LLM-Driven Knowledge Injection Advances Zero-Shot and Cross-Target Stance Detection

Zhao Zhang[1,3], Yiming Li[2,3], Jin Zhang[1,*], Hui Xu[1]
[1]CAS Key Laboratory of Network Data Science and Technology,
Institute of Computing Technology, Chinese Academy of Sciences
[2]Beijing Key Laboratory of Mobile Computing and Pervasive Device,
Institute of Computing Technology, Chinese Academy of Sciences
[3]University of Chinese Academy of Sciences
{zhangzhao22s,liyiming22s1,jinzhang,xuhui}@ict.ac.cn

Abstract

Stance detection aims at inferring an author's attitude towards a specific target in a text. Prior methods mainly consider target-related background information for a better understanding of targets while neglecting the accompanying input texts. In this study, we propose to prompt Large Language Models (LLMs) to explicitly extract the relationship between paired text and target as contextual knowledge. We then inject such LLM-driven knowledge into a generation model BART to exploit the rich contexts and semantics. Moreover, to further enhance the decoding capability of BART, a novel prototypical contrastive scheme is designed to align input contents with stance labels. Our experimental results demonstrate the state-of-the-art performance across several publicly available datasets, showcasing effectiveness in both zero-shot and cross-target stance detection scenarios. We publicly release our code to facilitate future research.[1]

1 Introduction

The objective of stance detection is to ascertain an individual's stance regarding a specified target in a text, which may either be explicitly mentioned or implied only (Küçük and Can, 2022). The significance of stance detection is evident in its role across various fields, such as predicting election and referendum outcomes (Kawintiranon and Singh, 2021), classifying rumors (Lin et al., 2021), detecting fake news (Hanselowski et al., 2018), and identifying instances of disinformation (Hardalov et al., 2022).

Stance detection can be divided into three categories according to the availability of test targets: 1) target-specific: stance detection for fixed targets (Hasan and Ng, 2014); 2) cross-target: stance detection for related targets (Augenstein et al., 2016); 3) zero-shot: stance detection for unseen targets (Allaway and McKeown, 2020). Among these tasks, the latter two are more challenging since test targets are inaccessible in the training data. One promising solution is to incorporate background knowledge as additional information, which previous works mainly focus on (He et al., 2022). Despite the understanding of specific targets being improved, such approaches solely consider target-related knowledge but ignore the relationship between texts and targets, which might mislead the stance detection model as shown in Figure 1.

Target: Joe Biden **Stance:** Favor
Text: #Correct! I am a #Republican and will vote for @JoeBiden.
Prior Target-Related Knowledge: Joe Biden was a member of the Democratic Party. ☹ contains confusing information that the author and Joe Biden belong to different parties
Our LLM-Driven Knowledge: Despite his Republican affiliation, the author has made a decision to vote for Joe Biden. ☺ resolves the confusion by analyzing the relationship between the text and target

Figure 1: An example comparing our LLM-driven knowledge with previous target-related knowledge.

Recently, LLMs, such as ChatGPT released by OpenAI[2], have shown superior natural language understanding performance. In this work, we specially design the prompt to make the most of ChatGPT's internal knowledge and reasoning capabilities for extracting the relationship between texts and targets explicitly. We refer to the consequent response as LLM-driven knowledge, a snippet of which can also be found in Figure 1. Compared to prior ones, LLM-driven knowledge is more specific and context-rich for stance detection.

In order to bridge the semantic gap between LLM-driven knowledge and stance labels, we adopt a bidirectional autoregressive model - BART (Lewis et al., 2020; Wen and Hauptmann, 2023) as our backbone. As a result, the detection scheme is reformulated as a conditional stance label generation task, where the condition consists of a partially filled template, the input text, and knowledge derived from LLMs. This transformation allows us

*Corresponding author.
[1]https://github.com/zhangzhao219/LKI-BART
[2]https://openai.com/chatgpt

Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers), pages 371–378
June 16-21, 2024 ©2024 Association for Computational Linguistics

to go beyond direct classification and leverage rich semantics for stance decoding.

Based on the above framework, we further devise a prototypical contrastive learning scheme to align stance representations with label semantics. Specifically, we adopt class-wise prototypes to model the feature space. Then a contrastive loss is optimized, forcing stance representations to be appealed to the corresponding prototype and repelled to prototypes of other stances. As a result, per-class representations can be clustered compactly on the feature space, making it easier to learn a mapping from stance representations to label semantics.

We refer to our method as **L**LM-Driven **K**nowledge **I**njection BART (LKI-BART), and an overview is depicted in Figure 2. Extensive experiments on two benchmark datasets suggest that LKI-BART achieves state-of-the-art performance on zero-shot and cross-target stance detection tasks. Additional ablation studies also indicate the effectiveness of each technique, namely LLM-driven knowledge, BART backbone, and prototypical contrastive learning, in LKI-BART.

2 Related Work

Zero-Shot and Cross-Target Stance Detection with External Knowledge. In addition to designing algorithms to learn transferable features, several works introduce external knowledge for a deeper understanding of unseen targets. Liu et al. (2021) involved commonsense knowledge based on the graph structure, Zhu et al. (2022) incorporated external knowledge from Wikipedia. In recent studies, He et al. (2022); Wen and Hauptmann (2023); Hanley and Durumeric (2023) have reached a consensus on the usage of external knowledge. However, none of them considers explicitly acquiring the relationship between texts and targets, which may result in redundant and confusing information.

Utilize LLMs for Stance Detection. Recently, some works have directly used LLMs for stance detection. Zhang et al. (2023a,b) explores various methods to prompt ChatGPT for stance reasoning. Lan et al. (2023) adopts distinct LLM-based Agents to create a collaborative stance detection system. Also, some studies have expressed a negative opinion on using LLMs for stance detection. Zhu et al. (2023) uses ChatGPT to annotate stance labels but obtains inferior results. Cruickshank and Ng (2023) conclude that overall accuracy is not much better than supervised models.

3 Approach

3.1 Problem Formulation

Formally, given a text x and a target t, the stance detection task aims to identify the stance y that x expresses towards t. y is basically in the collection $S = \{favor, against, neutral\}$, which may vary in different datasets. The detection model is trained to infer y given x and t with parameter θ. If knowledge k is involved, the formulation will be:

$$f(x, t, k; \theta) = y$$

3.2 LLM-Driven Knowledge

To acquire LLM-driven knowledge for stance detection, we employ a partially filled zero-shot prompt for each input text in every dataset, as shown in Appendix A. Specifically, our prompt is designed with three aspects in mind. First, we instruct the LLM to *list keywords*. In that case, the subsequent detection model may pay more attention to these words. Second, we prompt the LLM to *analyze implied emotions and rhetorical devices*, as these may be strong pieces of evidence for the expressed stance. Finally, unlike previous studies, we ask the LLM to *briefly analyze the stance* rather than produce exact answers directly, reducing the risks of intrinsic hallucination in LLMs, which may mislead the subsequent detection model into generating incorrect predictions. Besides, we add a brief description of the dataset at the beginning, so that LLM may find more background information concerning the dataset by retrieving its internal knowledge.

3.3 BART Backbone

As LLM-driven knowledge may cover rich contextual information, it is vital to associate such information with stance semantics. Recently, generative modeling has shown great potential by leveraging pre-training objectives to decode the answer (Radford et al., 2019). Inspired by prior works (Wen and Hauptmann, 2023), we inject LLM-driven knowledge into BART (Lewis et al., 2020), an autoregressive transformer, for stance detection.

In that case, the task is reframed as a denoising one, which takes $h(x, t, k)$ as input and generates an output sequence $\boldsymbol{u}$ containing stance labels. Specifically, $h(x, t, k)$ is a combination of input text x, target t and LLM-driven knowledge k with special tokens "<s> <stance> is the stance for the target t </s></s> x </s></s> k </s>", and $\boldsymbol{u}$ is formulated as "<s> <stance> is the stance

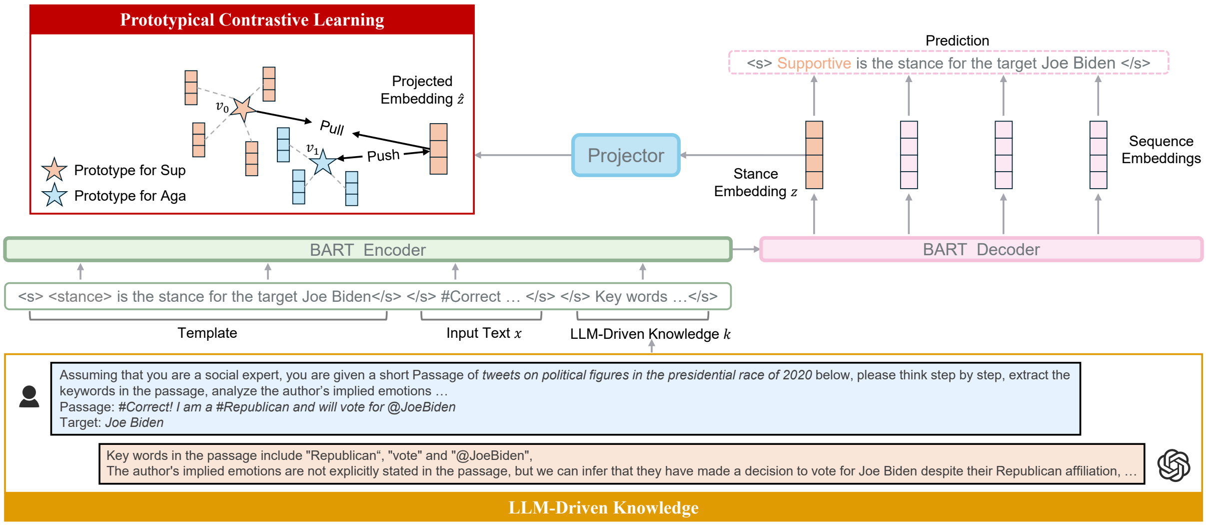

Figure 2: Overview of our proposed LKI-BART.

`for the target t </s>"`. Note that the `<stance>` placeholder is kept in input while being replaced by the predicted stance in output.

Finally, the BART model is trained by maximizing the log-likelihood over the whole sequence.

$$\mathcal{L}_{gen} = -\sum_{i=1}^{|\boldsymbol{u}|} \log p(u_i|, \boldsymbol{u}_{<i}, h(x,t,k); \theta)$$

where $p(u_i|, \boldsymbol{u}_{<i}, h(x,t,k); \theta)$ is the probability to select a token u_i at step i given the input $h(x,t,k)$ and previously generated tokens $\boldsymbol{u}_{<i}$.

3.4 Prototypical Contrastive Learning

One may notice that $\mathcal{L}_{gen}$ maximizes the likelihood over the entire sentence. In fact, more optimizations may be beneficial especially when decoding the stance label. To this end, we decouple the stance embedding $\boldsymbol{z} \in \mathbb{R}^{embed\_size}$ (the embedding used to generate the stance token) from BART decoder outputs. We then project it into a low-dimensional vector $\hat{\boldsymbol{z}} \in \mathbb{R}^{low\_embed\_size}$ to prevent $\boldsymbol{z}$ losing much semantics and being overcorrected. To regularize $\hat{\boldsymbol{z}}$ to be more discriminative in the latent space, inspired by Li et al. (2021a), we introduce the concept of prototypes, they are widely adopted in data-efficient learning Li et al. (2024) and can be viewed as the representatives of class-wise embeddings. However, different from Li et al. (2021a), we view prototypes as the representatives of class-wise projected embeddings instead of class-agnostic ones. By interacting with prototypes, a contrastive loss is employed to increase the intra-class similarity but decrease the inter-class similarity of projected stance embeddings. In the following, we will detail how to estimate class-wise prototypes and formulate the contrastive loss.

Online Prototype Update For each stance class c, we randomly initialize a vector $\boldsymbol{v}_c$ as its prototype before normalizing it into a unit one at the beginning. Along the training progress, we update $\boldsymbol{v}_c$ at each step in a moving average manner by,

$$\boldsymbol{v}_c \leftarrow \text{Normalize}(\beta \boldsymbol{v}_c + (1-\beta)\boldsymbol{v}'_c)$$

where β is a momentum coefficient, Normalize$(\cdot)$ is the normalization function, and $\boldsymbol{v}'_c$ is the centroid of embeddings belonging to class c in the batch.

Prototypical Contrastive Loss To compute the loss, we firstly obtain the embedding-to-class cosine similarity score $s_j = < \hat{\boldsymbol{z}}, \boldsymbol{v}_j >$. Then, we optimize the following loss,

$$\mathcal{L}_{con} = -\sum_{c=1}^{C} y_c \log \frac{\exp(\frac{s_c}{\gamma})}{\sum_{j=1}^{C} \exp(\frac{s_j}{\gamma})}$$

where γ is a scalar temperature parameter and $\boldsymbol{y}$ is the one-hot label for the current sample. If it belongs to class c, optimizing $\mathcal{L}_{con}$ will maximize s_c but minimize s_j ($j = 1, \cdots, C$ and $j \neq c$), thereby pulling together $\hat{\boldsymbol{z}}$ and class c's prototype $\boldsymbol{v}_c$ while pushing away $\hat{\boldsymbol{z}}$ from prototypes of other classes. Consequently, a well-structured feature space is modeled with stance embeddings from the same class clustering together, making it easier to reach semantic alignment.

Relation with JointCL Although previous work, namely JointCL (Liang et al., 2022b), also introduced the concept of prototypes, our approach is fundamentally distinctive from it. As for the purpose of prototypes, JointCL obtains class-agnostic prototypes to build a graph dedicated to adapting and refining representations of unseen targets, while we rely on class-specific prototypes to regularize the feature space. Besides, JointCL still follows the traditional supervised contrastive loss (Khosla et al., 2020), forcing larger cosine similarity between samples that share the same stance label, while prototypes are not involved in contrastive loss at all. By contrast, our method encourages large similarity scores between a sample's embedding and its corresponding prototype, which is less computationally intensive and more robust, since potential mis-calibration can be alleviated by viewing prototypes as representatives of class-wise embeddings.

The overall loss function of LKI-BART considering both $\mathcal{L}_{gen}$ and $\mathcal{L}_{con}$ is defined as follows:

$$\mathcal{L} = \lambda_l \cdot \mathcal{L}_{gen} + (1 - \lambda_l) \cdot \mathcal{L}_{con}$$

where λ_l is involved to balance the optimization.

4 Experiments

We evaluate our LKI-BART on VAST (Allaway and McKeown, 2020) and P-Stance (Li et al., 2021b), please refer to Appendix B and Appendix C for details of the training datasets as well as our experimental setup.

4.1 Results

4.1.1 Zero-Shot Stance Detection

We test our model on VAST, where the model is trained on thousands of targets and evaluated on targets not present in the training data.

Model	VAST			
	Sup	Aga	Neu	Avg
StSQA	-	-	-	68.9
GDA-CL	59.8	62.3	89.3	70.5
PT-HCL	61.7	63.5	89.6	71.6
JointCL	64.9	63.2	88.9	72.3
BS-RGCN	60.8	67.4	89.5	72.6
COLA	73.4	77.2	-	73.4
TarBK-BERT	65.7	63.9	91.2	73.6
WS-BERT-Single	-	-	-	75.3
CondGen	-	-	-	76.4
KASD-BERT	-	-	-	76.8
TATA	69.5	71.1	90.5	77.1
LKI-BART	75.1	72.9	90.7	**79.6**

Table 1: Experimental results on VAST.

Baselines. We compare our model with StSQA (Zhang et al., 2023b), GDA-CL (Li and Yuan, 2022), PT-HCL (Liang et al., 2022a), JointCL (Liang et al., 2022b), BS-RGCN (Luo et al., 2022), COLA (Lan et al., 2023), TarBK-BERT (Zhu et al., 2022), WS-BERT-Single (He et al., 2022), CondGen (Wen and Hauptmann, 2023), KASD-BERT (Li et al., 2023) and TATA (Hanley and Durumeric, 2023).

Results and Analysis. The results of LKI-BART on VAST are presented in Table 1. Notably, LKI-BART exhibits a significant performance improvement over prior approaches, validating its effectiveness on the zero-shot stance detection task.

Ablation Study on LLM-Driven Knowledge. To validate the effectiveness of our proposed LLM-driven knowledge, we integrate different components of it into our model as well as target-related knowledge from Wikipedia (He et al., 2022). The results are presented in Table 2, showing that each component contributes to the performance gain and the combination of them can complement each other and generally yields a superior result beyond Wiki knowledge only. Besides, we observe a significant improvement with the "Analyse stance" part. This is because instructing LLMs to analyze the stance directly seems to be the most straightforward way. Nevertheless, the other components also play essential roles. In cases where "Analyse stance" leads to an incorrect answer, our LKI-BART can still correct it with the guidance of other parts, suggesting its robustness.

Component	VAST			
	Sup	Aga	Neu	Avg
None	65.9	68.4	90.1	74.8
Wiki	71.4	71.8	87.4	76.9
Keywords	67.2	70.3	89.6	75.7
Implied emotions	67.6	72.5	90.6	76.9
Rhetorical devices	67.7	71.0	89.9	76.2
Analyse stance	71.2	71.9	89.3	77.4
All	**75.1**	**72.9**	**90.7**	**79.6**

Table 2: Ablation study on different knowledge types or components as contextual information for BART model.

Ablation Study on BART Backbone. As LLM-driven knowledge may cover rich contextual information, it is vital to associate such information with stance semantics. Therefore, we choose a seq2seq generative model BART as our backbone. To further verify this choice, we try two alternatives of BART to examine its effects in Table 3: 1) feed the LLM-driven knowledge into a BERT (Devlin et al., 2019) classification model (denoted as BERT); 2) feed the LLM-driven knowledge into another LLM and use its few-shot in-context abilities to directly infer the stance (denoted as few-shot LLM). As shown, training a seq2seq BART separately performs best, as it can learn how to leverage the knowledge provided by LLMs more effectively.

Backbone	**VAST**			
	Sup	Aga	Neu	Avg
few-shot LLM	68.7	67.8	72.4	69.6
BERT	73.5	**74.3**	85.2	77.7
BART	**75.1**	72.9	**90.7**	**79.6**

Table 3: Ablation study on different backbones.

4.1.2 Cross-Target Stance Detection

We adopt P-Stance for cross-target stance detection evaluation, where our model is trained on one target and tested on another related target.

Baselines. We compare our model with BiCE (Augenstein et al., 2016), CrossNet (Xu et al., 2018), BERTweet (Li et al., 2021b) and WS (He et al., 2022).

Results and Analysis. The results of our model on P-Stance are shown in Table 4. We can observe that LKI-BART outperforms the previous best model by 10-15 F1-pts. Interestingly, we note that among the three targets, our model performs best on JB, followed by DT and BS. We speculate that this discrepancy arises from JB being the current president and DT being the former president. In that case, LLMs possess more internal information about JB and DT thereby producing more informative knowledge on them after prompting.

Target	**P-Stance**			
	CrossNet	BERTweet	WS	LKI-BART
DT→ JB	56.67	58.88	68.30	**85.02**
DT→ BS	50.08	56.50	64.40	**79.57**
JB→ DT	60.43	63.64	67.70	**80.74**
JB→ BS	60.81	67.04	69.00	**79.65**
BS→ DT	52.99	58.75	63.60	**80.91**
BS→ JB	62.51	72.99	76.80	**85.56**

Table 4: Experimental results on P-Stance, where JB, DT, BS are short for Joe Biden, Donald Trump, and Bernie Sanders, respectively.

Feature Visualizations. Figure 3 shows the t-SNE (van der Maaten and Hinton, 2008) visualization of stance embeddings (the definition is given in Section 3.4) from the vanilla BART model and our LKI-BART model on the test set. Specifically, the vanilla BART model is based on the generation framework but trained without LLM-driven knowledge or prototypical contrastive loss. As shown, the visualization of the vanilla BART shows basically no clusters, while embeddings with our LKI-BART are gathered according to their labels.

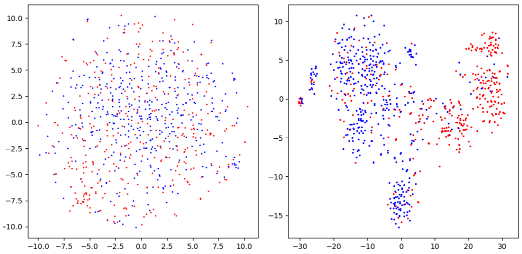

Figure 3: The t-SNE visualization of stance embeddings from the vanilla BART (left) and our LKI-BART (right). We use random initialization with perplexity as 50, blue stands for AGAINST label, and red stands for FAVOR label.

Ablation Study on Training Strategies. To validate the effects of our proposed strategies, we conduct several experiments. The corresponding results are listed in Table 5. As seen, LK greatly boosts the detection ability by explicitly modeling relationships between texts and targets while PCL can further bring about improvements as it helps to regularize the feature space. Besides, our PCL surpasses its variant SCL, which may be easily influenced by in-batch outliers due to its sample-to-sample contrastive formula.

Target	**P-Stance**			
	-	LK	LK + SCL	LK + PCL
DT→ JB	73.63	84.09	84.55	**85.02**
DT→ BS	62.36	78.71	**79.69**	79.57
JB→ DT	65.25	79.62	80.13	**80.74**
JB→ BS	66.74	78.68	78.41	**79.65**
BS→ DT	59.32	79.60	80.22	**80.91**
BS→ JB	72.67	84.83	84.98	**85.56**

Table 5: Ablation study for training strategies on P-Stance, where LK represents the LLM-driven knowledge, SCL denotes the supervised contrastive loss adopted in (Liang et al., 2022b), and PCL represents the proposed prototypical contrastive learning.

5 Conclusion

In this paper, we propose to collect LLM-driven knowledge to incorporate connections between input texts and unseen targets for zero-shot and cross-target stance detection. A generation framework BART is adopted to better leverage LLM-driven knowledge for detection and a prototypical contrastive loss is optimized for better alignment between input materials and stance semantics. Combining all the above techniques, our LKI-BART finally achieves state-of-the-art performance on VAST and P-Stance datasets.

Limitations

LKI-BART relies on knowledge generated by LLMs. Due to constraints in budget and time, we only experiment with GPT-3.5-turbo from Azure. We encourage further exploration by researchers to compare various LLMs and prompt formats.

Ethical Statement

In this research, it's crucial to acknowledge the potential limitations of LLMs. Although Azure has made significant progress to guard against abuse and unintended harm, ChatGPT may also produce biased information on certain targets as many other LLMs, especially on targets related to people. However, we do not adopt any additional processing for LLM-driven knowledge, while other parts of the training data come from publicly available datasets that are commonly employed in prior research. We keep fair and honest in our analysis of experimental results. Additionally, our LKI-BART is extremely lightweight and allows the reproduction of the experiments on common GPUs. We have made our code accessible for future investigations.

References

Emily Allaway and Kathleen McKeown. 2020. Zero-Shot Stance Detection: A Dataset and Model using Generalized Topic Representations. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8913–8931, Online. Association for Computational Linguistics.

Isabelle Augenstein, Tim Rocktäschel, Andreas Vlachos, and Kalina Bontcheva. 2016. Stance detection with bidirectional conditional encoding. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 876–885, Austin, Texas. Association for Computational Linguistics.

Iain J. Cruickshank and Lynnette Hui Xian Ng. 2023. Use of large language models for stance classification.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Ivan Habernal, Henning Wachsmuth, Iryna Gurevych, and Benno Stein. 2018. The argument reasoning comprehension task: Identification and reconstruction of implicit warrants. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1930–1940, New Orleans, Louisiana. Association for Computational Linguistics.

Hans W. A. Hanley and Zakir Durumeric. 2023. Tata: Stance detection via topic-agnostic and topic-aware embeddings.

Andreas Hanselowski, Avinesh PVS, Benjamin Schiller, Felix Caspelherr, Debanjan Chaudhuri, Christian M. Meyer, and Iryna Gurevych. 2018. A retrospective analysis of the fake news challenge stance-detection task. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1859–1874, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Momchil Hardalov, Arnav Arora, Preslav Nakov, and Isabelle Augenstein. 2022. A survey on stance detection for mis- and disinformation identification. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 1259–1277, Seattle, United States. Association for Computational Linguistics.

Kazi Saidul Hasan and Vincent Ng. 2014. Why are you taking this stance? identifying and classifying reasons in ideological debates. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 751–762, Doha, Qatar. Association for Computational Linguistics.

Zihao He, Negar Mokhberian, and Kristina Lerman. 2022. Infusing knowledge from Wikipedia to enhance stance detection. In *Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis*, pages 71–77, Dublin, Ireland. Association for Computational Linguistics.

Kornraphop Kawintiranon and Lisa Singh. 2021. Knowledge enhanced masked language model for stance detection. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4725–4735, Online. Association for Computational Linguistics.

Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. 2020. Supervised contrastive learning. *Advances in neural information processing systems*, 33:18661–18673.

Dilek Küçük and Fazli Can. 2022. A tutorial on stance detection. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, WSDM '22, page 1626–1628, New York, NY, USA. Association for Computing Machinery.

Xiaochong Lan, Chen Gao, Depeng Jin, and Yong Li. 2023. Stance detection with collaborative role-infused llm-based agents.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online. Association for Computational Linguistics.

Ang Li, Bin Liang, Jingqian Zhao, Bowen Zhang, Min Yang, and Ruifeng Xu. 2023. Stance detection on social media with background knowledge. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 15703–15717, Singapore. Association for Computational Linguistics.

Junnan Li, Pan Zhou, Caiming Xiong, and Steven C. H. Hoi. 2021a. Prototypical contrastive learning of unsupervised representations.

Yang Li and Jiawei Yuan. 2022. Generative data augmentation with contrastive learning for zero-shot stance detection. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 6985–6995, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Yiming Li, Xiangdong Wang, Hong Liu, Rui Tao, Long Yan, and Kazushige Ouchi. 2024. Semi-supervised sound event detection with local and global consistency regularization. In *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 271–275.

Yingjie Li, Tiberiu Sosea, Aditya Sawant, Ajith Jayaraman Nair, Diana Inkpen, and Cornelia Caragea. 2021b. P-stance: A large dataset for stance detection in political domain. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 2355–2365, Online. Association for Computational Linguistics.

Bin Liang, Zixiao Chen, Lin Gui, Yulan He, Min Yang, and Ruifeng Xu. 2022a. Zero-shot stance detection via contrastive learning. In *Proceedings of the ACM Web Conference 2022*, WWW '22, page 2738–2747, New York, NY, USA. Association for Computing Machinery.

Bin Liang, Qinglin Zhu, Xiang Li, Min Yang, Lin Gui, Yulan He, and Ruifeng Xu. 2022b. JointCL: A joint contrastive learning framework for zero-shot stance detection. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 81–91, Dublin, Ireland. Association for Computational Linguistics.

Hongzhan Lin, Jing Ma, Mingfei Cheng, Zhiwei Yang, Liangliang Chen, and Guang Chen. 2021. Rumor detection on Twitter with claim-guided hierarchical graph attention networks. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 10035–10047, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Rui Liu, Zheng Lin, Yutong Tan, and Weiping Wang. 2021. Enhancing zero-shot and few-shot stance detection with commonsense knowledge graph. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 3152–3157, Online. Association for Computational Linguistics.

Ilya Loshchilov and Frank Hutter. 2019. Decoupled weight decay regularization.

Yun Luo, Zihan Liu, Yuefeng Shi, Stan Z. Li, and Yue Zhang. 2022. Exploiting sentiment and common sense for zero-shot stance detection. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 7112–7123, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners.

Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86):2579–2605.

Haoyang Wen and Alexander Hauptmann. 2023. Zero-shot and few-shot stance detection on varied topics via conditional generation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 1491–1499, Toronto, Canada. Association for Computational Linguistics.

Chang Xu, Cécile Paris, Surya Nepal, and Ross Sparks. 2018. Cross-target stance classification with self-attention networks. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 778–783, Melbourne, Australia. Association for Computational Linguistics.

Bowen Zhang, Daijun Ding, and Liwen Jing. 2023a. How would stance detection techniques evolve after the launch of chatgpt?

Bowen Zhang, Xianghua Fu, Daijun Ding, Hu Huang, Yangyang Li, and Liwen Jing. 2023b. Investigating chain-of-thought with chatgpt for stance detection on social media.

Qinglin Zhu, Bin Liang, Jingyi Sun, Jiachen Du, Lanjun Zhou, and Ruifeng Xu. 2022. Enhancing zero-shot stance detection via targeted background knowledge. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '22, page 2070–2075, New York, NY, USA. Association for Computing Machinery.

Yiming Zhu, Peixian Zhang, Ehsan-Ul Haq, Pan Hui, and Gareth Tyson. 2023. Can chatgpt reproduce human-generated labels? a study of social computing tasks.

A Prompt Details

Assuming that you are a social expert, you are given a short Passage of **{dataset_description}** below, please think step by step, extract the keywords in the passage, analyze the author's implied emotions, rhetorical devices, etc., finally briefly analyze the author's stance on the Target, paying attention to giving the process of analysis without giving the conclusion.

Passage:**{text}**

Target:**{target}**

B Datasets

Allaway and McKeown (2020) construct a dataset named VAST with varied topics for evaluating zero-shot stance detection, the original examples of VAST are collected from (Habernal et al., 2018) under Apache-2.0 license[3], while P-Stance by Li et al. (2021b) is a commonly adopted benchmark for cross-target stance detection, under MIT license[4].

1) VAST contains 18,548 comments from New York Times "Room for Debate" section with 5,630 different targets for zero-shot and few-shot stance detection. Each instance can be classified as Favor, Against, or Neutral. The statistics are summarized in Table 6.

	Train	Dev	Test
# Examples	13,477	2,062	3,006
# Unique Comments	1,845	682	786
# Zero-shot Topics	4,003	383	600
# Few-shot Topics	638	114	159

Table 6: Statistics of VAST dataset.

2) P-Stance contains 21,574 political tweets with stance annotations for "Donald Trump", "Joe Biden", and "Bernie Sanders". Each tweet is annotated with a stance label "Favor" or "Against". The statistics are summarized in Table 7.

		Trump	Biden	Sanders
Train	Favor	2,937	2,552	2,858
	Against	3,425	3,254	2,198
Dev	Favor	365	328	350
	Against	430	417	284
Test	Favor	361	337	343
	Against	435	408	292

Table 7: Statistics of P-Stance dataset.

C Experimental Setup

We use unmodified bart-base from huggingface.co[5]. All our experiments are carried out on a single NVIDIA A100 40G with 50 epochs, which generally take about 3 hours for training. We use a base learning rate of 5e-6 with a warm-up proportion of 0.1 and AdamW (Loshchilov and Hutter, 2019) is adopted as the optimizer. The training batch size is defined as 64. For prototypical contrastive learning, hyper-parameters τ, λ_l, and β are set to 0.1, 0.8, and 0.99 respectively. Test results are reported based on the best overall F1 performance on the development set, using the averaged results from 5 different random seeds. Following the experimental setup by previous works, we use macro-F1 as the evaluation metric.

[3] https://github.com/UKPLab/argument-reasoning-comprehension-task/blob/master/LICENSE

[4] https://github.com/chuchun8/PStance/blob/main/LICENSE

[5] https://huggingface.co/facebook/bart-base