

Article

Implicit Stance Detection with Hashtag Semantic Enrichment

Li Dong ^{1,†} , Zinao Su ^{1,†}, Xianghua Fu ¹ , Bowen Zhang ^{1,2} and Genan Dai ^{1,2,*}

¹ College of Big Data and Internet, Shenzhen Technology University, Shenzhen 518118, China; dongli@sztu.edu.cn (L.D.); 202100202149@stumail.sztu.edu.cn (Z.S.); fuxianghua@sztu.edu.cn (X.F.); zhangbowen@sztu.edu.cn (B.Z.)

² Guangdong Key Laboratory for Intelligent Computation of Public Service Supply, Guangzhou 510006, China

* Corresponding: daigenan@sztu.edu.cn

† These authors contributed equally to this work.

Abstract: Stance detection is a crucial task in natural language processing and social computing, focusing on classifying expressed attitudes towards specific targets based on the input text. Conventional methods predominantly view stance detection as a task of target-oriented, sentence-level text classification. On popular social media platforms like Twitter, users often express their opinions through hashtags in addition to textual content within tweets. However, current methods primarily treat hashtags as data retrieval labels, neglecting to effectively utilize the semantic information they carry. In this paper, we propose a large language model knowledge-enhanced stance detection framework (LKESD) for stance detection. LKESD contains three main components: an instruction-prompted background knowledge acquisition module (IPBKA) that retrieves background knowledge of hashtags by providing handcrafted prompts to large language models (LLMs); a graph convolutional feature-enhancement module (GCFEM) is designed to extract the semantic representations of words that frequently co-occur with hashtags in the dataset by leveraging textual associations; an a knowledge fusion network (KFN) is proposed to selectively integrate graph representations and LLM features using a prompt-tuning framework. Extensive experimental results on three benchmark datasets demonstrate that our LKESD method outperforms 2.7% on all setups over compared methods, validating its effectiveness in stance detection tasks.

Keywords: stance detection; hashtag representation; knowledge-integrated methods

MSC: 68T50



Citation: Dong, L.; Su, Z.; Fu, X.; Zhang, B.; Dai, G. Implicit Stance Detection with Hashtag Semantic Enrichment. *Mathematics* **2024**, *12*, 1663. <https://doi.org/10.3390/math12111663>

Academic Editors: Iraklis Varlamis

Received: 30 April 2024

Revised: 22 May 2024

Accepted: 23 May 2024

Published: 26 May 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 an important task in the domains of natural language processing (NLP) and social computing, focused on classifying the expressed attitude towards a particular target based on the input text [1]. Early stance detection research primarily concentrates on evaluating data from online debate platforms, political analysis documents, and related sources. In recent years, the rapid development of the internet has led to a significant increase in the popularity of platforms such as Twitter, prompting researchers to investigate stance detection tasks for social media [2,3]. As a result, stance detection on social media has become a significant area of research.

Current stance detection approaches are typically framed as sentence-level classification tasks based on a specific target. These approaches can be divided into non-pretrained or pretrained language model (PLM) approaches. Non-pretrained models mainly employ such as recurrent neural networks, graph convolutional networks (GCNs), and traditional attention-based architectures for stance classification. For instance, Du et al. [4] employed an attention method utilizing target features. Sun et al. [5] developed hierarchical attention for modeling text representations through linguistic knowledge, and Liang et al. [6] proposed a GCN approach to differentiate target-specific and invariant features. Inspired

by the promising performance of PLMs, fine-tuning strategies have been developed to enhance the accuracy of stance detection [7]. These methods involve adapting pretrained models, e.g., BERT [8] and RoBERTa [9], utilizing stance detection datasets, thus adapting the models to this particular task. Typically, these methods mainly view stance detection as a target-oriented, sentence-level text classification task.

In recent social media stance detection methods, a persistent challenge remains despite their progress. Specifically, on popular social media platforms like Twitter, users frequently express their opinions through the use of hashtags in addition to textual content within tweets. However, current methods mainly treat hashtags as data retrieval labels, neglecting to effectively utilize the semantic information they carry. For instance, datasets such as SemEval-2016 Task 6 (SEM16) [10] and Pstance [11] employ hashtags as keywords for data collection. Consequently, these hashtags are prevalent across multiple instances and are often challenging to effectively represent using sentence-level text classification approaches.

Recently, Zhang et al. [12] introduced a challenging task, implicit stance detection (ISD), and proposed an ISD dataset, where hashtags play a crucial role as discriminative features within sentences. Examples include stance indicators such as “#voteTrump” and background knowledge related to the target like “#MAGA” and “#BLM”, among others. This approach closely aligns with real-world social media scenarios, where accurate stance detection necessitates a comprehensive understanding of the knowledge encapsulated within stance-related hashtags.

To date, several studies have explored the ISD task. Given the vast number of hashtags, early work employed unsupervised methods, such as k-nearest neighbors, to learn text representations of hashtags, subsequently integrating them into classifiers [13]. Building upon the characteristics of social media content, Huang et al. [12] proposed the biterm topic model (BTM) method to learn vector representations of hashtags from unsupervised data. However, these methods face limitations: they require large numbers of data for unsupervised algorithm learning, which is impractical in rapidly evolving social media scenarios. Additionally, while some work leverages the knowledge stored in large models to enhance social media text retrieval, the knowledge in these models may be prone to errors due to factors such as training sample timing, rendering them unsuitable for the rapidly changing landscape of social media.

To address the aforementioned challenges, we propose LKESD, a large language model knowledge-enhanced stance detection framework for hashtags. LKESD comprises three main components as follows. An instruction-prompted background knowledge acquisition module (IPBKA) that retrieves background knowledge by providing handcrafted prompts to large language models (LLMs). A graph convolutional feature enhancement module (GCFEM) is designed to mine the semantic representations of words that frequently co-occur with hashtags in the dataset text by leveraging textual associations. Subsequently, a knowledge fusion network (KFN) is constructed to selectively integrate graph representations and LLM features using a prompt-tuning framework. The prompt-tuning framework involves constructing a prompt fine-tuning method based on pre-trained language models (PLMs) for accurate stance detection.

We summarize our contributions as follows:

- We propose a LKESD framework for stance detection that can learn the semantic information of hashtags from both LLMs and corpora, thereby enhancing its applicability in real-world social media scenarios.
- We investigate stance detection from a novel perspective by exploring the semantic expressions of hashtags. We propose a novel KFN to achieve dynamic fusion of different semantic representation features.
- To validate the effectiveness of the LKESD model for stance detection on social media, we perform comprehensive experiments on widely used benchmarks. The experimental results demonstrate the effectiveness of the proposed method.

The subsequent sections of this paper are structured as follows. Section 2 reviews relevant literature on traditional and recent methods for stance detection. Section 3 presents the proposed model in detail. Section 4 outlines the experimental setup, including the datasets, baseline methods, and quantitative results. Lastly, Section 5 concludes the paper and discusses potential avenues for future work.

2. Related Work

2.1. Stance Detection

Stance detection aims to classify the perspective expressed in a text towards a given target and is closely related to argument mining, fact-checking, and aspect-level sentiment analysis [14,15].

As shown in Table 1, existing in-target stance detection methods can be categorized into two types: non-pretrained and pretrained approaches. Non-pretrained methods often utilize deep neural networks like attention-based methods and GCN to train stance classifiers. Attention-based methods utilize target-relevant information and implement attention mechanisms to determine stance polarity [4,5,16]. GCN-based methods utilize GCN to model relations between the target and text, enabling nuanced analysis of their connections [17–19].

For cross-target stance detection, existing methods can be categorized into two main types: word-level transfer and concept-level transfer. Word-level transfer methods leverage the commonality of words across targets to bridge knowledge gaps [20]. Concept-level transfer methods address cross-target challenges by utilizing shared concepts between targets to enable understanding and analysis [21–23].

Zero-shot stance detection poses a particular challenge, requiring models to deduce the stance towards unseen targets. To enable zero-shot learning, Allaway et al. [24] constructed a human-annotated dataset tailored for this setting. Allaway et al. [25] further applied adversarial learning to derive target-invariant features and used a target-specific dataset. Liu et al. [7] proposed a graph-based model integrating intra and extra semantic knowledge and common sense using BERT. Liang et al. [6] identified target-specific and target-invariant characteristics to obtain transferable features.

2.2. Incorporating Background Knowledge

Incorporating background knowledge to improve stance detection performance on social media has gained considerable attention in recent years. Earlier methods focused on enhancing the understanding of words in the text. For example, Zhang et al. [22] proposed a framework that extracted semantic and emotional word-level knowledge from lexicons to enable knowledge transfer across targets. Another common approach is to conduct pre-training on a corpus specific to the target domain, such as BERTweet [26] or COVID-Twitter-BERT [27]. Kazuaki et al. [28] introduced a method for extracting relevant concepts and events from Wikipedia articles and incorporating them into stance detection. Current retrieval methods typically employ keyword-based filtering [29] for knowledge retrieval.

Despite effective progress, an important challenge when applying these methods to social media is their inability to effectively address the semantic expressiveness of hashtags. To the best of our knowledge, semantic expressiveness is an emerging area of interest. Ghosh et al. [30] first proposes that to split hashtag into individual words and employ substitute vocabulary to clarify these expressions. However, since hashtags may contain informal text and this approach fails to incorporate contextual semantics, it yields suboptimal results. Zhang et al. [31] proposed constructing an unsupervised topic model and using the clustered topic words as semantic representations of hashtags. The low accuracy of clustering risks propagating errors, and the reliance on abundant unlabeled data makes unsupervised methods less adaptable to cross-target and zero-shot settings. Li et al. [32] proposed incorporating LLM to generate the explanation of hashtags and

enhance model performance. However, directly utilizing the knowledge from large models is restricted to their training corpus and may propagate errors.

Table 1. Stance detection methods.

Method	Year	Brief Description
Dey et al. [16]	2018	Attention-based stance polarity detection using target-relevant information.
Du et al. [4]	2017	Attention mechanisms for determining stance polarity.
Sun et al. [5]	2018	Employing attention methods for stance classification.
Li et al. [17]	2022	GCN for modeling relations between target and text.
Cignarella et al. [18]	2022	GCN-based analysis of text-target connections.
Conforti et al. [19]	2021	GCN utilization for nuanced stance analysis.
Augenstein et al. [20]	2016	Word-level transfer leveraging word commonality.
Wei et al. [21]	2019	Concept-level transfer using shared concepts.
Zhang et al. [22]	2020	Enhancing cross-target understanding with concepts.
Cambria et al. [23]	2018	Concept-based transfer for stance detection.
Allaway et al. [24]	2020	Human-annotated dataset for zero-shot stance detection.
Allaway et al. [25]	2021	Adversarial learning for target-invariant features.
Liu et al. [7]	2021	Graph-based model integrating semantic knowledge.
Liang et al. [6]	2022	Identifying transferable features for zero-shot learning.
Zhang et al. [22]	2020	Extracting semantic and emotional word-level knowledge from lexicons.
Nguyen et al. [26]	2020	Pre-training on a corpus specific to the target domain (BERTweet).
Mueller et al. [27]	2020	Pre-training on COVID-19 related tweets (COVID-Twitter-BERT).
Hanawa et al. [28]	2019	Extracting relevant concepts and events from Wikipedia articles.
Zhu et al. [29]	2022	Keyword-based filtering for knowledge retrieval.
Ghosh et al. [30]	2019	Splitting hashtags into individual words and using substitute vocabulary.
Zhang et al. [31]	2022	Using an unsupervised topic model for semantic representation of hashtags.
Li et al. [32]	2023	Incorporating LLM to generate explanations of hashtags.

3. LKESD Framework

We give the task definition and the overview of our model in Section 3.1 and Section 3.2, respectively. Then, we describe the details of the LKESD in Sections 3.3–3.5.

3.1. Problem Definition

The goal of stance detection is to predict the stance polarity of an input sentence x^t towards a specified target q^t using a model trained on a labeled dataset X . Here, $X = \{x_i, q_i\}_{i=1}^N$ represents the collection of labeled data, where x denotes the input text, q corresponds to the source target, and N is the total number of instances in X . Each sentence-target pair $(x, q) \in X$ is assigned a stance label y . The superscript t indicates test data.

3.2. Framework Overview

As depicted in Figure 1, LKESD consists of three main components: IPBKA, GCFEM, and KFN. IPBKA proposes an instruct-based zero-shot prompting method that acquires the knowledge for hashtags from LLMs. Since they are data from outside the training set, we refer to them as extra knowledge. The GCFEM first constructs a semantic graph containing hashtags from the input text and subsequently learns the vector representation of hashtags through a GCN network. Contrary to extra knowledge, we call it intra knowledge. Finally, the KFN is a prompt-tuning network that fuses extra and intra knowledge for accurate stance detection. This is achieved by creating a customized template for the PLM and integrating extra and intra knowledge.

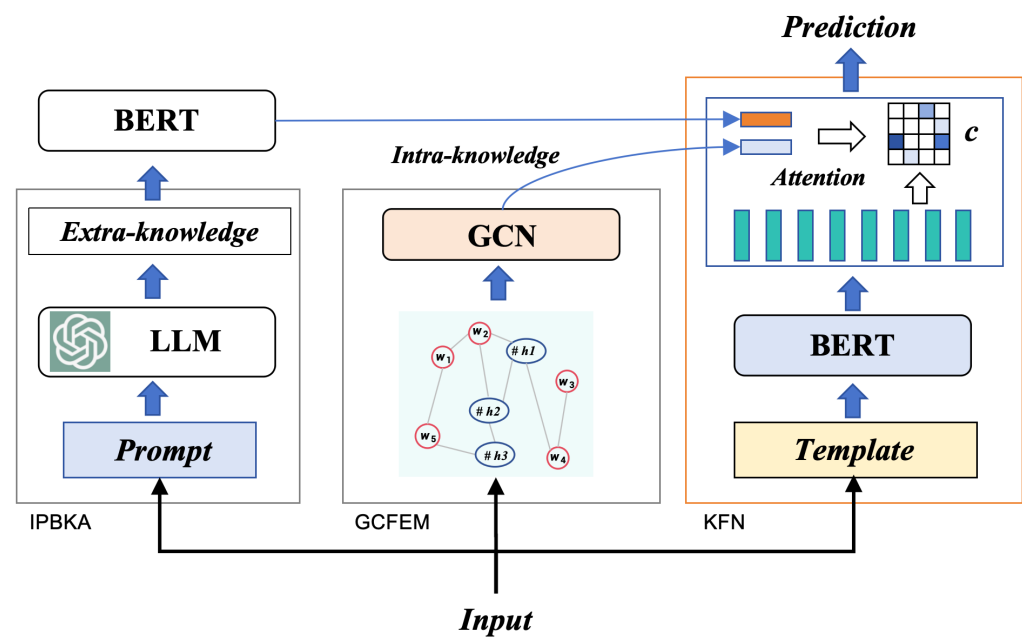


Figure 1. Framework overview of LKESD. The input examples for IPBKA can be found in Section 3.3. The GCFEM module takes tweets as input, while the KFN module uses pre-constructed prompt templates as input (see prompt design part in Section 3.5).

3.3. IPBKA

IPBKA is used to extract extra knowledge of hashtags from LLM. Inspired by the effectiveness of zero-shot instruction prompting in current LLMs [33], we construct an instruction template that is directly fed into the LLM to obtain the background knowledge of the hashtag. The specific template can be represented as *Prompt*:

Prompt: Given the following text, identify and analyze the semantic meaning of any hashtags present. Provide insights into the context and potential implications of these hashtags. Text: “[GIVEN TEXT]”

Subsequently, we input the obtained extra knowledge into the BERT model to generate the embedding vector of extra knowledge. Specifically, we use the average of hidden states as the representation of extra knowledge, denoted as r .

3.4. GCFEM

The GCFEM is employed to learn hashtag representations from the input text (intra knowledge). Compared to the knowledge obtained from IPBKA, the hashtag representations obtained from the text are closer to the input domain.

Specifically, to represent word-hashtag relationships, we first construct a semantic graph. The semantic graph employs words or hashtags as nodes and builds weighted edges between words or hashtags based on their co-occurrence frequency. We use G to represent the constructed graph.

Subsequently, we employ GCN to learn the embeddings of each node in the graph to fully leverage the multi-hop semantic connections between nodes. Given the semantic locality between words, we extract a λ -hop subgraph from the constructed graph for each hashtag, which is then input into a GCN to learn the graph representation. GCN is adopted due to its advantage of effectiveness and efficiency in learning graph embeddings.

In formal terms, let $E \in \mathbb{R}^{v \times d}$ represent a matrix containing all v nodes in the graph and their respective features, where d is the size of the node embedding. For each node, we extract a λ -hop subgraph G' from the entire graph G , which has a degree matrix D and an

adjacency matrix A . The normalized symmetric adjacency matrix of subgraph G' can be calculated as:

$$\tilde{A} = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}. \quad (1)$$

The subgraph representation $L \in \mathbb{R}^{n \times c}$ with n nodes can be computed by feeding the subgraph G' into a two-layer GCN as follows:

$$L = \sigma(\tilde{A} \sigma(\tilde{A} E W_i) W_j), \quad (2)$$

where σ denotes the *sigmoid* function, W_i and W_j are learnable parameters. After obtaining the graph representation L , the vector corresponding to the hashtag is retrieved from the graph and represented as k .

3.5. Knowledge Fusion Network

Knowledge Fusion Network is a prompt-tuning framework that takes input text information and feeds it into a pre-trained model through the construction of a template. The fusion layer then combines the input with intra and extra knowledge for fusion.

Prompt-tuning is a transformative framework that reformulates the original classification task as a masked language modeling task. In particular, prompt-tuning utilizes a natural language template p that is integrated into the given text x and the target q . The combined input is denoted as follows: “ $x_p = x$. The attitude to q is [MASK]”. Let M denote the BERT model, which gives the probability of each word v in the vocabulary being filled in [MASK] given $P_M([\text{MASK}] = v | x_p)$. In this case, v denotes the defined label word in the verbalizer. To transform the probabilities of these words to the probabilities of the labels, a verbalizer is employed 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., $f : V \rightarrow Y$.

The probability $P(y|x_p)$ of label y is formally computed as follows:

$$P(y|x_p) = \delta(P_M([\text{MASK}] = vs. | x_p) | vs. \in V). \quad (3)$$

where δ plays a pivotal role in transforming the probability distribution over label words to the probability distribution over labels.

Prompt design. The crucial aspect of a prompt-based method for stance detection is the construction of an appropriate prompt. In this paper, following the work of [12], our template is defined as follows:

Template (x_p): [Given input: x]. The attitude to [given target: q] is [MASK].

Fusion layer. Upon building the template, we employ a novel method that relates mapping labels onto continuous vectors, referred to as stance vectors, rather than explicit words or phrases. Specifically, we input x_p into BERT to obtain H , which represents the hidden vector generated by BERT. In this case, H is the input text representation vector. Further, we extract the vector at the “[MASK]” position from H as the stance vector, denoted as s . Subsequently, given vectors s , k and r , we employ the attention mechanism, enabling the learning of knowledge-enhanced textual representations.

Formally, the attention coupling factors c for each query can be computed as follows:

$$c^r = H r^T, \quad c^k = H k^T, \quad c^s = H s^T. \quad (4)$$

Subsequently, we normalize the three factors using the softmax function to obtain the attention weight:

$$c = \text{Softmax}(c^r + c^k + c^s). \quad (5)$$

With the attention weight c , the final representation can be computed as follows:

$$e = c(r + k + s) + \gamma H, \quad (6)$$

where γ denotes the scaling factor.

Given the defined label words from the verbalizer, we generate the probability that the token v can be chosen as the label words:

$$\omega = \frac{\exp(v_i \cdot e)}{\sum_{v_j \in V} \exp(v_j \cdot e)}, \quad (7)$$

where v represents the embedding of the token in the Verbalizer. Subsequently, we aggregate the probabilities of each label from ω for these words, denoted as $\hat{y}$. Finally, the loss function of the ensemble network can be computed using standard cross-entropy methods:

$$\mathcal{L} = - \sum_{i=1}^N \sum_{j=1}^C y_{ij} \log \hat{y}_{ij}. \quad (8)$$

Here, N denotes the number of training samples, C denotes the number of stance classes, and y_i represents the one-hot represented ground-truth label for the i -th sample. Ultimately, the attention layer is optimized using the standard gradient descent algorithm.

4. Experiments

4.1. Experimental Data

We present empirical evaluations on several benchmark datasets, containing ISD [12], SemEval-2016 Task 6 (SEM16) [10], COVID-19 [34]. The dataset statistics are summarized in Table 2.

Table 2. Statistics of datasets.

Dataset	Target	Favor	Against	Neutral
SEM16	SEM16-h	598	1620	648
COVID19	COV-h	1990	1928	2215
ISD	DT	875	1096	1134
	JB	1046	525	912

- **ISD.** The ISD dataset [12] is proposed for the stance detection task on social media, which presents a challenge as it consists of texts lacking explicit sentiment words. Therefore, understanding the relationship between the text and contextual knowledge, including the target and hashtag knowledge, is crucial for predicting stance polarity. ISD includes two targets: Trump (DT) and Biden (JB).
- **SEM16.** The original SEM16 dataset includes 4870 texts and annotated with one of three stance labels: “favor”, “against”, or “neutral”. To validate the efficacy of our hashtag fusion approach, we reorganized the original dataset. Hashtags containing only crawled user data were removed, and the remaining data were consolidated into a single dataset (SEM16-h). SEM16-h contains the same four targets as in previous work [21].
- **COVID-19.** The COVID-19 dataset contains 6133 tweets, each reflecting user positions on four specific targets associated with COVID-19 health mandates. Similar to SEM16, we process the dataset, and the remaining data are consolidated into a single task (COV-h). COV-h contains the same four targets as [34].

4.2. Compared Baseline Methods

To assess the efficacy of our proposed model, we performed an extensive analysis and comparative study with established baseline models, which are outlined as follows: Statistics-based methods:

- **BiLSTM** [20] adopts a bidirectional LSTM framework to encode the text and target separately, enabling the extraction of independent semantic features.
- **BiCond** [20] employs a bidirectional LSTM framework to simultaneously encode the text and target, thereby capturing their shared semantic features.

- **CrossNet** [35] builds upon the BiCond architecture by integrating a self-attention mechanism, which selectively highlights salient textual features.
- **AoA** [36] employs a dual-LSTM architecture, wherein two separate LSTM networks are dedicated to modeling the target and context, respectively, and an interactive attention mechanism is integrated to facilitate the examination of their relationships.
- **TPDG** [37] proposes a target-adaptive convolutional graph framework, which boosts stance detection accuracy by leveraging shared features from similar targets and capitalizing on their inherent relationships.

Fine-tuning based methods:

- **BERT** [8] leverages a pre-trained BERT architecture for stance detection, reformulating the input format to “[CLS] + text + [SEP] + target + [SEP]” to optimize the model’s training and fine-tuning procedures.
- **PT-HCL** [6] exploits contrastive learning to enhance the detection of subtle stance variations.

Prompt-tuning based methods:

- **MPT** [38] introduces a knowledge-infused prompt-tuning method for stance detection, which exploits a verbalizer carefully crafted by human experts to enhance the detection of subtle stance variations.
- **KPT** [39] leverages external lexicals to initialize the verbalizer component, which is embedded within the prompt framework, to facilitate the integration of domain-specific knowledge.
- **KEprompt** [12] employs a topic model to acquire hashtag representations and then performs prompt-tuning methods for stance detection.

Knowledge-enhanced methods:

- **SEKT** [22] presents a GCN framework that incorporates semantic knowledge to enhance stance detection capabilities.
- **TarBK** [29] integrates the target-related wiki knowledge from Wikipedia for stance detection.
- **Ts-CoT** [40] first proposes CoT methods with LLMs for stance detection.
- **KASD** [32] proposes to augment the hashtag by utilizing LLMs.

4.3. Implementation Details and Evaluation Metrics

In the experimental configuration, we selected to use the BERT-base uncased architecture as the PLM. The Adam optimizer with a learning rate of 0.0002 was used to train the model, and the mini-batch size was set to 32. For the LLM, we employ GPT-3.5 as the foundational architecture for knowledge elicitation.

Following previous research [6,22], we utilize the macro F1 score as the metric, which is the average F1 score across favor and against labels:

$$\begin{aligned}
 F1_{favor} &= \frac{2P_{favor}R_{favor}}{P_{favor} + R_{favor}} \\
 F1_{against} &= \frac{2P_{against}R_{against}}{P_{against} + R_{against}} \\
 F1_{none} &= \frac{2P_{none}R_{none}}{P_{none} + R_{none}}
 \end{aligned} \tag{9}$$

The F1 score can be computed based on precision and recall.

$$F1 = \frac{F1_{favor} + F1_{against}}{2} \tag{10}$$

4.4. Overall Performance

4.4.1. In-Target Setup

Table 3 presents the results of in-target stance detection in comparison to widely used benchmarks. The results demonstrate that our LKESD methods outperform most of the baseline methods across all datasets, thereby highlighting the performance of LKESD in stance detection.

Table 3. Comparative results of F1 score for in-target stance detection.

	Methods	DT	JB	SEM16-h	COV-h
glove	Bilstm	28.6	35.0	35.6	38.8
	Bicond	55.2	50.5	37.2	39.1
	MemNet	53.5	52.2	37.9	40.3
	AoA	55.9	57.6	38.1	40.9
	TPDG	64.2	60.0	45.4	47.7
BERT	BERT-FT	69.1	65.6	45.7	50.2
	MPT	69.0	65.9	46.8	52.2
	KPT	69.4	66.4	49.2	54.6
	KEPrompt	70.5	67.4	50.3	56.9
	TarBK	69.5	66.6	50.1	54.1
	Ts-CoT	69.4	69.1	53.9	57.4
	KASD	70.2	68.4	55.1	59.3
	LKESD	71.3[†]	72.0[†]	58.9[†]	61.5[†]

The experimental results show that our LKESD model outperforms most baseline models across all datasets, thereby validating the effectiveness of our proposed stance detection method. Furthermore, significance tests conducted on LKESD, with p -value < 0.05 (indicated as [†]), reveal statistically significant enhancements over the best-performing competitors across most evaluation metrics. Specifically, the experiments show that when using statistical-based embedding methods, all results perform poorly due to the inability of statistical word vector initialization methods to effectively represent hashtags. In contrast, pre-trained models (e.g., BERT) achieve improvements in accuracy, possibly due to the ability of pre-trained models like BERT to effectively leverage large-scale knowledge.

Finally, our LKESD outperforms the KASD method, which is enhanced with LLM, by an average of 2.7% across all tasks. This may be due to our knowledge fusion mechanism that effectively leverages both extra and intra knowledge for hashtags.

4.4.2. Cross-Target Setup

Acquiring large-scale annotated datasets requires substantial time and resources. Therefore, we further validate the effectiveness of LKESD within a cross-target setup. The goal of this setup is to use labeled data from the source target to predict the stance toward the destination target. The result can be found in Table 4. From the result, we can find that our LKESD significantly outperforms the best-performing baseline competitors. Specifically, methods leveraging large-scale models (KASD and LKESD) significantly outperform traditional knowledge enhancement methods, indicating that LLMs can effectively generate external knowledge to improve predictive performance. Furthermore, the F1 score of LKESD is on average 0.3% higher than that of KASD on average. This performance improvement may be due to the automatic selection of large-scale model knowledge by the fusion mechanism, which can effectively learn transferable important knowledge.

Table 4. Comparative results of F1 score for cross-target stance detection.

	Methods	JB→DT	DT→JB	COV-h→SEM16-h	SEM16-h→COV-h
glove	BiCond	47.9	43.6	30.2	33.8
	CrossNet	48.0	44.2	36.5	34.8
	SEKT	51.2	52.3	39.1	38.7
BERT	BERT-FT	55.7	57.3	40.2	41.5
	MPT	58.9	60.1	44.4	48.1
	KEprompt	60.4	64.2	45.3	50.3
	KASD	62.6	65.3	51.2	53.1
	LKESD	64.3	67.0	50.8	55.2

4.4.3. Zero-Shot Stance Detection

To further evaluate the generalizability of the model, we conduct a zero-shot setup for stance detection. The result is shown in Table 5. Following previous work [6,32,41], we select a specific target as the test set, with the remaining task data as the training set. For example, we use →DT to denote DT as the test set, with the remaining targets (JB, SEM16-h and COV-h) as the training data. From the experiments, we observe that LKESD can still achieve effective performance improvements in the zero-shot scenario. Specifically, we find that methods requiring large amounts of unlabeled samples to construct hashtag representations, compared to in-target and cross-target, have insufficient accuracy improvements in the zero-shot setting. This is due to the unseen target domain, which cannot effectively acquire hashtag background knowledge. In contrast, methods leveraging LLMs (KASD and LKESD) both achieve good performance. This is consistent with our expectation that zero-shot prompt learning methods can obtain hashtag background knowledge in unseen target domains.

Table 5. Comparative results of F1 score for zero-shot stance detection.

	Methods	→DT	→JB	→SEM16-h	→COV-h
glove	BiCond	44.5	41.7	31.2	31.7
	SEKT	46.8	49.1	34.0	33.4
BERT	BERT-FT	54.7	58.0	35.9	38.1
	MPT	59.2	58.9	36.5	40.3
	KEprompt	59.3	58.7	40.2	42.6
	PT-HCL	60.1	59.6	41.7	43.3
	KASD	61.9	63.3	46.3	48.2
	LKESD	62.0	64.2	46.9	50.1

4.5. Ablation Study

To evaluate the effect of each component in our model, we perform ablation studies by individually removing the IPBKA model (denoted as w/o IPBKA), GCFEM (denoted as w/o GCN), and KFN (denoted as w/o KFN). In particular, for w/o KFN, we directly concatenate the intra and extra knowledge to replace the attention-based fusion mechanism.

The ablation study results are presented in Figure 2. The results indicate that IPBKA, GCN, and KFN all make significant contributions to improving the performance of the proposed method. More specifically, the performance significantly decreases when IPBKA and GCN are removed. This may be due to the importance of hashtag semantic representation for stance detection. Finally, as expected, integrating all components results in the best performance across all experimental settings.

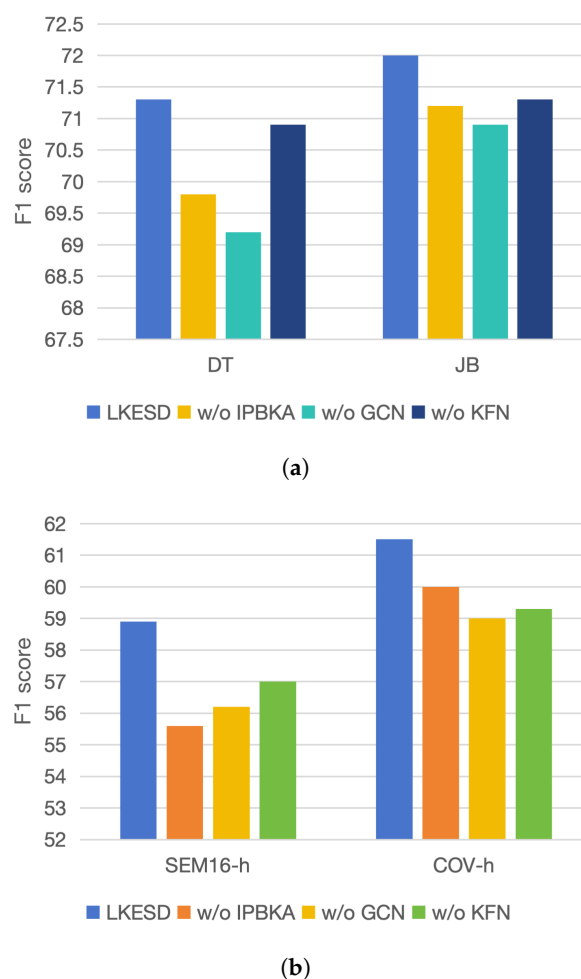


Figure 2. Ablation test results. (a) Ablation study with DT and JB targets. (b) Ablation study with SEM-h and COV-h targets.

5. Conclusions

In this paper, we propose a large language model knowledge-enhanced stance detection framework (LKESD) for stance detection that can learn the semantic information of hashtags, thereby enhancing its applicability in real-world social media scenarios. LKESD comprises three main components: An instruction-prompted background knowledge acquisition module (IPBKA) that retrieves background knowledge for hashtags by providing handcrafted prompts to LLMs. A graph convolutional feature enhancement module (GCFEM) aims to extract the semantic representations of words that frequently co-occur with hashtags in the dataset by leveraging textual associations. A knowledge fusion network (KFN) is proposed to selectively integrate graph representations and LLM features using a prompt-tuning framework. Experiments on three benchmark datasets demonstrate that our LKESD method outperforms comparison methods, validating its effectiveness in the stance detection task. In future work, we will further investigate the impact of biases in individual LLMs on our method. Additionally, we plan to explore stance detection methods that integrate graph-based LLMs.

Author Contributions: Conceptualization, L.D. and Z.S.; methodology, L.D.; software, Z.S.; validation, Z.S. and G.D.; formal analysis, X.F.; writing—original draft preparation, L.D. and Z.S.; writing—review and editing, G.D., X.F. and B.Z.; visualization, B.Z.; supervision, G.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by National Nature Science Foundation of China (No. 62306184), Natural Science Foundation of Top Talent of SZTU (grant no. GDRC202320), the Research Promotion

Project of Key Construction Discipline in Guangdong Province (2022ZDJS112), and University Stability Support Program of Shenzhen (20231129211559001).

Data Availability Statement: The data presented in this article and codes used are available upon request to the corresponding author.

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

References

- Küçük, D.; Can, F. Stance detection: A survey. *ACM Comput. Surv. CSUR* **2020**, *53*, 1–37. [\[CrossRef\]](#)
- 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. [\[CrossRef\]](#) [\[PubMed\]](#)
- 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. [\[CrossRef\]](#) [\[PubMed\]](#)
- Du, J.; Xu, R.; He, Y.; Gui, L. Stance classification with target-specific neural attention networks. In Proceedings of the 26th International Joint Conference on Artificial Intelligence, Melbourne, Australia, 19–25 August 2017.
- 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, NW, USA, 21–25 August 2018; pp. 2399–2409.
- 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.
- 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, 1–6 August 2021; pp. 3152–3157.
- 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; Burstein, J., Doran, C., Solorio, T., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; Volume 1 (Long and Short Papers), pp. 4171–4186.
- Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. Roberta: A robustly optimized bert pretraining approach. *arXiv* **2019**, arXiv:1907.11692.
- 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@NAACL-HLT, San Diego, CA, USA, 16–17 June 2016; pp. 31–41.
- 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.
- 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. [\[CrossRef\]](#)
- Darwish, K.; Stefanov, P.; Aupetit, M.; Nakov, P. Unsupervised user stance detection on Twitter. In Proceedings of the International AAAI Conference on Web and Social Media, Atlanta, GA, USA, 8 June 2020; Volume 14, pp. 141–152.
- Jain, R.; Jain, D.K.; Dharana, Sharma, N. Fake News Classification: A Quantitative Research Description. *ACM Trans. Asian Low Resour. Lang. Inf. Process.* **2022**, *21*, 1–17. [\[CrossRef\]](#)
- Rani, S.; Kumar, P. Aspect-based Sentiment Analysis using Dependency Parsing. *ACM Trans. Asian Low Resour. Lang. Inf. Process.* **2022**, *21*, 1–19. [\[CrossRef\]](#)
- 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, 26–29 March 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 529–536.
- 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. [\[CrossRef\]](#) [\[PubMed\]](#)
- 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; Association for Computational Linguistics: Stroudsburg, PA, USA, 2022; pp. 10–17.
- 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; Association for Computational Linguistics: Stroudsburg, PA, USA, 2021; pp. 181–187.
- 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.
- 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.
- 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.

23. 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, LA, USA, 2–7 February 2018.
24. 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; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 8913–8931.
25. 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; Association for Computational Linguistics: Stroudsburg, PA, USA, 2021; pp. 4756–4767.
26. Nguyen, D.Q.; Vu, T.; Nguyen, A.T. BERTweet: A pre-trained language model for English Tweets. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, EMNLP 2020—Demos, Online, 16–20 November 2020; Liu, Q., Schlangen, D., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 9–14. [\[CrossRef\]](#)
27. Müller, M.; Salathé, M.; Kummervold, P.E. COVID-Twitter-BERT: A Natural Language Processing Model to Analyse COVID-19 Content on Twitter. *arXiv* **2020**, arXiv:2005.07503.
28. Hanawa, K.; Sasaki, A.; Okazaki, N.; Inui, K. Stance Detection Attending External Knowledge from Wikipedia. *J. Inf. Process.* **2019**, *27*, 499–506. [\[CrossRef\]](#)
29. 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 SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, 11–15 July 2022; Amigó, E., Castells, P., Gonzalo, J., Carterette, B., Culpepper, J.S., Kazai, G., Eds.; ACM: New York, NY, USA, 2022; pp. 2070–2075. [\[CrossRef\]](#)
30. Ghosh, S.; Singhanian, P.; Singh, S.; Rudra, K.; Ghosh, S. Stance Detection in Web and Social Media: A Comparative Study. In Proceedings of the Experimental IR Meets Multilinguality, Multimodality, and Interaction—10th International Conference of the CLEF Association, CLEF 2019, Lugano, Switzerland, 9–12 September 2019; Springer: Berlin/Heidelberg, Germany, 2019; Lecture Notes in Computer Science; Volume 11696, pp. 75–87. [\[CrossRef\]](#)
31. Zhang, Z.; Zhang, A.; Li, M.; Smola, A. Automatic chain of thought prompting in large language models. *arXiv* **2022**, arXiv:2210.03493.
32. Li, A.; Liang, B.; Zhao, J.; Zhang, B.; Yang, M.; Xu, R. Stance detection on social media with background knowledge. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, Singapore, 6–10 December 2023; pp. 15703–15717.
33. Ding, D.; Chen, R.; Jing, L.; Zhang, B.; Huang, X.; Dong, L.; Zhao, X.; Song, G. Cross-target Stance Detection by Exploiting Target Analytical Perspectives. *arXiv* **2024**, arXiv:2401.01761.
34. Glandt, K.; Khanal, S.; Li, Y.; Caragea, D.; Caragea, C. Stance Detection in COVID-19 Tweets. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, Virtual Event, 1–6 August 2021; Association for Computational Linguistics: Stroudsburg, PA, USA, 2021; Volume 1: Long Papers, pp. 1596–1611. [\[CrossRef\]](#)
35. 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, Melbourne, Australia, 15–20 July 2018; Volume 2: Short Papers, pp. 778–783.
36. 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.
37. 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.
38. 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.
39. Shin, T.; Razeghi, Y.; IV, R.L.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; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 4222–4235.
40. Zhang, B.; Fu, X.; Ding, D.; Huang, H.; Li, Y.; Jing, L. Investigating Chain-of-thought with ChatGPT for Stance Detection on Social Media. *arXiv* **2023**, arXiv:2304.03087.
41. 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, Dublin, Ireland, 22–27 May 2022; Association for Computational Linguistics: Stroudsburg, PA, USA, 2022; Volume 1: Long Papers, pp. 81–91.

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.