


Knowledge-Augmented Interpretable Network for Zero-Shot Stance Detection on Social Media

Bowen Zhang , Daijun Ding, Zhichao Huang, Ang Li, Yangyang Li, Baoquan Zhang, and Hu Huang

Abstract—Stance detection on social media has become increasingly important for understanding public opinions on controversial issues. Existing methods often require large amounts of labeled data to learn target-independent transferable knowledge, which is infeasible under zero-shot settings where the target is unseen. Furthermore, most current stance detection models, primarily based on end-to-end deep learning architectures, lack transparency and may produce counter-intuitive and uninterpretable predictions. In this article, we propose a novel knowledge-augmented interpretable network (KAI) to enable zero-shot stance detection (ZSSD). First, we introduce an unsupervised approach based on large language models (LLM-KE) to elicit analysis perspectives, which is target-independent knowledge shared across different targets. This transferable knowledge bridges connections between seen and unseen targets. Second, we develop a bidirectional knowledge-guided neural production system (Bi-KGNPS) that effectively integrates such transferable knowledge through an iterative knowledge-variable binding process to guide stance predictions. Extensive experiments on benchmark datasets demonstrate KAI achieves new state-of-the-art performance on ZSSD. Moreover, our approach also delivers strong results on conventional in-target and cross-target stance detection. With the dual benefits of knowledge-augmented accuracy and model interpretability, this work represents an important advance toward practical stance detection systems that can generalize to emerging topics of interest. The proposed KAI framework provides an interpretable approach to effectively transfer knowledge across domains for zero-shot learning.

Index Terms—Attention mechanism, chain-of-thought (CoT), neural production system (NPS), zero-shot stance detection (ZSSD).

Manuscript received 9 December 2023; revised 23 March 2024; accepted 2 April 2024. This work was supported in part by the National Nature Science Foundation of China under Grant 62306184, in part by the Natural Science Foundation of Top Talent of SZTU under Grant GDRC202320, and in part by the Research Promotion Project of Key Construction Discipline in Guangdong Province under Grant 2022ZDJS112. (*Corresponding author: Hu Huang.*)

Bowen Zhang and Daijun Ding are with the College of Big Data and Internet, Shenzhen Technology University, Shenzhen 518000, China.

Zhichao Huang, Ang Li, and Baoquan Zhang are with the Department of Computer Science, Harbin Institute of Technology, Shenzhen 518000, China.

Yangyang Li is with the Academy of Cyber, Beijing 100000, China.

Hu Huang is with Peking University Shenzhen Graduate School, Shenzhen 518000, China (e-mail: h.huang@pku.edu.cn).

Digital Object Identifier 10.1109/TCSS.2024.3388723

I. INTRODUCTION

SOCIAL media platforms have increasingly served as a common forum for users to express their viewpoints on controversial issues. By aggregating and analyzing these perspectives, we can identify prevailing trends and opinions on contentious topics, ranging from abortion to epidemic prevention. This wealth of data offers significant potential for web mining and content analysis. The insights gathered from such analysis can be crucial for various decision-making processes, including advertising strategies and political elections. Consequently, automatic stance detection on social media has gained importance in opinion mining, facilitating a more nuanced understanding of user attitudes toward a wide array of issues.

Stance detection, a core task in natural language processing (NLP), aims to categorize attitudes toward a specific target based on opinionated texts [1]. Early research focused on two types of stance detection: in-target and cross target. In in-target detection, both training and test data involve the same targets, whereas in cross-target detection, the test targets are different but related to the training ones. However, it is practically challenging to pinpoint all potential in-target or corresponding cross-target topics before model development. Consequently, zero-shot stance detection (ZSSD), a technique designed to determine attitudes toward entirely new targets during inference, has recently gained considerable research interest.

So far, several works have been conducted for ZSSD, which can be classified into three categories: attention mechanisms, graph-based model, and contrastive learning [2]. For example, Allaway et al. [3], [4] constructed a comprehensive VAST dataset for ZSSD, which includes numerous targets spanning diverse domains including politics and education. Liu et al. [5] developed a commonsense knowledge-enhanced model that captures relational knowledge between texts and targets at both structural and semantic levels. Liang et al. [6] designed a contrastive learning approach that establishes connections between target-invariant and target-specific features. Liang et al. [7] proposed a joint contrastive learning technique utilizing both stance and target-aware prototypical graph contrastive learning. Li et al. [8] introduced a teacher–student framework leveraging augmented data to improve ZSSD. Generally, these works mainly focus on extracting target-independent word-level features as transferable knowledge.

Despite the effectiveness of previous studies, there are still several challenges for ZSSD, which are not addressed well

in prior work. First, current methods primarily extract target-independent, word-level knowledge from observed data. However, on the one hand, the expressive power of word-level features is limited when applied to unseen domains. On the other hand, existing methods focus mainly on extracting target-specific features, rarely considering the general knowledge shared among targets. Different targets may share common perspectives for analysis; for example, when discussing politicians, people often evaluate them based on their party affiliation, political stance, and personal qualities. By introducing these cross-target concept-level perspectives, knowledge transfer between different targets can be achieved, enhancing the generalization ability of the model. Second, most of the current stance detection models, primarily end-to-end deep learning constructs, suffer from a lack of transparency, functioning as black-box systems. As a result, these models may generate predictions that are not only unintuitive but also difficult to interpret. Furthermore, their inability to elucidate the analytical process limits their applicability in scenarios necessitating explanations.

In response, this study aims to develop a ZSSD method that can effectively extract target-invariant features and provide interpretability. Specifically, our first goal is to develop a ZSSD method that can extract conceptual-level target-independent knowledge in an unsupervised manner. This knowledge serves as target-invariant features to bridge and transfer knowledge between seen and unseen targets. Second, to enable interpretability, our goal is to develop a ZSSD model that can effectively integrate the aforementioned target-independent features and provide an interpretable analysis process while achieving accurate results.

To achieve this aim, we propose a novel neural network called knowledge-augmented interpretable network (KAI) for ZSSD. KAI consists of two components: large language model-driven knowledge elicitation module (LLM-KEM) and bidirectional knowledge-guided neural production system (Bi-KGNPS). Specifically, LLM-KEM is the unsupervised approach based on chain-of-thought (CoT) to extract conceptual, target-independent transferable knowledge. We propose incorporating targets' analytical perspectives as such features. This is motivated by the observation that humans tend to analyze different targets from similar perspectives, serving as bridges for knowledge transfer. For example, when assessing stances on "Hillary Clinton" and "Donald Trump," people may consider perspectives such as "gender, ideology, and personality." Our model builds on using these shared angles to harness common knowledge. Further, to extract such features unsupervised, we propose a CoT strategy for LLMs. This can effectively elicit analysis perspectives in a parameter-free way, hence usable for unseen targets.

Bi-KGNPS is a stance detection model based on the neural production system (NPS), which was previously proposed for computer vision and enables interpretable analysis through knowledge-variable binding coupling factors. We are the first to apply NPS to the stance detection task to enhance model interpretability. We posit that modeling the coordination between perspectives and tweets is key to improving stance detection

performance. Thus, Bi-KGNPS introduces a bidirectional iterative knowledge-variable binding mechanism to enhance stance detection. Through analyzing the bindings between perspectives and textual content, Bi-KGNPS jointly learns perspective representations tailored to the current text (content-oriented perspective features), as well as salient word features under each perspective (perspective-oriented text representations).

This work is a substantial extension of our previous work NPS4SD [9]. We propose two key extensions to NPS4SD: first, we expand the task to the more realistic zero-shot setting and pioneer the unsupervised acquisition of target-independent transferable knowledge without annotations; second, we devise a Bi-KGNPS that infuses transferable knowledge more effectively than the original NPS4SD. KAI offers two key advantages over NPS4SD: 1) it leverages LLMs to automatically generate multidimensional perspective knowledge, reducing manual effort and broadening coverage compared to the singular, manually defined rules in NPS4SD; and 2) KAI's Bi-KGNPS enables dynamic, iterative fusion, and reasoning between knowledge and text for stance prediction, surpassing the simpler, less collaborative knowledge binding mechanism in NPS4SD. Finally, experiments on strong baselines validate that our proposed KAI achieves state-of-the-art performance in ZSSD. Moreover, we demonstrate the generalizability of KAI in both in-target and cross-target setups.

In summary, this article presents several contributions as follows.

- 1) We propose a KAI framework that first introduces an analytical perspective as common target-relevant knowledge to bridge feature transfer across targets.
- 2) We also propose a Bi-KGNPS network that utilizes perspective features through variable binding to facilitate knowledge-guided stance detection. To our knowledge, we are the first to introduce NPSs into stance detection, advancing interpretable stance detection.
- 3) We carry out extensive experiments on publicly available benchmark datasets, and the results demonstrate our proposed model's superiority over state-of-the-art competitors.

II. RELATED WORK

A. Stance Detection

1) *In-Target and Cross-Target Stance Detection*: Stance detection aims to detect the attitude of a context (e.g., comment or review) according to the given target, which is critical to many scenarios such as argumentation mining, fake news detection, and fact-checking [10], [11], [12], [13]. Earlier work in stance detection primarily focused on in-target stance detection, where the training and testing targets are the same. For in-target stance setup, conventional methods can be classified into non- and pretrained methods. The non-pretrained methods mainly conduct deep neural networks, such as traditional attention-based models and graph-based models to train a stance classifier. The attention-based methods mainly utilize target-specific information as the attention query, and deploy an attention mechanism

for inferring the stance polarity [14]. Another approach is the graph-based text classification methods, which propose a graph convolutional network (GCN) to model the relation between target and text [15], [16]. Subsequently, several studies are also being conducted for cross-target stance detection tasks, which can be classified into two categories. The first class of methods is word-level transfer, which uses the common words shared by two targets to bridge the knowledge gap [17]. Second, some approaches handle this cross-target problem with concept-level knowledge shared by two targets [18], [19].

2) *Zero-Shot Stance Detection*: In practical applications, predefining and annotating complete targets is infeasible, thus spurring interest in ZSSD among researchers. Allaway and McKeown [20] first provided a large-scale human-labeled stance detection dataset for zero-shot scenarios. Subsequently, Allaway and McKeown [20] developed a target-specific stance detection dataset for ZSSD and applied adversarial learning to extract target-invariant information. Liu et al. [5] proposed a common sense knowledge-enhanced graph model grounded on BERT to leverage both inter- and extra-semantic information. Liang et al. [6] presented an effective approach to distinguish target-invariant or target-specific features to improve learning of transferable stance features. Several studies have explored the application of prompt tuning in stance detection [21], [22]. Jiang et al. [21], for example, pioneered the target-aware prompt distillation (TAPD) framework, applying prompt tuning for stance detection. In this framework, a verbalizer function maps each label to a hidden vector, enabling label prediction. Hardalov et al. [22] proposed a prompt-based framework specifically designed for cross-lingual stance detection.

B. Chain-of-Thought Model (CoT)

Recent works have explored enhancing CoT prompting to elicit impressive multihop reasoning from LLMs [23], [24], [25]. For example, Cai et al. [26] proposed a human-in-the-loop system augmented with CoT prompting, investigating how manual correction of sublogic in rationales can refine LLM reasoning. Fei et al. [27] introduced a multistep CoT approach that decomposes downstream tasks into multiple stages to improve prediction effectiveness. Ling et al. [28] presented a new CoT technique, which iteratively infers tasks via deductive reasoning and verification. Inspired by these multistep CoT techniques, we propose a novel knowledge elicitation method that effectively generates the perspectives and rationales of the targets.

C. Interpretable Neural Network for NLP

In recent years, developing interpretable models has attracted considerable interest in NLP. For tasks closely related to stance detection such as sentiment analysis and fact checking, interpretable models are widely studied, as they can effectively analyze the prediction process and incorporate knowledge such as prior knowledge.

For instance, Ito et al. [29] proposed an interpretable sentiment model that explicitly demonstrates the analysis process for sentiment analysis. Huang et al. [30] presented a sentiment interpretable logic tensor network, which encodes the prediction process following first-order logic rules. Yadav et al. [31]

developed the Teslin machine to achieve human-level interpretable learning. Guo et al. [32] proposed the Interpretable fake news detection with graph evidence (IKA) model for fake news detection that represents the rationale as a directed graph.

The NPS provides an architectural framework to efficiently identify and infer entity representations in input texts, while governing the interactions between these entities [33]. By dynamically selecting rules and establishing bindings between rules and entities, the control flow is determined. The variable binding method enables acquiring diverse rule-entity patterns. Ultimately, this combination of rule-entity patterns facilitates constructing the model analysis process.

III. METHODS

As shown in Fig. 1, KAI primarily consists of two components: LLM-KE and Bi-KGNPS. We give the task definition and the overview of our model in Sections III-A and III-B, respectively. Then, we describe the details of the LLM-KE and Bi-KGNPS in Sections III-C and III-D, respectively.

A. Problem Definition

Let $X = \{x_i, q_i\}_{i=1}^N$ represent the labeled data collection, where x refers to the input text and q corresponds to the source target. N denotes the total number of instances in X . Each sentence-target pair $(x, q) \in X$ is assigned a stance label y . ZSSD aims to predict the stance polarity of the input sentence x^u toward the given target q^u using a model trained on the labeled data X , where x^u and q^u are unseen during the training process.

B. Framework Overview

As shown in Fig. 1, KAI consists of two main components: LLM-KE and Bi-KGNPS. LLM-KE proposes a method based on CoT prompting that elicits target-relevant analytical perspectives and predicts the rationale behind each perspective from LLMs in two stages. Such knowledge can serve as target-independent transferable knowledge. Bi-KGNPS is a dual-branch network where information dynamically interacts between the two branches to enable knowledge-variable binding for elucidating the prediction process.

C. LLM-KE

LLM-KE is designed to elicit target analysis perspectives and rationale for each perspective by formulating instructions for the LLM. Specifically, LLM-KE involves the following steps.

1) *Step 1*: We first input the constructed instruction $S'1$ into the LLM to elicit the perspective v for the LLM's prediction. Here, γ is a hyperparameter that determines the number of perspectives to be acquired.

$S'1$: Please enumerate γ distinct perspectives from which stances towards the provided target [given target: q] may be stated.

2) *Step 2*: After obtaining γ perspectives, we use the LLM to generate rationales, denoted as r , for the stance analysis corresponding to each perspective.

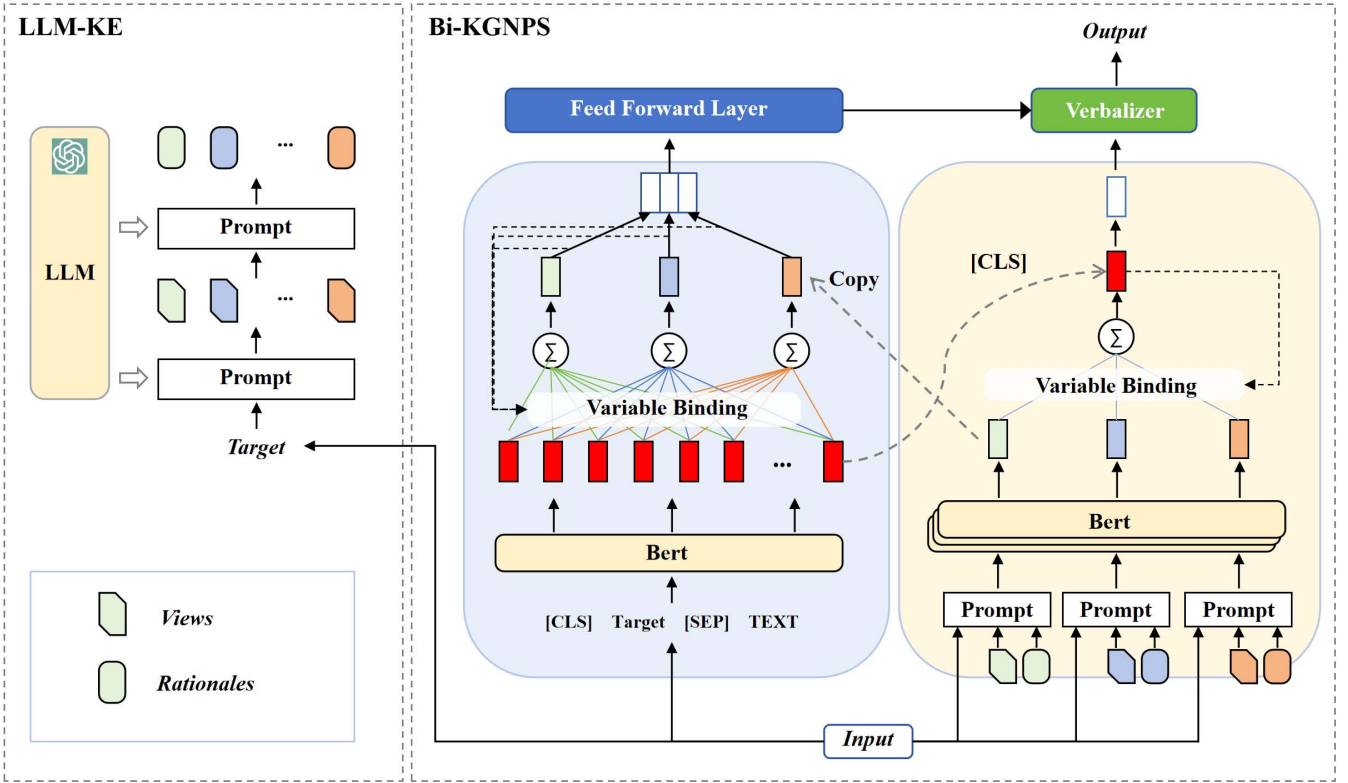


Fig. 1. Overall framework of KAI network for ZSSD.

TABLE I
EXAMPLE OF PERSPECTIVES AND RATIONALES FROM LLM-KE

Target	Hillary Clinton
Tweet	Hilary has lied, deleted Benghazi emails, and betrayed the trust of Americans scandal after scandal. <i>[Against]</i>
Perspective	Personal characteristics
Rationale	... This negative portrayal of her personal characteristics suggests that she may not be trustworthy and may have a tendency to deceive people. ...
Perspective	Political views
Rationale	... From a political views angle, this could be seen as a negative reflection on her ability to lead America ...

S'2: Oriented to the [given target: q], given the input [given input: x], and under the [given perspective: $v \in \gamma$], give the stance analysis thinking or explanation. Give a stance judgment (favor, against, none) at the end.

Examples of LLM-KE results are provided in Table I. For the target “Hillary Clinton,” when two perspectives are selected, LLM-KE generates the aspects of “personal characteristics” and “political views.” LLM-KE can then provide stance reasoning from both perspectives.

D. Bi-KGNPS

As shown in Fig. 2, Bi-KGNPS consists of two branches. The left branch takes the input text and its corresponding stance target, treating perspective features as knowledge to select important text features and to learn perspective-oriented text representations. The right branch processes the perspectives and

rationales, using the text content as knowledge to construct variable bindings for selecting the most relevant perspectives to the text and to learn content-oriented perspective features. Finally, a feedforward network layer is employed to fuse both sets of features for the final stance polarity prediction.

1) Left Branch:

a) *Embedding layer*: The left branch first learns vector representations of the inputs using a BERT model. Specifically, we format the target and text into the following input pattern for BERT: “[CLS] + target + [SEP] + text,” where “target” and “text” represent their respective texts. After inputting this pattern into BERT, we obtain the textual representation vector $H = \{h_{[\text{cls}]}, h_1, \dots, h_t^m\}$, where $h_{[\text{cls}]}$ is the hidden vector of the [CLS] token, h_1 to h_a denote the vectors of the target, and h_{a+1} to h_t^m denote the vectors of the text.

b) *Variable binding layer*: Subsequently, we treat the perspective representation vectors as query vectors for an attention mechanism. By constructing an attention mechanism, we can learn perspective-oriented textual representations. Since multiple perspectives simultaneously build attention, inspired by the work of [34], we propose a novel multihop attention mechanism. Formally, assuming we have three perspectives with feature vectors denoted as R_i , R_j , and R_k (details of obtaining them are described in the right branch), the attention coupling factors c can be computed as

$$c_i^1 = R_i^1(H_i^1)^T, \quad c_j^1 = R_j^1(H_j^1)^T, \quad c_k^1 = R_k^1(H_k^1)^T. \quad (1)$$

Initially, the input text is identical across the three attention queries, such that $H_i^1 = H_j^1 = H_k^1 = H$. Then, the next iteration

can then be computed as follows:

$$R_i^2 = c_i^1 H_i^1, \quad R_j^2 = c_j^1 H_j^1, \quad R_k^2 = c_k^1 H_k^1 \quad (2)$$

where the query $H_{i,j,k}^2$ preserves the same dimensionality as the previous query $R_{i,j,k}^1$. The knowledge-variable binding pattern representation for the next round is calculated as follows:

$$\begin{aligned} H_i^2 &= \lambda \mathcal{LN}(\sigma(R_i^2)) + H_i^1 \\ H_j^2 &= \lambda \mathcal{LN}(\sigma(R_j^2)) + H_j^1 \\ H_k^2 &= \lambda \mathcal{LN}(\sigma(R_k^2)) + H_k^1 \end{aligned} \quad (3)$$

where \mathcal{LN} performs the standard layer normalization. σ denotes the Sigmoid function. After t iterations, where t indicates the number of steps in the process, the representation q can be calculated as

$$q_i = H_i^t \mathcal{S}\left(\sum_{\gamma} c_i^t\right), \quad q_j = H_j^t \mathcal{S}\left(\sum_{\gamma} c_j^t\right), \quad q_k = H_k^t \mathcal{S}\left(\sum_{\gamma} c_k^t\right) \quad (4)$$

where $\mathcal{S}(z) = e^z / \sum_k e^{z_k}$ denotes the softmax function.

After obtaining representations q_i , q_j , and q_k , the final perspective-oriented text representation is calculated by concatenating these patterns. We use a straight-through Gumbel SoftMax approach [35] for a learnable hard decision during training

$$\begin{aligned} U &= \{q_i \oplus q_j \oplus q_k\} \\ \alpha &= \operatorname{argmax}_{\tau}(U_{\tau} W + \phi), \quad \phi \sim \text{Gumbel}(0, 1) \\ \mathcal{E} &= \sum_{t=1}^n \alpha_t U_{\tau} \end{aligned} \quad (5)$$

where \oplus denotes the concatenation operator, and W is a learnable weight matrix. The Gumbel SoftMax distribution is introduced to allow for differentiability of the ‘‘argmax’’ operation during backpropagation by sampling from a Gumbel distribution parameterized by location 0 and scale 1. Ultimately, the weighted sum \mathcal{E} is passed through a fully connected layer to derive the final perspective-oriented text representation, denoted by e_p .

2) *Right Branch*: The right branch is dedicated to integrating knowledge from multiple perspectives. We employ a soft-prompt network to encapsulate data across these varying perspectives. This network utilizes tailored templates that blend perspectives and predicted rationales, which, when fed into a BERT model, yield vector representations for perspective features. One key benefit of this approach is the improved sentence encoding that can be achieved using hidden vectors from [MASK] token positions, as demonstrated by prior research [36].

In practice, we construct a text template p that embeds the perspectives v and target q into a single prompt. Each prompt is formally represented as follows: ‘‘ $x_p =$ From the perspective of [given view v], [given rationale r], the attitude toward [target] is [MASK].’’ After formulating the template, we introduce a technique that maps the perspective knowledge onto a series of continuous vectors. For each template x_p^i , we input it into

an LLM M to extract the hidden vector corresponding to the [MASK] token, denoted by R_i . Subsequently, our approach generates a set of γ distinct vectors ($R = \{R_i\}^{\gamma}$), which are associated with the hidden vectors produced by the [MASK] token of each template x_p . These vectors are treated as trainable parameters and are refined during the model’s optimization process. To maintain compatibility with the token embeddings in the BERT, we ensure these stance vectors match the dimension of the BERT’s hidden embeddings.

We then employ the hidden representation $h_{[\text{cls}]}$ as the attention query to determine the attention weight α_t for the t th vector in R

$$\alpha_t = \operatorname{softmax}(h_{[\text{cls}]}^T R_t) \quad (6)$$

$$e_o = \sum_{t=1}^n \alpha_t R_t. \quad (7)$$

Here, e_o is the aggregated feature vector that represents the content-oriented perspective knowledge, a composite of the weighted perspective vectors.

E. Stance Classification Layer

To improve the accuracy of stance detection, we utilize a verbalizer-based approach. A verbalizer is a component that maps each label or stance category to a set of semantically related words. This mapping allows the model to evaluate the semantic similarity between the concatenated text representation vectors and the verbalizer’s label vectors.

The embedding of a word selected by the verbalizer is denoted as w . For each word provided by the verbalizer, we estimate the probability that token w represents the label word. Here, the probability calculation is as follows:

$$\delta = \frac{\exp(w_i^T (e_o \oplus e_p))}{\sum_{w_j \in V_{\text{erb}}} \exp(w_j^T (e_o \oplus e_p))} \quad (8)$$

where w_i is the embedding of the i th token in the verbalizer, and V_{erb} represents the vocabulary of the verbalizer’s tokens. Here, e_o is the feature vector representing the content-oriented perspective, while e_p represents the perspective-oriented text representation. We then calculate the sum of the probabilities for the words that correspond to each label, which is denoted as \hat{y} . Specifically, for each label, we aggregate the probabilities assigned to all words mapped to that label by the verbalizer.

The model is trained to minimize the discrepancy between the predicted and actual stance labels using a cross-entropy loss function

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

where N is the number of samples in the training set, and C is the number of stance classes, and y_{ij} is the ground-truth label in one-hot encoded form for the i th sample and j th class. The optimization of the attention layer and the rest of the model parameters are achieved using the gradient descent algorithm.

TABLE II
STATISTICS OF SEM16 AND P-STANCE DATASETS

Dataset	Target	Favor	Against	Neutral
Sem16	HC	163	565	256
	FM	268	511	170
	LA	167	544	222
P-Stance	Biden	3217	4079	-
	Sanders	3551	2774	-
	Trump	3663	4290	-

TABLE III
STATISTICS OF VAST DATASET

	Train	Dev	Test
#Examples	13 477	2062	3006
#Unique comments	1845	682	786
#Zero-shot topics	4003	383	600
#few-shot topics	638	114	159

IV. EXPERIMENTS

In this section, we introduce the evaluation metrics employed in the experiments and the baseline methods used during the experiments.

A. Experimental Data

To evaluate the effectiveness of our method, we conduct extensive experiments on SemEval-2016 Task 6 (SEM16) [37] and VAST [3]. SEM16 contains six predefined targets across multiple domains, including Donald Trump (DT), Hillary Clinton (HC), feminist movement (FM), legalization of abortion (LA), atheism (A), and climate change (CC). Each instance could be classified as “favor, against, or neutral.” Following [18], we remove targets A and CC due to data quality issues and regard one target as the zero-shot testing target while training on the other targets. We randomly selected 15% of the training set as the development data to tune hyperparameters. P-stance [38] comprises tweets pertaining to three prominent politicians: Joe Biden (Biden), Bernie Sanders (Sanders), and Donald Trump (Trump). As observed by the authors, the annotation consistency for samples labeled “none” is low. Consistent with previous work, we exclude these “none”-labeled samples from our analysis. VAST contains a large number of diverse targets. Each data instance is comprised of a sentence, a target, and a stance polarity toward the target, which can be “pro, con, or neutral.” Following [8], we evaluate our model’s performance on zero-shot topics with 100% and 10% training sizes, respectively. The statistics of three datasets are shown in Tables II and III.

B. Evaluation Metrics

We use F_{avg} as the evaluation metric to evaluate the performance of our baseline models, the same as most previous works [9], [19], [38]. First, the F1-score of label “favor” and “against” is calculated as follows:

$$F_{\text{favor}} = \frac{2 \cdot P_{\text{favor}} \cdot R_{\text{favor}}}{P_{\text{favor}} + R_{\text{favor}}} \quad (10)$$

$$F_{\text{against}} = \frac{2 \cdot P_{\text{against}} \cdot R_{\text{against}}}{P_{\text{against}} + R_{\text{against}}} \quad (11)$$

where P and R are the precision and recall, respectively. After that, the F_{avg} is calculated as

$$F_{\text{avg}} = \frac{F_{\text{favor}} + F_{\text{against}}}{2}. \quad (12)$$

C. Baseline Methods

To gain a fundamental understanding of how existing stance detection models perform on our dataset, we employed the following models.

- 1) *Bicond* [17] utilizes two BiLSTM models to individually encode the underlying sentences and their associated targets.
- 2) *SEKT* [19] leverages external knowledge, particularly semantic and emotion lexicons, to facilitate knowledge transfer across diverse targets for cross-target stance detection.
- 3) *CrossNet* [39] encodes text and topic using identical bidirectional architectures as BiLSTM and incorporates a target-specific attention layer prior to classification for cross-target stance detection.
- 4) *TPDG* [40] presents a novel technique that autonomously distinguishes and calibrates the target-dependent and target-independent roles of a term regarding a specific target within stance expressions.
- 5) *TOAD* [4] employs adversarial learning to accomplish generalization across various topics in ZSSD.
- 6) *TGA Net* [3] enables implicit construction and utilization of associations between training and evaluation topics without supervision. It utilizes BERT for encoding texts and targets, followed by classification through two fully connected layers.
- 7) *BERT-GCN* [5] presents a commonsense knowledge-enhanced model harnessing both structural and semantic relational knowledge to augment generalization and reasoning capabilities under zero-shot and few-shot settings.
- 8) *BERT-Joint* [41] refers to bidirectional encoder representations from transformers, which are pretrained language models encoded on large unlabeled corpora to represent sentences and tokens as dense vectors.
- 9) *JoinCL* [7] comprises stance contrastive learning and target-aware prototypical graph contrastive learning for generalizing target-dependent stance features to unseen targets.
- 10) *PT-HCL* [6] proposes a contrastive learning approach utilizing both semantic and sentiment knowledge to improve cross-domain transferability of features.
- 11) *TarBK* [42] incorporates specific background knowledge from Wikipedia to reduce the semantic gap between known and unseen targets, enhancing generalization and reasoning.
- 12) *SDAgu* [43] presents a self-supervised data augmentation method leveraging coreference resolution to mitigate limited labeled data and implicit stance expression.
- 13) *MPT* [44] develops prompt-tuning-based PLM to perform stance detection, where humans define the verbalizer.

TABLE IV
EXPERIMENTAL RESULTS ON TWO ZSSD DATASETS

	Model	SEM16				P-Stance			VAST (100%)			VAST (10%)		
		HC	FM	LA	DT	Biden	Sanders	Trump	Pro	Con	All	Pro	Con	All
Sta.	Bicond	32.7 [‡]	40.6 [‡]	34.4 [‡]	30.5 [‡]	50.1	52.6	50.4	44.6 [‡]	47.4 [‡]	42.8 [‡]	29.8 [◇]	40.1 [◇]	34.8 [◇]
	CrossNet	38.3 [‡]	41.7 [‡]	38.5 [‡]	35.6 [‡]	52.8	53.1	52.8	46.2 [‡]	43.4 [‡]	43.4 [‡]	37.3 [◇]	32.9 [◇]	36.2 [◇]
	SEKT	50.1	44.2	44.6	46.8	64.4	61.3	60.8	50.4 [‡]	44.2 [‡]	41.8 [‡]	-	-	-
	TPDG	50.9 [‡]	53.6 [‡]	46.5 [‡]	47.3 [‡]	64.4	62.0	60.1	53.7 [‡]	49.6 [‡]	51.9 [‡]	-	-	-
	TOAD	51.2 [‡]	54.1 [‡]	46.2 [‡]	49.5 [‡]	-	-	-	42.6 [‡]	36.7 [‡]	41.0 [‡]	-	-	-
BERT	TGA Net	49.3 [‡]	46.6 [‡]	45.2 [‡]	40.7 [‡]	74.7	69.9	64.4	53.7	62.0	67.4	47.6 [◇]	58.2 [◇]	64.1 [◇]
	BERT-Joint	50.1	42.1	44.8	41.0	75.0	71.1	67.2	56.8	67.6	71.0	50.7	56.3	64.9
	BERT-GCN	50.0 [‡]	44.3 [‡]	44.2 [‡]	42.3 [‡]	74.9	71.1	70.5	58.3 [‡]	60.6 [‡]	68.6 [‡]	-	-	-
	JointCL	54.4	54.0	50.0	50.5	-	-	-	64.9	63.2	71.2	53.8 [◇]	57.1 [◇]	65.5 [◇]
	TarBK	55.1 ^N	53.8 ^N	48.7 ^N	50.8 ^N	75.5	70.5	65.8	65.7 ^N	63.9 ^N	73.6 ^N	-	-	-
	PT-HCL	54.5 [‡]	54.6 [‡]	50.9 [‡]	50.1 [‡]	-	-	-	61.7 [‡]	63.5 [‡]	71.6 [‡]	-	-	-
	SDAgu	-	-	-	-	-	-	-	60.1	65.2	71.3	-	-	-
	Openstance*	-	-	-	-	-	-	-	63.1	66.4	73.1	61.4	60.1	70.2
	TTS*	-	-	-	-	-	-	-	59.5	65.2	71.4	55.8	65.5	70.3
	NPS4SD	60.1	56.7	51.0	51.4	76.4	72.9	70.4	69.2	67.6	74.4	68.7	64.3	72.0
	Ours (KAI) [†]	76.4	73.7	69.4	72.1	85.7	80.5	75.9	66.7	73.0	76.3	63.5	74.0	75.2

Note: Results with [‡], [◇], and ^N are retrieved from [6], [8], and [43], respectively. * indicates that we utilize BERT as the backbone classifier for a fair comparison. [†] refers to a p -value < 0.05 . Best scores are in bold.

- 14) *KEPrompt* [45] proposes an automatic verbalizer to automatically define the label words and a background knowledge injection method to integrate the external background knowledge.
- 15) *Ts-Cot* [46] develops a CoT method for LLMs to perform stance detection. Due to the evolution of LLM model versions, the foundation model is upgraded to the GPT-3.5 baseline.

V. EXPERIMENTAL RESULTS

A. Result

To evaluate the stability of the model, following [19], we ran the method three times and reported the F1-score. The main experimental results of ZSSD on two benchmark datasets are reported in Table IV. We observe that our method consistently outperforms all baselines on all datasets, which verifies the effectiveness of our proposed approach in ZSSD. Furthermore, compared with previous models, our method shows statistically significant improvements (p -value < 0.05) in most evaluation metrics, validating the effectiveness of our proposed approach in ZSSD.

Specifically, we first observe that most neural-network-based methods (statistics-based and fine-tuning-based methods) perform poorly under zero-shot settings, owing to their reliance on seen targets and samples. Notably, directly introducing pre-trained BERT did not improve performance, likely because finetuning-based BERT methods cannot capture implicit stance expressions absent from the text, such as background knowledge and topics related to the target. After incorporating background knowledge, prediction accuracy significantly increased. For instance, compared with the best BERT finetuning model BERT-Joint, the knowledge-infused approach TarBK improved by 2.6% on average. This validates the efficacy of incorporating background knowledge. Additionally, we observed that contrastive learning methods (e.g., JointCL) outperformed finetuning models. This is because contrastive learning can effectively

capture domain-invariant representations, thereby improving model transferability across domains.

The proposed KAI method yields better performance than all the baselines in all the tasks. For instance, our proposed approach achieves substantial gains in average $F1_{avg}$ over the current state-of-the-art (NPS4SD), increasing by 18.1% on SEM16, 7.5% on P-Stance, and 3.2% on the VAST dataset (10%). The advantage of KAI comes from its two characteristics: i) we propose perspectives as target-independent knowledge that bridges target differences; ii) we devise a perspective acquisition mechanism via COT applicable under zero-shot settings; and iii) we devise a Bi-KGNPS network to jointly capture the associative relationships between text and perspective, to better leverage relevance for improved performance.

We further benchmarked our KAI model against competing methods that utilize CoT for LLMs, such as GPT-3.5 and Llama2. The results are illustrated in Fig. 2. The results demonstrate that the proposed KAI method can significantly enhance prediction accuracy over both LLM frameworks, thoroughly validating the efficacy of KAI integration for improving LLM performance. In particular, with Llama2 as the base, KAI attained consistent, considerable gains across all settings, averaging a 3.48% boost over multiple zero-shot task configurations with Ts-Cot-Llama2. When built on GPT-3.5, KAI also achieved marked improvements across most tasks.

B. Ablation Study

To study the influence of each component of our model, we implement an ablation test for KAI by removing the external perspective knowledge (denoted as w/o KEM), the NPS (denoted as w/o NPS), the left branch (denoted as w/o LB), and the right branch (denoted as w/o RB).

Specifically, for w/o KEM, we remove the multiperspective knowledge, so the structure becomes a conventional interactive attention-based model backbone. For w/o NPS, we concatenate the perspectives with the original input in the format of “[SEP] text [SEP] perspectives” and feed it into the BERT model. For

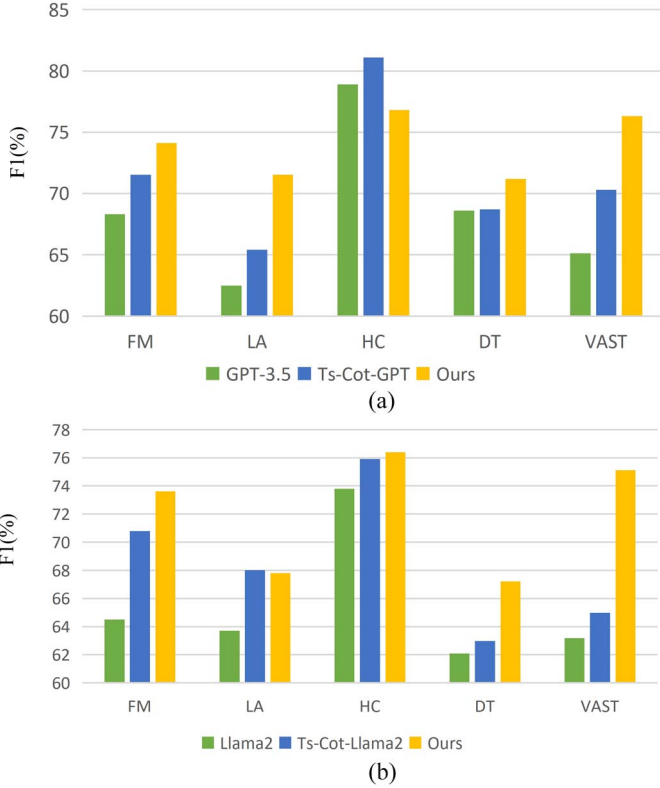


Fig. 2. Comparative evaluation of LLM experimental performance. (a) In comparison with GPT-3.5. (b) In comparison with Llama2.

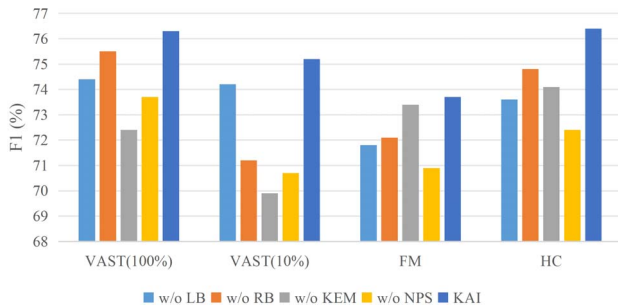


Fig. 3. Performance of ablation study.

w/o LB and w/o RB, we construct separate knowledge-guided variable binding methods, similar to those discussed in [48]. The former selects text features based on analyzed perspectives as queries, while the latter uses the text as queries to select perspective features.

The results are summarized in Fig. 3 and show that all proposed components significantly improve the performance of KAI. Notably, perspective knowledge has the greatest impact on the effectiveness of KAI. Discarding KEM severely decreases classification accuracy, as expected. This decrease occurs because perspectives serve as domain-invariant knowledge, effectively bridging the gap between different targets. Incorporating perspective features leads to substantial performance gains from both LB and RB components. These gains are attributed to

TABLE V
PERFORMANCE COMPARISON OF CROSS-TARGET STANCE DETECTION

	Methods	FM→LA	LA→FM	HC→DT	DT→HC
Sta.	BiCond	45.0	41.6	29.7	35.8
	CrossNet	45.4	43.3	43.1	36.2
	SEKT	53.6	51.3	47.7	42.0
	TPDG	58.3	54.1	50.4	52.9
	BERT-Joint	48.2	34.4	45.9	42.7
BERT	MPT	42.1	47.6	47.1	58.7
	KEPROMPT	49.1	54.2	54.6	60.9
	JointCL	58.8	54.5	52.8	54.3
	PT-HCL	59.3	54.6	53.7	55.3
	TarBK	59.1	54.6	53.1	54.2
	NPS4SD	58.9	57.2	64.6	66.2
	Ours (KAI)	71.9	74.8	73.6	75.4

Note: Best scores are in bold.

TABLE VI
PERFORMANCE COMPARISON OF
IN-TARGET STANCE DETECTION

Methods	SEM16		
	FM	LA	HC
TPDG	67.3	74.7	73.4
BERT-FT	62.3	62.4	67.0
STANCY	61.7	63.4	64.7
MPT	63.3	63.5	71.3
KEprompt	72.1	69.1	74.4
NPS4SD	68.1	69.6	79.1
Ts-Cot	67.2	62.5	83.2
Ours (KAI)	74.4	74.9	86.3

Note: Best scores are in bold.

the components' abilities to learn text features from specific perspectives and perspective features from specific content, providing more target-agnostic transferable knowledge. Text features from specific perspectives contribute to representing target-invariant lexical knowledge, while perspective features from specific content assist in selecting the most suitable analytical perspectives, thereby enhancing the model's overall performance. As anticipated, integrating all components yields the best results across all experimental datasets.

C. Cross-Target Stance Detection Scenario

To evaluate the generalizability of our KAI model, we assessed its cross-target performance on the SEM16 dataset. As shown in Table V, KAI outperforms all baselines under cross-target settings, validating the efficacy and generality of KAI for stance detection. Notably, when comparing the results of Table V with those in Table IV, we observe that KAI shows more improvement under cross-target conditions than in zero-shot settings. This is likely because targets in the cross-target setting are more domain-related; for example, perspectives largely overlap for Trump-Clinton (FM-HC). Consequently, greater gains are achieved when there is higher domain relevance, as the model can leverage the commonalities in perspectives between related targets to make more informed predictions.

D. In-Target Stance Detection Scenario

Our evaluation extended to the in-target stance detection scenario using the SEM16 dataset. The experimental results, as

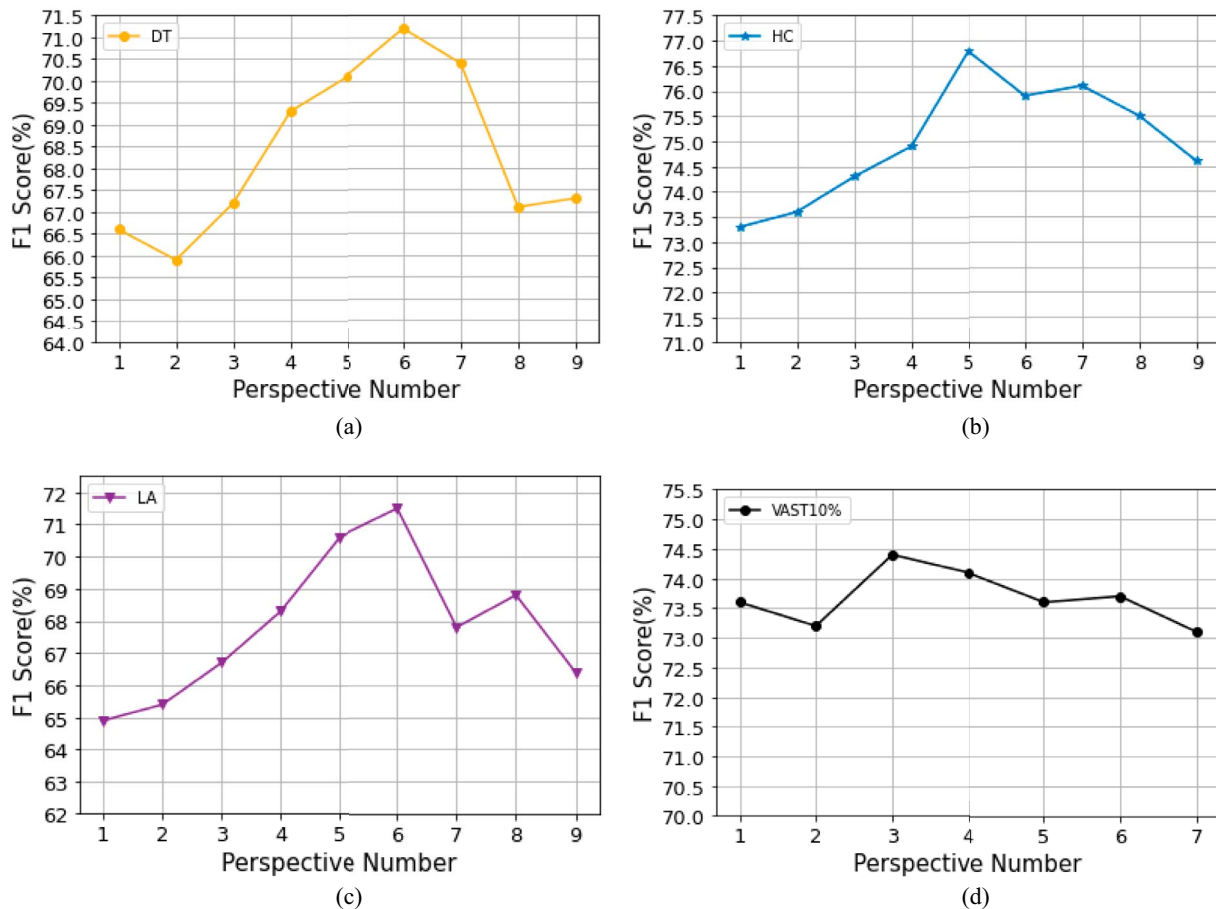


Fig. 4. Performance of perspective numbers. (a) Donald Trump. (b) Hillary Clinton. (c) Legalization of abortion. (d) VAST (10%).

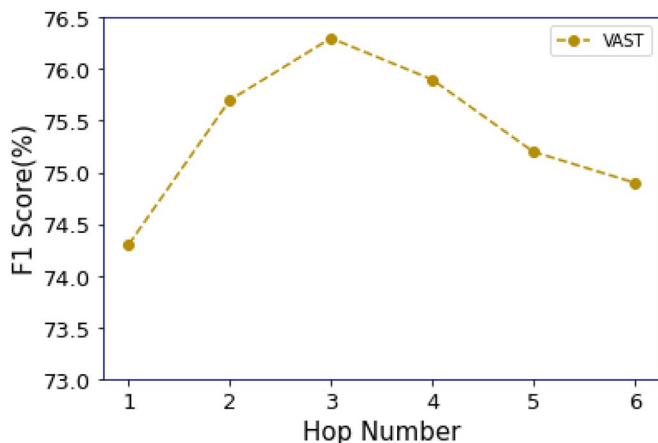


Fig. 5. Performance of hop numbers with VAST dataset.

detailed in Table VI, reveal that our KAI model outperforms all compared methods. In the in-target setup, the enhanced performance gains from KAI can likely be attributed to its multiperspective modeling approach. This approach equips the model with the ability to assimilate and integrate background knowledge from various dimensions related to the stance target, thus enriching the model’s analytical depth. In summary, the

results from the in-target scenario—alongside those from zero-shot and cross-target conditions—reinforce the effectiveness and generalizability of our KAI framework in diverse stance detection contexts.

E. Impact of Perspective Numbers

Perspective information constitutes a pivotal component within KAI, with the number of perspectives profoundly impacting stance detection accuracy. Hence, we conduct further analyses on the effects of varying perspective counts on stance detection performance. Experiments are shown in Fig. 4. The results demonstrate that under different targets, knowledge integration across multiple perspectives substantially influences stance detection accuracy. It can be observed that the number of perspectives holds important ramifications for stance detection capabilities. Initially, accuracy climbs gradually as perspectives augment, owed to the target-invariant essence of perspective-based analysis that effectively bridges feature transfer across targets. However, as the perspective parameter scales, performance peaks then deteriorates, potentially due to excessively fine-grained perspective partitioning causing perspective traits to become overly target-specific. For instance, given three perspectives for Hillary target with granularity of “gender, politics,” etc., six perspectives lead to finer divisions such as

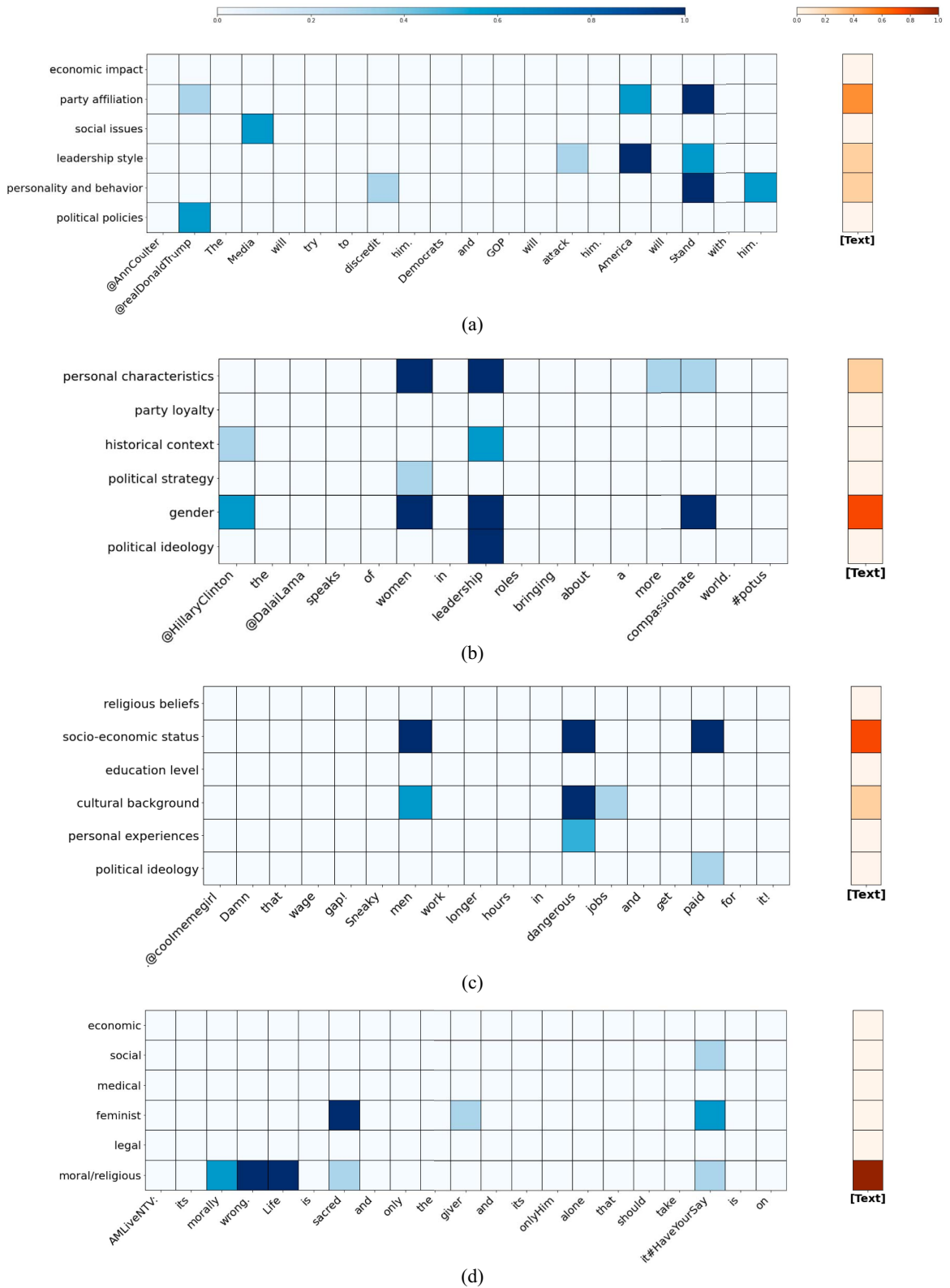


Fig. 6. Case study. (a) Donald Trump. (b) Hillary Clinton. (c) Feminist Movement. (d). Legalization of Abortion.

“woman, presidential candidate” which resemble target-specific knowledge. It is foreseeable that as granularity becomes increasingly specific, performance declines in zero-shot scenarios as perspective knowledge becomes more target-specific.

F. Analysis of Routing Iterations

The number of routing iterations constitutes a pivotal hyperparameter within KAI that is unique to the NPS architecture and profoundly impacts model performance and runtime. In

this experiment, we conduct an analysis to ascertain the optimal routing iteration quantity across all three dataset configurations. The results are shown in Fig. 5. Experiments are undertaken with routing iterations spanning (1, 2, 3, 4, 5, and 6). The empirical results demonstrate that peak performance is attained with iteration counts between 1 and 5. Beyond three iterations, overfitting phenomena emerge leading to deteriorating accuracy.

G. Case Study

To better illustrate how our method works, we conduct four case studies on examples where KAI accurately predicts the target while contrastive methods mostly fail. Specifically, we visualize the coupling factors c of LB and RB. As shown in Fig. 6, the left part (blue) indicates which words are selected by the viewpoint, and the right part (orange) indicates which viewpoint is selected for the full text. Visualization of the combined attention coupling factors enables analysis of the model’s decision-making process.

The first sample examines subtopics such as media, the Democratic Party, and the Republican Party’s positions toward the target “Donald Trump,” expressing support. Conventional BERT models struggle to effectively understand the relationships between subtopics and the target, incorrectly classifying the stance as “against.” However, Fig. 6(a) shows that the KAI model focuses on “party affiliation” as the main viewpoint and “leadership style” and “personality and behavior” as secondary viewpoints, which are highly relevant to the text. KAI’s attention to these salient viewpoints allows it to correctly determine the stance as supportive of Donald Trump.

The second sample, see Fig. 6(b), explores the idea that having women in leadership roles helps to create a more compassionate world, expressing support for Hillary Clinton. Traditional methods struggle to deeply understand the nuances of the tweet content, incorrectly classifying the stance as “neutral.” In contrast, our KAI model effectively attends to the “gender” and “personal characteristics” viewpoints, and under these viewpoints, focuses on salient keywords such as “women, leadership” and “compassionate.” This allows KAI to accurately identify the stance. By attending to relevant viewpoints and keywords, KAI captures the underlying semantics and makes accurate predictions.

The third sample discusses the reasons behind wage disparities, stating that men’s jobs are more dangerous and time-consuming, expressing opposition to the feminist movement. Conventional models struggle to recognize implicit expressions and misclassify the stance as “neutral.” Our model primarily adopts a “socio-economic status” perspective, with “cultural background” as a secondary perspective, concentrating on keywords such as “men, dangerous, and paid” to provide an accurate classification.

The fourth sample morally criticizes “abortion” from the perspective of emphasizing values such as life and choice, despite not explicitly mentioning the term “abortion.” Existing methods struggle to deeply understand the connection between the text and target, failing to effectively link the text to abortion

and incorrectly classifying the stance as “neutral.” In contrast, KAI accurately captures salient keywords such as “moral” and “life” and links them to the legalization of abortion through the “moral/religious” viewpoint. Meanwhile, the viewpoint also appropriately ignores distracting words such as “mistake,” thus correctly identifying the stance polarity. By attending to relevant keywords and viewpoints, KAI makes the connection between the implicit text and abortion, enabling accurate stance detection.

VI. CONCLUSION

This article presented a novel KAI network to advance stance detection, particularly for the challenging zero-shot setting. The key innovation is eliciting target-independent conceptual knowledge as transferable features using LLMs. Specifically, we introduced analysis perspectives shared by different targets as such common knowledge. This allows bridging connections between seen and unseen targets to enable knowledge transfer. Moreover, we developed a Bi-KGNPS that integrates this knowledge through an iterative knowledge-variable binding process to guide stance predictions. Extensive experiments demonstrate KAI achieves new state-of-the-art accuracy on multiple ZSSD benchmarks. Our approach also delivers strong performance on conventional in-target and cross-target setups. The proposed framework provides an interpretable approach to effectively infuse domain knowledge for transfer learning, representing an important step toward practical stance detection systems that can generalize to new targets of interest.

REFERENCES

- [1] D. Küçük and F. Can, “Stance detection: A survey,” *ACM Comput. Surv.*, vol. 53, no. 1, pp. 1–37, 2020.
- [2] N. Yin et al., “Deal: An unsupervised domain adaptive framework for graph-level classification,” in *Proc. 30th ACM Int. Conf. Multimedia*, 2022, pp. 3470–3479.
- [3] E. Allaway and K. Mckeown, “Zero-shot stance detection: A dataset and model using generalized topic representations,” in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2020, pp. 8913–8931.
- [4] E. Allaway, M. Srikanth, and K. Mckeown, “Adversarial learning for zero-shot stance detection on social media,” in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Human Lang. Technol.*, 2021, pp. 4756–4767.
- [5] R. Liu, Z. Lin, Y. Tan, and W. Wang, “Enhancing zero-shot and few-shot stance detection with commonsense knowledge graph,” in *Proc. Findings Assoc. Comput. Linguistics: ACL-IJCNLP*, 2021, pp. 3152–3157.
- [6] B. Liang, Z. Chen, L. Gui, Y. He, M. Yang, and R. Xu, “Zero-shot stance detection via contrastive learning,” in *Proc. ACM Web Conf.*, 2022, pp. 2738–2747.
- [7] B. Liang et al., “JointCL: A joint contrastive learning framework for zero-shot stance detection,” in *Proc. 60th Annu. Meeting Assoc. Comput. Linguistics*, vol. 1: Long Papers, Dublin, Ireland: Association for Computational Linguistics 2022, pp. 81–91.
- [8] Y. Li, C. Zhao, and C. Caragea, “TTS: A target-based teacher-student framework for zero-shot stance detection,” in *Proc. ACM Web Conf.*, 2023, pp. 1500–1509.
- [9] B. Zhang, D. Ding, G. Xu, J. Guo, Z. Huang, and X. Huang, “Twitter stance detection via neural production systems,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Piscataway, NJ, USA: IEEE, 2023, pp. 1–5.
- [10] R. Jain, D. K. Jain, Dharana, and N. Sharma, “Fake news classification: A quantitative research description,” *ACM Trans. Asian Low Resour. Lang. Inf. Process.*, vol. 21, no. 1, pp. 3:1–3:17, 2022.
- [11] N. Yin et al., “Messages are never propagated alone: Collaborative hypergraph neural network for time-series forecasting,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 1, pp. 2333–2347, Apr. 2024.

- [12] S. Rani and P. Kumar, "Aspect-based sentiment analysis using dependency parsing," *ACM Trans. Asian Low Resour. Lang. Inf. Process.*, vol. 21, no. 3, pp. 56:1–56:19, 2022.
- [13] N. Yin, L. Shen, M. Wang, X. Luo, Z. Luo, and D. Tao, "OMG: Towards effective graph classification against label noise," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 12, pp. 12873–12886, Dec. 2023.
- [14] J. Du, R. Xu, Y. He, and L. Gui, "Stance classification with target-specific neural attention networks," in *Proc. Int. Joint Conf. Artif. Intell.*, 2017, pp. 3988–3994.
- [15] C. Li et al., "Joint stance and rumor detection in hierarchical heterogeneous graph," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 6, pp. 2530–2542, Jun. 2022.
- [16] A. T. Cignarella, C. Bosco, and P. Rosso, "Do dependency relations help in the task of stance detection?" in *Proc. 3rd Workshop Insights Negative Results NLP (Insights@ACL)*, Dublin, Ireland: Association for Computational Linguistics, May 26, 2022, pp. 10–17.
- [17] I. Augenstein, T. Rocktaeschel, A. Vlachos, and K. Bontcheva, "Stance detection with bidirectional conditional encoding," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Austin, Texas, USA, Nov. 2016.
- [18] P. Wei and W. Mao, "Modeling transferable topics for cross-target stance detection," in *Proc. 42nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, New York, NY, USA: ACM, 2019, pp. 1173–1176.
- [19] B. Zhang, M. Yang, X. Li, Y. Ye, X. Xu, and K. Dai, "Enhancing cross-target stance detection with transferable semantic-emotion knowledge," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 3188–3197.
- [20] E. Allaway and K. R. McKeown, "Zero-shot stance detection: A dataset and model using generalized topic representations," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, Online. Association for Computational Linguistics, Nov. 16–20, 2020, pp. 8913–8931.
- [21] Y. Jiang, J. Gao, H. Shen, and X. Cheng, "Few-shot stance detection via target-aware prompt distillation," in *Proc. 45th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2022, pp. 837–847.
- [22] M. Hardalov, A. Arora, P. Nakov, and I. Augenstein, "Few-shot cross-lingual stance detection with sentiment-based pre-training," in *Proc. AAAI Conf. Artif. Intell.*, vol. 36, no. 10, 2022, pp. 10729–10737.
- [23] J. Wei et al., "Chain-of-thought prompting elicits reasoning in large language models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 35, pp. 24824–24837, 2022.
- [24] D. Zhou et al., "Least-to-most prompting enables complex reasoning in large language models," 2022, *arXiv:2205.10625*.
- [25] Z. Zhang, A. Zhang, M. Li, and A. Smola, "Automatic chain of thought prompting in large language models," 2022, *arXiv:2210.03493*.
- [26] Z. Cai, B. Chang, and W. Han, "Human-in-the-loop through chain-of-thought," 2023, *arXiv:2306.07932*.
- [27] H. Fei, B. Li, Q. Liu, L. Bing, F. Li, and T.-S. Chua, "Reasoning implicit sentiment with chain-of-thought prompting," 2023, *arXiv:2305.11255*.
- [28] Z. Ling et al., "Deductive verification of chain-of-thought reasoning," 2023, *arXiv:2306.03872*.
- [29] T. Ito, K. Tsubouchi, H. Sakaji, T. Yamashita, and K. Izumi, "Word-level contextual sentiment analysis with interpretability," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 04, 2020, pp. 4231–4238.
- [30] H. Huang, B. Zhang, L. Jing, X. Fu, X. Chen, and J. Shi, "Logic tensor network with massive learned knowledge for aspect-based sentiment analysis," *Knowl.-Based Syst.*, vol. 257, no. C, 2022, Art. no. 109943.
- [31] R. K. Yadav, L. Jiao, O.-C. Granmo, and M. Goodwin, "Human-level interpretable learning for aspect-based sentiment analysis," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 16, 2021, pp. 14203–14212.
- [32] H. Guo, W. Zeng, J. Tang, and X. Zhao, "Interpretable fake news detection with graph evidence," in *Proc. 32nd ACM Int. Conf. Inf. Knowl. Manage.*, 2023, pp. 659–668.
- [33] A. G. Alias Parth Goyal et al., "Neural production systems," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 25673–25687.
- [34] B. Zhang, X. Li, X. Xu, K.-C. Leung, Z. Chen, and Y. Ye, "Knowledge guided capsule attention network for aspect-based sentiment analysis," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 28, pp. 2538–2551, 2020.
- [35] E. Jang, S. Gu, and B. Poole, "Categorical reparametrization with Gumbel-Softmax," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, OpenReview.net, 2017, pp. 1–13.
- [36] T. Jiang et al., "PromptBERT: Improving BERT sentence embeddings with prompts," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2022, pp. 8826–8837.
- [37] S. M. Mohammad, S. Kiritchenko, P. Sobhani, X. Zhu, and C. Cherry, "SemEval-2016 task 6: Detecting stance in tweets," in *Proc. Int. Workshop Semantic Eval. (SemEval)*, San Diego, CA, USA, Jun. 2016.
- [38] Y. Li, T. Sosea, A. Sawant, A. J. Nair, D. Inkpen, and C. Caragea, "P-stance: A large dataset for stance detection in political domain," in *Proc. Findings Assoc. Comput. Linguistics: ACL/IJCNLP*, Online Event, Aug. 1–6, 2021, pp. 2355–2365.
- [39] J. Du, R. Xu, Y. He, and L. Gui, "Stance classification with target-specific neural attention," in *Proc. 26th Int. Joint Conf. Artif. Intell. (IJCAI)*, 2017, pp. 3988–3994. [Online]. Available: <https://doi.org/10.24963/ijcai.2017/557>
- [40] B. Liang et al., "Target-adaptive graph for cross-target stance detection," in *Proc. Web Conf. WWW*, Virtual Event, Ljubljana, Slovenia, Apr. 19–23, 2021, pp. 3453–3464.
- [41] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Human Lang. Technol.*, vol. 1: Long and Short Papers, Jun. 2019, pp. 4171–4186.
- [42] Q. Zhu, B. Liang, J. Sun, J. Du, L. Zhou, and R. Xu, "Enhancing zero-shot stance detection via targeted background knowledge," in *Proc. 45th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2022, pp. 2070–2075.
- [43] J. Zhang, S. Wu, X. Zhang, and Z. Feng, "Task-specific data augmentation for zero-shot and few-shot stance detection," in *Companion Proc. ACM Web Conf.*, 2023, pp. 160–163.
- [44] S. Hu et al., "Knowledgeable prompt-tuning: Incorporating knowledge into prompt verbalizer for text classification," in *Proc. 60th Annu. Meeting Assoc. Comput. Linguistics*, vol. 1: Long Papers, 2022, pp. 2225–2240.
- [45] H. Huang et al., "Knowledge-enhanced prompt-tuning for stance detection," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 22, no. 6, pp. 1–20, 2023.
- [46] B. Zhang, X. Fu, D. Ding, H. Huang, Y. Li, and L. Jing, "Investigating chain-of-thought with ChatGPT for stance detection on social media," 2023, *arXiv:2304.03087*.
- [47] X. Zhang, J. Yuan, Y. Zhao, and B. Qin, "Knowledge enhanced target-aware stance detection on tweets," in *Proc. China Conf. Knowl. Graph Semantic Comput.*, New York, NY, USA: Springer, 2021, pp. 171–184.