

Article

Leveraging Chain-of-Thought to Enhance Stance Detection with Prompt-Tuning

Daijun Ding ¹, Xianghua Fu ², Xiaojiao Peng ², Xiaomao Fan ², Hu Huang ^{3,*} and Bowen Zhang ^{2,*}¹ College of Applied Science, Shenzhen University, Shenzhen 518052, China; dingdaijun2021@email.szu.edu.cn or ding_dai_jun@outlook.com² College of Big Data and Internet, Shenzhen Technology University, Shenzhen 518118, China; fuxianghua@sztu.edu.cn (X.F.); pengxiaojiao@sztu.edu.cn (X.P.); fanxiaoma@sztu.edu.cn (X.F.)³ Shenzhen Graduate School, Peking University, Shenzhen 518055, China

* Correspondence: h.huang@pku.edu.cn (H.H.); zhangbowen@sztu.edu.cn or zhang_bo_wen@foxmail.com (B.Z.)

Abstract: Investigating public attitudes towards social media is crucial for opinion mining systems to gain valuable insights. Stance detection, which aims to discern the attitude expressed in an opinionated text towards a specific target, is a fundamental task in opinion mining. Conventional approaches mainly focus on sentence-level classification techniques. Recent research has shown that the integration of background knowledge can significantly improve stance detection performance. Despite the significant improvement achieved by knowledge-enhanced methods, applying these techniques in real-world scenarios remains challenging for several reasons. Firstly, existing methods often require the use of complex attention mechanisms to filter out noise and extract relevant background knowledge, which involves significant annotation efforts. Secondly, knowledge fusion mechanisms typically rely on fine-tuning, which can introduce a gap between the pre-training phase of pre-trained language models (PLMs) and the downstream stance detection tasks, leading to the poor prediction accuracy of the PLMs. To address these limitations, we propose a novel prompt-based stance detection method that leverages the knowledge acquired using the chain-of-thought method, which we refer to as PSDCOT. The proposed approach consists of two stages. The first stage is knowledge extraction, where instruction questions are constructed to elicit background knowledge from a VLPLM. The second stage is the multi-prompt learning network (M-PLN) for knowledge fusion, which learns model performance based on the background knowledge and the prompt learning framework. We evaluated the performance of PSDCOT on publicly available benchmark datasets to assess its effectiveness in improving stance detection performance. The results demonstrate that the proposed method achieves state-of-the-art results in in-domain, cross-target, and zero-shot learning settings.

Keywords: stance detection; prompt-tuning; chain-of-thought**MSC:** 68T50

Citation: Ding, D.; Fu, X.; Peng, X.; Fan, X.; Huang, H.; Zhang, B. Leveraging Chain-of-Thought to Enhance Stance Detection with Prompt-Tuning. *Mathematics* **2024**, *12*, 568. <https://doi.org/10.3390/math12040568>

Academic Editor: Faheim Sufi

Received: 4 January 2024

Revised: 5 February 2024

Accepted: 8 February 2024

Published: 13 February 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/>).

1. Introduction

Stance detection is a fundamental task in natural language processing (NLP), where the goal is to classify attitudes towards a particular target given opinionated input texts [1]. This task has gained significant attention in recent years due to its importance in various applications, such as political analysis, social media monitoring, and customer feedback analysis. In its early stages, research on stance detection was primarily centered on online debates that adhere to a standardized sentence structure, and where the user's attitude is typically straightforwardly expressed [2,3]. However, with the rapid growth of the internet, social media platforms such as X have become more popular, and researchers have started exploring the mining of social media for stance detection [4,5].

Conventional methods for stance detection can be viewed as target-based sentence-level classification tasks, which can be classified into non-pretrained and pretrained language models (PLMs). Non-pretrained models employ deep neural networks (DNNs), such as long short-term memory (LSTM), attention-based models (Att), and graph convolutional networks (GCN), to build stance classification models. For instance, Du et al. [6] employed an attention-based approach that leverages target-specific information. Dey et al. [7] employed two independent LSTMs to sieve non-neutral text and classify attitudes separately. Sun et al. [8] devised a hierarchical attention mechanism that learned text representation by utilizing linguistic features. Liang et al. [9] introduced an effective GCN-based approach that distinguished between target-invariant and target-specific features. Inspired by the recent success of PLMs, fine-tuning techniques have been developed to improve the accuracy of stance detection [10]. Fine-tuning techniques have been developed to adapt pre-trained language models (PLMs) to specific tasks. One such technique involves constructing a stance classification head at the top of the special token denoted as “<cls>” and fine-tuning the entire model accordingly. In stance detection, the model is exposed to many input text–stance label pairs during fine-tuning. The <cls> token learns semantic and syntactic patterns that correlate with different stances. In sum, these methods typically regard stance detection as a target-oriented sentence-level text classification task. Nevertheless, the efficacy of social media data analysis methods is impeded by the sparsity problem arising from the concise and informal expressions commonly encountered on these platforms. Such content typically lacks context, details, or elaboration and often incorporates abbreviations and slang for which pre-trained language models (PLMs) lack corresponding background knowledge. Consequently, PLMs struggle to comprehend the semantics conveyed in the text, leading to erroneous judgments.

Recently, some pioneering studies have been conducted to address the sparsity problem by utilizing external knowledge to enhance the performance and interpretability of stance detection. For example, He et al. [11] improved the performance of text classifiers by introducing target-related Wikipedia documents as content supplements. Diaz et al. [12] constructed a stance tree by retrieving external knowledge from a knowledge base and used it as evidence to support stance prediction, thus enhancing the accuracy of stance detection. Zhang et al. [13] utilize external knowledge from semantic and emotion lexicons as a bridge to enable knowledge transfer across different targets. Nasiri et al. [14] addressed the issue of a lack of annotated datasets in Persian pose detection tasks through data augmentation and transfer learning. Hardalov et al. [15] proposed a novel semi-supervised approach to address the issue of scarce data in cross-language scenarios. Khiabani et al. [16] enhanced stance detection performance in low-shot cross-target scenarios through multi-modal embeddings derived from both textual and network features of the data. Although these works have achieved improvements in the performance and interpretability, these methods still face the following challenges in practical applications: (1) Most existing methods require the design of complex attention mechanisms to filter out noise and extract task-related background knowledge. However, such methods require a large number of annotated samples, which is clearly time-consuming and labor-intensive. In order to ease the applicability of knowledge-enhanced stance detection, it would be highly desirable to develop knowledge-acquired algorithms that are less dependent on feature engineering and with high-quality task-related background knowledge. (2) Most of these knowledge fusion mechanisms rely on fine-tuning models. However, the fine-tuning approach creates a gap between the pre-training phase of PLM and downstream stance detection tasks, resulting in the reduced prediction accuracy of the PLMs.

To tackle the challenges mentioned above, in this paper, we propose a prompt-based stance detection method by leveraging the knowledge acquired by the chain-of-thought method (PSDCOT). The proposed model is motivated by two considerations. First, the advancements in large language models, such as GPT-3.5, etc., have demonstrated their powerful knowledge generation capabilities, and COT methods can effectively mine knowledge from these models to support prediction with evidence. Second, prompt learning

methods can improve prediction performance by fitting the downstream task to the upstream training process. The proposed PSDCOT consists of two stages. The first stage is knowledge extraction, where instruction questions are constructed to elicit background knowledge from a VLPLM. The second stage is the multi-prompt learning network (M-PLN) for knowledge fusion, which learns model performance based on the background knowledge and the prompt learning framework. Extensive tests were conducted on publicly available benchmark datasets to evaluate the performance of the proposed PSDCOT method. The results demonstrate that the proposed method effectively improves stance detection performance, achieving state-of-the-art results in in-domain, cross-target, and zero-shot learning settings.

In summary, this paper presents several significant contributions:

- A PSDCOT framework is proposed, which improves prompt-based stance detection models by incorporating the background knowledge into prediction.
- A M-PLN approach is proposed, which enhances the performance of stance detection by carefully constructing prompt templates and synthesizing analysis from multiple perspectives.
- Extensive experiments are carried out on publicly available benchmark datasets, and the results demonstrate the superiority of the proposed model over state-of-the-art competitors. The source code of this paper is available at <https://github.com/SzuDdj/PSDCOT> accessed on 3 January 2024.

The present paper is structured as follows. Section 2 offers an overview of related research, including traditional and recent methods of stance detection, as well as methods for prompt tuning. Section 3 outlines the details of our innovative method. Section 4 presents the findings of our experimental analysis. The paper concludes with Section 5, which summarizes our key findings.

2. Related Work

2.1. Stance Detection

The aim of stance detection is to identify and analyze the perspective of a given text regarding a particular target [17,18]. (1) In the scenario of an in-target setting, conventional techniques can be broadly categorized as non-pretrained and pretrained methods. Deep neural networks, say Att and GCN, are commonly utilized by non-pretrained techniques for the purpose of training stance classifiers. Att methods focus primarily on target-specific data as the attention query, and employ an attention mechanism to obtain the stance polarity [6–8,19]. The GCN methods introduce a graph convolutional network to model the correlation between the target and the text [20–22]. (2) Various methods have focused on cross-target stance detection tasks, which can be broadly classified into two categories. The first category is related to word-level transfer, which makes use of shared words between two targets as a means to bridge knowledge gaps [23]. The second category addresses cross-target issues by utilizing concept-level knowledge that is common between two targets [13,24,25]. (3) Zero-shot stance detection is a particularly challenging scenario, where a trained stance detection model is required to infer the stance of an unseen target. In response to this challenge, Allaway and McKeown [26] have developed a large-scale dataset for stance detection that has been labeled by human annotators. The dataset is specifically designed for the zero-shot scenario. Moreover, Allaway et al. [27] have utilized adversarial learning in order to extract target-invariance information and have used a stance detection dataset that is specific to the target to conduct zero-shot stance detection. Liu et al. [10] have proposed a graph model that incorporates both intra- and extra-semantic information, in addition to common sense knowledge based on BERT. This approach is aimed at enhancing the semantic information obtained. Additionally, Liang et al. [9] have introduced a robust method for detecting target-specific or target-invariant features to help acquire transferable stance features.

2.2. Background Knowledge Enhanced Stance

The use of background knowledge to enhance the performance of stance detection has garnered attention as an effective approach to improving performance [28]. For instance, He et al. [11] introduced target-related background knowledge, such as Wikipedia knowledge, and proposed a fine-tuning learning method to improve the model's learning ability. Similarly, Luo et al. [10] constructed background knowledge as a knowledge graph and utilized graph neural network methods to develop a stance predictor. Additionally, Huang et al. [29] introduced the use of *#hashtag* background knowledge to improve content learning. Furthermore, Luo et al. [30] incorporated sentiment knowledge to better learn attitudes.

2.3. Prompt-Tuning Methods

Prompt tuning has gained widespread popularity in diverse natural language processing (NLP) domains, for example, text classification [31], natural language understanding [32], and sentiment analysis [33]. The verbalizer plays a critical role in prompt tuning and significantly impacts its effectiveness [34]. The methods for designing verbalizers can be categorized into human-designed and automatic verbalizers. Human-designed verbalizers rely primarily on the personal expertise of the creator and may lack sufficient coverage [32]. Automatic verbalizers are designed using search methods, but they require a significant number of training and validation sets to optimize [35]. Previous studies on prompt-based models have concentrated on stance detection [15,36]. Jiang et al. [36] presented TAPD, a prompt-tuning framework designed for stance detection. TAPD utilizes a verbalizer that maps labels to hidden vectors to facilitate label prediction. Likewise, Hardalov et al. [15] developed a prompt-based approach for cross-lingual stance detection. Furthermore, Huang et al. [29] proposed the use of SenticNet to construct an atomic verbalizer. In conclusion, the prompt learning framework has shown remarkable progress in detecting stances.

3. Our Methodology

To represent the labeled dataset, we utilize $X = \{x_i, q_i\}_{i=1}$, where x and q , respectively, denote the input text and the corresponding target. Each (x, q) pair in X is assigned a stance label y . The objective of stance detection is to infer a stance label for the input sentence in the context of a given target q .

3.1. Model Overview

As illustrated in Figure 1, our PSDCOT consists of the chain-of-thought module for knowledge extraction (KE) and a multi-Prompt learning Network (M-PLN) two main components. Here, KE aims to extract the external knowledge for enhancing stance detection via COT methods. In M-PLN, we design an attention-based network for background knowledge integration for stance detection.

3.2. Knowledge Extraction

To elicit background knowledge effectively, we design the chain-of-thought prompt method. The proposed approach is motivated by the observation that the emerging capabilities of very large models enable them to generatively generate an understanding as background knowledge. Therefore, we aim to leverage the background knowledge of a large model to enhance the performance of stance detection.

Specifically, we propose the step-by-step question-answering strategy to elicit knowledge. Such a method teaches language models to solve the stance detection by providing a one-shot example. First, we construct the question-answer pair (QAP), and then feed the constructed question into the VLPLM and acquire the explanation of the reason for prediction. For example, given the following input: "RT GunnJessica: Because i want young American women to be able to be proud of the 1st woman president #SemST", the question for ChatGPT input is as follows: "What is the attitude of the sentence: "RT GunnJessica: Because i

want young American women to be able to be proud of the 1st woman president #SemST” to the target “Hillary Clinton” select from “favor, against or neutral”. For this particular example, ChatGPT returns a correct result. Second, we further inquire as to why the model predicts a certain stance polarity. As shown in Table 1, large language models have the capability to fill in missing information in sentences, such as subjects, and decipher hashtags.

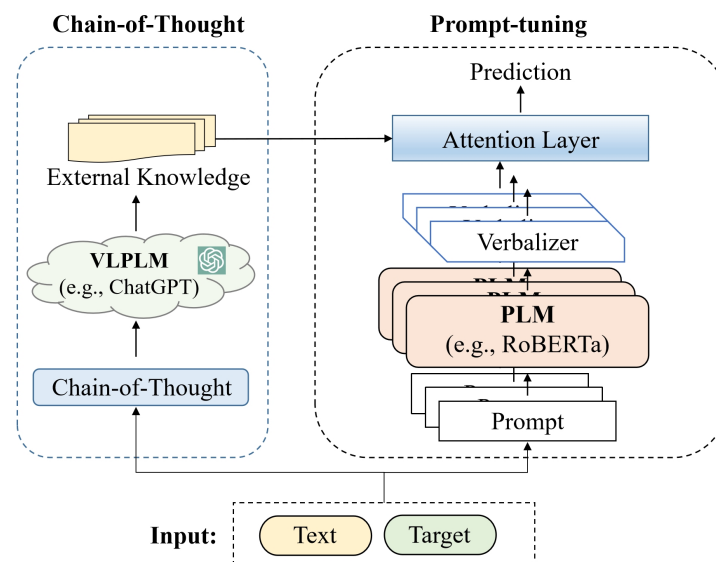


Figure 1. The overall structure of the proposed PSDCOT.

Table 1. Example of ChatGPT Output.

Tweet and Target	Background Knowledge
Tweet: KoriWasinger JButtawipo she also voted for the Iraq war and was responsible for gathering TheDemocrats votes to support it. Target: Hillary Clinton [Against]	The sentence mentions Clinton’s role in gathering votes from Democrats to support the war suggests that the speaker believes Clinton bears responsibility for the decision to go to war and holds her accountable for it. [Against]
Tweet: You know you’re in a patriarchy when women are the ones who are blamed for prostitution #whatisjustice #inequality. Target: Feminist Movement [Favor]	The sentence seems to express a belief that women are unfairly blamed for engaging in prostitution and suggests that the society in which this occurs is a patriarchy, which is a system that favors men and oppresses women. The use of hashtags such as #whatisjustice and #inequality also suggests that the speaker is supportive of the feminist movement and its goals of promoting gender equality. [Favor]
Tweet: Let’s agree that it’s not ok to kill a 7lbs baby in the uterus DWStweets #DNC #Clinton2016 HillaryforIA #ProCompromise Target: Legalization of Abortion [Against]	The use of hashtags such as “#DNC” and “#Clinton2016” suggests a negative sentiment towards the legalization of abortion, and the mention of it being “not ok to kill a 7lbs baby in the uterus” may be seen as a criticism of the idea that some lives are more valuable than others. [Against]

3.3. Multi-Prompt Learning Network (M-PLN)

Preliminary: Prompt-tuning with PLM. Prompt-tuning is a transformative approach that reframes the stance detection task as a masked language modeling task. Specifically, prompt-tuning methodology adopts a text template p which is thoughtfully incorporated into the given text x and the target q . For example, to classify sentence x as being in favor or against, the prompt-tuning process envelops the sentence x with a predefined text template to yield a novel text representation $x_p = \text{“We should support this. The attitude to the } \langle \text{Target } q \rangle \text{ is [MASK].”}$ Let M be the pre-trained language model, which provides the probability of each word v in the vocabulary being filled in [MASK] given $P_M([\text{MASK}] = v | x_p)$.

In this context, v represents the defined label word in the verbalizer. To map the probabilities of these words to the probabilities of the labels, a verbalizer is utilized as a mapping function f from the defined words in the vocabulary, which form the label word set V , to the label space Y , i.e., $f : V \rightarrow Y$. Formally, the probability $P(y|x_p)$ of label y , is computed as follows:

$$P(y|x_p) = \mu(P_M([MASK] = v|x_p)|v \in y). \quad (1)$$

where μ serves as a crucial component in transforming the probability distribution over label words to the probability distribution over labels. To illustrate, in the aforementioned example, prompt tuning can set V_1 to represent the words “support” and “agree”, and V_2 to represent the word “opposition”. Additionally, μ can be defined as an identity function. The instance is then categorized under the *favor* class if the average likelihood of the terms in V_1 exceeds that of the terms in V_2 . In prompt tuning, the objective of learning is to minimize

$$l(y|x_p) = -\log P_M([MASK] = v|x_p). \quad (2)$$

Prompt Design. The key to the prompt-based method for stance detection is to construct the appropriate prompt. Previous research has demonstrated that the performance of different prompts varies significantly, and this issue is further compounded for stance detection. The expressions and topics exhibit a wide range of diversity among distinct target groups, thereby rendering the formulation of a universal prompt for the entirety of these targets infeasible. To account for this heterogeneity, our approach involves the creation of multiple prompts derived from varied perspectives. Based on prior research, we have addressed stance detection by considering not only sentiment polarity in text, but also stance-aware words and target–text relations. Therefore, we design prompt templates from three perspectives, as shown by T_1 , T_2 , and T_3 , respectively.

$T_1 = \langle \text{TEXT} \rangle$. According to the sentiment expression in the text, the stance polarity is [MASK].

$T_2 = \langle \text{TEXT} \rangle$. The stance polarity of this text is [MASK].

$T_3 = \langle \text{TEXT} \rangle$. The stance polarity toward the $\langle \text{Target} \rangle$ is [MASK].

We employ three RoBERTa models as our pre-trained language model (PLM). The [MASK] and [SEP] tokens are sourced directly from the RoBERTa vocabulary. Our prompts are easily customizable for the pre-training tasks of other PLMs.

Target-aware Verbalizer. In prompt-based fine-tuning, a verbalizer, which is an injective function $f: Y \rightarrow V$, is typically defined to map each label to a single token from the PLM’s vocabulary. The efficacy of prompt-based methods heavily relies on the design of the verbalizer, and a straightforward approach of assigning a fixed concrete word to each label may not result in optimal outcomes. To address this issue, previous studies, such as that of Schick et al. [32], have suggested mapping each label to a phrase that can better represent the semantic meaning of the label, e.g., using “in favor of” instead of “favor”. However, predicting consecutive [mask] tokens poses a new challenge. In an effort to tackle this problem, Gao et al. [34] proposed generating the verbalized word for each label via a pruned set of the top- k vocabulary words that are highly probable according to the PLM. Nonetheless, this approach involves a computationally demanding and time-consuming brute-force search for each label. Moreover, given the wide array of expressions used across various targets, we assert that a single phrase or token may be inadequate for capturing the stance information. To tackle this concern, we utilize a novel solution that involves the mapping of labels onto continuous vectors, called stance vectors, instead of explicit words or phrases. These vectors are amenable to be trained during optimization. Our modality revolves around the generation of three distinct vectors that correspond to the ones generated by the [MASK] of diverse templates. The stance vectors from T_1 , T_2 , and T_3 are $V_{T,1}$, $V_{T,2}$, and $V_{T,3}$, respectively. To ensure coherence with token embeddings within the PLM, the stance vectors have been dimensionally aligned with the size of said embeddings.

Attention Layer. The attention layer is proposed to integrate background knowledge with the prompt-based model. Specifically, we utilize three soft stance vectors, $V_{T,1}$, $V_{T,2}$, and $V_{T,3}$, as three queries to guide the attention in an iterative manner. The hidden state of the attention mechanism is acquired by feeding the background knowledge into the independent PLM. Here, the hidden state is denoted as H . By computing the attention queries and hidden states, the coupling coefficient matrix k can be computed as follows:

$$\begin{aligned} k_{T,1}^1 &= V_{T,1}^1 (H_{T,1}^1)^\top \\ k_{T,2}^1 &= V_{T,2}^1 (H_{T,2}^1)^\top \\ k_{T,3}^1 &= V_{T,3}^1 (H_{T,3}^1)^\top \end{aligned} \quad (3)$$

where $k_{\{T,1;T,2;T,3\}} \in \mathbb{R}^{n \times n}$. Then, the query of next iteration V^2 can be updated as follows:

$$\begin{aligned} V_{T,1}^2 &= k_{T,1}^1 H_{T,1}^1 \\ V_{T,2}^2 &= k_{T,2}^1 H_{T,2}^1 \\ V_{T,3}^2 &= k_{T,3}^1 H_{T,3}^1 \end{aligned} \quad (4)$$

where the dimension of the new query V^2 is the same as that of the initial query V^1 . Subsequently, the input of the next iteration can be updated by

$$\begin{aligned} H_{T,1}^2 &= \lambda \text{LayerNorm}(\sigma(V_{T,1}^2)) + H_{T,1}^1 \\ H_{T,2}^2 &= \lambda \text{LayerNorm}(\sigma(V_{T,2}^2)) + H_{T,2}^1 \\ H_{T,3}^2 &= \lambda \text{LayerNorm}(\sigma(V_{T,3}^2)) + H_{T,3}^1 \end{aligned} \quad (5)$$

where *LayerNorm* performs the standard layer normalization.

After t iterations, the output hidden state e can be found as follows:

$$\begin{aligned} \mathbf{q}_{T,1} &= H_{T,1}^t (\text{softmax} \sum_i k_{T,1}^t), \\ \mathbf{q}_{T,2} &= H_{T,2}^t (\text{softmax} \sum_i k_{T,2}^t), \\ \mathbf{q}_{T,3} &= H_{T,3}^t (\text{softmax} \sum_i k_{T,3}^t), \\ e &= \text{avg}(\mathbf{q}_{T,1} + \mathbf{q}_{T,2} + \mathbf{q}_{T,3}) \end{aligned} \quad (6)$$

where $\text{softmax}(f_i) = \frac{e^{f_i}}{\sum_j e^{f_j}}$, $\oplus$ is the concatenation operator.

The detailed process is presented in Algorithm 1.

Algorithm 1 PSDCOT

Input: $X = \{x_i^{\text{train}}, q_i^{\text{train}}\}_{i=1}^{N_{\text{train}}}$, T_1 , T_2 , T_3 , V

Output: e

- 1: Utilize 1-shot example COT to teach language models and acquire background knowledge from X .
 - 2: Initialize $V_{T,1}$, $V_{T,2}$, $V_{T,3}$
 - 3: **for** t in T iterations **do**
 - 4: Obtain coupling coefficients: $k_{T,1}^{t-1}$, $k_{T,2}^{t-1}$, $k_{T,3}^{t-1}$
 - 5: Update queries: $V_{T,1}^t$, $V_{T,2}^t$, $V_{T,3}^t$
 - 6: Update queries: $H_{T,1}^t$, $H_{T,2}^t$, $H_{T,3}^t$
 - 7: **end for**
 - 8: Obtain the $q_{T,1}^t$, $q_{T,2}^t$, $q_{T,3}^t$
 - 9: Obtain the e
 - 10: **return** e
-

3.4. Stance Classification

We classify the stance expressed in the text by assessing the semantic similarity between the target-aware stance vectors and the average of the label vector (which is defined in the verbalizer). To integrate the background knowledge, we concatenate the representation of background knowledge e with the target-aware stance vectors to enhance the stance detection performance, which is denoted as follows:

$$\gamma = e \oplus \frac{1}{m} \sum_{i=1}^m V_{T,i}^1 \quad (7)$$

Based on the words provided by the verbalizer, we calculate the probability of selecting token v as the label word.

$$\delta = \frac{\exp(v_i \cdot \gamma)}{\sum_{v_j \in V} \exp(v_j \cdot \gamma)}, \quad (8)$$

where v is the embedding of the token in verbalizer. Then, we sum the words' probabilities of each label, which is denoted as $\hat{y}$.

Finally, the loss function can be effectively implemented through the utilization of the standard cross-entropy method:

$$\mathcal{L} = - \sum_{i=1}^N \sum_{j=1}^C y_{ij} \log \hat{y}_{ij}, \quad (9)$$

Here, N denotes the magnitude of the training set and C denotes the number of stance classes. Every ground-truth label, y_i , pertaining to the i -th individual sample, is represented in the one-shot format. To optimize the attention layer, the standard method of the gradient descent algorithm is employed.

4. Experiments

4.1. Experimental Data

This paper presents experimental results on robust benchmark datasets, including SemEval-2016 Task 6 (SemEval2016) [37], P-stance [38], ISD [29], and VAST [26].

- **SemEval-2016.** The dataset of SemEval2016 comprises 4870 tweets with diverse targets, with each tweet being labeled with a label of “favor”, “against”, or “neutral”. In accordance with the proposed configuration by [24], four targets—Donald Trump (D), Hillary Clinton (H), Legalization of Abortion (L), and Feminist Movement (F)—are deemed appropriate for evaluating the efficacy of the stance detection task, and hence have been chosen for our study. Specifically, for the cross-target setup [9,13,24], we construct eight cross-target stance detection tasks ($D \rightarrow H$, $H \rightarrow D$, $F \rightarrow L$, $L \rightarrow F$), where the source target is represented by the left side of the arrow, and the destination target is represented by the right side.
- **P-stance.** To enhance the data volume for performance evaluation, the P-stance dataset comprises 21,574 tweets, targeting “Donald Trump (DT_p)”, “Joe Biden (JB_p)”, and “Bernie Sanders (BS_p)”. For cross-domain setup, we construct six settings: $DT \rightarrow JB$, $DT \rightarrow BS$, $JB \rightarrow DT$, $JB \rightarrow BS$, $BS \rightarrow DT$, and $BS \rightarrow JB$.
- **VAST.** The VAST dataset, as presented by Allaway and Mckeown [26], encompasses a diverse range of targets that span across various themes, such as politics, education, and public health. The dataset comprises three distinct stance labels, with the label set being defined as “Pro”, “Neutral”, and “Con”. The training set comprises 4003 samples, while the dev and test sets consist of 383 and 600 samples, respectively. As per Liang et al. [9], we evaluate our model's performance on zero-shot topics.
- **ISD.** The ISD dataset, proposed by Huang et al. [29], poses a challenge as it consists of texts without explicit sentiment words. Therefore, for predicting stance polarity, it

is crucial to comprehend the interplay between the text and contextual knowledge, including knowledge of the target and hashtags. The target of ISD are “Donald Trump (DT_i)” and “Joe Biden (JB_i)”.

4.2. Compared Baseline Methods

In order to assess the efficacy of our proposed model, we conducted a thorough evaluation and comparison with a range of established baselines. The details of these baseline models are presented below for reference:

Statistics-based methods:

- **BiLSTM** [23]. The BiLSTM methodology utilizes a bidirectional Long Short-Term Memory (LSTM) network to encode the underlying sentence and the corresponding target independently.
- **MemNet** [39]. The MemNet architecture embraces a memory network, enhanced with a multi-hop attention mechanism, to effectively encode textual data.
- **AOA** [40]. The AOA model employs two Long Short-Term Memory (LSTM) networks to model the target and context separately, and incorporates an interactive attention mechanism for modeling their interrelation.
- **ASGCN** [41]. The ASGCN approach leverages a dependency tree for modeling dependencies and leverages Graph Convolutional Networks (GCN) to learn compact and expressive text representations.
- **TAN** [6]. The TAN model introduces target-specific attention in conjunction with a Long-Short Term Memory for the task of stance detection.
- **TPDG** [42]. The TPDG model presents a novel solution for stance detection through the utilization of a target-adaptive graph convolutional network. The proposed framework integrates shared features from analogous targets, thereby enhancing the model’s effectiveness in accurately delineating the stance towards a given target.
- **AT-JSS-Lex** [43]. The AT-JSS-Lex model suggests a target-adaptive graph convolutional network for the purpose of stance detection. This mechanism draws inspiration from the practice of utilizing common features from analogous targets.
- **TOAD** [27]. The TOAD uses adversarial learning to generalize across topics.

Fine-tuning based methods:

- **RoBERTa-FT** [44]. These methods employ a pretrained BERT or RoBERTa model for stance detection, with the given context and target converted to the format of “[CLS] + text + [SEP] + target + [SEP]” to adapt to the training and fine-tuning of the model.
- **PT-HCL** [9]. The PT-HCL model presents a novel approach to cross-target and zero-shot stance detection using contrastive learning. To achieve this, the model leverages a BERT-based architecture to establish a shared representation space for diverse targets.

Prompt-tuning based methods:

- **MPT**. MPT has devised a prompt-tuning based PLM for stance detection, which employs a verbalizer defined by human experts.
- **AutoPT** [31]. AutoPT introduced an innovative approach for stance detection, which involves the generation of label words derived from the given data corpus via an auto-prompt method.
- **KPT** [35]. KPT introduced external lexicons to define the verbalizer for the prompt framework.

Knowledge-enhanced methods:

- **WS-BERT-Dual** [27]. The WS-BERT-Dual introduces target-related wiki knowledge to enhance stance detection ability.

4.3. Implementation Details

In our experimental setup, we opted for pre-trained language models utilizing RoBERTa-large architecture. For training the model, we employed the Adam optimizer while using a mini-batch size of 32 and a learning rate of 0.0002. To advance the current state-of-the-art, we detail, in a comprehensive manner, the templates leveraged to stimulate pre-trained language models throughout this paper.

As per the recommendations of previous works [9,13], we employ the micro-average F1 score as our primary evaluation metric. Our first step in this process involves calculating the F1 scores for each of the categories, namely “Favor” and “Against”:

$$\begin{aligned} F1_{favor} &= \frac{2P_{favor}R_{favor}}{P_{favor} + R_{favor}} \\ F1_{against} &= \frac{2P_{against}R_{against}}{P_{against} + R_{against}} \end{aligned} \quad (10)$$

The F1-score can be computed based on P and R , which, respectively, stand for precision and recall.

$$F1_{avg} = \frac{F1_{favor} + F1_{against}}{2} \quad (11)$$

Second, because the targets in the dataset are unbalanced, we compute the micro-F1 as another evaluation metric:

$$F1_m = \frac{2TP}{2TP + FP + FN} \quad (12)$$

4.4. Overall Performance

4.4.1. In-Domain Setup

The results of in-domain stance detection using different robust benchmarks are presented in Tables 2 and 3. Based on the obtained outcomes, several conclusions can be drawn. (1) The pretrained models exhibit a remarkable enhancement in the performance of stance detection for most configurations when compared to statistic-based methods. For instance, RoBERTa-FT demonstrates an average improvement of 10.3% in comparison with the top-performing statistic-based method (TPDG) on the ISD dataset. This finding provides further validation on the effectiveness of utilizing pretrained models in stance detection. (2) Prompt-based PLM methods exhibit consistent improvement in multiple tasks when compared to fine-tuning PLM. For instance, PSDCOT achieves a 10.35% improvement in $F1_{avg}$ and 8.15% in $F1_m$ on average in ISD datasets, in contrast to RoBERTa-FT. This result indicates that the utilization of a prompt framework can significantly enhance the effectiveness of PLMs at tapping into their true capabilities. (3) The utilization of external knowledge serves as an indispensable factor in the completion of stance detection assignments in social media. By incorporating external knowledge into the procedure, a noteworthy enhancement in the performance of stance detection is observed. For example, after integrating background knowledge of the target, WS-BERT-Dual improves by 2.35% in $F1_{avg}$ on average of ISD and P-stance datasets compared with RoBERTa-FT. (4) The PSDCOT method proposed in this paper surpasses all the established baselines across a majority of the evaluation tasks. Our experimental results demonstrate a notable improvement of 11.86% in $F1_{avg}$ over the most effective neural network-based model (TPDG), 5.24% in $F1_{avg}$, and 4.72% in $F1_m$ over the top-performing fine-tuned PLM model (RoBERTa-FT), and 2.7% in $F1_{avg}$ and 3.28% in $F1_m$ over the best-performing prompt-tuning approach (KPT), when averaging across seven distinct tasks. Furthermore, when compared to the current state-of-the-art external knowledge augmentation technique (WS-BERT-Dual), the PSDCOT method also achieves an average improvement of 6.35% in $F1_{avg}$ on both the ISD and P-stance datasets. The advantage of PSDCOT comes from its two characteristics: (i) We propose a COT method to extract the background knowledge behind text and targets. This

knowledge can effectively improve the performance of position detection from the results.
(ii) We propose a multi-prompt learning network, which can effectively fuse background knowledge with the predictor.

Table 2. Performance comparison on $F1_{avg}$. The results with † are retrieved from [13]; ‡ are retrieved from [36]. The $\mathbb{I}$ mark refers to a p -value < 0.05 . The best scores are in bold. Note that, to evaluate the stability of the model, following [13], we evaluated the stability of our proposed PSDCOT by running the method three times and reporting the average score.

No.	Embedding	Methods	SemEval2016			ISD		P-Stance	
			F	L	H	DT_i	JB_i	DT	JB
1	Statistic.	BiLSTM †	51.6	59.1	55.8	28.6	35.0	69.7	68.6
2		BiCond †	52.9	61.2	56.1	55.2	50.5	70.6	68.4
3		TAN †	55.8	63.7	65.4	50.4	52.5	77.1	77.6
4		AT-JSS-Lex ‡	61.5	68.4	68.3	-	-	-	-
5		MemNet †	51.1	58.9	52.3	53.5	52.2	76.8	77.2
6		AoA †	55.4	58.3	51.6	55.9	57.6	77.2	77.6
7		ASGCN †	56.2	59.5	62.2	55.1	58.2	76.8	78.2
8		TPDG	67.3	74.7	73.4	64.2	60.0	76.8	78.1
9	PLM	RoBERTa-FT	72.3	67.6	81.5	72.4	72.4	88.1	86.5
10		WS-BERT-Dual	-	-	-	76.4	76.2	88.2	88.0
11		MPT	73.3	71.4	81.3	72.8	72.1	86.0	87.0
12		AutoP	72.8	72.6	81.4	71.2	70.8	86.3	86.4
13		KPT	75.2	74.2	82.6	75.7	74.4	88.4	88.1
14		PSDCOT	75.4 $\mathbb{I}$	73.4	83.9 $\mathbb{I}$	83.3 $\mathbb{I}$	82.2 $\mathbb{I}$	89.7 $\mathbb{I}$	89.6 $\mathbb{I}$

Table 3. Performance comparison of stance detection ($F1_m$). The $\mathbb{I}$ mark refers to a p -value < 0.05 . The best scores are in bold.

	Methods	F	L	H	DT_i	JB_i
Statistic.	BiLSTM	60.9	56.49	68.2	43.6	35.4
	BiCond	65.3	56.64	67.6	55.3	50.6
	CrossNet	71.7	59.95	68.2	43.2	43.7
PLM	RoBERTa-FT	76.6	74.8	77.3	72.8	70.9
	MPT	77.1	76.1	76.7	67.2	70.1
	KPT	78.1	77.0	80.0	70.2	74.3
	PSDCOT	79.3 $\mathbb{I}$	77.3 $\mathbb{I}$	79.4	79.9 $\mathbb{I}$	80.1 $\mathbb{I}$

4.4.2. Cross-Target Setup

Obtaining a vast dataset that has been adequately annotated demands a substantial investment of time and resources. Hence, our proposal is to examine the efficacy of our approach within a cross-target framework. The objective of the cross-target framework is to predict the stance of the target destination by leveraging labeled data from the source target. The results of SenEval-2016 and P-Stance are reported in Tables 4–6. Based on the results, our proposed method outperforms the other baselines by a significant margin. Specifically, compared with the previous promising statistical method (TPDG), PSDCOT achieves an average improvement of 16.15% in $F1_{avg}$ and 17.43% in $F1_m$ on average, which confirms the effectiveness of utilizing a prompt-tuning framework in cross-target setup. In comparison to fine-tuning based methods (e.g., RoBERTa-FT), PSDCOT achieves an average improvement of 9.73% in $F1_{avg}$ and 6.6% in $F1_m$. These results further emphasize the crucial role of using the knowledge-enhanced network in cross-target stance detection. Furthermore, PSDCOT achieves superior stability compared to KPT and MPT. For instance, PSDCOT achieves an average improvement of 1.45% in $F1_{avg}$ and 2.78% in $F1_m$ over MPT, 0.82% in $F1_{avg}$, and 0.95% in $F1_m$ KPT, respectively, across all four setups from Tables 4 and 5.

Table 4. Performance comparison of cross-target stance detection ($F1_{avg}$) on SemEval-2016. The results with † are retrieved from [13]. The ‡ mark refers to a p -value < 0.05. The best scores are in bold.

Embedding	Methods	F→L	L→F	H→D	D→H
Statistic.	BiLSTM †	44.8	41.2	29.8	35.8
	BiCond †	45.0	41.6	29.7	35.8
	CrossNet †	45.4	43.3	43.1	36.2
	VTN	47.3	47.8	47.9	36.4
	SEKT †	53.6	51.3	47.7	42.0
	TPDG	58.3	54.1	50.4	52.9
PLM	RoBERTa-FT	49.1	55.3	65.9	71.1
	MPT	63.4	68.9	68.2	74.0
	KPT	65.0	68.5	69.3	74.2
	PSDCOT	66.7 ‡	68.4	70.3 ‡	74.9 ‡

Table 5. Performance comparison of cross-target stance detection ($F1_m$) on SemEval-2016. The results with † are retrieved from [13]. The ‡ mark refers to a p -value < 0.05. The best scores are in bold.

	Methods	F→L	L→F	H→D	D→H
Statistic.	BiLSTM †	35.4	34.6	56.8	44.4
	BiCond †	35.6	36.8	58.7	45.8
	CrossNet †	43.0	42.9	49.1	47.4
	SEKT †	51.0	50.7	44.9	44.4
PLM	TPDG	66.5	57.7	51.6	49.1
	RoBERTa-FT †	58.7	61.1	66.5	72.7
	MPT †	67.2	68.1	65.4	73.6
	KPT †	70.2	69.5	65.1	76.8
	PSDCOT	71.9 ‡	69.8 ‡	67.9 ‡	75.8

Table 6. Performance comparison of cross-target stance detection ($F1_{avg}$) on P-Stance. The ‡ mark refers to a p -value < 0.05. The best scores are in bold.

Model	DT→JB	DT→BS	JB→DT	JB→BS	BS→DT	BS→JB
BiCond	55.8	51.8	58.2	60.2	51.4	57.7
CrossNet	56.7	50.1	60.4	60.8	53.0	62.6
BERT	58.8	56.3	63.6	67.0	58.8	73.0
WS-BERT-Dual	68.3	64.4	67.7	69.0	63.6	76.8
PSDCOT	80.2 ‡	75.4 ‡	83.2 ‡	77.2 ‡	78.2 ‡	80.9 ‡

4.4.3. Zero-Shot Stance Detection

In certain scenarios, the target of a particular text may be absent from the training dataset; therefore, we compare it against the most advanced competitors in the field. The results of our experiments are reported in Table 7. Notably, due to the inherent limitations and challenges of zero-shot stance detection, all methods underperform in comparison to the in-target and cross-target setups. In particular, methods that focus only on statistics, without leveraging external background knowledge, perform poorly. On the other hand, approaches based on fine-tuning, such as PT-HCL, BERT-FT, and RoBERTa-FT, consistently outperform statistical-based methods. This outcome validates the remarkable benefits of leveraging knowledge acquired from a vast corpus. Despite the challenging nature of zero-shot stance detection, our PSDCOT model exhibits considerable potential, surpassing all benchmark approaches on the VAST dataset. Consequently, our findings imply that PSDCOT represents a promising strategy for addressing the demanding task of zero-shot stance detection by effectively incorporating background knowledge and adopting a prompt-tuning framework.

Table 7. Performance comparison of zero-shot stance detection. The $\P$ mark refers to a p -value < 0.05 . The best scores are in bold.

	Model	VAST	D	H	F	L
1	BiCond	42.8	30.5	32.7	40.6	34.4
2	CrossNet	43.4	35.6	38.3	41.7	38.5
3	SEKT	41.8	42.7	44.1	46.0	45.3
4	TPDG	51.9	47.3	50.9	53.6	46.5
5	TOAD	41.0	49.5	51.2	54.1	46.2
6	BERT	66.1	40.1	49.6	41.9	44.8
7	RoBERTa	69.7	-	-	-	-
8	TGA-NET	66.6	40.7	49.3	46.6	45.2
9	BERT-GCN	68.6	42.3	50.0	44.3	44.2
10	CKE-Net	70.2	-	-	-	-
11	PT-HCL	71.6	50.1	54.5	54.6	50.9
12	COT	68.9	71.6	78.9	68.7	61.5
13	Ours	74.7 $\P$	71.6	79.1 $\P$	69.3 $\P$	64.2 $\P$

4.5. Ablation Study

In order to investigate the influence of each individual component in the PSDCOT method, we conducted an ablation test by removing the proposed component, which is denoted as “w/o”.

The variants of PSDCOT are as follows:

- **w/o P:** PSDCOT without the prompt-tuning framework; instead, we use a standard fine-tuning strategy. Specifically, the stance vector is the hidden state of the <cls> vector.
- **w/o COT:** We discard the Knowledge Extraction and use commonsense knowledge as the background knowledge following [11].
- **w/o t:** We removed the multi-template and only kept the commonly used single-view T_3 template.

The findings of our ablation study are depicted in Figure 2. Our results indicate that the chain-of-thought (COT) method significantly enhances the performance of our PSDCOT method. Specifically, we found that the removal of background knowledge acquired by the COT approach led to a substantial deterioration in performance. This observation underscores the importance of leveraging external knowledge-enriched models to facilitate a deeper understanding of the given stance. Furthermore, our study reveals that fine-tuning leads to a considerable drop in performance when compared to prompt-tuning. This finding highlights the efficacy of prompt learning in bridging the gap between large model pre-training and downstream tasks, such as stance detection and its capability to improve performance. Notably, our analysis exhibited that the best performance was achieved by combining all the aforementioned factors across all the experiments.

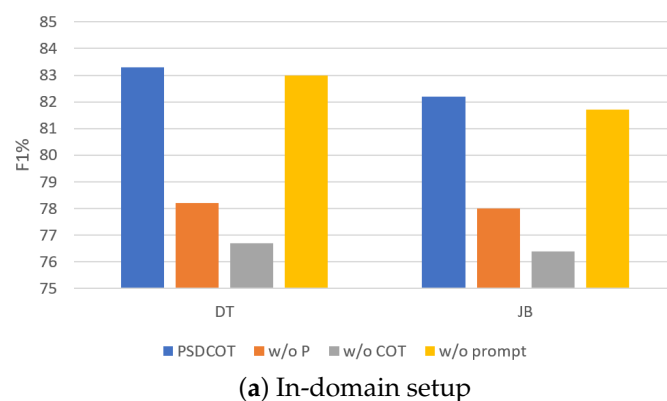


Figure 2. Cont.

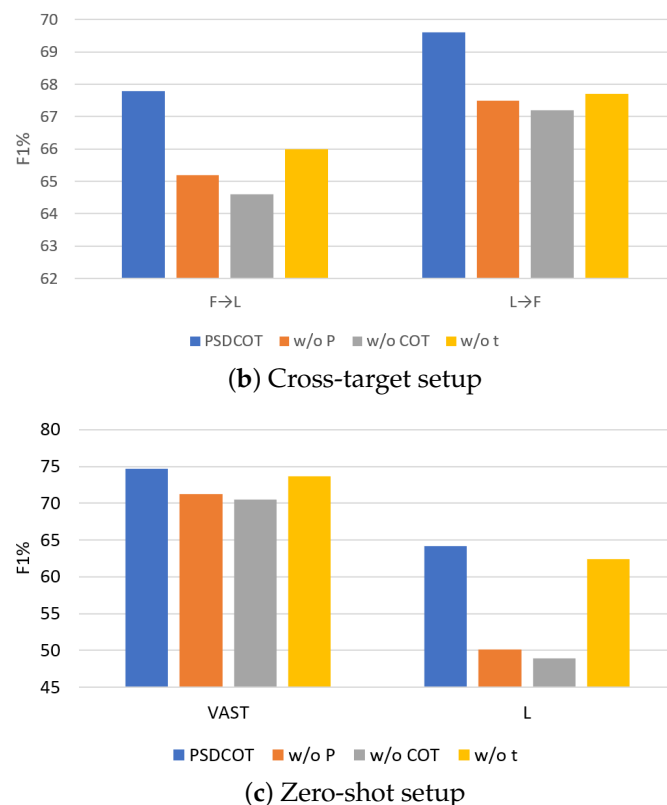


Figure 2. $F1_{avg}$ of the ablation test.

5. Conclusions

In this paper, we propose a novel prompt-based stance detection approach, referred to as PSDCOT, which utilizes a chain-of-thought method to elicit knowledge and fuses knowledge through a multi-prompt learning network. The experimental results demonstrated that PSDCOT achieves state-of-the-art performance in in-target, cross-target, and zero-shot settings. In future work, we plan to elicit and prune knowledge from Large Language Models (LLMs) to enhance background knowledge accuracy and eliminate irrelevant information. Furthermore, we may dedicate efforts in constructing virtual text contexts to alleviate the challenge of social media data sparsity.

Author Contributions: Conceptualization, H.H.; Methodology, D.D. and B.Z.; Software, X.F. (Xiao-mao Fan); Validation, B.Z.; Formal analysis, X.F. (Xianghua Fu); Resources, X.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by the Natural Science Foundation of Top Talent of SZTU 452 (grant no. GDRC202320) and the Research Promotion Project of Key Construction Discipline 453 in Guangdong Province (2022ZDJS112).

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The author declare no conflict of interest.

References

1. Küçük, D.; Can, F. Stance detection: A survey. *ACM Comput. Surv. CSUR* **2020**, *53*, 1–37. [\[CrossRef\]](#)
2. Walker, M.A.; Anand, P.; Abbott, R.; Grant, R. Stance classification using dialogic properties of persuasion. In Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, Montreal, QC, Canada, 3–8 June 2012; pp. 592–596.
3. Somasundaran, S.; Wiebe, J. Recognizing stances in online debates. In Proceedings of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, Association for Computational Linguistics, Singapore, 2–7 August 2009; pp. 226–234.

4. Yang, M.; Zhao, W.; Chen, L.; Qu, Q.; Zhao, Z.; Shen, Y. Investigating the transferring capability of capsule networks for text classification. *Neural Netw.* **2019**, *118*, 247–261. [[CrossRef](#)] [[PubMed](#)]
5. Zhang, Y.; Tiwari, P.; Song, D.; Mao, X.; Wang, P.; Li, X.; Pandey, H.M. Learning interaction dynamics with an interactive LSTM for conversational sentiment analysis. *Neural Netw.* **2021**, *133*, 40–56. [[CrossRef](#)] [[PubMed](#)]
6. 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.
7. Dey, K.; Shrivastava, R.; Kaushik, S. Topical Stance Detection for Twitter: A Two-Phase LSTM Model Using Attention. In Proceedings of the European Conference on Information Retrieval, Grenoble, France, 27–28 March 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 529–536.
8. Sun, Q.; Wang, Z.; Zhu, Q.; Zhou, G. Stance detection with hierarchical attention network. In Proceedings of the 27th International Conference on Computational Linguistics, Santa Fe, NM, USA, 20–26 August 2018; pp. 2399–2409.
9. Liang, B.; Chen, Z.; Gui, L.; He, Y.; Yang, M.; Xu, R. Zero-Shot Stance Detection via Contrastive Learning. In Proceedings of the ACM Web Conference 2022, Lyon, France, 25–29 April 2022; pp. 2738–2747.
10. 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, Bangkok, Thailand, 1–6 August 2021; pp. 3152–3157.
11. He, Z.; Mokherberian, N.; Lerman, K. Infusing Wikipedia Knowledge to Enhance Stance Detection. *arXiv* **2022**, arXiv:2204.03839.
12. Diaz, G.A.; Chesñevar, C.I.; Estevez, E.; Maguitman, A. Stance Trees: A Novel Approach for Assessing Politically Polarized Issues in Twitter. In Proceedings of the 15th International Conference on Theory and Practice of Electronic Governance, Guimaraes, Portugal, 4–6 October 2022; pp. 19–24.
13. 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, Online, 5–10 July 2020; pp. 3188–3197.
14. Nasiri, H.; Analoui, M. Persian stance detection with transfer learning and data augmentation. In Proceedings of the 2022 27th International Computer Conference, Computer Society of Iran (CSICC), Tehran, Iran, 23–24 February 2022; IEEE: New York, NY, USA, 2022; pp. 1–5.
15. Hardalov, M.; Arora, A.; Nakov, P.; Augenstein, I. Few-shot cross-lingual stance detection with sentiment-based pre-training. In Proceedings of the AAAI Conference on Artificial Intelligence, Online, 22 February–1 March 2022; Volume 36, pp. 10729–10737.
16. Khiabani, P.J.; Zubiaga, A. Few-shot learning for cross-target stance detection by aggregating multimodal embeddings. *IEEE Trans. Comput. Soc. Syst.* **2023**. [[CrossRef](#)]
17. Jain, R.; Jain, D.K.; Dharana, S.; Sharma, N. Fake News Classification: A Quantitative Research Description. *ACM Trans. Asian Low Resour. Lang. Inf. Process.* **2022**, *21*, 3. [[CrossRef](#)]
18. Rani, S.; Kumar, P. Aspect-based Sentiment Analysis using Dependency Parsing. *ACM Trans. Asian Low Resour. Lang. Inf. Process.* **2022**, *21*, 56. [[CrossRef](#)]
19. Wei, P.; Lin, J.; Mao, W. Multi-target stance detection via a dynamic memory-augmented network. In Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, Ann Arbor, MI, USA, 8–12 July 2018; ACM: New York, NY, USA, 2018; pp. 1229–1232.
20. Li, C.; Peng, H.; Li, J.; Sun, L.; Lyu, L.; Wang, L.; Yu, P.S.; He, L. Joint Stance and Rumor Detection in Hierarchical Heterogeneous Graph. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *33*, 2530–2542. [[CrossRef](#)] [[PubMed](#)]
21. Cignarella, A.T.; Bosco, C.; Rosso, P. Do Dependency Relations Help in the Task of Stance Detection? In Proceedings of the Third Workshop on Insights from Negative Results in NLP, Insights@ACL 2022, Dublin, Ireland, 26 May 2022; pp. 10–17.
22. Conforti, C.; Berndt, J.; Pilehvar, M.T.; Giannitsarou, C.; Toxvaerd, F.; Collier, N. Synthetic Examples Improve Cross-Target Generalization: A Study on Stance Detection on a Twitter corpus. In Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, WASSA@EACL 2021, Online, 19 April 2021; pp. 181–187.
23. Augenstein, I.; Rocktaeschel, T.; Vlachos, A.; Bontcheva, K. Stance Detection with Bidirectional Conditional Encoding. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Austin, TX, USA, 1–5 November 2016.
24. Wei, P.; Mao, W. Modeling Transferable Topics for Cross-Target Stance Detection. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, Paris, France, 21–25 July 2019; ACM: New York, NY, USA, 2019; pp. 1173–1176.
25. Cambria, E.; Poria, S.; Hazarika, D.; Kwok, K. SenticNet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, AK, USA, 2–7 February 2018.
26. Allaway, E.; McKeown, K. 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), Punta Cana, Dominican Republic, 16–20 November 2020; pp. 8913–8931.
27. Allaway, E.; Srikanth, M.; McKeown, K. 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 2021, Online, 6–11 June 2021; pp. 4756–4767.

28. Zhu, Q.; Liang, B.; Sun, J.; Du, J.; Zhou, L.; Xu, R. Enhancing Zero-Shot Stance Detection via Targeted Background Knowledge. In Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, 11–15 July 2022; pp. 2070–2075.
29. Huang, H.; Zhang, B.; Li, Y.; Zhang, B.; Sun, Y.; Luo, C.; Peng, C. Knowledge-enhanced Prompt-tuning for Stance Detection. *ACM Trans. Asian Low Resour. Lang. Inf. Process.* **2023**, *22*, 1–20. [[CrossRef](#)]
30. Luo, Y.; Liu, Z.; Shi, Y.; Li, S.Z.; Zhang, Y. Exploiting Sentiment and Common Sense for Zero-shot Stance Detection. In Proceedings of the 29th International Conference on Computational Linguistics, Gyeongju, Republic of Korea, 12–17 October 2022; pp. 7112–7123.
31. Hu, S.; Ding, N.; Wang, H.; Liu, Z.; Li, J.; Sun, M. Knowledgeable prompt-tuning: Incorporating knowledge into prompt verbalizer for text classification. *arXiv* **2021**, arXiv:2108.02035.
32. 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; pp. 255–269.
33. Li, C.; Gao, F.; Bu, J.; Xu, L.; Chen, X.; Gu, Y.; Shao, Z.; Zheng, Q.; Zhang, N.; Wang, Y.; et al. Sentiprompt: Sentiment knowledge enhanced prompt-tuning for aspect-based sentiment analysis. *arXiv* **2021**, arXiv:2109.08306.
34. 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, Online, 1–6 August 2021; pp. 3816–3830.
35. Shin, T.; Razeghi, Y.; Logan IV, R.L.; 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; pp. 4222–4235.
36. Jiang, Y.; Gao, J.; Shen, H.; Cheng, X. Few-Shot Stance Detection via Target-Aware Prompt Distillation. In Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, 11–15 July 2022; pp. 837–847.
37. Mohammad, S.; 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-2016), San Diego, CA, USA, 16–17 June 2016; pp. 31–41.
38. Li, Y.; Sosea, T.; Sawant, A.; Nair, A.J.; Inkpen, D.; Caragea, C. P-Stance: A Large Dataset for Stance Detection in Political Domain. In Proceedings of the Findings of the Association for Computational Linguistics: ACL/IJCNLP, Online, 1–6 August 2021.
39. Tang, D.; Qin, B.; Liu, T. Aspect Level Sentiment Classification with Deep Memory Network. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, TX, USA, 1–4 November 2016.
40. Huang, B.; Ou, Y.; Carley, K.M. Aspect level sentiment classification with attention-over-attention neural networks. In Proceedings of the International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation, Washington, DC, USA, 10–13 July 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 197–206.
41. Zhang, C.; Li, Q.; Song, D. Aspect-based Sentiment Classification with Aspect-specific Graph Convolutional Networks. 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), Hong Kong, China, 3–7 November 2019; pp. 4560–4570.
42. 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.
43. Li, Y.; Caragea, C. Multi-task stance detection with sentiment and stance lexicons. 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), Hong Kong, China, 3–7 November 2019; pp. 6299–6305.
44. 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; pp. 4171–4186.

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.