

# Enhancing Zero-Shot Stance Detection with Contrastive and Prompt Learning

Zhenyin Yao, Wenzhong Yang \* and Fuyuan Wei

School of Computer Science and Technology, Xinjiang University, Urumqi 830046, China; 107552103571@stu.xju.edu.cn (Z.Y.); wfv@stu.xju.edu.cn (F.W.)

\* Correspondence: yangwenzhong@xju.edu.cn

**Abstract:** In social networks, the occurrence of unexpected events rapidly catalyzes the widespread dissemination and further evolution of network public opinion. The advent of zero-shot stance detection aligns more closely with the characteristics of stance detection in today's digital age, where the absence of training examples for specific models poses significant challenges. This task necessitates models with robust generalization abilities to discern target-related, transferable stance features within training data. Recent advances in prompt-based learning have showcased notable efficacy in few-shot text classification. Such methods typically employ a uniform prompt pattern across all instances, yet they overlook the intricate relationship between prompts and instances, thereby failing to sufficiently direct the model towards learning task-relevant knowledge and information. This paper argues for the critical need to dynamically enhance the relevance between specific instances and prompts. Thus, we introduce a stance detection model underpinned by a gated multilayer perceptron (gMLP) and a prompt learning strategy, which is tailored for zero-shot stance detection scenarios. Specifically, the gMLP is utilized to capture semantic features of instances, coupled with a control gate mechanism to modulate the influence of the gate on prompt tokens based on the semantic context of each instance, thereby dynamically reinforcing the instance–prompt connection. Moreover, we integrate contrastive learning to empower the model with more discriminative feature representations. Experimental evaluations on the VAST and SEM16 benchmark datasets substantiate our method's effectiveness, yielding a 1.3% improvement over the JointCL model on the VAST dataset.



**Citation:** Yao, Z.; Yang, W.; Wei, F. Enhancing Zero-Shot Stance Detection with Contrastive and Prompt Learning. *Entropy* **2024**, *26*, 325. <https://doi.org/10.3390/e26040325>

Academic Editors: Zoran H. Perić, Vlado Delić, Vladimir Despotovic and Zoran Ognjanović

Received: 4 March 2024

Revised: 30 March 2024

Accepted: 9 April 2024

Published: 11 April 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** zero-shot stance detection; social networking; prompt learning; contrastive learning

## 1. Introduction

Stance detection aims to automatically identify an individual's stance or attitude (e.g., favor, against, or neutral) expressed in text towards a specific proposition, topic, or target [1–4]. Traditionally, this task has focused on learning classifiers to predict stances on the same topic. However, in real-world scenarios, with the continuous emergence of new topics, it is impractical to train a classifier for each topic due to the time-consuming and expensive nature of the process. Therefore, zero-shot stance detection, which seeks to recognize stances towards unseen targets based on knowledge of visible targets, has gradually gained attention.

To tackle the issue of zero-shot stance detection, existing research has endeavored to explore attention mechanisms [5], contrastive learning [6], adversarial learning [7], or graph architectures based on external commonsense knowledge [8]. However, these methods often face limitations in capturing capabilities and a dependency on external resources, thereby failing to fully leverage the intrinsic information contained within datasets. On the other hand, with the widespread adoption of pretrained models such as BERT [9] and GPT [10] in the field of natural language processing, prompt learning has emerged as a novel technique, thus demonstrating significant potential in addressing zero-shot stance detection tasks. This approach transforms text classification tasks into cloze tests, where

the pretrained language model is responsible for selecting the appropriate word from a set of candidates to fill in the blanks, thereby ensuring semantic alignment with another piece of text. For example, in identifying the sentiment of a social media post like “I missed the bus today”, we might continue with a prompt like “I felt so\_\_” and ask the PLM to choose from a set of emotion-related words to complete the sentence. In this manner, by selecting suitable prompts, we can manipulate the model’s behavior to predict the desired output using the pretrained LM itself, sometimes without the need for any additional task-specific training [11].

Brown et al. [12] first introduced the concept of prompts in the context method. Subsequently, Schick and Schütze [13] proposed PET, which achieves improvements by leveraging patterns in natural language understanding. Some studies [14–16] have automated the search for prompts to reduce the dependence on manual pattern design by human experts. All these methods utilize natural language as prompts; hence, they are referred to as discrete prompts. Other methods such as Ppt [17], Prefix-tuning [18], and P-tuningV2 [19] replace natural language prompts with trainable continuous tokens, thereby automatically searching for the optimal prompts in high-dimensional space. Accordingly, these methods are known as continuous prompts. Current prompt-based learning methods typically train models targeting specific task objectives, thus seldom considering the applicability of samples to prompts. Although some recent works [20–22] have attempted to generate prompts using contextual information, they often overlook how samples influence prompts, thus focusing instead on how prompts contribute to instances. Such methods usually apply the same prompt pattern across all instances, thereby leading to an inability to fully explore the specific associations between instances and prompts, as well as to guide the model to learn knowledge and information most relevant to the task.

To effectively address the issue of insufficient relevance between instances and prompts in the field of natural language processing, this paper proposes an innovative solution. We introduce a gated mechanism at the core, the gated multilayer perceptron (gMLP) [23], to capture and refine the relevance between instances and prompts precisely. Through this mechanism, we can calculate a relevance score that is used to dynamically adjust the influence of prompts on instances. This not only strengthens the association between instances and prompts but also achieves effective control of the information flow, thereby enhancing the precision and efficiency of the processing.

Furthermore, this study ingeniously integrates strategies from prompt learning and contrastive learning. Prompt learning stimulates the model’s sensitivity to specific tasks by designing appropriate prompts, while contrastive learning enhances the model’s discriminative power by comparing differences between various instances. This combination not only improves the model’s ability to capture subtle differences but also enhances its understanding of complex relationships.

The main contributions of this paper are as follows:

- We propose a novel stance detection model that combines the advantages of prompt learning and contrastive learning, thus enabling effective stance detection in zero-shot scenarios.
- We introduce a gating mechanism that can dynamically adjust the influence of the gate on prompt tokens based on the semantic features of the instance, thereby enhancing the relevance between instances and prompts.
- We conducted experiments on two benchmark datasets, VAST and SEM16, and the results demonstrate that our model outperforms existing state-of-the-art methods on both datasets.

## 2. Related Work

### 2.1. Zero-Shot Stance Detection

Early research on zero-shot stance detection methods largely focused on stance detection within a set of targets, i.e., detection tasks where the training and test sets share the same targets [24]. Crosstarget stance detection is a task similar to zero-shot stance detection, in which a classifier trained on a known target is used to predict stances on data

for an unknown target [25]. Existing crosstarget stance detection studies typically utilize models based on attention mechanisms [26] or graph networks [27], thereby learning target-associated features from the training set's targets and then applying them to predict test sets that are closely related to the target dataset. Unlike crosstarget stance detection tasks, zero/few-shot stance detection aims to automatically determine the stance outcomes for various unknown target data. Under this task requirement, Conforti et al. [28] constructed a large-scale expert-annotated stance detection dataset, where the test set's targets were invisible relative to the training set. Allaway et al. [5] built a zero-shot stance detection dataset with a wide range of topics, thus covering a broad spectrum of related topic categories. Based on this dataset, Allaway et al. [5] proposed a topic grouping attention model to capture the relationship between targets and general topic representations, but they used a fixed BERT model without further fine-tuning, which significantly limited the model's performance. In another study, Allway et al. [7] applied a dataset for intratarget stance detection to zero-shot stance detection and employed adversarial learning to extract sample-independent transferable features. However, it required a large amount of unlabeled data from the target, which is not feasible for zero-shot stance detection tasks. Liu et al. [8] introduced relevant commonsense knowledge from both structural and semantic perspectives, thereby proposing a commonsense-enhanced graph model based on BERT to address zero/few-shot stance detection tasks, but they overlooked the relationships between targets. Liang et al. [29] solved this problem using a joint contrastive learning framework and conducting contrastive learning from both context-aware and target-aware perspectives, but their focus was on the contrast between classes, thus ignoring the connections between targets within the same class.

## 2.2. Prompt Learning

Prompt learning is commonly defined as a method that transforms downstream learning tasks into text generation tasks by incorporating prompt information into the text input. Petroni F. et al. [30] introduced the LAMA dataset to test language models' comprehension of factual and commonsense knowledge. This dataset comprises a set of data sources, each containing a set of facts, which could be in the form of triples or answer pairs. Brown T. et al. [12] created manually crafted prefix prompts for various tasks, including question answering, machine translation, and commonsense reasoning. These prefix prompts demonstrated strong performance across many NLP tasks and benchmarks in zero-shot, one-shot, and few-shot settings. Schick T. et al. [13] targeted text classification and conditional text generation tasks by converting original texts into a "cloze" format using predefined templates, thereby aiding language models in understanding downstream tasks, which is especially challenging in small sample learning settings with only a few samples. Jiang Z. et al. [31], in the MINE method, adopted a mining approach to automatically discover templates from texts containing input  $x$  and output  $y$ . This method scrapes data from text corpora (like Wikipedia) and then looks for dependency paths between inputs and outputs. Yuan W. et al. [32] used phrase replacements from a thesaurus to translate prompts back and forth between different languages. Wang et al. [33] utilized conceptual knowledge as prompts, thereby enabling models to more effectively understand the nuances of the text and achieving heightened classification accuracy in zero-shot scenarios. Zhu et al. [34] integrated soft knowledge into the prompt tuning process; this strategy markedly improved the model's grasp of short text contexts, thereby substantially enhancing text classification performance. Goswami et al. [35] introduced a novel lightweight prompt-based method that adapts language models trained on broad domain datasets to various low-resource fields. This method employs domain-specific keywords and trainable gated prompts, thus providing targeted guidance for the intended domain. These studies demonstrate that model guidance through prompt content modification is effective; however, most techniques employ identical prompts across instances, thereby neglecting the specific relationships between instances and prompts. Motivated by these approaches, our study introduces a mechanism using a gated multilayer perceptron (gMLP) to dynamically

adjust the impact of prompts on instances, thereby significantly optimizing the model's performance in stance detection tasks for particular instances.

### 3. Methodology

In this section, we introduce the prompt learning method for zero-shot stance detection. We then present the architecture of EZSD-CP, which adds an additional layer between the embedding layer and encoder of pretrained language models (PLMs). This architecture is depicted in Figure 1. The architecture of the EZSD-CP framework mainly consists of six parts: (1) Setting prompt templates to insert prompts between the target and comment text, thus better stimulating the potential of pretrained language. (2) BERT word embeddings, where the target, comment text, and prompt sentences are fed into the BERT model for word embedding to obtain a semantic representation of the text. (3) The gMLP module, which uses gMLP to capture the semantic relevance between instances and prompts and then utilizes this relevance as a gating mechanism to dynamically adjust the influence of prompts on instances. (4) Stance contrastive learning, which performs contrastive learning based on the supervisory signal of stance labels to better generalize stance features and improve the model's generalization ability. (5) Concat is an integration module that fuses the rich semantic vectors provided by BERT word embeddings with context-sensitive prompt tokens obtained through a carefully designed gating mechanism to generate a comprehensive enhanced feature representation. (6) The encoder module, where we use the deep network architecture of BERT to process word embeddings further, thus obtaining vector representations that include deeper contextual relations.

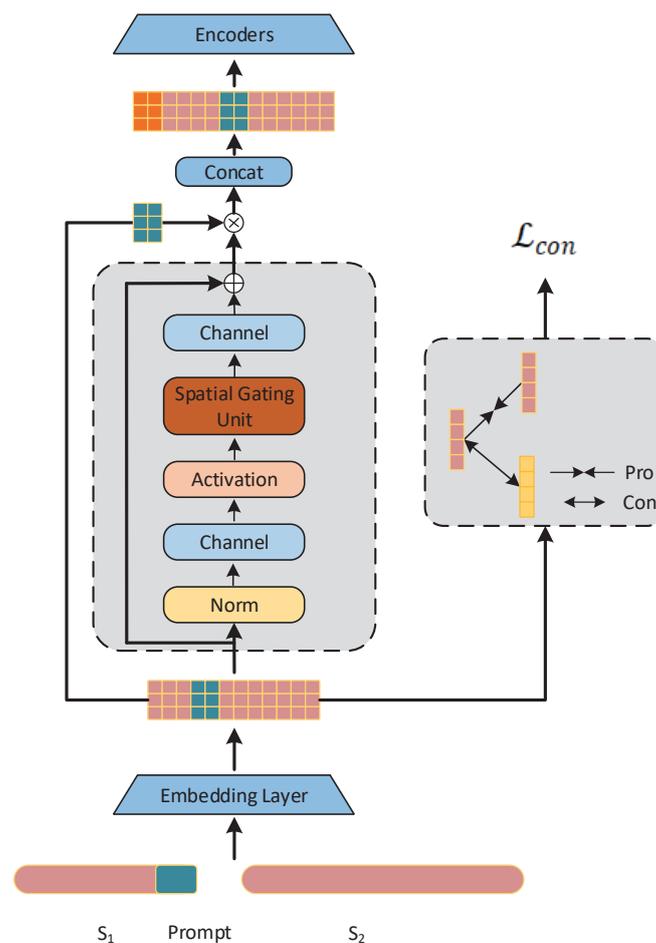


Figure 1. Overall model diagram.

### 3.1. Task Description

Let  $\mathcal{M}$  be a pretrained language model (PLM) with a vocabulary  $\mathcal{V}$ . For a zero-shot stance detection instance  $(s_1, s_2)$ , our goal is to predict the stance of  $s_2$  towards  $s_1$ , where  $s_1$  and  $s_2$  represent the target and comment text, respectively. In prompt learning,  $s_1$  and  $s_2$  are typically placed within a specific pattern consisting of special tokens, text pairs, and external prompt tokens. For example, in our task, the instance  $(s_1, s_2)$  is inserted into a pattern with prompt tokens—[CLS],  $p_1, s_1$ , [MASK],  $p_2, s_2$ , [SEP]—and then  $\mathcal{M}$  is used to select the appropriate word  $w \in \mathcal{V}^*$ , where  $p_1, p_2 \in \mathcal{V}_p$  are prompt tokens, and  $\mathcal{V}^*$  is the set of candidate label words. Finally, the label word  $w \in \mathcal{V}^*$  is mapped onto the actual labels. In our task, the mapping function is “neutral”  $\rightarrow 2$ , “favor”  $\rightarrow 1$ , and “against”  $\rightarrow 0$ .

$$\mathcal{P}(y|(s_1, s_2)) = P(w|\mathcal{M}((s_1, s_2), p)) \quad (1)$$

Here,  $\mathcal{P}$  represents the probability distribution of  $y$  given the input text pair  $(s_1, s_2)$ , where  $p = p_1, p_2, \dots, p_k$ , and  $k$  is the length of the prompt. Generally, prompt learning is divided into two main categories: discrete and continuous. Discrete prompt learning methods search for human-understandable prompt tokens, meaning that the prompt tokens are a subset of the vocabulary of the pretrained language model (PLM). In contrast, continuous prompt learning methods use pseudo tokens in the pattern, which, during training or inference, are projected into differentiable high-dimensional vectors.

### 3.2. Encoding Module

We use BERT as our pretrained language model and use the coding layer in BERT for word embedding of instances  $(s_1, s_2)$  and prompts:

$$E = [E_1; E_p; E_2] = \text{Embed}([s_1, p, s_2]) \quad (2)$$

where  $E \in \mathbb{R}^{L \times d}$  is the input embedding matrix,  $E_1 \in \mathbb{R}^{L_1 \times d}$  and  $E_2 \in \mathbb{R}^{L_2 \times d}$  are the embedding matrices of  $s_1$  and  $s_2$ , respectively,  $E_p \in \mathbb{R}^{k \times d}$  is the embedding matrix of the prompt markers,  $L$  is the sequence length,  $L_1$ ,  $L_2$ , and  $k$  are the lengths of  $s_1$ ,  $s_2$ , and  $p$ , respectively, and  $d$  is the dimension of the embedding.

### 3.3. gMLP Module

In the EZSD-CP model, the process of extracting semantic information from instances primarily focuses on effectively extracting information from multiple tokens that compose the prompt. The gMLP model, with its unique structure, such as the spatial gating unit, efficiently processes this semantic information, thereby particularly excelling in understanding and analyzing complex relationships between tokens, as shown in Figure 2. Consequently, we attempt to utilize gMLP to generate channelwise gating signals.

$$W = \sigma(\text{Dense}(\text{gMLP}(E))) \quad (3)$$

where  $W \in \mathbb{R}^{k \times d}$  is a weight matrix.

The gMLP consists of a stack of  $L$  blocks with the same size and structure. In our model, the input to the gMLP is  $E \in \mathbb{R}^{L \times d}$ . Each block is defined as

$$Z = \sigma(EU), \quad \tilde{Z} = s(Z), \quad Y = \tilde{Z}V \quad (4)$$

where  $\sigma$  is the activation function, and  $U$  and  $V$  define linear projections along the channel dimensions—the same as the FFNs of transformers (e.g., they have shapes of  $768 \times 3072$  and  $3072 \times 768$ ).

One of the key components in the above formulation is  $s(\cdot)$ , which is a layer capturing spatial interactions, as shown in Figure 2. When  $s$  is a constant mapping, the above transformation degenerates into a regular feedforward neural network (FFN) in which individual tokens are processed independently without any communication across tokens.

For gMLP, a major concern is designing an excellent system that captures complex spatial interactions across tokens. Unlike transformers, the model does not require a positional embedding, as this information will be captured in  $s(\cdot)$ .

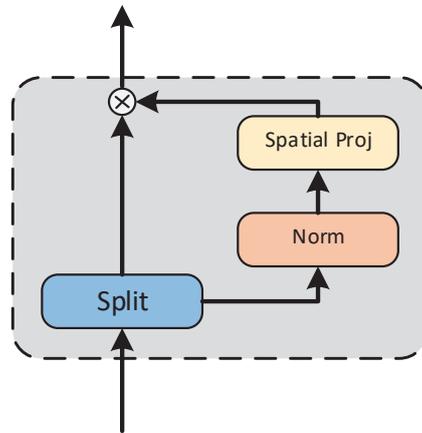


Figure 2. Spatial gating unit.

### 3.4. Stance Contrastive Learning

To enhance the generalization ability of stance learning, Gunel et al. [36] proposed a method that defines stance comparison loss on a hidden vector of examples with supervised stance labeling information. The purpose of this loss function is to capture the similarities between examples within the same category and compare them with examples from other categories. Specifically, given a hidden vector  $\{h_i\}_{i=1}^{N_b}$  in a small batch  $H$  ( $N_b$  is the size of the small batch), take one of the data  $h_i$  as an anchor. Among them,  $h_i, h_j \in H$ . The same label in the same batch is considered a positive pair, i.e.,  $y_i = y_j$ , where  $y_i$  and  $y_j$  are the labels of samples  $h_i$  and  $h_j$ , respectively. Those with different labels in the same batch are considered negative samples, and then the loss of all positive pairs  $(h_i, h_j)$  and  $(h_i, h_j)$  is calculated as

$$\mathcal{L}_{con} = -\frac{1}{N} \sum_{h_i \in H} \ell(h_i). \tag{5}$$

$$\ell(h_i) = \log \frac{\sum_{j \in H \setminus i | [y_i=y_j]} \exp(f(h_i, h_j) / \tau_s)}{\sum_{j \in H \setminus i} \exp(f(h_i, h_j) / \tau_s)}. \tag{6}$$

where  $\|[y_i = y_j] \in \{0, 1\}$  is an indicator function; here, its value is 1,  $f(h_i, h_j)$  is the cosine similarity function for computing  $h_i$ , and  $h_j, f(h_i, h_j) = \text{sim}(u, v) = u^T v / \|u\| \|v\|$ .  $\tau_s$  is the temperature coefficient for comparison learning.

### 3.5. Concat Module

Finally, EZSD-CP multiplies the prompt embedding and gate weights channelwise and concatenates the new prompt embedding  $E'_p$  with  $E_1$  and  $E_2$ .

$$E'_p = \text{Gate}(E_p; E) = W \odot E_p \tag{7}$$

$$E' = [E'_p, E_1, E_2] \tag{8}$$

where  $\odot$  stands for channel multiplication, so the continuous prompted learning method in Equation (1) translates to the EZSD-CP:

$$\mathcal{P}(y|(s_1, s_2)) = P(w|\mathcal{M}(E')) \tag{9}$$

### 3.6. Training

The learning objective of our proposed model is to train the model by uniting a supervised stance detection loss  $\mathcal{L}_{CE}$  and a contrast learning loss  $\mathcal{L}_{con}$ . The total loss consists of the sum of the two losses:

$$\mathcal{L}_{Loss} = \lambda_c \mathcal{L}_{CE} + \lambda_n \mathcal{L}_{con} \quad (10)$$

$\lambda_c$ ,  $\lambda_n$  are tuning hyperparameters, where  $\mathcal{L}_{CE}$  is the crossentropy loss.  $\mathcal{L}_{Loss}$  is calculated as shown in Algorithm 1.

---

#### Algorithm 1 Calculation of the stance contrastive and crossentropy losses

---

```

1: Input:  $s_1, s_2$ 
2: Output:  $\mathcal{L}_{Loss}$ 
3:  $E \leftarrow [E_1; E_p; E_2] \leftarrow BERT\_Embedding(s_1, p, s_2)$  ▷  $p$  is the prompt token
4:  $\mathcal{L}_{con} \leftarrow Contrastive(E)$  ▷ Compute contrastive learning loss
5:  $W \leftarrow \sigma(Dense(gMLP(E)))$  ▷ Get channel gating
6:  $E'_p \leftarrow W \odot E_p$  ▷ Control prompt weights
7:  $E' \leftarrow [E'_p, E_1, E_2]$  ▷ Concatenate processed embeddings with original
8:  $\hat{y} \leftarrow \mathcal{M}[E']$  ▷ Classification probability predicted by PLM
9:  $\mathcal{L}_{CE} \leftarrow CrossEntropy(\hat{y}, y)$  ▷ Compute classification loss
10:  $\mathcal{L}_{Loss} \leftarrow \mathcal{L}_{con} + \mathcal{L}_{CE}$  ▷ Total loss
11: return  $\mathcal{L}_{Loss}$ 

```

---

## 4. Experiments

### 4.1. Datasets and Evaluation Indicators

Our model was evaluated using the zero/few-shot dataset released in 2020 and the SEM16 dataset published in 2016.

The Varied Stance Topics (VAST) [5] is specifically designed for zero/few-shot stance detection and includes comments from the New York Times “Room for Debate” section, thereby covering a wide range of topics. There are over ten thousand data entries comprising more than 6000 targets. The statistics for VAST are shown in Table 1.

SEM16 contains six predefined targets, including Donald Trump (DT), Hillary Clinton (HC), the feminist movement (FM), the legalization of abortion (LA), atheism (A), and climate change (CC). The statistics for SEM16 are shown in Table 2.

Consistent with previous work, we use the macro average of the F1 scores for each target as the evaluation metric. First, the F1 values for the three categories were calculated, and then the average of the F1 values for all categories was taken.

**Table 1.** Detailed statistics of VAST. # denotes “number of” or “count”, indicating the quantity for each category listed.

Statistics	Train	Dev	Test
# Examples	13,477	2062	3006
# Documents	1845	682	786
# Zero-shot Topics	4003	383	600
# Few-shot Topics	638	114	159

**Table 2.** Data statistics for SEM16. DT: Donald Trump, HC: Hillary Clinton, FM: Feminist Movement, LA: Legalization of Abortion, CC: Climate Change is a Real Concern, A: Atheism. # denotes “number of” or “count”, indicating the quantity for each category listed.

Topic	# Ex	# Unlabeled	Keywords
DT	707	2194	trump, Trump
HC	984	1898	hillary, clinton
FM	949	1951	femini
LA	933	1899	aborti
CC	564	1900	climate
A	733	1900	atheism, atheist

#### 4.2. Experimental Implementation

Our experiments were all encoded using case-insensitive BERTbase with a 12-layer transformer encoder, where each word token was mapped to a 768-dimension embedding. We optimized our model using the Adam optimizer, with all dropout rates set to 0.1. learning rates were chosen from  $(1, 2, 3, 4, 5) \times 10^{-5}$ ; the training batch size was set to 8, the step size to 0.1, and the final choice of all hyperparameters was based on the performance on the validation set.  $\lambda c$  and  $\lambda n$  were set to 0.5 and 1, respectively. The learning rates were set to  $1 \times 10^{-5}$ . The median comparative learning loss was chosen from 0.14 to 0.07, both using an A40 graphics card for the experiments.

#### 4.3. Baseline Method

To demonstrate viability, we compared the proposed model with the following state-of-the-art models:

- BERT-joint [5]: Contextual conditional encoding followed by a two-layer feedforward neural network.
- TGA Net [5]: The model using contextual conditional encoding and topic-grouped attention.
- BERT-GCN [8]: The model applies the conventional GCN [37] only considering the node information aggregation.
- CKE-Net [8]: A model based on BERT using the CompGCN [38] to obtain the commonsense information.
- DTCL [39]: The model introduces a latent topic cluster embedding and a discrete latent topic variable to build a bridge between various targets.
- ST-PL [40]: The model designs an agent task framework that combines self-supervised learning and cue learning for automatically identifying and exploiting goal-irrelevant gestural expression features while excluding goal-relevant expression features through a data augmentation strategy.
- JointCL [29]: The model consists of stance contrastive learning and target-aware prototypical graph contrastive learning.

#### 4.4. Main Results

The overall results of our model compared to the baseline are presented in Table 3. To assess the efficacy of our approach across various scenarios, we conducted experiments on the VAST and SEM16 datasets. Our model significantly outperformed all baselines, thereby affirming the effectiveness of our gate mechanism for controlling the influence of prompts on instances and our supervised contrastive learning method. To show our experimental results more clearly, as shown in Figure 3, we compare the experimental results of JointCL and EZSD-CP\_bert in a bar chart.

Specifically, our model's performance on the VAST dataset was two percentage points higher than the ST-PL model. This notable enhancement in performance can be attributed to the combined effect of the adopted gating mechanism strategy and contrastive learning approach. While the ST-PL model, grounded in prompt-based learning, demonstrated commendable capability, our model further refined the dynamic interplay between instances and prompts during the learning process. Moreover, by reinforcing the model's ability to distinguish features via contrastive learning, we achieved even more impressive results in the challenging zero-shot stance detection task, thereby validating the efficacy of our method.

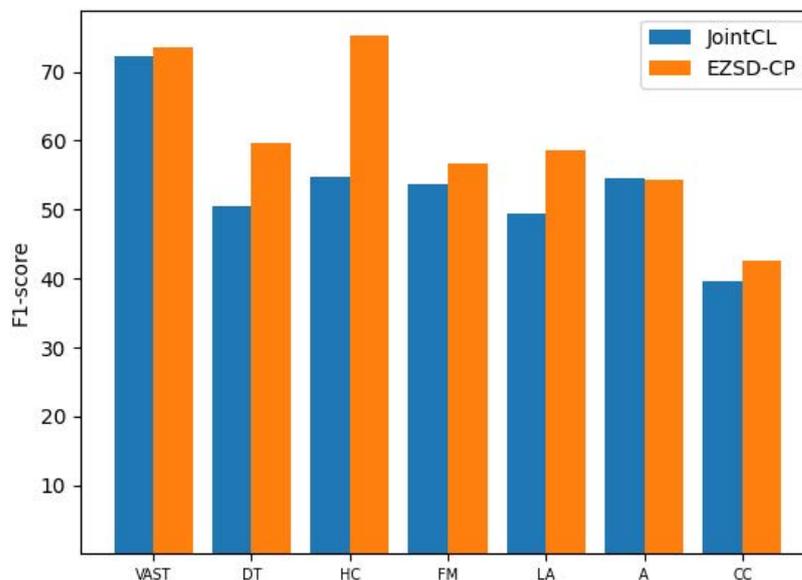
In the realm of zero-shot stance detection, the JointCL model is regarded as the current best practice due to its enhancement of intercategory connections through clustering. Our model surpassed JointCL in performance, thereby highlighting the significance of introducing gating mechanisms and contrastive learning strategies. The CKE-Net model attempts to strengthen the link between targets and texts by integrating the ConceptNet common sense knowledge graph. In contrast, our model, which capitalizes on the potential

of pretrained models through prompt learning, yielded superior outcomes, thus further confirming the effectiveness of prompt-based learning methodologies.

To comprehensively evaluate the generalizability of our model, we experimented with replacing BERT with RoBERTa and compared the results across both datasets. Although the findings on RoBERTa were also encouraging, we noticed a slight decline in performance on the VAST dataset, by 0.6 percentage points compared to BERT. This observation might stem from the multithemed nature of the VAST dataset, where RoBERTa exhibits a finer-grained focus on language comprehension within similar or identical themes. Conversely, on the SEM16 dataset, RoBERTa’s overall performance generally exceeded that of BERT, thus potentially illustrating the inherent advantages of larger-scale pretrained models in zero-shot tasks.

**Table 3.** Experimental results on VAST dataset and SEM16 dataset.

Model	VAST (%)				SEM16 (%)						
	Pro	Con	Neu	All	DT	HC	FM	LA	A	CC	
BERT-joint [5]	54.6	58.4	85.3	66.1	-	-	-	-	-	-	
TGA Net [5]	55.4	58.5	85.8	66.6	41.5	48.7	46.6	45.3	54.2	35.4	
BERT-GCN [8]	58.3	60.6	86.9	68.6	42.3	50.0	44.3	44.2	53.6	35.5	
CKE-Net [8]	61.2	61.2	88.0	70.2	-	-	-	-	-	-	
DTCL [39]	60.0	64.7	87.6	70.8	-	-	-	-	-	-	
ST-PL [40]	-	-	-	-	48.4	53.7	51.2	48.1	52.2	35.2	
JointCL [29]	64.9	63.2	88.9	72.3	50.5	54.8	53.8	49.5	<b>54.5</b>	39.7	
EZSD-CP_bert (ours)	<b>65.4</b>	<b>64.5</b>	<b>90.6</b>	<b>73.6</b>	59.58	75.2	56.7	58.5	54.48	<b>42.5</b>	
EZSD-CP_roberta (ours)	65.2	64.3	89.5	73.0	<b>68.8</b>	<b>76.3</b>	<b>62.2</b>	<b>64.4</b>	54.44	37.3	



**Figure 3.** Comparative bar graphs of experimental results of EZSD-CP and JointCL on VAST and SEM16 datasets.

#### 4.5. Ablation Experiments

In this study, to delve into the role of each component within the EZSD-CP model, we designed a series of ablation experiments to evaluate the contribution of each component, and the corresponding experimental results are detailed in Table 4. The results of the ablation experiments clearly indicate that once the gMLP gating mechanism was removed, the model suffered significant performance losses across all evaluation metrics. This phenomenon strongly underscores the importance of the gMLP gating mechanism, namely

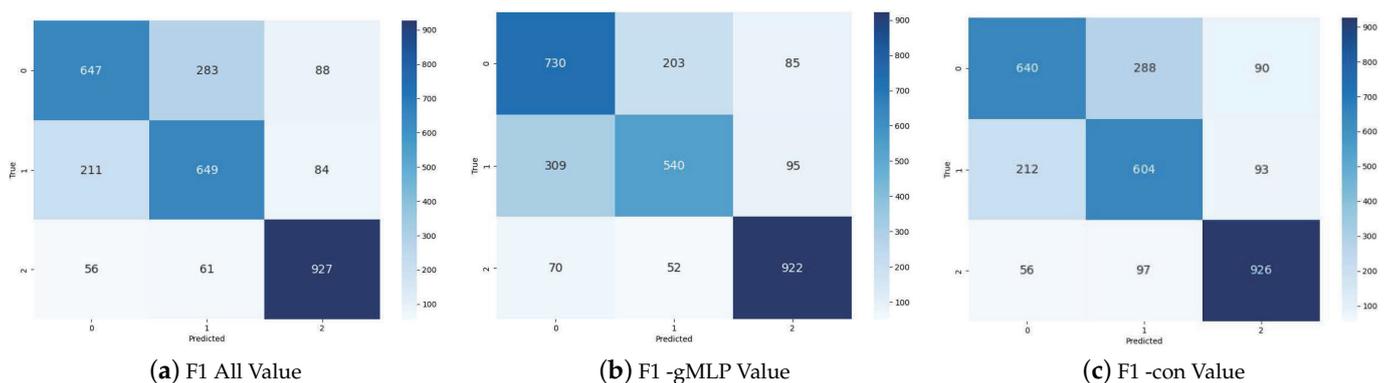
its role in enabling the model to flexibly adjust the response intensity to prompt tokens based on the semantic features of input instances. Furthermore, when we removed the stance contrastive learning (con) component from the model, we observe a significant decrease of nearly six percentage points in the overall performance. This decline reveals the importance of stance contrastive learning within the model, particularly its effectiveness in learning the similarities of stance features within the same category, thereby enhancing the model's generalization ability to similar targets.

**Table 4.** Results of the ablation experiment. The gMLP indicates a gating mechanism, and the con indicates stance contrastive learning.

Model	VAST (%)				SEM16 (%)					
	Pro	Con	Neu	All	DT	HC	FM	LA	A	CC
EZSD-CP (ours)	65.4	64.5	90.6	73.6	59.58	75.2	56.7	58.5	54.48	42.5
gMLP	62.1	68.6	85.9	72.2	56.2	74.0	54.9	56.4	35.7	32.0
con	63.6	66.5	88.3	71.6	53.3	73.2	53.4	55.6	37.1	32.9

#### 4.6. Confounding Experiment

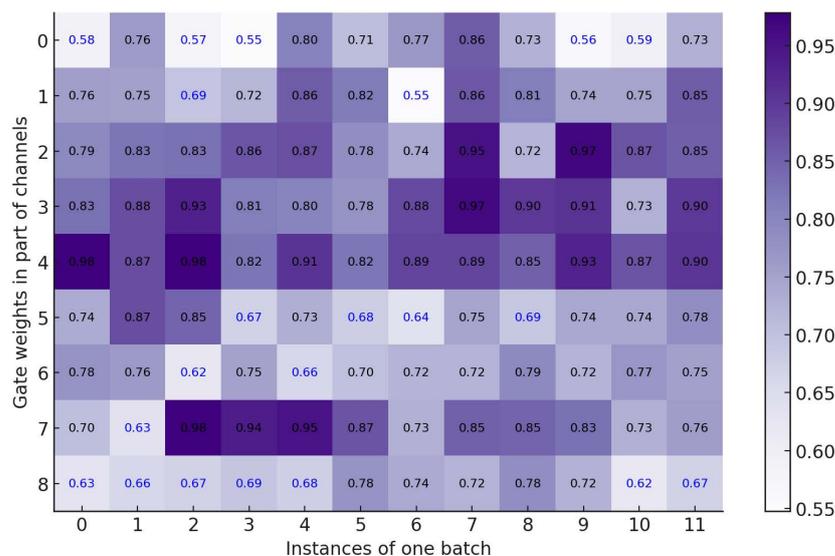
In this experiment, to visually evaluate the model's performance on the test dataset and accurately reveal the model's classification efficacy across different categories, we conducted an in-depth confusion matrix analysis for each newly added component of the model. As shown in Figure 4, after the removal of the gMLP module, the recognition accuracy for categories other than the "against" category experienced a decline. This result powerfully indicates the significant role of the gMLP module in enhancing the overall classification accuracy of the model. Furthermore, by combining the data from Figure 4a,c, it can be observed that the prediction accuracy for all categories decreased after the removal of the con module, with the "favor" label experiencing a significant reduction for 40 cases in its prediction accuracy. This change clearly points out the positive contribution of the con module within the model, especially in improving the recognition capability for the "favor" stance. Therefore, these experimental results not only verify the effectiveness of the gMLP and con modules but also highlight their value in constructing an efficient stance detection system.



**Figure 4.** The confusion matrices over different modules. F1 All denotes our proposed model EZSD-CP, F1 -gMLP denotes the removal of the gMLP module, and F1 -con denotes the removal of the contrast learning module. '0' represents against, '1' indicates favor, and '2' denotes neutral.

#### 4.7. Analysis and Discussion

We meticulously examined how the gating mechanism influenced the EZSD-CP. We randomly selected 12 instances and analyzed the gating signal weights generated on certain channels for different instances. The visualization results recorded in Figure 5 indicate that the gating mechanism can obtain varying gate weights for different instances.



**Figure 5.** Gate weights in different channels.

## 5. Conclusions

In this paper, we elucidate the main problems currently faced by zero-shot stance detection in the context of prompt learning and validate the importance of using a gating mechanism to regulate the influence of prompts on different instances. We proposed an instance-guided prompt learning method, EZSD-CP. EZSD-CP constructs prompts using a weight matrix extracted from instances. Thus, during the training and inference process, the prompts are constrained by the semantic information of the instances. At the same time, contrastive learning was introduced into the model, thereby enabling it to learn more discriminative feature representations. This straightforward approach achieved state-of-the-art performance on both the VAST dataset and the SEM16 dataset.

In the future, we plan to explore a better gating mechanism that can more effectively adjust the influence of prompts on instances based on the semantic information of sentences.

**Author Contributions:** Conceptualization, Z.Y. and W.Y.; methodology, Z.Y.; software, Z.Y.; validation, Z.Y. and F.W.; formal analysis, Z.Y.; investigation, Z.Y.; resources, Z.Y.; data curation, Z.Y.; writing—original draft preparation, Z.Y.; writing—review and editing, Z.Y.; visualization, Z.Y.; supervision, Z.Y.; project administration, W.Y.; funding acquisition, W.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the “Tianshan Talent” Research Project of Xinjiang (No. 2022TSYCLJ0037), the National Natural Science Foundation of China (No. 62262065), the Autonomous Region Science and Technology Program (No. 2022B01008), and the Autonomous Region Science and Technology Program (No. 2020A02001).

**Data Availability Statement:** The VAST dataset is available at <https://github.com/emilyalloway/zero-shot-stance>, accessed on 16 August 2023. The SEM16 dataset is available at <http://alt.qcri.org/semeval2016/task6/>, accessed on 12 September 2023.

**Acknowledgments:** We thank all anonymous reviewers for their constructive comments.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Somasundaran, S.; Wiebe, J. Recognizing Stances in Online Debates. In Proceedings of the ACL 2009, Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the AFNLP, Singapore, 2–7 August 2009; Su, K., Su, J., Wiebe, J., Eds.; The Association for Computer Linguistics: Stroudsburg, PA, USA, 2009; pp. 226–234.

2. Augenstein, I.; Rocktäschel, T.; Vlachos, A.; Bontcheva, K. Stance Detection with Bidirectional Conditional Encoding. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, TX, USA, 1–4 November 2016; Su, J., Carreras, X., Duh, K., Eds.; The Association for Computational Linguistics: Stroudsburg, PA, USA, 2016; pp. 876–885. [[CrossRef](#)]
3. Mohammad, S.M.; Kiritchenko, S.; Sobhani, P.; Zhu, X.; Cherry, C. 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, 16–17 June 2016; Bethard, S., Cer, D.M., Carpuat, M., Jurgens, D., Nakov, P., Zesch, T., Eds.; The Association for Computer Linguistics: Stroudsburg, PA, USA, 2016; pp. 31–41. [[CrossRef](#)]
4. Sobhani, P.; Inkpen, D.; Zhu, X. A Dataset for Multi-Target Stance Detection. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017, Valencia, Spain, 3–7 April 2017; Volume 2: Short Papers; Lapata, M., Blunsom, P., Koller, A., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2017; pp. 551–557. [[CrossRef](#)]
5. Allaway, E.; McKeown, K.R. 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 2020, Online, 16–20 November 2020; Webber, B., Cohn, T., He, Y., Liu, Y., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 8913–8931. [[CrossRef](#)]
6. Liang, B.; Chen, Z.; Gui, L.; He, Y.; Yang, M.; Xu, R. Zero-Shot Stance Detection via Contrastive Learning. In Proceedings of the WWW'22: The ACM Web Conference 2022, Lyon, France, 25–29 April 2022; Laforest, F., Troncy, R., Simperl, E., Agarwal, D., Gionis, A., Herman, I., Médini, L., Eds.; ACM: New York, NY, USA, 2022; pp. 2738–2747. [[CrossRef](#)]
7. Allaway, E.; Srikanth, M.; McKeown, K.R. 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, 6–11 June 2021; Toutanova, K., Rumshisky, A., Zettlemoyer, L., Hakkani-Tür, D., Beltagy, I., Bethard, S., Cotterell, R., Chakraborty, T., Zhou, Y., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2021; pp. 4756–4767. [[CrossRef](#)]
8. Liu, R.; Lin, Z.; Tan, Y.; Wang, W. Enhancing Zero-shot and Few-shot Stance Detection with Commonsense Knowledge Graph. In Proceedings of the Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021, Online Event, 1–6 August 2021; Zong, C., Xia, F., Li, W., Navigli, R., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2021; Volume ACL/IJCNLP 2021, Findings of ACL, pp. 3152–3157. [[CrossRef](#)]
9. Devlin, J.; Chang, M.; Lee, K.; Toutanova, K. 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, 2–7 June 2019; Volume 1 (Long and Short Papers); Burstein, J., Doran, C., Solorio, T., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 4171–4186. [[CrossRef](#)]
10. Radford, A.; Narasimhan, K.; Salimans, T.; Sutskever, I. Improving Language Understanding by Generative Pre-Training. 2018. Available online: <https://gwern.net/doc/www/s3-us-west-2.amazonaws.com/d73fdc5ffa8627bce44dcda2fc012da638ffb158.pdf> (accessed on 3 March 2024).
11. Liu, P.; Yuan, W.; Fu, J.; Jiang, Z.; Hayashi, H.; Neubig, G. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Comput. Surv.* **2023**, *55*, 1–35. [[CrossRef](#)]
12. Brown, T.B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language Models are Few-Shot Learners. In Proceedings of the Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, Virtual, 6–12 December 2020.
13. Schick, T.; Schütze, H. 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, 19–23 April 2021; Merlo, P., Tiedemann, J., Tsarfaty, R., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2021; pp. 255–269. [[CrossRef](#)]
14. Gao, T.; Fisch, A.; Chen, D. Making Pre-trained Language Models Better Few-shot Learners. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, 1–6 August 2021; Zong, C., Xia, F., Li, W., Navigli, R., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2021; pp. 3816–3830. [[CrossRef](#)]
15. Shin, T.; Razeghi, Y.; Logan, R.L., IV; Wallace, E.; Singh, S. AutoPrompt: Eliciting Knowledge from Language Models with Automatically Generated Prompts. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, 16–20 November 2020; Webber, B., Cohn, T., He, Y., Liu, Y., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 4222–4235. [[CrossRef](#)]
16. Zhong, Z.; Friedman, D.; Chen, D. Factual Probing Is [MASK]: Learning vs. Learning to Recall. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021, Online, 6–11 June 2021; Toutanova, K., Rumshisky, A., Zettlemoyer, L., Hakkani-Tür, D., Beltagy, I., Bethard, S., Cotterell, R., Chakraborty, T., Zhou, Y., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2021; pp. 5017–5033. [[CrossRef](#)]

17. Gu, Y.; Han, X.; Liu, Z.; Huang, M. PPT: Pre-trained Prompt Tuning for Few-shot Learning. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Dublin, Ireland, 22–27 May 2022; pp. 8410–8423.
18. Li, X.L.; Liang, P. Prefix-Tuning: Optimizing Continuous Prompts for Generation. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, 1–6 August 2021; Zong, C., Xia, F., Li, W., Navigli, R., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2021; pp. 4582–4597. [[CrossRef](#)]
19. Liu, X.; Ji, K.; Fu, Y.; Tam, W.; Du, Z.; Yang, Z.; Tang, J. P-Tuning: Prompt Tuning Can Be Comparable to Fine-tuning Across Scales and Tasks. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Dublin, Ireland, 22–27 May 2022; pp. 61–68.
20. Zhou, Y.; Maharjan, S.; Liu, B. Scalable Prompt Generation for Semi-supervised Learning with Language Models. In Proceedings of the Findings of the Association for Computational Linguistics: EACL 2023, Dubrovnik, Croatia, 2–6 May 2023; Vlachos, A., Augenstein, I., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2023; pp. 758–769. [[CrossRef](#)]
21. Deng, B.; Wang, W.; Feng, F.; Deng, Y.; Wang, Q.; He, X. Attack Prompt Generation for Red Teaming and Defending Large Language Models. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2023, Singapore, 6–10 December 2023; Bouamor, H., Pino, J., Bali, K., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2023; pp. 2176–2189.
22. Gao, J.; Xiang, L.; Wu, H.; Zhao, H.; Tong, Y.; He, Z. An Adaptive Prompt Generation Framework for Task-oriented Dialogue System. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2023, Singapore, 6–10 December 2023; Bouamor, H., Pino, J., Bali, K., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2023; pp. 1078–1089.
23. Liu, H.; Dai, Z.; So, D.R.; Le, Q.V. Pay Attention to MLPs. In Proceedings of the Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, Virtual, 6–14 December 2021; pp. 9204–9215.
24. Du, J.; Xu, R.; He, Y.; Gui, L. Stance classification with target-specific neural attention networks. In Proceedings of the International Joint Conferences on Artificial Intelligence, Melbourne, Australia, 19–25 August 2017.
25. Liang, B.; Fu, Y.; Gui, L.; Yang, M.; Du, J.; He, Y.; Xu, R. Target-adaptive Graph for Cross-target Stance Detection. In Proceedings of the WWW'21: The Web Conference 2021, Ljubljana, Slovenia, 19–23 April 2021. pp. 3453–3464. [[CrossRef](#)]
26. Xu, C.; Paris, C.; Nepal, S.; Sparks, R. 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, 15–20 July 2018; Volume 2: Short Papers; Gurevych, I., Miyao, Y., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2018; pp. 778–783. [[CrossRef](#)]
27. Zhang, B.; Yang, M.; Li, X.; Ye, Y.; Xu, X.; Dai, K. 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, 5–10 July 2020; Jurafsky, D., Chai, J., Schluter, N., Tetreault, J.R., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 3188–3197. [[CrossRef](#)]
28. Conforti, C.; Berndt, J.; Pilehvar, M.T.; Giannitsarou, C.; Toxvaerd, F.; Collier, N. 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, ACL 2020, Online, 5–10 July 2020; Jurafsky, D., Chai, J., Schluter, N., Tetreault, J.R., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 1715–1724. [[CrossRef](#)]
29. Liang, B.; Zhu, Q.; Li, X.; Yang, M.; Gui, L.; He, Y.; Xu, R. JointCL: A Joint Contrastive Learning Framework for Zero-Shot Stance Detection. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2022, Dublin, Ireland, 22–27 May 2022; Muresan, S., Nakov, P., Villavicencio, A., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2022; pp. 81–91. [[CrossRef](#)]
30. Petroni, F.; Rocktäschel, T.; Riedel, S.; Lewis, P.S.H.; Bakhtin, A.; Wu, Y.; Miller, A.H. Language Models as Knowledge Bases? 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, 3–7 November 2019; Inui, K., Jiang, J., Ng, V., Wan, X., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 2463–2473. [[CrossRef](#)]
31. Jiang, Z.; Xu, F.F.; Araki, J.; Neubig, G. How Can We Know What Language Models Know. *Trans. Assoc. Comput. Linguist.* **2020**, *8*, 423–438. [\\_A\\_00324](#) [[CrossRef](#)]
32. Yuan, W.; Neubig, G.; Liu, P. BARTScore: Evaluating Generated Text as Text Generation. In Proceedings of the Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, Virtual, 6–14 December 2021. pp. 27263–27277.
33. Wang, Y.; Wang, W.; Chen, Q.; Huang, K.; Nguyen, A.; De, S. Prompt-based Zero-shot Text Classification with Conceptual Knowledge. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, ACL 2023, Toronto, ON, Canada, 9–14 July 2023; Padmakumar, V., Vallejjo, G., Fu, Y., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2023; pp. 30–38. [[CrossRef](#)]
34. Zhu, Y.; Wang, Y.; Mu, J.; Li, Y.; Qiang, J.; Yuan, Y.; Wu, X. Short text classification with Soft Knowledgeable Prompt-tuning. *Expert Syst. Appl.* **2024**, *246*, 123248. [[CrossRef](#)]

35. Goswami, K.; Lange, L.; Araki, J.; Adel, H. SwitchPrompt: Learning Domain-Specific Gated Soft Prompts for Classification in Low-Resource Domains. In Proceedings of the Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2023, Dubrovnik, Croatia, 2–6 May 2023; Vlachos, A., Augenstein, I., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2023; pp. 2681–2687. [[CrossRef](#)]
36. Gunel, B.; Du, J.; Conneau, A.; Stoyanov, V. Supervised Contrastive Learning for Pre-trained Language Model Fine-tuning. In Proceedings of the 9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, 3–7 May 2021.
37. Kipf, T.N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. In Proceedings of the 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, 24–26 April 2017.
38. Vashishth, S.; Sanyal, S.; Nitin, V.; Talukdar, P.P. Composition-based Multi-Relational Graph Convolutional Networks. In Proceedings of the 8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, 26–30 April 2020.
39. Liu, R.; Lin, Z.; Fu, P.; Liu, Y.; Wang, W. Connecting Targets via Latent Topics And Contrastive Learning: A Unified Framework For Robust Zero-Shot and Few-Shot Stance Detection. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2022, Virtual and Singapore, 23–27 May 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 7812–7816. [[CrossRef](#)]
40. Chen, Z.; Liang, B.; Xu, R. A Topic-based Prompt Learning Method for Zero-Shot Stance Detection. In Proceedings of the 21st Chinese National Conference on Computational Linguistics, Nanchang, China, 14–16 October 2022; pp. 535–544.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.