	48.96
-TCR&TCL	48.92	51.34	55.42
-TCR&TCL + SCL	49.32	51.40	55.92
-ST&TCR&TCL	41.12	44.45	49.72

Table 2: Ablation results on SemEval-2016. "ST" means self-training procedure, and "TCR" means target-aware consistency regularization.

only using ST), the performance further improves after equipping TCR, indicating the effectiveness of our proposed target-aware consistency training. The comparison among "-ST&TCL" (i.e., only using TCR), "-TCR&TCL" (i.e., only using ST) and "-TCL" (i.e., using ST&TCR) further validates that the model can achieve an acceptable performance merely with the self-training procedure. In addition, TCL is replaced by SCL (i.e., "-TCR&TCL + SCL") to verify the indispensability of TCL under the semi-supervised framework. The drop in performance compared to TCL (i.e., "-TCR") confirms that target-specific information is of vital significance in semi-supervised learning. The proposed method of TCL adapts well to few-shot stance detection with or without labeled data.

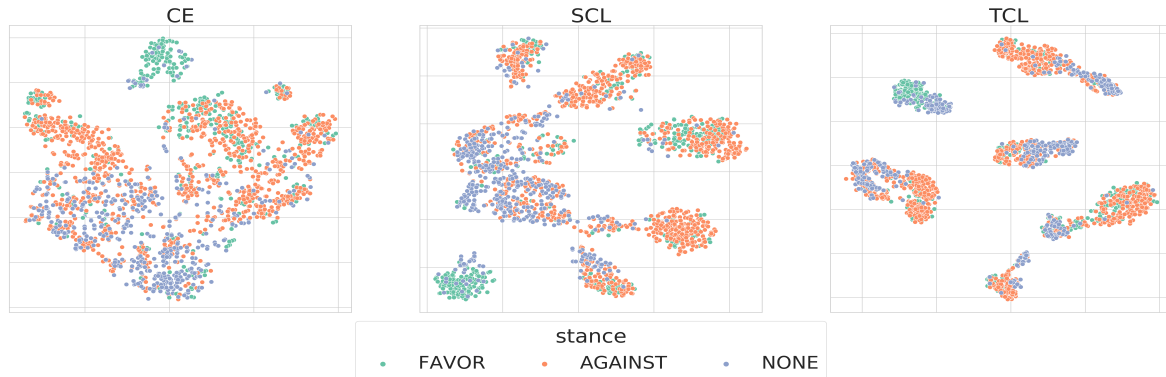To demonstrate the effectiveness of our proposed data augmentation, i.e., "masking the target", we

Figure 2: 2D t-SNE plots of the learned [CLS] representations on the unlabeled data of the SemEval2016 from the BERT models only trained by 20-shot labeled samples for every target, which are fine-tuned on different objectives CE (left), SCL (middle), and TCL (right).

Model	5-shot	10-shot	20-shot
NT	52.84	55.00	57.11
BT	52.13	52.21	55.96
SR	52.26	54.28	55.23
RD	50.92	52.35	55.99

Table 3: The performance of different data augmentation in semi-supervised learning on SemEval2016. "NT": masking the target, "BT": back translation, "SR": synonym replacement, "RD": random deletion.
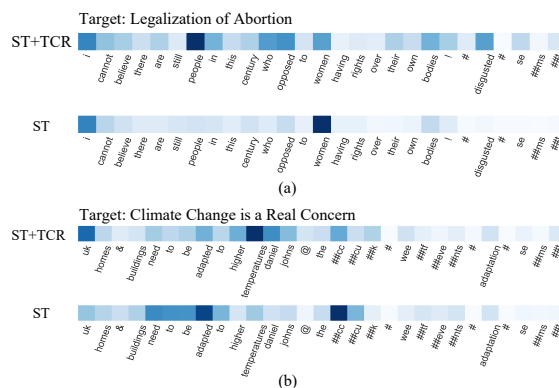

Figure 3: The heatmap of the attention weights of [CLS] towards each subword in the content for "ST+TCR" and "ST" under semi-supervised setting. The attention weights are averaged from the multi-heads in the top layer. The darker the color, the greater the weight.

replace it with multiple common data augmentation methods in NLP, and the results are shown in Table 3. Evidently, our method performs best for the current task. Back translation is implemented by the toolbox of nlpaug[4], while synonym replacement and random deletion are from EDA (Wei and Zou, 2019) method[5]. Note that although common data augmentation methods improve the results as well, none of the improvements is as significant as ours.

3.6 Visualization

Visualization for TCL In Figure 2, the t-SNE plots from the representations of [CLS] are shown. Such representations are the output from BERT-base fine-tuned by different objectives, under the setting of 20 available labeled samples. As observed, for the case of using cross-entropy loss only, all samples are promiscuously scattered. There is a similar but better distribution for the model trained by SCL, where most samples are mixed. For TCL, there are six obviously independent clusters, as there are six targets in SemEval2016. Furthermore,

even inside a cluster itself, representations from the same target but of different labels can be identified and are much more separate than that in the other two diagrams. Such a phenomenon verifies TCL helps to learn better representations, increasing intra-target compactness and inter-target discrepancy, which improves the performance of the model.

Visualization for TCR In Figure 3, we show the heatmap of attention weights to demonstrate the effectiveness of our proposed target-aware consistency regularization. As the classifier deals with the representations from "[CLS]", we pick the attention-weight matrix of "[CLS]" from the top layer towards each subword in the content. In order to avoid possible influence from TCL, the model is trained by "ST" and "ST+TCR" respectively. Take (a) in Figure 3 for an example, whose target is "Legalization of Abortion", the model trained

[4]https://github.com/makcedward/nlpaug
[5]https://github.com/jasonwei20/eda_nlp

by "ST+TCR" notices target-relevant words like "people, woman, opposed and disgusted", while "ST" merely pays attention to "woman". Such a phenomenon confirms that the proposed method of target-aware consistency regularization improves the ability of digging target-relevant information for models.

4 Related Work

Stance Detection Stance detection aims to identify the attitude from an opinion towards a certain target. Incipient studies focus on in-target stance detection (Augenstein et al., 2016; Siddiqua et al., 2019; Mohammad et al., 2016b; Du et al., 2017; Wei et al., 2019), which only train the model and perform the prediction on a single target. Li et al. (2021) investigated the multi-target training and knowledge distillation in the stance detection task. Data augmentation (Li and Caragea, 2021) has been used for in-target setting to improve performance in fully supervised learning. Cross-target stance detection (Zhang et al., 2020b; Wei and Mao, 2019; Allaway et al., 2021) hopes to transfer knowledge between related targets, which attempts to mitigate the lack of labeled training data for a new target. Dutta et al. (2022) focused on semi-supervised user stance detection using the information from tweets posted by users and their followers, whereas we do not consider specific user information. Hardalov et al. (2021) studied few-shot cross-lingual stance detection, which transferred the knowledge from English resources to non-English scenarios. Recently, Allaway and McKeown (2020) defined zero-shot and few-shot stance detection, according to which the targets have no or very few training examples. And, they present a new dataset VAST, which consists of thousands different targets. However, VAST includes a wide range of similar expressions for one target (e.g., "guns on campus" versus "firearms on campus"). The situation above makes the source of the model's benefit too ambiguous to trace. Therefore, SemEval-2016 and COVID-19-Stance are relatively much more accessible for studying in the few-shot setting, compared with VAST.

Contrastive Learning In recent years, contrastive learning has made significant progress in self-supervised representation learning, both in the CV (Chen et al., 2020) and NLP (Gao et al., 2021) domains. Khosla et al. (2020) introduced supervised contrastive learning (SCL), which further ex-tended the self-supervised contrastive learning to the fully-supervised setting by leveraging label information. Gunel et al. (2021) integrated the SCL objective for fine-tuning pre-trained language models, which significantly improves the performance in the few-shot learning settings. In stance detection, Liang et al. (2022) proposed a hierarchical contrastive learning loss to take both the data types and the stance labels into account. They subdivided the data types by judging whether a sample is sensitive to its corresponding target based on the self-supervised learning pretext task. However, this strategy is not suitable for the few-shot setting, which further reduces the target-related information availability to the model.

Semi-supervised Learning Consistency regularization, pseudo-labeling, and self-training are all important components of semi-supervised learning. Consistency regularization (Xie et al., 2020a; Sohn et al., 2020) constrains the model to make consistent predictions of the same example under varied noises. And, pseudo-labeling (Lee, 2013) selects those unlabeled data with high confidence as a form of entropy minimization. These methods use the model being trained to generate pseudo-labels instead of a separate teacher model pre-trained on labeled data. Self-training (Xie et al., 2020b) allows a teacher model pre-trained on labeled data, and then applies the combination of labeled and pseudo-labeled data to retrain a student model. Recent, semi-supervised learning methods have combined those techniques to some extent. Cascante-Bonilla et al. (2021); Rizve et al. (2021) combine the self-training process with pseudo-labeling, while using curriculum labeling and uncertainty-aware techniques to improve the filter process for unlabeled data. Sohn et al. (2020) unify the consistency regularization and pseudo-labeling. However, these task-agnostic methods cannot adequately mine target-relevant stance features in the opinion content for stance detection.

5 Conclusion

In this paper, we focus on in-target few-shot stance detection to alleviate the demand for human supervision, and propose a novel target-aware semi-supervised framework with contrastive learning and consistency regularization. The target-aware contrastive learning objective performs well on both labeled and unlabeled data, which promotes the model's ability to distinguish various classes

and targets. Moreover, our proposed target-aware consistency regularization is validated to be more efficient in mining target-relevant stance features in the content. Experiments on two popular benchmarks demonstrate the effectiveness and consistent improvements over baselines.

Acknowledgments

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 R. 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, NAACL-HLT 2021, Online, June 6-11, 2021*, pages 4756–4767. 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.

Paola Cascante-Bonilla, Fuwen Tan, Yanjun Qi, and Vicente Ordonez. 2021. Curriculum labeling: Revisiting pseudo-labeling for semi-supervised learning. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, pages 6912–6920. AAAI Press.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey E. Hinton. 2020. A simple framework for contrastive learning of visual representations. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pages 1597–1607. PMLR.

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, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics.

Jiachen Du, Ruifeng Xu, Yulan He, and Lin Gui. 2017. Stance classification with target-specific neural attention. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pages 3988–3994.

Subhabrata Dutta, Samiya Caur, Soumen Chakrabarti, and Tanmoy Chakraborty. 2022. Semi-supervised stance detection of tweets via distant network supervision. In *WSDM '22: The Fifteenth ACM International Conference on Web Search and Data Mining, Virtual Event / Tempe, AZ, USA, February 21 - 25, 2022*, pages 241–251. ACM.

Tianyu Gao, Xingcheng Yao, and Danqi Chen. 2021. Simcse: Simple contrastive learning of sentence embeddings. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*, pages 6894–6910. Association for Computational Linguistics.

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)*, Online. Association for Computational Linguistics.

Beliz Gunel, Jingfei Du, Alexis Conneau, and Veselin Stoyanov. 2021. Supervised contrastive learning for pre-trained language model fine-tuning. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net.

Momchil Hardalov, Arnav Arora, Preslav Nakov, and Isabelle Augenstein. 2021. Few-shot cross-lingual stance detection with sentiment-based pre-training. *CoRR*, abs/2109.06050.

Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. 2020. Supervised contrastive learning. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*.

Dong-Hyun Lee. 2013. Pseudo-label : The simple and efficient semi-supervised learning method for deep neural networks. *ICML 2013 Workshop : Challenges in Representation Learning (WREPL)*.

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*, Online. Association for Computational Linguistics.

Yingjie Li, Chenye Zhao, and Cornelia Caragea. 2021. Improving stance detection with multi-dataset learning and knowledge distillation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Bin Liang, Zixiao Chen, Lin Gui, Yulan He, Min Yang, and Ruifeng Xu. 2022. Zero-shot stance detection via contrastive learning. In *WWW '22: The ACM Web Conference 2022, Virtual Event, Lyon, France, April 25 - 29, 2022*, pages 2738–2747. ACM.

Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016a. 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, Svetlana Kiritchenko, Parinaz Sobhani, Xiao-Dan Zhu, and Colin Cherry. 2016b. Semeval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2016, San Diego, CA, USA, June 16-17, 2016*, pages 31–41. The Association for Computer Linguistics.

Mamshad Nayeem Rizve, Kevin Duarte, Yogesh Singh Rawat, and Mubarak Shah. 2021. In defense of pseudo-labeling: An uncertainty-aware pseudo-label selection framework for semi-supervised learning. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net.

Timo Schick and Hinrich Schütze. 2021. Exploiting cloze-questions for few-shot text classification and natural language inference. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, EACL 2021, Online, April 19 - 23, 2021*, pages 255–269. 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, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 1868–1873. Association for Computational Linguistics.

Jake Snell, Kevin Swersky, and Richard S. Zemel. 2017. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 4077–4087.

Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. 2020. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*.

Jason W. Wei and Kai Zou. 2019. EDA: easy data augmentation techniques for boosting performance on text classification tasks. 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 2019, Hong Kong, China, November 3-7, 2019*, pages 6381–6387. Association for Computational Linguistics.

Penghui Wei and Wenji Mao. 2019. Modeling transferable topics for cross-target stance detection. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2019, Paris, France, July 21-25, 2019*, pages 1173–1176. ACM.

Penghui Wei, Wenji Mao, and Guandan Chen. 2019. A topic-aware reinforced model for weakly supervised stance detection. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pages 7249–7256. AAAI Press.

Qizhe Xie, Zihang Dai, Eduard H. Hovy, Thang Luong, and Quoc Le. 2020a. Unsupervised data augmentation for consistency training. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*.

Qizhe Xie, Minh-Thang Luong, Eduard H. Hovy, and Quoc V. Le. 2020b. Self-training with noisy student improves imagenet classification. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*, pages 10684–10695. Computer Vision Foundation / IEEE.

Chang Xu, Cécile Paris, Surya Nepal, and Ross Sparks. 2018. Cross-target stance classification with self-attention networks. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 2: Short Papers*, pages 778–783. Association for Computational Linguistics.

Bowen Zhang, Min Yang, Xutao Li, Yunming Ye, Xi-aofei Xu, and Kuai Dai. 2020a. Enhancing cross-target stance detection with transferable semantic-emotion knowledge. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online. Association for Computational Linguistics.

Bowen Zhang, Min Yang, Xutao Li, Yunming Ye, Xi-aofei Xu, and Kuai Dai. 2020b. Enhancing cross-target stance detection with transferable semantic-emotion knowledge. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 3188–3197. Association for Computational Linguistics.

A Hyper-parameters setting

As few-shot learning is extremely sensitive to hyper-parameters, we conduct a grid hyper-parameters search based on the performance on the validation data for learning rate lr $\in \{$2e-5, 3e-5$\}$, temperature $\tau \in \{0.05, 0.1, 0.2, 0.3, 0.4\}$ and scalar weighting $\lambda_b \in \{0.05, 0.1, 0.2, 0.5\}$ on two different datasets across different shot sizes. Here, we give the hyperparameter settings of our model in Table 4 and Table 5. Experiments are conducted on NVIDIA RTX TITAN GPUs.

Model	5	10	20
BERT	3e-5	3e-5	2e-5
SCL	(3e-5,0.5,0.2)	(3e-5,0.5,0.3)	(3e-5,0.05, 0.3)
TCL	(3e-5,0.2,0.05)	(3e-5,0.2,0.05)	(3e-5,0.2,0.05)
STCC	(3e-5,0.5,0.2)	(2e-5,0.5,0.2)	(3e-5,0.5,0.4)

Table 4: Hyper-parameter configurations for SemEval-2016. In parentheses from left to right are learning rate, scalar weighting λ_b, and temperature τ.

Model	5	10	20
BERT	3e-5	3e-5	3e-5
SCL	(2e-5,0.5,0.3)	(3e-5,0.5,0.4)	(3e-5,0.5,0.4)
TCL	(2e-5,0.5,0.2)	(2e-5,0.5,0.3)	(3e-5,0.5,0.3)
STCC	(2e-5,0.5,0.3)	(3e-5,0.5,0.1)	(2e-5,0.2,0.3)

Table 5: Hyper-parameter configurations for COVID-19. In parentheses from left to right are learning rate, scalar weighting λ_b, and temperature τ.