

# JointCL: A Joint Contrastive Learning Framework for Zero-Shot Stance Detection

Bin Liang<sup>1,2\*</sup>, Qinglin Zhu<sup>1\*</sup>, Xiang Li<sup>1\*</sup>, Min Yang<sup>3</sup>, Lin Gui<sup>4</sup>, Yulan He<sup>4,5</sup>, Ruifeng Xu<sup>1,6†</sup>

<sup>1</sup>School of Computer Science and Technology,  
Harbin Institute of Technology, Shenzhen, China

<sup>2</sup>Joint Lab of HITSZ and China Merchants Securities, Shenzhen, China

<sup>3</sup>SIAT, Chinese Academy of Sciences, Shenzhen, China

<sup>4</sup>Department of Computer Science, University of Warwick, U.K

<sup>5</sup>The Alan Turing Institute, UK, <sup>6</sup>Peng Cheng Laboratory, Shenzhen, China

{bin.liang, zhuqinglin, xiangli}@stu.hit.edu.cn  
min.yang@siat.ac.cn, {lin.gui, Yulan.He}@warwick.ac.uk  
xuruifeng@hit.edu.cn

## Abstract

Zero-shot stance detection (ZSSD) aims to detect the stance for an unseen target during the inference stage. In this paper, we propose a joint contrastive learning (JointCL) framework, which consists of stance contrastive learning and target-aware prototypical graph contrastive learning. Specifically, a stance contrastive learning strategy is employed to better generalize stance features for unseen targets. Further, we build a prototypical graph for each instance to learn the target-based representation, in which the prototypes are deployed as a bridge to share the graph structures between the known targets and the unseen ones. Then a novel target-aware prototypical graph contrastive learning strategy is devised to generalize the reasoning ability of target-based stance representations to the unseen targets. Extensive experiments on three benchmark datasets show that the proposed approach achieves state-of-the-art performance in the ZSSD task<sup>1</sup>.

## 1 Introduction

Stance detection aims to automatically identify one’s opinionated standpoint/attitude (e.g. *Pro*, *Con*, or *Neutral*, etc.) expressed in text towards a specific proposition, topic, or target (Somasundaran and Wiebe, 2010; Augenstein et al., 2016; Mohammad et al., 2016; Sobhani et al., 2017). For example, a text “*Everyone is able to believe in whatever they want.*” expresses a stance of “*Pro*” towards the target “*Atheism*”.

Existing methods achieved promising performance in in-target stance detection when trained and tested on the datasets towards the same set of

targets (Mohtarami et al., 2018; Graells-Garrido et al., 2020), and in cross-target stance detection that identifies the stance of a destination target using models trained on a related source target in a one-to-one way (Xu et al., 2018; Zhang et al., 2020; Liang et al., 2021a). In practice, however, it is infeasible to enumerate all possible (in-target) or related (cross-target) targets beforehand for training stance detection models. Hence, zero-shot stance detection (ZSSD) (Allaway and McKeown, 2020), which aims to detect the stance for unseen targets during the inference stage is a promising scenario forward.

To deal with ZSSD, intuitively, we can either reason the target-based stance features from the learned stance information based on the context (i.e., from the *context-aware* perspective), or identify stance information that is potentially relevant with unseen targets from the learned target-related stance expressions (i.e., from the *target-aware* perspective). Existing research attempted to explore attention mechanism (Allaway and McKeown, 2020), adversarial learning (Allaway et al., 2021), or graph architecture based on external commonsense knowledge (Liu et al., 2021) to learn the stance representations from the context regarding the known targets, aiming to generalize the learned stance features to the unseen targets for ZSSD. But they tend to ignore that the stance information of an unseen target can be represented in the light of the known targets from the *target-aware* perspective.

In this paper, to generalize the stance features to the unseen targets, we propose a joint contrastive learning (JointCL) framework to leverage the stance features of known targets from both the *context-aware* and the *target-aware* perspectives. On the one hand, from the *context-aware* perspective, we explore a **Stance Contrastive Learning**

\* Equal contribution

† Corresponding Author

<sup>1</sup>The source code of this work is released at <https://github.com/HITSZ-HLT/JointCL>

strategy, which effectively improves the quality of stance features by leveraging the similarity of training instances in a stance class while pushing away instances from other stance classes. This essentially allows the exploitation of target-based contextual stance features to better generalize to the unseen targets. On the other hand, from the *target-aware* perspective, we propose a feasible solution to capture the relationships between the known targets and the unseen ones. Specifically, inspired by (Li et al., 2021), we explore a clustering method to generate prototypes from all training instances. We then build prototypical graphs linking the prototypes with the target-based representations, in which each prototype is regarded as a bridge that allows the sharing of the graph structures between known targets and unseen ones. Based on the prototypical graphs, we devise a novel **Target-Aware Prototypical Graph Contrastive Learning** strategy to learn the correlation and difference among the target-based representations. Specifically, a novel edge-oriented graph contrastive loss is deployed to make the graph structures similar for similar target-based representations, and different for dissimilar ones. This essentially generalizes the graph structures learned from the known targets to the unseen ones, so as to better derive target-aware stance information for the unseen targets by the graph representations.

The main contributions of our work are summarized as follows:

- The ZSSD task is approached from a new perspective for detecting stance of an unseen target via reasoning the target-based stance features from the learned stance information based on the context or devising the target-aware stance information that is potentially relevant with the unseen target from the learned ones.
- We propose a novel joint contrastive learning (`JointCL`) framework, which consists of stance contrastive learning and target-aware prototypical graph contrastive learning, to generalize the target-based stance features to the unseen targets.
- Extensive experiments on three benchmark datasets show that the proposed `JointCL` framework outperforms state-of-the-art baselines in the ZSSD task. Further, the proposed `JointCL` framework can be easily extended

to the few-shot and cross-target stance detection and achieves outstanding performance.

## 2 Related Work

### 2.1 Zero-Shot Stance Detection

Zero-shot stance detection (ZSSD) aims to detect stance for destination unseen targets by learning stance features from known targets (Allaway and McKeown, 2020). To deal with zero-shot stance detection, Allaway and McKeown (2020) created a new dataset consisting of a large range of topics covering broad themes, called Varied Stance Topics (VAST). Based on it, they proposed a topic-grouped attention model to implicitly capture relationships between targets by using generalized topic representations. Allaway et al. (2021) adopted a target-specific stance detection dataset (Mohammad et al., 2016) and deployed adversarial learning to extract target-invariant transformation features in ZSSD. More recently, to exploit both the structural-level and semantic-level information of the relational knowledge, Liu et al. (2021) proposed a common-sense knowledge enhanced graph model based on BERT (Devlin et al., 2019) to tackle ZSSD.

### 2.2 Contrastive Learning

Contrastive learning in the latent space has recently shown great promise, which aims to make the representation of a given anchor data to be similar to its positive pairs and dissimilar to its negative pairs (Hadsell et al., 2006; Wu et al., 2018; Tian et al., 2020; Chen et al., 2020a; Khosla et al., 2020; Chen et al., 2020b; Zhang et al., 2021; Wang et al., 2021; Gunel et al., 2021). Various contrastive learning approaches have been developed to deal with natural language processing tasks (Kachuee et al., 2021; Qin et al., 2021; Yang et al., 2021; Liu and Liu, 2021; Liang et al., 2021b), including unsupervised text representation learning (Giorgi et al., 2021), text classification (Qiu et al., 2021), and text clustering (Zhang et al., 2021).

More recently, Li et al. (2021) presented prototypical contrastive learning and a ProtoNCE loss to encourage representations to be closer to their assigned prototypes. However, this method only models the relationship between an anchor instance and its nearest prototype. On the other hand, You et al. (2020) proposed a graph contrastive learning framework based on graph data augmentation, which improves the graph representations for better generalizability and robustness. However, their ap-

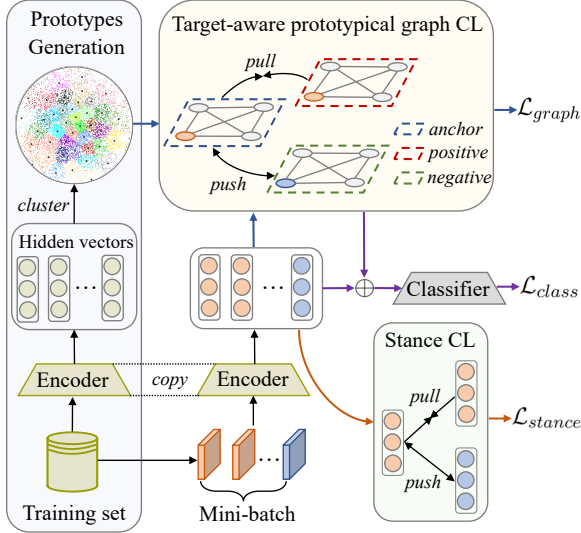


Figure 1: The architecture of our JointCL framework.  $\oplus$  is vector concatenation. In the graphs, the gray ellipses denote prototypes, others denote hidden vectors. Vectors with the same color hold the same stance.

proach ignores the relationships of edges regarding the graph structures. In our (JointCL) framework, we devise a novel edge-oriented graph contrastive loss to learn the contrastive information of the relationships between prototypes and the targets, thus generalizing the graph structures to the unseen targets for learning target-aware stance information.

### 3 Methodology

In this section, we describe the proposed Joint Contrastive Learning (JointCL) framework for zero-shot stance detection in detail. As demonstrated in Figure 1, the architecture of the JointCL framework contains four main components: 1) *stance contrastive learning*, which performs contrastive learning based on the supervised signal of stance labels for better generalization of stance features; 2) *prototypes generation*, which derives the prototypes of the training data by a clustering method; 3) *target-aware prototypical graph contrastive learning*, which performs the edge-oriented graph contrastive learning strategy based on the target-aware prototypical graphs for sharing the graph structures between known targets and unseen ones; 4) *classifier*, which detects the stances of targets based on the hidden vectors and graph representations.

#### 3.1 Task Description

Formally, let  $\mathcal{D}_s = \{(r_s^i, t_s^i, y_s^i)\}_{i=1}^{N_s}$  be the training set for the source targets, where  $t_s^i$  and  $y_s^i$  are

the training target and the stance label towards the context  $r_s^i$  respectively.  $N_s$  is the number of the training instances. Further, let  $\mathcal{D}_d = \{(r_d^i, t_d^i)\}_{i=1}^{N_d}$  be the testing set for the targets which are unseen in the training set. Here,  $t_d^i$  is the testing target in the context  $r_d^i$ . The goal of ZSSD is to predict a stance label (e.g. “Pro”, “Con”, or “Neutral”) of each testing instance by training a model on the training set.

#### 3.2 Encoder Module

Given a sequence of words  $r = \{w_i\}_{i=1}^n$  and the corresponding target  $t$ , where  $n$  is the length of the sentence  $r$ , we adopt a pre-trained BERT (Devlin et al., 2019) as the Encoder Module and feed “[CLS] $r$ [SEP] $t$ [SEP]” as input into the encoder module to obtain a  $d_m$ -dimensional hidden representation  $\mathbf{h} \in \mathbb{R}^{d_m}$  of each input instance:

$$\mathbf{h} = \text{BERT}([\text{CLS}]r[\text{SEP}]t[\text{SEP}])_{[\text{CLS}]} \quad (1)$$

Here, we use the vector of the [CLS] token to represent the input instance. For the training set  $\mathcal{D}_s$ , the hidden representations of the training instances can be represented as  $\mathcal{H} = \{\mathbf{h}_i\}_{i=1}^{N_s}$ .

#### 3.3 Stance Contrastive Learning

As previously discussed in Gunel et al. (2021), good generalization requires capturing the similarity between examples in one class and contrasting them with examples in other classes. To improve the generalization ability of stance learning, we define a stance contrastive loss on the hidden vectors of instances with the supervised stance label information. Given the hidden vectors  $\{\mathbf{h}_i\}_{i=1}^{N_b}$  in a mini-batch  $\mathcal{B}$  (here,  $N_b$  is the size of mini-batch), and an *anchor* of hidden vector  $\mathbf{h}_i$ ,  $\mathbf{h}_i, \mathbf{h}_j \in \mathcal{B}$  with the same stance label is considered as a *positive* pair, i.e.  $y^i = y^j$ , where  $y^i$  and  $y^j$  are the stance labels of  $\mathbf{h}_i$  and  $\mathbf{h}_j$ , respectively, while the samples  $\{\mathbf{h}_k \in \mathcal{B}, k \neq i\}$  are treated as *negative* representations with respect to the *anchor*. Then the contrastive loss is computed across all *positive* pairs, both  $(\mathbf{h}_i, \mathbf{h}_j)$  and  $(\mathbf{h}_j, \mathbf{h}_i)$  in a mini-batch:

$$\mathcal{L}_{stance} = \frac{-1}{N_b} \sum_{\mathbf{h}_i \in \mathcal{B}} \ell^s(\mathbf{h}_i) \quad (2)$$

$$\ell^s(\mathbf{h}_i) = \log \frac{\sum_{j \in \mathcal{B} \setminus i} \mathbb{1}_{[y^i=y^j]} \exp(f(\mathbf{h}_i, \mathbf{h}_j)/\tau_s)}{\sum_{j \in \mathcal{B} \setminus i} \exp(f(\mathbf{h}_i, \mathbf{h}_j)/\tau_s)} \quad (3)$$

where  $\mathbb{1}_{[i=j]} \in \{0, 1\}$  is an indicator function evaluating to 1 iff  $i = j$ .  $f(\mathbf{u}, \mathbf{v}) = \text{sim}(\mathbf{u}, \mathbf{v}) = \mathbf{u}^\top \mathbf{v} / \|\mathbf{u}\| \|\mathbf{v}\|$  denotes the cosine similarity between vectors  $\mathbf{u}$  and  $\mathbf{v}$ .

### 3.4 Prototypes Generation

In the Prototypical Networks for few-shot learning, Snell et al. (2017) derived the prototype of each class by computing the mean vector of the embedded support points belonging to the class. However, in the ZSSD data, the distribution of targets is usually imbalanced. Therefore, inspired by (Li et al., 2021), we perform  $k$ -means clustering on the hidden vectors of the training instances  $\mathcal{H} = \{\mathbf{h}_i\}_{i=1}^{N_s}$  to generate  $k$  clusters as the prototypes  $\mathcal{C} = \{\mathbf{c}_i\}_{i=1}^k$  with respect to the target-based representations of training set. Here, a prototype is defined as a *representative embedding for a group of semantically similar instances* (Li et al., 2021). Clustering is performed at each training epoch to update the prototypes.

### 3.5 Prototypical Graph

Once the prototypes are generated, a prototypical graph is constructed to capture the relationships between the prototypes and the known targets. This enables the learning of the representation of a target-based instance by modeling the different weights of edges between its corresponding target and various prototypes, so as to generalize the learned graph information to the unseen targets. Here, the prototypes and the target-based representations are updated in an alternative manner. For a hidden vector  $\mathbf{h}_i$  of a training instance  $i$ , we first treat the prototypes  $\mathcal{C}$  and the hidden vector  $\mathbf{h}_i$  as nodes of the prototypical graph:  $\mathbf{X} = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_k, \mathbf{h}_i]$ , and then construct the adjacency matrix  $\mathcal{G} \in \mathbb{R}^{(k+1) \times (k+1)}$  of the fully-connected graph,  $\mathcal{G}_{i,j} = \mathcal{G}_{j,i} = 1$ .

Next, we feed the nodes  $\mathbf{X}$  and the corresponding adjacency matrix  $\mathcal{G}$  into a graph attention network (GAT) (Velickovic et al., 2018) to derive the attention scores  $\alpha_i$  and the graph representation  $\mathbf{z}_i$  for the target-based instance  $i$ :

$$\alpha_i = a(\text{GAT}(\mathbf{X}; \mathcal{G})) \quad (4)$$

$$\mathbf{z}_i = f(\text{GAT}(\mathbf{X}; \mathcal{G})) \quad (5)$$

where  $\text{GAT}(\cdot)$  represents GAT operation.  $a(\cdot)$  denotes retrieving the attention score matrix from the GAT operation,  $f(\cdot)$  denotes retrieving the graph representation for  $\mathbf{h}_i$ .

### 3.6 Target-Aware Prototypical Graph Contrastive Learning

From the target-aware perspective, we further explore a Target-Aware Prototypical Graph Contrastive Learning strategy, aiming at generalizing the graph structures learned from the known targets to the unseen ones. Specifically, for the attention matrices  $\{\alpha_i\}_{i=1}^{N_b}$  in each mini-batch  $\mathcal{B}$ , we devise a novel edge-oriented prototypical graph contrastive loss, making the graph structure of similar target-based representations to be similar. This essentially allows the model to learn the representations of (unseen) targets through the prototypes, thus generalizing the target-aware stance information to the unseen targets.

For an *anchor* instance  $i$  with edge weights (i.e., the attention score matrix)  $\alpha_i$ , we construct a *positive* pair  $(\alpha_i, \alpha_j)$  by retrieving the attention score matrix of instance  $j$  which is either about the same target or has been assigned to the same prototype, and expresses the same stance as  $i$ . We also construct negative pairs,  $(\alpha_i, \alpha_k), \alpha_k \in \mathcal{B}, k \neq i$ . Then, the edge-oriented graph contrastive loss is defined as<sup>2</sup>:

$$\mathcal{L}_{graph} = \frac{-1}{N_b} \sum_{\alpha_i \in \mathcal{B}} \ell^g(\alpha_i) \quad (6)$$

$$\ell^g(\alpha_i) = \log \frac{\sum_{j \in \mathcal{B} \setminus i} \Phi(i, j) \exp(f(\alpha_i, \alpha_j) / \tau_g)}{\sum_{j \in \mathcal{B} \setminus i} \exp(f(\alpha_i, \alpha_j) / \tau_g)} \quad (7)$$

$$\Phi(i, j) = \begin{cases} 1 & \text{if } y^i = y^j \text{ and } p^i = p^j \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

where  $p^i = p^j$  represents the instances  $i$  and  $j$  correspond to the same target or belong to the same prototype, and express the same stance.

The calculation of the stance and edge-oriented prototypical graph contrastive losses for each mini-batch  $\mathcal{B}$  is illustrated in Algorithm 1.

### 3.7 Stance Detection

For each instance  $i$ , we first concatenate the hidden vector  $\mathbf{h}_i$  and the graph representation  $\mathbf{z}_i$  to get the output representation  $\mathbf{v}_i$  towards the instance  $i$ :

$$\mathbf{v}_i = \mathbf{h}_i \oplus \mathbf{z}_i \quad (9)$$

Then the output representation  $\mathbf{v}_i$  is fed into a classifier with a softmax function to produce the pre-

<sup>2</sup>Here, to compute the cosine similarity, we flatten each matrix  $\alpha_i$  into a one-dimensional array.

**Algorithm 1:** Calculation of the stance and edge-oriented prototypical graph contrastive losses for each mini-batch  $\mathcal{B}$ .

---

**Input:**  $\mathcal{B} = \{\mathbf{h}_i, \boldsymbol{\alpha}_i\}_{i=1}^{N_b}$ ,  $\ell^s, \ell^g \leftarrow 0, 0$   
**Output:**  $\mathcal{L}_{stance}, \mathcal{L}_{graph}$

```

1 for  $i = 1$  to  $N_b$  do
2    $\mathbf{h}_i, \boldsymbol{\alpha}_i \leftarrow \mathcal{B}$ 
3    $\ell^s(\mathbf{h}_i)_{pos}, \ell^s(\mathbf{h}_i)_{neg} \leftarrow 0, 0$ 
4    $\ell^g(\boldsymbol{\alpha}_i)_{pos}, \ell^g(\boldsymbol{\alpha}_i)_{neg} \leftarrow 0, 0$ 
5   for  $j = 1$  to  $N_b$  and  $j \neq i$  do
6      $\mathbf{h}_j, \boldsymbol{\alpha}_j \leftarrow \mathcal{B}$ 
7     if  $y^i == y^j$  then
8        $\ell^s(\mathbf{h}_i)_{pos+} = \exp(f(\mathbf{h}_i, \mathbf{h}_j)/\tau_s)$ 
9     if  $p^i == p^j$  and  $y^i == y^j$  then
10       $\ell^g(\boldsymbol{\alpha}_i)_{pos+} = \exp(f(\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j)/\tau_g)$ 
11       $\ell^s(\mathbf{h}_i)_{neg+} = \exp(f(\mathbf{h}_i, \mathbf{h}_j)/\tau_s)$ 
12       $\ell^g(\boldsymbol{\alpha}_i)_{neg+} = \exp(f(\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j)/\tau_g)$ 
13     ▷ Computing stance contrastive loss for each  $\mathbf{h}_i$ 
14      $\ell^s+ = \ell^s(\mathbf{h}_i)_{pos+} / \ell^s(\mathbf{h}_i)_{neg+}$ 
15     ▷ Computing edge-oriented prototypical graph
        contrastive loss for each  $\boldsymbol{\alpha}_i$ 
16      $\ell^g+ = \ell^g(\boldsymbol{\alpha}_i)_{pos+} / \ell^g(\boldsymbol{\alpha}_i)_{neg+}$ 
17   ▷ Stance contrastive loss for a mini-batch  $\mathcal{B}$ 
18    $\mathcal{L}_{stance} = -\ell^s / N_b$ 
19   ▷ Edge-oriented prototypical graph contrastive loss
        for a mini-batch  $\mathcal{B}$ 
20    $\mathcal{L}_{graph} = -\ell^g / N_b$ 

```

---

dicted stance distribution  $\hat{\mathbf{y}}_i \in \mathbb{R}^{d_y}$ :

$$\hat{\mathbf{y}}_i = \text{softmax}(\mathbf{W}\mathbf{v}_i + \mathbf{b}) \quad (10)$$

where  $d_y$  is the dimensionality of stance labels.  $\mathbf{W} \in \mathbb{R}^{d_y \times d_m}$  and  $\mathbf{b} \in \mathbb{R}^{d_y}$  are trainable parameters. We adopt a cross-entropy loss between predicted distribution  $\hat{\mathbf{y}}_i$  and ground-truth distribution  $\mathbf{y}_i$  of instance  $i$  to train the classifier:

$$\mathcal{L}_{class} = -\sum_{i=1}^{N_b} \sum_{j=1}^{d_y} y_i^j \log \hat{y}_i^j \quad (11)$$

### 3.8 Learning Objective

The learning objective of our proposed model is to train the model by jointly minimizing the three losses generated by stance detection, stance contrastive learning, and target-aware prototypical graph contrastive learning. The overall loss  $\mathcal{L}$  is formulated by summing up three losses:

$$\mathcal{L} = \gamma_c \mathcal{L}_{class} + \gamma_s \mathcal{L}_{stance} + \gamma_g \mathcal{L}_{graph} + \lambda \|\Theta\|^2 \quad (12)$$

where  $\gamma_c$ ,  $\gamma_s$  and  $\gamma_g$  are tuned hyper-parameters.  $\Theta$  denotes all trainable parameters of the model,  $\lambda$  represents the coefficient of  $L_2$ -regularization.

	Train	Dev	Test
# Examples	13477	2062	3006
# Unique Comments	1845	682	786
# Zero-shot Topics	4003	383	600
# Few-shot Topics	638	114	159

Table 1: Statistics of VAST dataset.

Dataset	Target	Favor	Against	Neutral	Unrelated
SEM16	DT	148	299	260	-
	HC	163	565	256	-
	FM	268	511	170	-
	LA	167	544	222	-
	A	124	464	145	-
	CC	335	26	203	-
WT-WT	CA	2469	518	5520	3115
	CE	773	253	947	554
	AC	970	1969	3098	5007
	AH	1038	1106	2804	2949

Table 2: Statistics of SEM16 and WT-WT datasets.

## 4 Experimental Setup

### 4.1 Datasets

We conduct experiments on three datasets to evaluate the proposed JointCL framework. 1) **VAST** (Allaway and McKeown, 2020), which contains a large variety of targets. Each instance consists of a sentence  $r$ , a target  $t$ , and a stance label  $y$  (“Pro”, “Con”, or “Neutral”) towards  $t$ . To show the generalizability of coping with few-shot stance detection, following (Allaway and McKeown, 2020), we also conduct experiments on few-shot condition. The statistics of VAST dataset are shown in Table 1. 2) **SEM16**, which contains 6 pre-defined targets, including *Donald Trump* (DT), *Hillary Clinton* (HC), *Feminist Movement* (FM), *Legalization of Abortion* (LA), *Atheism* (A), and *Climate Change* (CC). Each instance can be classified as *Favor*, *Against* or *Neutral*. 3) **WT-WT**, which contains 5 pre-defined company pairs (target), including *CVS\_AET* (CA), *CI\_ESRX* (CE), *ANTM\_CI* (AC), and *AET\_HUM* (AH). Each instance refers to a stance label from *Support* (corresponding to *Favor*), *Refute* (corresponding to *Against*), *Comment* (corresponding to *Neutral*), or *Unrelated*. The statistics of WT-WT and SEM16 datasets are shown in Table 2. Following (Allaway et al., 2021) and (Conforti et al., 2020), for **SEM16** and **WT-WT** datasets, we use the leave-one-target-out evaluation setup.

### 4.2 Implementation Detail

**Training Settings** The pre-trained uncased BERT-base (Devlin et al., 2019) is used as the

embedding module in which each word token is mapped to a 768-dimensional embedding. The learning rate is set to  $3e-5$ . Following (Xu et al., 2018), the coefficient  $\lambda$  is set to  $1e-5$ . Adam is utilized as the optimizer. The mini-batch size is set to 16, considering the trade-off between computational resource and evaluation performance. For contrastive losses, both the temperature parameters  $\tau_s$  and  $\tau_g$  are set to 0.07. For clustering, the number of clusters are set to  $k = 100$  for the VAST dataset and  $k = 10$  for the WT-WT and SEM16 datasets respectively. Corresponding to the number of  $k$ , we set  $\gamma_c = 0.8$ ,  $\gamma_s = 1$ , and  $\gamma_g = 0.1$  for VAST dataset and  $\gamma_g = 0.5$  for WT-WT and SEM16 datasets, respectively. They are the optimal hyper-parameters in the pilot studies. We apply early stopping in training process and the patience is 5. We report averaged scores of 10 runs to obtain statistically stable results.

**Evaluation Metric** For the VAST dataset, following (Allaway and McKeown, 2020), we calculate Macro-averaged F1 of each label to measure the testing performance of the models. For the SEM16 dataset, following (Allaway et al., 2021), we report  $F_{avg}$ , the average of F1 on *Favor* and *Against*. For the WT-WT dataset, following (Conforti et al., 2020), we report the Macro F1 score of each target.

### 4.3 Comparison Models

We compare the proposed JointCL with a series of strong stance detection baselines, including neural network-based method: BiCond (Augenstein et al., 2016), attention-based models: CrossNet (Xu et al., 2018) and SiamNet (Santosh et al., 2019), knowledge-based method: SEKT (Zhang et al., 2020), graph network method: TPDG (Liang et al., 2021a), adversarial learning method: TOAD (Allaway et al., 2021), and BERT-based methods: BERT (Devlin et al., 2019), TGA Net (Allaway and McKeown, 2020), BERT-GCN (Liu et al., 2021), and CKE-Net (Liu et al., 2021).

In addition, we provide variants of our proposed JointCL in the ablation study:

- (1) “w/o  $\mathcal{L}_{stance}$ ” denotes without stance contrastive learning.
- (2) “w/o  $\mathcal{L}_{graph}$ ” denotes without prototypical graph contrastive learning.
- (3) “w/o *graph*” denotes that this model performs the target-aware contrastive learning on the hidden representations of the instances with the

supervised information from target labels. That is, the contrastive loss functions of Eq. 6 and Eq. 7 are replaced by:

$$\mathcal{L}_{graph} = \frac{-1}{N_b} \sum_{\mathbf{h}_i \in \mathcal{B}} \ell^g(\mathbf{h}_i) \quad (13)$$

$$\ell^g(\mathbf{h}_i) = \log \frac{\sum_{j \in \mathcal{B} \setminus i} \mathbb{1}_{[t^i=t^j]} \exp(f(\mathbf{h}_i, \mathbf{h}_j)/\tau)}{\sum_{j \in \mathcal{B} \setminus i} \exp(f(\mathbf{h}_i, \mathbf{h}_j)/\tau)} \quad (14)$$

(4) “w/o *cluster*” denotes without using clustering to generate prototypes. That is, this model simply takes the mean of target-based hidden representations as a prototype.

(5) “w/o *edge*” denotes without considering edge information, i.e., it performs the prototypical graph contrastive learning on the graph representations of the instance nodes. The contrastive loss functions of Eq. 6 and Eq. 7 are replaced by:

$$\mathcal{L}_{graph} = \frac{-1}{N_b} \sum_{z_i \in \mathcal{B}} \ell^g(z_i) \quad (15)$$

$$\ell^g(z_i) = \log \frac{\sum_{j \in \mathcal{B} \setminus i} \mathbb{1}_{[p^i=p^j]} \exp(f(z_i, z_j)/\tau)}{\sum_{j \in \mathcal{B} \setminus i} \exp(f(z_i, z_j)/\tau)} \quad (16)$$

## 5 Experimental Results

### 5.1 Main Results

The main comparison results of ZSSD on three benchmark datasets are reported in Table 3. It can be observed from the experimental results, our proposed JointCL framework performs consistently better than the non-BERT and the BERT-based comparison models on both the VAST and WT-WT datasets, and achieves overall better performance than the comparison baselines on the SEM16 dataset. This verifies the effectiveness of our JointCL framework in the ZSSD task. Furthermore, the significance tests of JointCL over the baseline models show that our JointCL significantly outperforms the baseline models (the results of  $p$ -value on most of the evaluation metrics are less than 0.05). More concretely, in comparison with the adversarial learning-based model (TOAD), our JointCL achieves significant improvement across all datasets. This indicates that exploring graph contrastive learning to model the relationships among targets can better generalize the target-based stance features to the unseen targets. In addition, the comparison results between our JointCL

Model	VAST (%)				SEM16 (%)						WT-WT (%)			
	Pro	Con	Neu	All	DT	HC	FM	LA	A	CC	CA	CE	AC	AH
BiCond	44.6 <sup>‡</sup>	47.4 <sup>‡</sup>	34.9 <sup>†</sup>	42.8 <sup>‡</sup>	30.5 <sup>‡</sup>	32.7 <sup>‡</sup>	40.6 <sup>‡</sup>	34.4 <sup>‡</sup>	31.0 <sup>‡</sup>	15.0 <sup>‡</sup>	56.5 <sup>‡</sup>	52.5 <sup>‡</sup>	64.9 <sup>‡</sup>	63.0 <sup>‡</sup>
CrossNet	46.2 <sup>‡</sup>	43.4 <sup>‡</sup>	40.4 <sup>†</sup>	43.4 <sup>‡</sup>	35.6	38.3	41.7	38.5	39.7	22.8	59.1 <sup>‡</sup>	54.5 <sup>‡</sup>	65.1 <sup>‡</sup>	62.3 <sup>‡</sup>
SiamNet	47.5	43.3	39.6	43.5	36.9	37.5	44.3	41.6	41.2	25.6	58.3 <sup>‡</sup>	54.4 <sup>‡</sup>	68.7 <sup>‡</sup>	67.7 <sup>‡</sup>
SEKT	50.4 <sup>†</sup>	44.2 <sup>†</sup>	30.8 <sup>†</sup>	41.8 <sup>†</sup>	-	-	-	-	-	-	-	-	-	-
TPDG	53.7	49.6	52.3	51.9	47.3	50.9	53.6	46.5	48.7	32.3	66.8 <sup>b</sup>	65.6 <sup>b</sup>	74.2 <sup>b</sup>	73.1 <sup>b</sup>
TOAD	42.6	36.7	43.8	41.0	49.5 <sup>‡</sup>	51.2 <sup>‡</sup>	<b>54.1<sup>‡</sup></b>	46.2 <sup>‡</sup>	46.1 <sup>‡</sup>	30.9 <sup>‡</sup>	55.3	57.7	58.6	61.7
BERT	54.6 <sup>‡</sup>	58.4 <sup>‡</sup>	85.3 <sup>f</sup>	66.1 <sup>‡</sup>	40.1 <sup>f</sup>	49.6 <sup>‡</sup>	41.9 <sup>f</sup>	44.8 <sup>‡</sup>	<b>55.2<sup>‡</sup></b>	37.3 <sup>‡</sup>	56.0 <sup>b</sup>	60.5 <sup>b</sup>	67.1 <sup>b</sup>	67.3 <sup>b</sup>
TGA Net	55.4 <sup>‡</sup>	58.5 <sup>‡</sup>	85.8 <sup>†</sup>	66.6 <sup>‡</sup>	40.7	49.3	46.6	45.2	52.7	36.6	65.7	63.5	69.9	68.7
BERT-GCN	58.3 <sup>†</sup>	60.6 <sup>†</sup>	86.9 <sup>†</sup>	68.6 <sup>†</sup>	42.3	50.0	44.3	44.2	53.6	35.5	67.8	64.1	70.7	69.2
CKE-Net	61.2 <sup>†</sup>	61.2 <sup>†</sup>	88.0 <sup>†</sup>	70.2 <sup>†</sup>	-	-	-	-	-	-	-	-	-	-
JointCL (ours)	<b>64.9*</b>	<b>63.2*</b>	<b>88.9*</b>	<b>72.3*</b>	<b>50.5*</b>	<b>54.8*</b>	53.8	<b>49.5*</b>	54.5	<b>39.7*</b>	<b>72.4*</b>	<b>70.2*</b>	<b>76.0*</b>	<b>75.2*</b>

Table 3: Experimental results on three ZSSD datasets. The results with <sup>‡</sup> are retrieved from (Allaway and McKeown, 2020), <sup>†</sup> from (Liu et al., 2021), <sup>‡</sup> from (Allaway et al., 2021), <sup>‡</sup> from (Conforti et al., 2020), and <sup>b</sup> from (Liang et al., 2021a). Best scores are in bold. Results with \* denote the significance tests of our JointCL over the baseline models at  $p$ -value  $< 0.05$ .

Model	VAST (%)				SEM16 (%)						WT-WT (%)			
	Pro	Con	Neu	All	DT	HC	FM	LA	A	CC	CA	CE	AC	AH
JointCL (ours)	<b>64.9</b>	<b>63.2</b>	<b>88.9</b>	<b>72.3</b>	<b>50.5</b>	<b>54.8</b>	<b>53.8</b>	<b>49.5</b>	<b>54.5</b>	<b>39.7</b>	<b>72.4</b>	<b>70.2</b>	<b>76.0</b>	<b>75.2</b>
w/o $\mathcal{L}_{stance}$	61.6	60.7	87.2	69.8	46.2	51.4	51.2	45.3	52.5	36.3	69.4	67.8	72.1	71.4
w/o $\mathcal{L}_{graph}$	62.5	62.1	87.8	70.7	48.8	52.7	51.5	48.2	53.2	38.1	70.5	68.3	74.7	73.6
w/o graph	60.8	62.3	87.7	70.3	46.5	50.3	49.7	45.6	52.3	37.4	69.8	68.7	73.2	71.7
w/o cluster	59.6	62.2	86.8	69.5	47.4	53.1	52.3	48.6	53.7	38.8	70.9	69.2	74.9	72.6
w/o edge	63.3	62.5	88.4	71.4	49.2	53.4	53.1	48.9	53.5	39.2	71.2	69.5	75.2	74.2

Table 4: Experimental results of ablation study.

and the previous BERT-based models demonstrate that the stance representations learned from known targets can be better generalized to the unseen targets with our proposed novel contrastive learning strategy.

## 5.2 Ablation Study

To analyze the impact of different components in our proposed JointCL on the performance, we conduct an ablation study and report the results in Table 4. We can observe that the removal of stance contrastive learning (“w/o  $\mathcal{L}_{stance}$ ”) sharply reduces the performance in all evaluation metrics and across all datasets. This indicates that performing contrastive learning based on stance information can improve the quality of stance representations for better generalizing the learned stance features to the unseen targets, and thus improve the performance of ZSSD. The removal of edge-oriented prototypical graph contrastive learning (“w/o  $\mathcal{L}_{graph}$ ”) leads to considerable performance degradation. This implies that performing target-based contrastive learning for prototypical graph can generalize the graph relations between known targets and prototypes to the unseen targets, which enables the model to derive better representation for the examples of unseen targets, and thus leads

to improved ZSSD performance.

In addition, from the results of “w/o graph” we can see that purely performing the target-based contrastive learning on the hidden representations slashes the learning ability of stance contrastive learning, and thus leads to poorer performance. This verifies the effectiveness of exploring prototypical graph contrastive learning in our JointCL. We also observe that the performance of “w/o cluster” drops consistently across datasets, which indicates that exploring clustering method can effectively relieve the problem of the imbalanced distribution of targets in the dataset. The removal of edge-oriented graph contrastive strategy (“w/o  $\mathcal{L}_{edge}$ ”) leads to noticeable performance degradation. This implies that, to represent the (unseen) targets with prototypes, we should pay more attention to the relationships between targets and prototypes, rather than simply drawing closer similar target-based representations in the graph.

## 5.3 Impact of the Values of $k$

To analyze the impact of using different values of  $k$  in  $k$ -means clustering on the performance, we conduct experiments on the three datasets, and show the results in Figure 2. Here, for VAST, we show the results of all labels. For the SEM16 and WT-

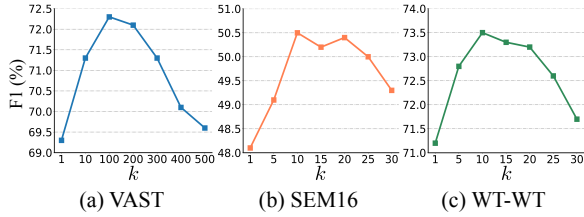


Figure 2: Experimental results of different values of  $k$ .

Model	Pro	Con	Neu	All
BiCond	45.4	46.3	25.9	39.2
Cross-Net	50.8	50.5	41.0	47.4
SEKT	51.0	47.9	21.5	47.4
BERT	54.4	59.7	79.6	64.6
TGA Net	58.9	59.5	80.5	66.3
BERT-GCN	62.8	63.4	83.0	69.7
CKE-Net	64.4	62.2	83.5	70.1
JointCL (ours)	63.2	<b>66.7</b>	<b>84.6</b>	<b>71.5</b>

Table 5: Experimental results of few-shot condition. Results of baselines are retrieved from (Liu et al., 2021).

WT, we show the average performance of all targets. We observe that for VAST that contains a large number of targets (more than 5,000 in the training set), the performance increases with the increasing value of  $k$  and peaks at  $k = 100$ . Further increasing the values of  $k$  results in worse performance. Similarly, for SEM16 and WT-WT, better performance is obtained in the region of  $k \in [10, 20]$  and peaks when  $k = 10$ . This implies that we can set an appropriate region for the value of  $k$  according to the number of targets in the dataset.

#### 5.4 Generalizability Analysis

**Analysis of Few-Shot Condition** To evaluate the generalizability of our JointCL framework in few-shot stance detection, following (Allaway and McKeown, 2020; Liu et al., 2021), we also evaluate JointCL in the few-shot condition on the VAST dataset. From the experimental results shown in Table 5, we can see that JointCL performs overall better than all the comparison methods under the few-shot condition. This verifies the effectiveness and generalizability of JointCL in dealing with both zero-shot and few-shot stance detection.

**Analysis of Cross-Target Scenario** We further conduct comparison experiments in the cross-target scenario on the SEM16 dataset. Cross-target stance detection trains on a source target and tests on an unseen but related one, which is a task related to ZSSD. We report the results in Table 6. It can be observed that JointCL achieves consistently better performance on all cross-target scenarios, which

Model	HC→DT	DT→HC	FM→LA	LA→FM
BiCond	29.7	35.8	45.0	41.6
CrossNet	43.1	36.2	45.4	43.3
BERT	43.6	36.5	47.9	33.9
SEKT	47.7	42.0	53.6	51.3
TPDG	50.4	52.9	58.3	54.1
JointCL (ours)	<b>52.8</b>	<b>54.3</b>	<b>58.8</b>	<b>54.5</b>

Table 6: Experimental results of cross-target condition. “HC→DT” denotes training on HC and testing on DT, etc. Results of baselines are retrieved from (Liang et al., 2021a).

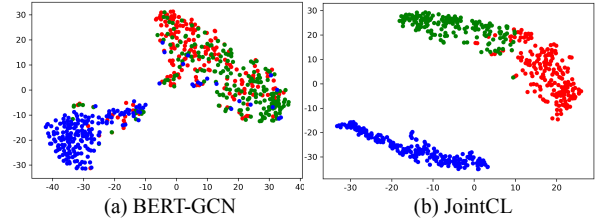


Figure 3: Visualization of intermediate embeddings. Red dots denote *Pro* examples, green dots denote *Con* examples, and blue dots denote *Neutral* examples.

verifies that our JointCL can generalize the learning ability to deal with cross-target scenarios. In addition, when compared with the results of Table 3, we see that the results of cross-target stance detection are generally better than ZSSD. This shows that recognizing the relationships among targets in advance can potentially improve the stance detection performance for the unseen targets, which illustrates the challenge of the ZSSD task from another angle.

#### 5.5 Visualization

To qualitatively demonstrate how the proposed JointCL captures good generalization of stance features for unseen targets in ZSSD, we randomly select 200 test instances for each label from VAST dataset and show the t-SNE (van der Maaten and Hinton, 2008) visualization of intermediate embeddings learned by BERT-GCN and our proposed JointCL on VAST in Figure 3. It can be seen that the distributions of representations derived from BERT-GCN largely overlap especially for the *Pro* and *Con* stances. But there are clear separations between different stances (including the *Pro* and *Con* stances) produced by our proposed JointCL. This verifies that the novel joint contrastive learning strategy in JointCL can better separate representations from different stances, so as to improve the performance of ZSSD.



## 6 Conclusion

In this paper, we propose a novel joint contrastive learning (JointCL) framework to deal with the zero-shot stance detection (ZSSD) task. On the one hand, we deploy a stance contrastive learning strategy to improve the quality of stance representations, so as to capture good generalization of stance features for the unseen targets. This is based on our observation that for some cases we can determine the stance towards a specific target from its associated context. On the other hand, we devise a target-aware prototypical graph contrastive learning strategy to generalize the learned graph information to the unseen targets by leveraging the prototypes as a bridge to model the relationships between known and unseen targets. This is for other cases when it is difficult to infer the stance for an unseen target from the context, but instead, could be relatively easier by exploiting the target-aware stance information from the learned associated targets. Experimental results on three benchmark datasets show that our JointCL achieves state-of-the-art performance in ZSSD. Further, the generalizability analysis shows that our JointCL can also perform outstandingly on few-shot and cross-target stance detection.

## Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (61876053, 62006062, 62176076, 62006060), UK Engineering and Physical Sciences Research Council (grant no. EP/V048597/1, EP/T017112/1), Natural Science Foundation of Guangdong Province of China (No. 2019A1515011705), Shenzhen Foundational Research Funding (JCYJ20200109113441941, JCYJ20210324115614039), Shenzhen Science and Technology Innovation Program (Grant No. KQTD20190929172835662), Joint Lab of Lab of HITSZ and China Merchants Securities. Yulan He is supported by a Turing AI Fellowship funded by the UK Research and Innovation (UKRI) (grant no. EP/V020579/1).

## References

- Emily Allaway and Kathleen McKeown. 2020. [Zero-Shot Stance Detection: A Dataset and Model using Generalized Topic Representations](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8913–8931, Online. Association for Computational Linguistics.
- Emily Allaway, Malavika Srikanth, and Kathleen McKeown. 2021. [Adversarial learning for zero-shot stance detection on social media](#). In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4756–4767, Online. Association for Computational Linguistics.
- Isabelle Augenstein, Tim Rocktäschel, Andreas Vlachos, and Kalina Bontcheva. 2016. [Stance detection with bidirectional conditional encoding](#). In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 876–885, Austin, Texas. Association for Computational Linguistics.
- Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey E. Hinton. 2020a. [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.
- Ting Chen, Simon Kornblith, Kevin Swersky, Mohammad Norouzi, and Geoffrey E. Hinton. 2020b. [Big self-supervised models are strong semi-supervised learners](#). In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*.
- Costanza Conforti, Jakob Berndt, Mohammad Taher Pilehvar, Chryssi Giannitsarou, Flavio Toxvaerd, and Nigel Collier. 2020. [Will-they-won't-they: A very large dataset for stance detection on Twitter](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1715–1724, Online. Association for Computational Linguistics.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of deep bidirectional transformers for language understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- John Giorgi, Osvald Nitski, Bo Wang, and Gary Bader. 2021. [DeCLUTR: Deep contrastive learning for unsupervised textual representations](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 879–895, Online. Association for Computational Linguistics.
- Eduardo Graells-Garrido, Ricardo Baeza-Yates, and Mounia Lalmas. 2020. Representativeness of abortion legislation debate on twitter: A case study in argentina and chile. In *Companion Proceedings of the Web Conference 2020*, page 765–774.

- Beliz Gunel, Jingfei Du, Alexis Conneau, and Veselin Stoyanov. 2021. [Supervised contrastive learning for pre-trained language model fine-tuning](#). In *International Conference on Learning Representations*.
- Raia Hadsell, Sumit Chopra, and Yann LeCun. 2006. Dimensionality reduction by learning an invariant mapping. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 1735–1742. IEEE.
- Mohammad Kachuee, Hao Yuan, Young-Bum Kim, and Sungjin Lee. 2021. [Self-supervised contrastive learning for efficient user satisfaction prediction in conversational agents](#). In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4053–4064, Online. Association for Computational Linguistics.
- Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschiot, 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*.
- Junnan Li, Pan Zhou, Caiming Xiong, and Steven Hoi. 2021. [Prototypical contrastive learning of unsupervised representations](#). In *International Conference on Learning Representations*.
- Bin Liang, Yonghao Fu, Lin Gui, Min Yang, Jiachen Du, Yulan He, and Ruifeng Xu. 2021a. Target-adaptive graph for cross-target stance detection. In *the Web Conference 2021 (WWW '21)*.
- Bin Liang, Wangda Luo, Xiang Li, Lin Gui, Min Yang, Xiaoqi Yu, and Ruifeng Xu. 2021b. [Enhancing Aspect-Based Sentiment Analysis with Supervised Contrastive Learning](#), page 3242–3247. Association for Computing Machinery, New York, NY, USA.
- Rui Liu, Zheng Lin, Yutong Tan, and Weiping Wang. 2021. [Enhancing zero-shot and few-shot stance detection with commonsense knowledge graph](#). In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 3152–3157, Online. Association for Computational Linguistics.
- Yixin Liu and Pengfei Liu. 2021. [SimCLS: A simple framework for contrastive learning of abstractive summarization](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 1065–1072, Online. Association for Computational Linguistics.
- Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. [SemEval-2016 task 6: Detecting stance in tweets](#). In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 31–41, San Diego, California. Association for Computational Linguistics.
- Mitra Mohtarami, Ramy Baly, James Glass, Preslav Nakov, Lluís Màrquez, and Alessandro Moschitti. 2018. [Automatic stance detection using end-to-end memory networks](#). In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 767–776, New Orleans, Louisiana. Association for Computational Linguistics.
- Yujia Qin, Yankai Lin, Ryuichi Takanobu, Zhiyuan Liu, Peng Li, Heng Ji, Minlie Huang, Maosong Sun, and Jie Zhou. 2021. [ERICA: Improving entity and relation understanding for pre-trained language models via contrastive learning](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 3350–3363, Online. Association for Computational Linguistics.
- Yao Qiu, Jinchao Zhang, and Jie Zhou. 2021. [Improving gradient-based adversarial training for text classification by contrastive learning and auto-encoder](#). In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1698–1707, Online. Association for Computational Linguistics.
- T. Y.S.S. Santosh, Srijan Bansal, and Avirup Saha. 2019. Can siamese networks help in stance detection? In *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data*, page 306–309.
- 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.
- 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, Valencia, Spain. Association for Computational Linguistics.
- Swapna Somasundaran and Janyce Wiebe. 2010. [Recognizing stances in ideological on-line debates](#). In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 116–124, Los Angeles, CA. Association for Computational Linguistics.
- Yonglong Tian, Dilip Krishnan, and Phillip Isola. 2020. Contrastive multiview coding. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16*, pages 776–794. Springer.

- Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86):2579–2605.
- Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. [Graph attention networks](#). In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net.
- Ziqi Wang, Xiaozhi Wang, Xu Han, Yankai Lin, Lei Hou, Zhiyuan Liu, Peng Li, Juanzi Li, and Jie Zhou. 2021. [CLEVE: Contrastive Pre-training for Event Extraction](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 6283–6297, Online. Association for Computational Linguistics.
- Zhirong Wu, Yuanjun Xiong, Stella X. Yu, and Dahua Lin. 2018. [Unsupervised feature learning via non-parametric instance discrimination](#). In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pages 3733–3742. IEEE Computer Society.
- 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 (Volume 2: Short Papers)*, pages 778–783, Melbourne, Australia. Association for Computational Linguistics.
- Nan Yang, Furu Wei, Binxing Jiao, Daxing Jiang, and Linjun Yang. 2021. [xMoCo: Cross momentum contrastive learning for open-domain question answering](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 6120–6129, Online. Association for Computational Linguistics.
- Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen. 2020. Graph contrastive learning with augmentations. *Advances in Neural Information Processing Systems*, 33:5812–5823.
- Bowen Zhang, Min Yang, Xutao Li, Yunming Ye, Xiaofei Xu, and Kuai Dai. 2020. [Enhancing cross-target stance detection with transferable semantic-emotion knowledge](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3188–3197, Online. Association for Computational Linguistics.
- Dejiao Zhang, Feng Nan, Xiaokai Wei, Shang-Wen Li, Henghui Zhu, Kathleen McKeown, Ramesh Nallapati, Andrew O. Arnold, and Bing Xiang. 2021. [Supporting clustering with contrastive learning](#). In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5419–5430, Online. Association for Computational Linguistics.