ARTICLE

Check for
updates

Cross-Target Stance Detection with Sentiments-Aware Hierarchical Attention Network

Kelan Ren, Facheng Yan, Honghua Chen, Wen Jiang, Bin Wei and Mingshu Zhang[*]

College of Cryptographic Engineering, Engineering University of PAP, Xi'an, 710086, China

*Corresponding Author: Mingshu Zhang. Email: zms2099@163.com

ABSTRACT

The task of cross-target stance detection faces significant challenges due to the lack of additional background information in emerging knowledge domains and the colloquial nature of language patterns. Traditional stance detection methods often struggle with understanding limited context and have insufficient generalization across diverse sentiments and semantic structures. This paper focuses on effectively mining and utilizing sentiment-semantics knowledge for stance knowledge transfer and proposes a sentiment-aware hierarchical attention network (SentiHAN) for cross-target stance detection. SentiHAN introduces an improved hierarchical attention network designed to maximize the use of high-level representations of targets and texts at various fine-grain levels. This model integrates phrase-level combinatorial sentiment knowledge to effectively bridge the knowledge gap between known and unknown targets. By doing so, it enables a comprehensive understanding of stance representations for unknown targets across different sentiments and semantic structures. The model's ability to leverage sentiment-semantics knowledge enhances its performance in detecting stances that may not be directly observable from the immediate context. Extensive experimental results indicate that SentiHAN significantly outperforms existing benchmark methods in terms of both accuracy and robustness. Moreover, the paper employs ablation studies and visualization techniques to explore the intricate relationship between sentiment and stance. These analyses further confirm the effectiveness of sentence-level combinatorial sentiment knowledge in improving stance detection capabilities.

KEYWORDS

Cross-target stance detection; sentiment analysis; commentary-level texts; hierarchical attention network

1 Introduction

Unlike target-specific stance detection, cross-target stance detection requires models with the ability to apply learned knowledge to unseen targets. In recent years, technologies such as graph neural networks, knowledge graphs, and attention mechanisms have been proposed for stance prediction [1–3], which identify social, linguistic, and political factors reflecting stance, and design new knowledge transfer architectures to efficiently capture the identified factors across targets. While these models show higher accuracy empirically compared to existing language models through selecting more representative data samples using domain knowledge, enhanced data augmentation, and transfer

learning methods, the reasons for their effectiveness in cross-target stance prediction have not been fully proven.

With the rise of pre-trained language models, such as Bidirectional Encoder Representations from Transformers (BERT), which have been extensively pre-trained on large corpora of text [4–6], the effectiveness of these models in processing short texts and social media data has been demonstrated. For instance, Bello et al. [7] utilized BERT for sentiment analysis of tweets, showcasing its effectiveness in handling short and informal texts. He et al. [8] demonstrated the robustness of BERT in handling noisy external knowledge through pre-training and fine-tuning methods, showing that it can infer the correct stance even in the presence of noisy external information. However, these models still fall short in addressing the challenges of cross-target stance detection, particularly when the stance inclination towards a target heavily relies on specific contextual and emotional nuances [9,10], and their generalization ability weakens when faced with new domains.

Early methods for cross-target stance detection bridged the knowledge gap between targets by leveraging shared common words or conceptual-level knowledge [1,11,12]. Nevertheless, these methods face challenges when dealing with short, informal texts from social media and distinguishing implied stances from users, resulting in limited effectiveness. Zhang et al. [4] addressed this issue by utilizing semantic and sentiment dictionaries to facilitate knowledge transfer. Their model constructed a semantic-sentiment heterogeneous graph and learned multi-hop semantic connections through graph convolutional networks, integrating this as prior knowledge into a Bi-directional Long Short-term Memory (Bi-LSTM) stance classifier. However, sentiment in text is expressed through interactions among multiple words, and word-level analysis fails to capture these interactions effectively. Consequently, word-level sentiment knowledge is limited in judging the sentiment and stance of sentences that contain implicit opinions or require deep understanding.

In light of this, the paper proposes a new approach for cross-target stance detection, termed the Sentiment-based Hierarchical Attention Model (SentiHAN). SentiHAN introduces hierarchical attention networks and sentiment analysis modules to account for stance at the phrase level, particularly in cases where there are significant sentimental expression differences across targets. By leveraging the sentiment analysis module, SentiHAN better perceives the emotional diversity in text. This hierarchical model not only captures emotional nuances more effectively but also helps mitigate some challenging linguistic phenomena, such as irony and nuanced expressions, that are difficult to understand semantically. Additionally, the model addresses the weak transferability issues of traditional multi-layer attention networks, maintaining a high level of generalization capability when encountering new targets. Moreover, sentiment analysis provides additional interpretability, aiding in understanding how the model non-linearly determines stances based on emotional cues.

2 Related Work

2.1 Sentiment Analysis

Among the existing methods for cross-target stance detection, researchers focus on improving the transferability of stance detection models by utilizing external knowledge sources such as Wikipedia and common sense, given the multidimensional characteristics of stance [4,8,13] which is especially more observed in tasks involving few-shot and zero-shot stance detection [10,14]. However, the selection and introduction of external knowledge are largely based on subjective judgment, leading to data redundancy or even the introduction of incorrect external knowledge, while models typically lack the ability to assess the validity of data. Consequently, these methods suffer from poor interpretability and inefficient training, creating challenges in enhancing model training accuracy under such approaches.

Some studies have integrated sentiment analysis into stance detection to increase the accuracy and applicability of the model [13,15,16]. These models are capable of learning rich sentimental context information from text, but they may face limitations in cross-target stance detection because they are not specifically designed for this purpose. Meanwhile, the interaction between stance and sentiment is still controversial, with some studies arguing that using sentiment as a feature is inefficient [17]. However, these studies often determine sentiment relevance based on the correlation between stance and sentiment or utilize sentiment dictionaries for sentiment analysis, which may not fully capture the complex relationship between sentiment and position.

A common sentiment feature model used for stance analysis is Sent2Vec [18], which learns word embeddings based on the sentiment similarity between words, better reflecting the sentiment nuances in the text. Although this model is effective in capturing local sentiment contexts of words, it may not fully reveal the complex sentiment attitudes of the entire text. A Transferable Transformer-Based Architecture for Compositional Sentiment Semantics (SentiBERT) [19] not only considers sentiment connections between words but also integrates global sentiment information at the text level. It uses a pre-trained Bi-LSTM model to understand the sentimental diversity dependent on the document context. Furthermore, a BERT variant for sentiment prediction, a language model pre-trained on a large amount of data, can be fine-tuned on specific domain data to adapt to the problem of cross-target stance detection in comment texts. The core idea of incorporating sentiment for hierarchical perception in cross-target stance detection is to utilize the combined sentiment between short sentiment, explicitly modeling the complex underlying relationships between short sentence sentiment, short sentence semantics, and text stances towards targets, refining the information that reflects the target sentimental tendency in cross-target stance detection tasks. Accordingly, through explicit modeling at the sentiment level, the model can not only enhance its deep understanding of text granularity but also promote its adaptability and generalizability in cross-target contexts, moving closer to an efficient way for humans to handle and understand stances.

2.2 Knowledge Transfer

To transcend the limitations of pure textual information, researchers have begun to shift their focus to approaches that consider both textual and additional information (e.g., sentiment, semantics, and target relationships) [20]. Research on few-shot and zero-shot stance detection provides valuable insights for cross-target stance detection. For example, Luo et al. [10] explored how sentiment and common sense can be utilized to improve zero-shot stance detection, demonstrating that integrating sentiment analysis and common sense reasoning can significantly enhance stance detection performance in the absence of explicit targets. Fu et al. [21] proposed enhancing stance detection accuracy by identifying the opinion within texts, helping models better understand and handle stance relationships across different targets. Wang et al.'s [14] research showed that combining adversarial networks with external knowledge can improve the generalization ability of models on unseen targets, which is particularly important in addressing the frequent emergence of new targets in cross-target detection. Stance detection methods based on transfer learning aim to transfer the knowledge from pre-trained models and datasets to new scenarios, thereby reducing the reliance on labeled data in new contexts. This approach primarily addresses cross-target stance detection and zero/low-shot stance detection tasks. By using sentiment knowledge as an auxiliary feature for stance representation in cross-target stance detection, the models can identify sentiment polarity (positive, negative, or neutral) in text, bridging the gap between different targets. This transfer of knowledge is particularly beneficial for the following reasons:

Firstly, sentiment knowledge serves as a universally applicable semantic feature, enriching the representation learning of both the text and the target. It can act as a bridge for cross-target knowledge transfer, even across different themes [4]. For instance, whether the stance concerns political events or product reviews, the sentiment features often exhibit similar forms. These shared sentiment features enable the model to accurately capture the stance of new targets.

Secondly, the transfer of sentiment knowledge helps alleviate the issue of data scarcity for specific targets. In cross-target tasks, certain targets may have limited data. By leveraging the sentiment knowledge from existing targets, the model can still perform well even with insufficient data. This effectiveness is due to the highly generalized nature of sentiment features extracted through sentiment analysis, which can be effectively transferred across different targets.

Finally, from an information theory perspective, the transfer of sentiment features enhances the efficiency of model in utilizing information. By integrating sentiment knowledge, the model can more accurately comprehend the underlying meaning of the text instance detection, thereby reducing information uncertainty and improving detection accuracy and robustness.

2.3 Cross-Target Stance Detection

Stance detection is often regarded as a special kind of text classification task. In this field, converting words into embedding vectors is a routine task, thus general language models can be applied [22]. The complexity of cross-target stance detection requires models that provide a deep understanding of the semantics and model the dependencies between multiple targets. Cross-target stance detection is more demanding due to the nature of social media texts [23]. For example, texts are short, informal, may contain abbreviations, and are more fragmented resulting in a lack of contextual information [24]. In recent research, Ding et al. [25] proposed a Multi-Perspective Prompt Tuning (MPPT), which uses instruction-based perspectives from Pre-trained Language Models as a bridge for knowledge transfer. However, MPPT tends to exhibit potential biases inherent in large language models when applied to short texts.

Existing studies have investigated joint learning for stance and sentiment detection [13]. Sun et al. [26] proposed an LSTM-based model to capture both the stance and sentiment information of posts; however, their model did not include an attention mechanism. The lack of attentional mechanisms in text classification models hinders their ability to capture contextual relationships, focus on important information, deal with ambiguity and provide interpretability, leading to sub-optimal performance and reduced transparency. To address these limitations, Li et al. [27] enhanced their models by introducing attention mechanisms along with sentiment and stance features, which significantly improved the performance of the SemEval-16 dataset. The attention mechanism demonstrated significant results in understanding crucial parts of the text and integrating information in a hierarchical structure [15]. Especially in short texts, it can assist the model in focusing on the information that is most crucial for stance. Further research has demonstrated the effectiveness of hierarchical attention mechanisms in processing complex textual structures. Ratmele et al. [28] combined Hierarchical Attention Network (HAN) with hierarchical networks that simulate the structure of documents, validating its effectiveness through the analysis of 29,604 discussion posts. Similarly, Chanaa [29] highlighted the effectiveness of HAN in extracting opinions from unstructured reviews, particularly in capturing multi-layered information within texts.

Given these insights, this paper introduces a hierarchical perceptual mechanism that meticulously examines and integrates various levels of textual information to enhance the recognition of stances in short texts. This approach not only alleviates the reliance of the model on external knowledge

and adapts to the diversity and complexity of short texts but also improves the ability of the model to recognize both implicit and explicit stance signals. Additionally, the high-level features extracted through sentiment analysis are generalizable, and by utilizing the weighted processing capabilities of attention mechanisms, the model can better adapt to new target domains, thereby enhancing its cross-target generalization capabilities.

3 SentiHAN Model

Inspired by the Knowledge-aware Hierarchical Attention Networks (KHAN) [15], this paper proposes the SentiHAN model, which makes several improvements by fully considering the informal nature and information sparsity of short text structures:

1) External Knowledge. Instead of injecting external knowledge through knowledge graphs, the model incorporates structural sentiment, an inherent attribute of text. This approach mitigates the issue of noise while supplementing shared features.

2) Injection Method. The simple combination of inputs is expanded to complementary learning of combined semantics and combined sentiment within the multi-head attention mechanism (text-level attention layer). This enhances the ability of the model to learn transferable stance knowledge across different text structures, semantic environments, and sentiment backgrounds.

3) Cross-Target Paradigm. The hierarchical structure is retained, particularly replacing the title-level attention layer with the target-level attention layer. This emphasizes the target-driven role in cross-target stance detection tasks, ensuring the applicability of model across different target pairs.

The hierarchical perception model structure proposed in this paper is illustrated in Fig. 1, encompassing: at the word level, key information is captured through word embeddings and attention mechanisms; at the sentence level, sentiment analysis and dependency relationship are combined to understand the syntactic and sentimental structure; at the target level, the focus is on identifying arguments relevant to the target. Through such multidimensional analysis, we aim to offer a comprehensive perspective for understanding and analyzing cross-target stances in short texts, significantly enhancing accuracy and reliability in practical applications.
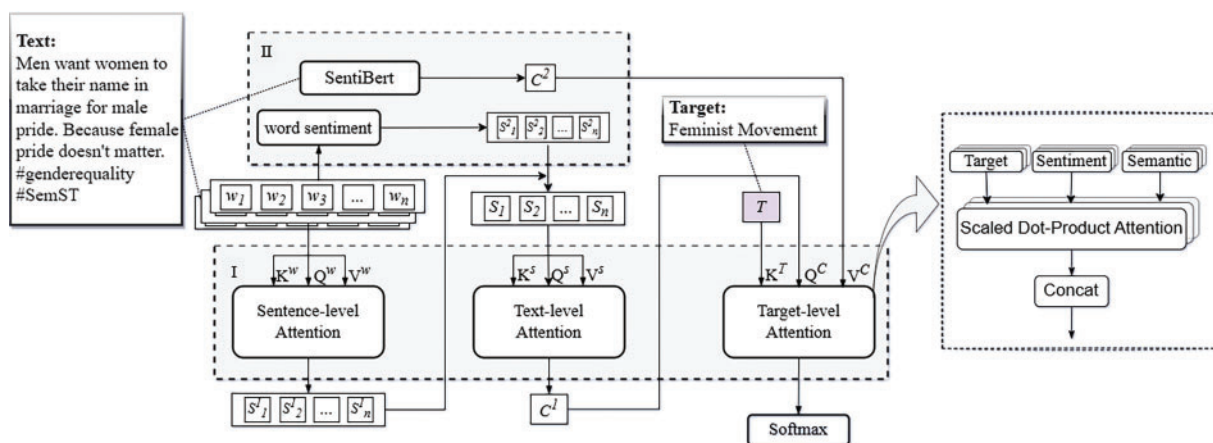

Figure 1: Structure of the SentiHAN: hierarchical attention network (I), sentiment analysis (II)

Short texts often come as unstructured data, merely represented as sequences of words (or symbols). We preprocess the input texts to represent them in a structured form. Initially, the input text

is converted into a set of sentences, each sentence is represented as $W_j = \{w_1, w_2, \ldots, w_n\}$. Where, W_j denotes the collection of word embedding vectors for the j-th sentence, w_i represents the d-dimensional embedding vector for the i word, with n being the maximum word count in a sentence. All word embedding vectors in the datasets are denoted by $\hat{W} = \{W_1, W_2, \ldots, W_N\}$, where N denotes the total count of words in the entire dataset.

3.1 Sentence-Level Attention Layer

In the sentence-level attention layer, our goal is to learn the "relationships between words" within the sentence to better capture the local context of each word. Specifically, we break down short texts into phrases and apply a multi-head self-attention block to the collection of word embedding vectors in each phrase, determining the contribution of each word to the stance. This step identifies the words with the greatest semantic significance in the text:

$$S_j = Multi\ Head\ \{Q^w,\ K^w,\ V^w\} \tag{1}$$

Here $Q^w = K^w = V^w = W_j$, S_j represents the set of word embedding vectors for the j-th sentence, where the vectors contain information about the relevance and contribution of each word within the given sentence. This process ensures that the importance and contextual relationships of each word in the sentence are effectively captured and expressed. Then, the output S_j^1 (context-aware for the j-th sentence) is passed to a feed-forward layer to represent the output of the sentence-level attention layer, thereby providing a foundation for higher-level text analysis. The purpose of this "sentence-type" attention layer is to capture the local context of each sentence by focusing only on the relationships between closely positioned words, without the interference of words located further away. Nonetheless, a significant drawback of this method is its difficulty in capturing the global context of news articles. This limitation is addressed in the text-level attention layer.

It is worth noting that the core advantage of the multi-head attention mechanism lies in its ability to learn different features from different subspaces. A small number of heads cannot achieve the desired effect, while a large number of heads can lead to model instability, especially when training data is limited, and significantly increases computational costs and memory consumption. Through our experiments, we found that the model performs best when the number of heads is set to four, which is consistent with the findings in the literature [15].

3.2 Text-Level Attention Layer

In this layer, the model uses text-level attention mechanisms to integrate "relationships between sentences" and capture the global semantics of the text. This enables the model to understand the text at a higher level of abstraction, providing foundational support for the final stance assessment. Initially, the word embedding vectors from the previous layer are averaged sentence by sentence to create sentence embedding vectors. Thus, the d-dimensional word embedding vectors of the j-th sentence are averaged to produce a d-dimensional embedding vector $\hat{S}_j$ for each sentence, where each embedding vector ($\hat{S}_j$) retains local context information captured from the previous layer. Subsequently, sentence-level multi-head self-attention blocks are applied to the collections of sentence embedding vectors for each text. Formally, the text-level attention layer for the k-th text is defined similarly to the Sentence-level attention layer:

$$C_k^1 = Muti\ Head\ \left(Q^s,\ K^s,\ V^s\right) \tag{2}$$

In this definition, $Q^S = K^S = V^S = \hat{S}_j$. In terms of parameter optimisation, different sets of parameters are used for sentence-level and word-level attention layers. Thus, the significance of Eq. (2) is to capture global semantic features by emphasising the key phrase part of the text in order to capture global semantic features. It also collects the phrase-level local semantic features captured by the word-level layer to achieve hierarchical contextual understanding, analysing the key parts in the focus with limited information, so that the model has a stronger overall comprehension ability. On the other hand, the hierarchical structure alleviates the problem of difficulty in learning informal structural knowledge typical of short texts, especially transferable grammatical knowledge containing key inflectional structures, and enhances the applicability and generalisation ability.

Sentiment Analysis. SentiBERT's training in sentiment analysis allows cross-target stance detection to require a model that learns in one domain and transfers to another, which captures generic sentiment indications as well as flexibly adapts to new, unseen targets and text types. This layer uses SentiBERT for combined sentiment analysis in short texts, converting word-level feature embeddings into sentiment embeddings that are responsible for extracting the sentiment features associated with individual words.

$$C_k^2 = SentiBert\,(S_k) \tag{3}$$

These affective embeddings complement semantic features as additional knowledge.

3.3 Target-Level Attention Layer

This layer is particularly crucial for cross-target detection. A target, as the reference for expressed opinions or stands, utilizes the attention mechanism to calculate the influence of different sentences on different targets. This ensures that the model generates accurate, personalized stance assessments for each target during cross-target analysis. In complex contexts, the stance impact of a specific sentence on a target may depend on its context within the entire text. The attention mechanism helps the model dynamically capture this dependency, assigning higher weights to sentences more relevant to a particular stance target. This ensures that the model, when detecting stances, considers the most crucial information instead of treating all information equally. Inspired by the significance of targeting, we apply a headline-level attention layer to sentence embedding. The goals of this headline attention layer are: (1) to reinforce the key ideas contained in the headline; (2) to capture fine-grained information; (3) to map sentiment relevancy. Set as the text's headline embedding vector as T_k, the target-level attention layer of the text is defined as follows:

$$C_k^T = Multihead\,\left(Q^t,\ K^t,\ V^t\right) \tag{4}$$

where $Q^t = T_k$, $V^t = C_k^1$, $K^t = C_k^2$. This process will be re-weighted according to the context of the objective, i.e., reinforcing the key ideas in the headline. Taking text sentiment as the value (V) means that once the information related to the target within the text is determined, the model will summarize the sentiment information related to these specific parts and use it for stance detection (i.e., sentiment relevancy mapping). Taking target semantics as the query (Q) allows the model to focus on information related to the target. When the key (K) comes from text semantics, the attention mechanism can calculate the relevance between the target and different parts of the text, determining the most relevant information parts in the text related to the target (i.e., fine-grained information capture). At the same time, the multi-head attention mechanism allows for learning the relationships between target, semantics, and sentiment from multiple perspectives, enhancing the complexity of the model and the depth of capturing stance.

However, using the target attention layer in the final stage of SentiHAN might cause the globally semantic-aware sentence embedding vectors to be overly biased towards target perception, resulting in poor robustness. To prevent this issue, we added a residual connection to the output of the target attention layer to preserve the two-dimensional semantic features previously learned from the sentence-level attention layer:

$$C_k = C_k^1 + C_k^T \qquad (5)$$

Wherein, C_k represents the set of final sentence embedding in text k, with the aggregated embedding then passed through the following output:

$$\hat{y} = \text{soft max}\,[\text{avg}\,(C_k)]$$

The model is trained by minimizing the multi-class cross-entropy loss in order to effectively predict the stance of the text.

$$\min_{\Theta} \ell = -\sum_{i=1}^{N}\sum_{j=1}^{C} y_i^j \log \hat{y}_i^j + \lambda \|\Theta\|^2 \qquad (6)$$

Here, N denotes the size of the training set, C represents the total number of stance categories, Θ denotes all trainable parameters of the model, and λ is the coefficient for the L2 regularization term.

4 Experimental Validation

4.1 Experimental Settings

4.1.1 Experimental Data

The experiments selected two datasets: SemEval16 [3] and P-Stance [23]. The statistical details of the datasets are shown in Table 1. Sem16 covers two topics and four targets: Donald Trump (DT), Hillary Clinton (HC), legalization of abortion (LA), and the feminist movement (FM), with instances categorized as support, against, or neutral. Four cross-target stance detection tasks were constructed on Sem16 (FM→LA, LA→FM, HC→DT, DT→HC). Meanwhile, the P-Stance dataset is an order of magnitude larger than other publicly available datasets and offers multiple targets under the same theme: "Donald Trump," "Joe Biden," and "Bernie Sanders," with instances divided into support and oppose. Six cross-target pairs were constructed in the P-Stance datasets: Trump→Sanders, Sanders→Trump, Sanders→Biden, Biden→Sanders, Trump→Biden, and Biden→Trump.

Table 1: Statistics of SemEval16 and P-Stance datasets

Dateset	Target	Favor	Against	None	Total	
Sem16	FM	268	511	170	949	3573
	LA	167	544	222	933	
	HC	163	565	256	984	
	DT	148	299	260	707	
P-Stance	Donald trump	3663	4290	–	7953	21574
	Joe biden	3217	4079	–	7296	
	Bernie sanders	3551	2774	–	6325	

This paper uses Macro-F1 (the macro-average F1 of "support" and "against") and $F1_{avg}$ (the average of micro-average F1 and macro-average F1) as the evaluation metrics to measure the classification performance of the model.

$$Macro\_F1 = \frac{F1_{support} + F1_{oppose}}{2} \tag{7}$$

$$F1 = \frac{2 * precision * recall}{precision + recall} \tag{8}$$

$$F1_{avg} = \frac{F1_{micro} + F1_{macro}}{2} \tag{9}$$

4.1.2 Baseline Methods

Benchmark Methods. This paper evaluates and compares the following reliable baseline methods against its model:

BiCond [11]: Utilizes bidirectional conditional encoding to simultaneously learn representations of sentences and targets for stance detection. This method emphasizes the importance of considering the target in understanding textual attitudes.

CrossNet [1]: Uses self-attention layers to extract important context words related to the target, learning specific target stance features. This approach highlights the role of self-attention mechanisms in extracting target-relevant information.

SEKT [4]: A knowledge-based Graph Convolutional Network (GCN) model that incorporates semantic-emotional knowledge into heterogeneous graph construction to bridge the gap between source and target objectives, achieving cross-target stance detection.

BERT [5]: Employs the pre-trained BERT-base model, using "[CLS] target [SEP] sentence [SEP]" as input for each instance. This reflects the strong capabilities of pre-trained models in various natural language processing tasks, while also emphasizing the importance of directly combining target and sentence information in the input.

JointCL [30]: A joint contrastive learning framework consisting of stance contrastive learning and target-aware prototype graph contrastive learning.

TarBK [31]: Combines background knowledge for cross-target stance detection.

TPDG [3]: Combines target information and sentence structure, proposing a model of target-adaptive pragmatic dependency graphs through the construction of dependency graphs to understand textual stances.

PT-HCL [32]: A stance detection method for unknown targets based on a hierarchical contrastive learning strategy.

KEprompt [33]: Expands the prompt tuning capabilities of language models by utilizing external semantic knowledge and injecting background knowledge.

TGA-Net [34]: Implicitly captures relationships between topics using generalized topic representations.

4.1.3 Experimental Details

The SentiHAN method was implemented on Ubuntu 18.04 OS (Operating System) using PyTorch 1.1.0. Experiments ran on a machine equipped with an Intel Xeon Platinum 8255C CPU (Central

Processing Unit), 72 GB of memory, and an NVIDIA RTX 2080 Ti(11 GB), with CUDA 10.0 and cuDNN v7.6.5 installed. The batch size was set to 64, using the Adam optimizer, with a learning rate of 1e-3, and a weight decay factor of 5e-2 for all datasets.

To ensure a fair and consistent comparison, we re-implemented and evaluated the results of certain benchmark models within the same experimental environment. Additionally, to assess the performance of our proposed method, we selected benchmark model results reported in related literature. The experimental settings and datasets in these studies are highly relevant to our research and are widely recognized as standard benchmarks in the field. While we have made every effort to ensure a fair comparison of all results, there may be uncontrollable factors due to differences in experimental setups in the cited literature. Our experimental results take these potential differences into account. The hyperparameters of the KHAN model were referenced [15], fine-tuned, and experimented with in our model using the following settings to achieve optimal performance: batch_size = 64, learning_rate = 0.001, embed_size = 256, dropout = 0.4, num_layer = 2, num_head = 4.

4.2 Experimental Results

Table 2 reports the experimental results of models on SemEval16. CrossNet, as an extension of BiCond, incorporates attention mechanisms, highlighting the role of self-attention in extracting target-related information. The model performs slightly better than its predecessor, demonstrating the feasibility of attention mechanisms in perceiving stance representations for specific entities. However, this word-level information-based knowledge transfer scheme still shows overall poor performance. BERT utilizes contextual semantic information of different targets to achieve good results in FM-LA and HC-DT, but due to the lack of any knowledge transfer strategy, its performance in other tasks is unsatisfactory, indicating instability. SEKT and TPDG improve the generalization ability of model through transfer learning methods, achieving better and more balanced results across all labels. JointCL, TarBK, and PT-HCL show similar results, indicating comparable performance in cross-target stance detection. KEprompt performs poorly in FM-LA, LA-FM, and HC-DT but significantly outperforms other models in DT-HC (60.9%), demonstrating strong performance in this specific setting. SentiBERT performs well in all four tasks, with an average Macro-F1 of 62.05%, improving by 6.32% over the fine-tuning-based method PT-HCL and by 6.8% over the leading knowledge-infused model TarBK. From the perspective of external knowledge, SEKT incorporates sentiment knowledge of words, TarBK introduces target background knowledge based on Wikipedia, KEprompt integrates prompt tuning with ConceptGraph background knowledge, and SentiHAN learns memory links of combined sentiment knowledge and semantic features. The $F1_{avg}$ scores of multiple target pairs, as shown in Fig. 2, indicate that SentiHAN outperforms other models. However, models based on sentiment knowledge (SEKT, SentiHAN) show less stability compared to those based on commonsense knowledge (TarBK, KEprompt). This may be influenced by the distribution of semantic knowledge in the dataset, as discussed in the case analysis.

Table 3 reports the experimental results of models on P-Stance. Compared to RoBERTa, Cross-Net, and TGA-Net, SentiHAN shows more balanced performance across all targets and achieves the best average Macro-F1 score without obvious weaknesses. SentiHAN's consistently high scores and balanced performance indicate strong generalization ability across different targets. This means the model can effectively learn and adapt to the data characteristics of different targets, performing well in various test scenarios.

Table 2: F1$_{avg}$ scores of models based on the SemEval16. † are retrieved using the same experimental environment of the proposed method in this paper

Model	FM-LA	LA-FM	HC-DT	DT-HC	Average
Bicond	45.0	41.6	29.7	35.8	38.03
CrossNet	45.4	43.3	43.1	36.2	42.00
BERT	47.9	33.9	43.6	36.5	40.48
SEKT	53.6	51.3	47.7	42.0	48.65
KEprompt	49.1	54.2	54.6	**60.9**	54.70
TPDG†	59.2	55.2	51.3	54.0	54.93
TarBK	59.1	54.6	53.1	54.2	55.25
JointCL†	59.4	55.7	53.4	54.9	55.70
PT-HCL†	59.8	54.7	54.2	55.1	56.0
SentiHAN	**67.7**	**67.5**	**57.1**	55.9	**62.05**
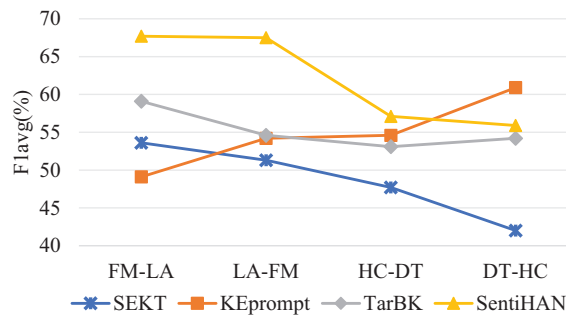
Figure 2: Comparison of external knowledge models

Table 3: Macro-F1 scores of models based on the P-Stance

Source Test	Trump Sanders	Sanders Trump	Sanders Biden	Biden Sanders	Trump Biden	Biden Trump	Average
RoBERTa	0.53	0.59	**0.78**	**0.66**	**0.77**	0.62	0.66
CrossNet	0.46	0.51	0.69	0.58	0.60	0.54	0.56
TGA-Net	0.57	0.60	0.69	0.60	0.69	0.59	0.62
SentiHAN	**0.67**	**0.66**	0.69	**0.64**	0.69	**0.66**	**0.67**

4.3 Ablation Study

The results of the ablation experiments are summarised in Table 4. In this experiment, we verify the effectiveness of sentiment knowledge (SEN), semantic knowledge (SEM), and target knowledge (TG). Experiment compare the three versions of SentiHAN:

w/o SEN: The performance significantly declined, particularly in the FM-LA task where the F1$_{avg}$ value dropped from 70.0 to 51.5. This decline may be attributed to the lack of sentiment knowledge,

which prevents the model from accurately identifying and interpreting sentiment features in the text, thus affecting stance detection. The absence of sentiment knowledge leaves the model without the necessary contextual information support when dealing with expressions laden with strong sentiment nuances.

Table 4: Ablation study results ($F1_{avg}$ values). w/o SEN: remove sentiment knowledge; w/o SEM: remove semantic knowledge; w/o TG: remove target knowledge

Model	FM-LA	LA-FM	HC-DT	DT-HC
w/o SEN	51.5	47.1	40.8	51.2
w/o SEM	68.4	57.6	53.0	57.7
w/o TG	66.4	56.0	53.4	57.6
SentiHAN	**70.0**	**68.9**	**59.5**	**58.3**

w/o SEM: There was a notable negative impact on capturing deep semantic relationships within the text. For instance, the $F1_{avg}$ value in the HC-DT task decreased from 59.5 to 53.0. This decline can be attributed to the crucial role of semantic knowledge in understanding the overall text and capturing complex semantic connections. The lack of semantic knowledge prevents the model from fully understanding the semantic relationships between sentences and paragraphs, thereby affecting the judgment of emotions and stance. This issue is particularly pronounced when dealing with long texts or sentences with complex structures, where SEM fails to effectively establish semantic associations between words.

w/o TG: The overall performance was poor, possibly due to the critical role of target knowledge in accurately conveying sentiment features across different targets and improving stance detection accuracy. The absence of target knowledge means the model cannot distinguish and associate sentiment information across multiple targets in sentiment analysis tasks.

In summary, sentiment knowledge, semantic knowledge, and target knowledge all play significant roles in enhancing the performance of the SentiBERT method. The integration of target knowledge and sentiment knowledge, along with a deep understanding of semantics, enables the SentiHAN model to exhibit superior performance in capturing complex semantic links between words, emotions, and stance labels. This further confirms the necessity and superiority of combining hierarchical perception and sentiment analysis strategies. It also enhances the adaptability and generalization ability of model in cross-target contexts, thereby more closely approximating the efficient manner in which humans process and understand stances.

5 Case Study

Fig. 3 displays the distribution of three types of stance labels and three types of sentiment labels within the datasets for DT, HC, FM, and LA, where the sum of "negative" and "positive" dominates, indicating the prevalence of sentimental tone in texts. Additionally, the proportion of negative sentiments generally exceeds that of positive sentiments, aligning with the rough ratio of support to opposition, which coincides with the customary alignment of sentimental expression and stance tendencies.
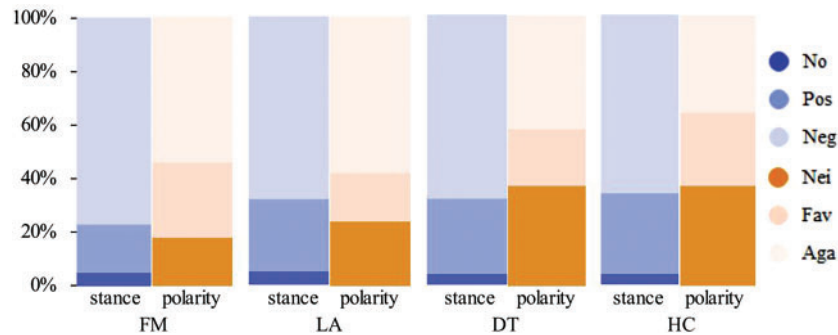
Figure 3: Sentiment and stance label distribution in the SemEval16. Stance: favor, against, and neither; sentiment: negative, positive, and neither (denoted as no). The vertical axis represents proportions

However, Table 5 presents the orthogonal relationship between stance and sentiment for DT, HC, FM, and LA, under which the alignment between sentiment and stance exhibits variability. Typically, positive sentiments are the main means of expressing support, while negative sentiments predominantly convey opposition. Nevertheless, in the "support" stance of FM and LA, the proportion of positive sentiment is noticeably lower than that of combined sentiments. This is contrary to the expected relationship between sentiment and stance, and the discrepancies arise mainly from three aspects: (1) supporters might also hold critical views on certain aspects of the issues or solutions; (2) cultural and social contexts, which may relate to the nature of the issue or the background of the discussion; and (3) implications, irony, or other forms of subtle expression techniques. This phenomenon reveals that, even in seemingly straightforward expressions of support or opposition, the usage and expression of sentiments are far more complex than they appear, underscoring the importance of considering sentimental complexity when studying stance expressions.

Table 5: Orthogonal relationship between sentiment and stance in the SemEval16. Highest proportions are in bold

Stance	FM			LA		
	Pos	Neg	No	Pos	Neg	No
Favor	31.72%	66.04%	2.24%	25.15%	70.66%	4.19%
Against	8.41%	88.85%	2.74%	25.18%	70.22%	4.60%
Neither	27.06%	58.24%	14.71%	29.73%	60.36%	9.91%
Stance	DT			HC		
	Pos	Neg	No	Pos	Neg	No
Favor	75.00%	21.62%	3.38%	74.85%	21.47%	3.68%
Against	15.72%	82.27%	2.01%	15.93%	82.12%	1.95%
Neither	13.46%	78.08%	8.46%	33.20%	58.20%	8.59%

Fig. 4 further intuitively reflects the differences in the aforementioned results and the superiority of our model. In the FM and DT instance visualizations, the stark color contrast widely demonstrates the presence of emotionalized contexts in texts. When dealing with texts that contain rich

sentimental expressions and clear sentimental tendencies, the model utilizes SentiBERT effectively to capture combined sentiments, managing to grasp both evident sentimental signals and structured information in the texts. Particularly by observing instance 7 in FM and instances 1 and 13 in DT, we notice that although the sentimental tendencies at the word level generally tilt positive, the overall sentimental polarity of the texts appears negative, explaining the unconventional distributions shown in Table 5. In these cases, the SentiHAN model still exhibits superior performance compared to other models, suggesting its advantages in handling texts with complex sentiments or implicit expressions, demonstrating its flexibility and robustness.
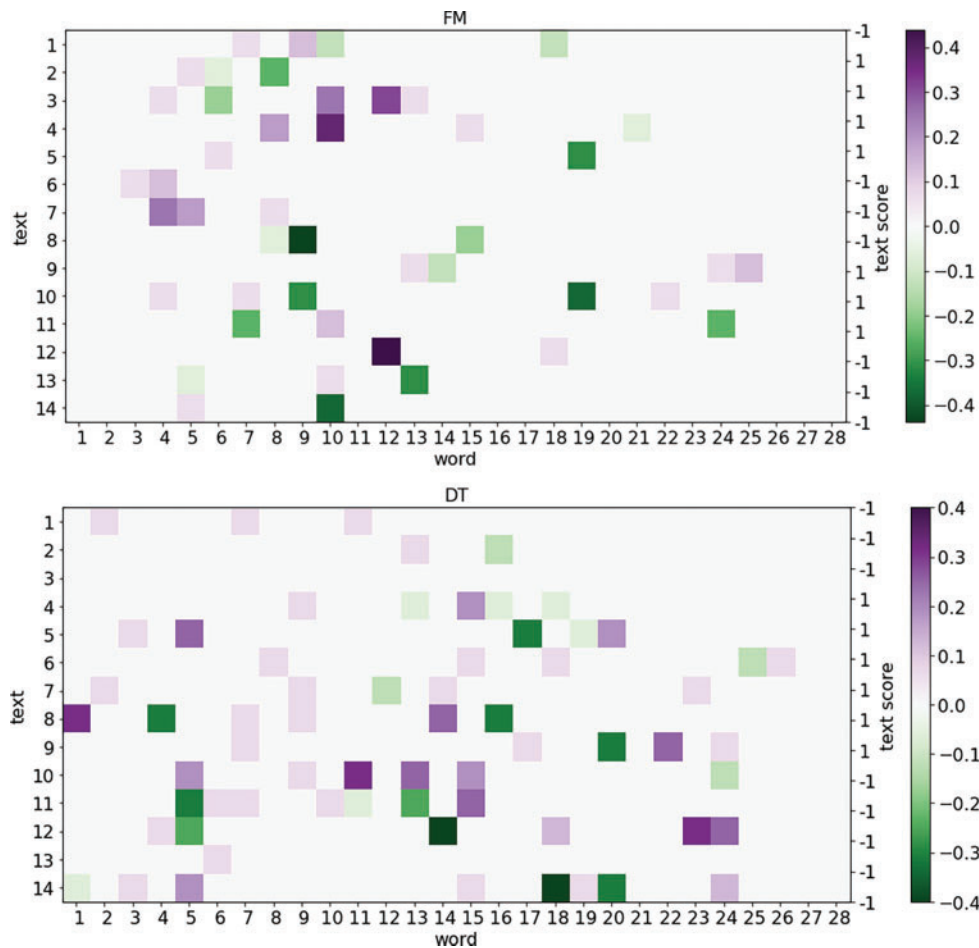
Figure 4: Visualisation of sentimental polarity. Randomly selected ten text instances targeted towards FM and DT (6 supporting, 6 opposing, and 2 neutral) for a visual analysis. The left vertical axis represents the instances, the horizontal axis indicates the position of words in sentences, and colors represent the sentimental polarity of words (learned from a sentiment lexicon) while the right vertical axis shows the sentence's sentimental polarity

To illustrate the inner workings of SentiHAN and to provide a deeper understanding of how sentiment knowledge contributes to stance detection, Fig. 5 presents a case study on the interpretability of compositional sentiment and stance. It is crucial to understand how SentiHAN simultaneously captures both sentiment and stance. In the sentence, "If feminists spent half the time reading papers

instead of reading tumblr, they would not be ignorant sexist bigots," words like "ignorant" and "bigots" are highlighted in strong negative colors, indicating a pronounced negative sentiment. The hierarchical structure of SentiHAN facilitates a better understanding of how sentiment words are utilized across multiple layers, revealing the complexity of the interplay between stance and sentiment. The phrase "not be ignorant sexist bigots" clearly conveys a critical stance towards the behavior of those labeled as "feminists." By analyzing sentiment knowledge, we can uncover the emotional inclinations conveyed by different words and phrases in the text, thus gaining a clearer understanding of how these emotions influence the expression of a stance. SentiHAN proves to be a powerful tool for elucidating this relationship, making the stance detection process more comprehensible.
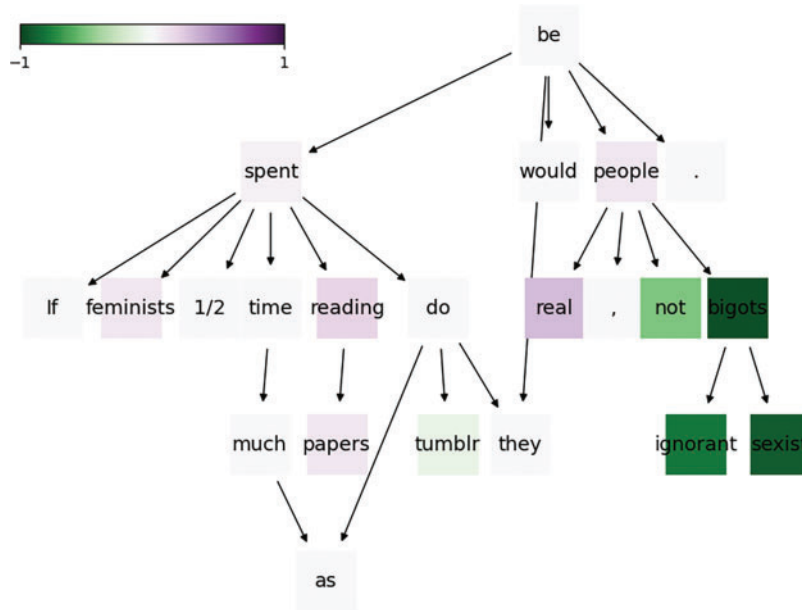

Figure 5: Case for interpretability of compositional sentiment stance

6 Error Analysis

We conducted a detailed analysis of the error predictions made by the SentiHAN model on the SemEval-2016 dataset and discussed the findings with reference to specific examples from the dataset. The main sources of errors identified are as follows:

Impact of Text Noise: The informal nature of social media and short texts can introduce noise into the text, such as spelling errors, abbreviations, and non-standard language expressions. Although SentiHAN mitigates noise effects through structural sentiment information, such noise can still lead to misclassification in certain cases. For example, in the tweet "Where is the childcare program @joanburton which you said would be in place? #loneparents #7istooyoung #SemST," the presence of multiple hashtags may cause the model to incorrectly handle the influence of these tags, resulting in an erroneous prediction.

Insufficient External Knowledge: While SentiHAN reduces noise by substituting knowledge graphs with structural sentiment injection, the model may still lack sufficient external knowledge to accurately understand the context in texts requiring extensive background knowledge. For instance, in the tweet "I get several requests with petitions to save whales but never to stop the slaughter of

babies in the wombs of mothers. #abortion. #SemST," the statement involves specific cultural and historical contexts. Petitions to save whales generally concern animal protection and environmental issues, which are focal points for many global environmental organizations. This statement expresses dissatisfaction or confusion about the imbalance in societal concerns. Due to the lack of relevant background knowledge, the model misinterprets the stance of the comment. This indicates that future research should focus on finding effective methods to leverage topic tags and external knowledge while minimizing the noise they introduce.

In addition, the SentiHAN model uses SentiBERT for sentiment analysis, which effectively integrates emotion knowledge for stance detection. However, SentiBERT, being a variant of the BERT architecture, has relatively high computational complexity. This can affect inference speed, particularly in real-time applications like social media monitoring, where large volumes of data are processed. Inference speed can vary depending on factors such as hardware setup (CPU *vs*. GPU (Graphic Processing Unit) /TPU (Tensor Processing Unit)), the length of input data, and the overall complexity of the model. While SentiHAN provides accurate stance detection, managing and optimizing its real-time inference speed is crucial for maintaining responsiveness in high-demand environments.

7 Conclusion

This paper proposes a novel sentiment-aware multi-granularity feature fusion cross-target stance detection model (SentiHAN). The model reveals that in cross-target stance detection, text, sentiment knowledge, and target information can be utilized from the perspective of different granularity features, a facet often neglected by most existing methods. Unlike previous cross-target stance detection methods, SentiHAN begins with key semantic analysis at the lexical level and then shifts focus to sentences and phrases with dense stance expressions at the intermediate level. Finally, it synthesizes this data at the highest level to obtain a comprehensive understanding of the entire document or cross-target stances. By incorporating sentiment knowledge, the model's understanding of text semantics is enhanced, enabling more effective learning of contextual semantic backgrounds. In SentiHAN, different parameter sets are used for different levels of the network, effectively increasing complementarity among different granularities, and thereby enhancing the cross-target stance analysis capability of the model. The application of the multi-head attention mechanism plays a crucial role in capturing complex dependencies among semantics, sentiment, and targets, contributing to the improved transferability of stance knowledge across targets.

Extensive experimental results on real-world public datasets not only demonstrate the superiority of SentiHAN but also showcase the advancement of multi-level feature fusion and the incorporation of sentiment knowledge methods in cross-target stance detection tasks. The experimental results confirm the efficiency and robustness of models, improving stance prediction in complex thematic and sentiment contexts, particularly in sentiment-driven expressions prevalent in social media and reviews.

However, there are still many areas worthy of exploration in this work. Firstly, the framework structure and learning strategies can be further optimized to enhance the framework's adaptability to different sentiment types. Secondly, various pre-trained models can be explored to enhance the ability framework to handle multi-source external knowledge. Finally, the research needs to find a balance between model performance and interpretability, with more focus on consistent performance evaluation across different datasets. In future work, we will continue to explore and research these issues.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Mingshu Zhang, Kelan Ren; data collection: Facheng Yan and Honghua Chen; analysis and interpretation of results: Bin Wei, Wen Jiang; draft manuscript preparation: Kelan Ren. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: No new data were created during this study. The study brought together existing data obtained upon request and subject to licence restrictions from a number of different sources.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] C. Xu, C. Paris, S. Nepal, and R. Sparks, "Cross-target stance classification with self-attention networks," in *ACL-Annu. Meet. Assoc. Comput. Linguist, Proc. Stud. Res. Workshop*, Australia, 2018, pp. 778–783.

[2] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *NAACL HLT-Conf. N. Am. Chapter Assoc. Comput. Linguistics: Hum. Lang. Technol.-Proc. Demonstr. Sess.*, Minneapolis, 2019, pp. 4171–4186.

[3] B. Liang *et al.*, "Target-adaptive graph for cross-target stance detection," in *Web Conf. -Proc. World Wide Web Conf.*, Ljubljana Slovenia, Apr. 19–23, 2021, pp. 3453 –3464.

[4] 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. Annu. Meet. Assoc. Comput. Linguist.*, Jul. 2020, pp. 3188–3197.

[5] Y. Zhang *et al.*, "Stance-level sarcasm detection with BERT and stance-centered graph attention networks," *ACM Trans. Internet Technol.*, vol. 23, no. 2, pp. 1–21, Jul. 2023. doi: 10.1145/3617123.

[6] Y. Sun and Y. Li, "Stance detection with knowledge enhanced BERT," in *Lect. Notes Comput.*, Jun. 2021, vol. 13070, pp. 239–250. doi: 10.1007/978-3-030-93049-3_20.

[7] A. Bello, S. -C. Ng, and M. -F. Leung, "A BERT framework to sentiment analysis of tweets," *Sensors*, vol. 23, no. 1, Jan. 2023, Art. no. 506. doi: 10.3390/s23010506.

[8] Z. He, N. R. Mokhberian, and K. Lerman, "Infusing knowledge from wikipedia to enhance stance detection," in *WASSA-Workshop Comput. Approaches Subj., Sentim. Soc. Media Anal., Proc. Workshop*, Dublin, Ireland, May 2022, pp. 71–77.

[9] A. Upadhyaya, M. Fisichella, and W. Nejdl, "A multi-task model for sentiment aided stance detection of climate change tweets," 2022, *arXiv:2211.03533*.

[10] Y. Luo, Z. Liu, and Y. Shi, "Exploiting sentiment and common sense for zero-shot stance detection," in *Proc. Main Conf. Int. Conf. Comput. Linguist.*, Coling, Gyeongju, Republic of Korea, Oct. 2022, pp. 7112–7123.

[11] I. Augenstein, T. Rocktäschel, A. Vlachos, and K. Bontcheva, "Stance detection with bidirectional conditional encoding," in *EMNLP-Conf. Empir. Methods Nat. Lang. Process.*, Austin, Texas, Nov. 2016, pp. 876–885.

[12] P. Wei and W. Mao, "Modeling transferable topics for cross-target stance detection," in *SIGIR-Proc. Int. ACM SIGIR Conf. Res. Dev. Inf. Retr.*, Paris, France, Jul. 2019, pp. 1173–1176.

[13] M. Hardalov, A. Arora, N. Nakov, and I. Augenstein, "Few-shot cross-lingual stance detection with sentiment-based pre-training," in *Proc. AAAI Conf. Artif. Intell., AAAI*, Jun. 2022, vol. 36, pp. 10729–10737. doi: 10.1609/aaai.v36i10.21318.

[14] C. Wang *et al.*, "Adversarial network with external knowledge for zero-shot stance detection," in *Lect. Notes Comput. Sci.*, Harbin, China, Sep. 2023, vol. 14232, pp. 824–835. doi: 10.1007/978-981-99-6207-5_26.

[15] Y. Ko *et al.*, "KHAN: Knowledge-aware hierarchical attention networks for political stance prediction," in *ACM Web Conf. -Proc. World Wide Web Conf.*, WWW, Austin, TX, USA, Apr. 2023, pp. 1572–1583.

[16] M. Hosseinia, E. Dragut, and A. Mukherjee, "Stance prediction for contemporary issues: Data and experiments," in *Proc. Eighth Int. Workshop Nat. Lang. Process. Soc. Med.*, Jul. 2020, pp. 32–40. doi: 10.18653/v1/2020.socialnlp-1.5.

[17] A. Aldayel and W. Magdy, "Assessing sentiment of the expressed stance on social media," in *Lect. Notes Comput. Sci.*, Doha, Qatar, Nov. 2019, pp. 277– 286. doi: 10.1007/978-3-030-34971-4_19.

[18] M. N. Moghadasi and Y. Zhuang, "Sent2Vec: A new sentence embedding representation with sentimental semantic," in *Proc. -IEEE Int. Conf. Big Data, Big Data*, Atlanta, GA, USA, Dec. 2020, pp. 4672–4680. doi: 10.1109/BigData50022.2020.9378337.

[19] D. Yin, T. Meng, and K. W. Chang, "SentiBERT: A transferable transformer-based architecture for compositional sentiment semantics," in *Proc. Annu. Meet. Assoc. Comput Linguist*, USA, Jul. 2020, pp. 3695– 3706.

[20] C. Li and D. Goldwasser, "Using social and linguistic information to adapt pretrained representations for political perspective identification," in *Find. Assoc. Comput. Linguist.: ACL-IJCNLP*, Aug. 2021, pp. 4569–4579. doi: 10.18653/v1/2021.findings-acl.401.

[21] Y. Fu *et al.*, "Incorporate opinion-towards for stance detection," *Know.-Based Syst.*, vol. 246, Jun. 2022, Art. no. 108657. doi: 10.1016/j.knosys.2022.108657.

[22] C. Li *et al.*, "MEAN: Multi-head entity aware attention network for political perspective detection in news media," in *Proc. Fourth Workshop NLP Internet Freedom: Censorship, Disinform.,* Propaganda, Jun. 2021, pp. 66–75. doi: 10.18653/v1/2021.nlp4if-1.10.

[23] N. Alturayeif, H. Luqman, and M. Ahmed, "A systematic review of machine learning techniques for stance detection and its applications," *Neural Comput. Appl.*, vol. 35, pp. 5113–5144, Jan. 2023. doi: 10.1007/s00521-023-08285-7.

[24] K. Cortis and B. Davis, "Over a decade of social opinion mining: A systematic review," *Artif. Intell. Rev.*, vol. 54, pp. 4873–4965, Oct. 2021. doi: 10.1007/s10462-021-10030-2.

[25] D. Ding *et al.*, "Cross-target stance detection by exploiting target analytical perspectives," in *ICASSP IEEE Int. Conf. Acoust. Speech Signal Process Proc.*, Seoul, Republic of Korea, Apr. 2024, pp. 10651–10655.

[26] Q. Sun, Z. Wang, S. Li, Q. Zhu, and G. Zhou, "Stance detection via sentiment information and neural network model," *Front Comput. Sci.*, vol. 13, pp. 127–138, Feb. 2019.

[27] Y. Li and C. Caragea, "Multi-task stance detection with sentiment and stance lexicons," in *EEMNLP-IJCNLP-Conf. Empir. Methods Nat. Lang. Process. Int. Jt. Conf. Nat. Lang. Process., Proc. Conf.*, Hong Kong, China, 2019, pp. 6299–6305.

[28] A. Ratmele and R. Thakur, "OpExHAN: Opinion extraction using hierarchical attention network from unstructured reviews," *Soc. Netw. Anal. Min.*, vol. 12, no. 1, Dec. 2022, Art. no. 148.

[29] A. Chanaa, "E-learning text sentiment classification using hierarchical attention network (HAN)," *Int. J. Emerg. Technol. Learn.*, vol. 16, no. 13, pp. 157– 167, Jul. 2021.

[30] B. Liang *et al.*, "JointCL: A joint contrastive learning framework for zero-shot stance detection," in *Proc. Annu. Meet. Assoc. Comput. Linguist*, Dublin, Ireland, May 2020, vol. 1, pp. 81–91.

[31] Q. Zhu *et al.*, "Enhancing zero-shot stance detection via targeted background knowledge," in *SIGIR-Proc. Int. ACM SIGIR Conf. Res. Dev. Inf. Retr.*, New York, NY, USA, Jul. 2022, pp. 2070–2075. doi: 10.1145/3477495.3531807.

[32] B. Liang *et al.*, "Zero-shot stance detection via contrastive learning," in *WWW-Proc. ACM Web Conf.*, Apr. 2022, pp. 2738–2747. doi: 10.1145/3485447.3511994.

[33] H. Huang *et al.*, "Knowledge-enhanced prompt-tuning for stance detection," *ACM Trans. Asian Low-Resour. Lang. Inf. Process*, vol. 22, no. 159, pp. 1–20, Jun. 2023. doi: 10.1145/3588767.

[34] P. J. Khiabani and A. Zubiaga, "Few-shot learning for cross-target stance detection by aggregating multimodal embeddings," *IEEE Trans. Computat. Soc. Syst.*, vol. 11, no. 2, pp. 2081–2090, Apr. 2024. doi: 10.1109/TCSS.2023.3264114.