Target-Oriented Relation Alignment for Cross-Lingual Stance Detection

Ruike Zhang[1,2]**, Nan Xu**[1,3]*****, Hanxuan Yang**[2,1]**, Yuan Tian**[1,2]**, Wenji Mao**[1,2]*
[1]Institute of Automation, Chinese Academy of Sciences
[2]School of Artificial Intelligence, University of Chinese Academy of Sciences
[3]Beijing Wenge Technology Co., Ltd.
{zhangruike2020,xunan2015,yanghanxuan2020,
tianyuan2021,wenji.mao}@ia.ac.cn

Abstract

Stance detection is an important task in text mining and social media analytics, aiming to automatically identify the user's attitude toward a specific target from text, and has wide applications in a variety of domains. Previous work on stance detection has mainly focused on monolingual setting. To address the problem of imbalanced language resources, cross-lingual stance detection is proposed to transfer the knowledge learned from a high-resource (source) language (typically English) to another low-resource (target) language. However, existing research on cross-lingual stance detection has ignored the inconsistency in the occurrences and distributions of targets between languages, which consequently degrades the performance of stance detection in low-resource languages. In this paper, we first identify the target inconsistency issue in cross-lingual stance detection, and propose a fine-grained Target-oriented Relation Alignment (TaRA) method for the task, which considers both target-level associations and language-level alignments. Specifically, we propose the Target Relation Graph to learn the in-language and cross-language target associations. We further devise the relation alignment strategy to enable knowledge transfer between semantically correlated targets across languages. Experimental results on the representative datasets demonstrate the effectiveness of our method compared to competitive methods under variant settings.

1 Introduction

Stance detection is an important task in public opinion mining and social media analytics, which aims to automatically identify the user's attitude (e.g., "*in favor of*" or "*against*") toward a specific target (e.g., *entity*, *topic*, or *claim*) from text. It has been widely applied to many domains such as veracity checking, market analysis, social security and government decision-making (Küçük and Can, 2020; AlDayel and Magdy, 2021).

	English	**French**
Target	Feminist Movement	légaliser l'avortement (*Legalization of Abortion)*
Text	I'm a feminist, I believe in equality for all. #EqualityForAll	C'est tellement génial que #lovewins - étende maintenant l'égalité des droits des femmes *(It's so awesome that #lovewins - now extends women's equal rights)*
Stance	Favor	Favor

Table 1: An example of cross-lingual stance detection. The original French text with the target is presented with English translation.

Existing studies on stance detection are mainly conducted in monolingual setting, focusing on English (Hardalov et al., 2021; Allaway et al., 2021; Liang et al., 2022). In contrast to the abundant corpora in English, the annotated data resources for stance detection in other languages are usually scarce. To address the imbalanced data resources between languages and support stance-related applications in low-resource languages, cross-lingual stance detection is proposed to transfer the knowledge learned from the high-resource (source) language to the low-resource (target) language (Küçük and Can, 2020). Two studies have been conducted for cross-lingual stance detection, by adopting contrastive language adaptation to align the representations across languages (Mohtarami et al., 2019) and pre-training the language model to acquire additional knowledge (Hardalov et al., 2022).

In addition to the problem of imbalanced language resources, another important issue in cross-lingual stance detection has been ignored by current research. Due to the differences in contextual information, language expressions and socio-cultural backgrounds, the occurrences and distributions of the concerned targets may vary considerably across languages. For example, even in the same domain such as "presidential election", the targets in the English corpus are distinct from those in the French one. Since the discrepancy in target distri-

*Corresponding author

Findings of the Association for Computational Linguistics: ACL 2023, pages 6391–6404
July 9-14, 2023 ©2023 Association for Computational Linguistics

butions pervasively exists between languages, the targets in the source and target language datasets cannot precisely align in cross-lingual stance detection. This brings about the target inconsistency issue, that is, the difficulty of knowledge transfer across languages caused by the misalignment, which inevitably leads to the performance decrease for cross-lingual stance detection.

To address the target inconsistency issue, semantic associations between different targets can be utilized for cross-lingual stance detection. Table 1 gives an example of a target relation. The target "*Feminist Movement*" in English and the target "*Legalization of Abortion*" in French are highly correlated, since both are highly associated with women's rights and they mention quite similar topics such as equality. Target relations reflect the semantic associations between targets on the shared background information or topics, which prevalently exist in textual expressions within and across languages.

In this paper, we model the target-level associations and propose a fine-grained **Ta**rget-oriented **R**elation **A**lignment (TaRA) method for cross-lingual stance detection. Our method considers both target-level associations and language-level alignments. In addition, to guarantee the cross-language performance on stance detection, the in-language target relation learning and relation contrastive alignments should be maintained first. Specifically, it first learns the relations between different targets via target relation graph within each language (i.e., in-language), and constructs the cross-lingual relation graph to compensate for target inconsistency. We then devise the in-language and cross-language relation alignment strategies to align the samples with highly correlated targets based on the relation graph, so as to enable knowledge transfer between semantically correlated targets across languages.

The contributions of our work are as follows:

- We identify the target inconsistency issue in cross-lingual stance detection for the first time, and propose a computational method TaRA to tackle this problem via target-level correlation learning and relation alignment across languages.

- Our method learns the associations between targets via target relation graphs within and across languages, and designs relation alignment strategies that enable in-language knowledge enhancement and cross-lingual knowledge transfer among semantically correlated targets.

- We conduct experiments on the representative multilingual stance datasets with variant settings, and the results demonstrate the effectiveness of our method compared to the competitive methods.

2 Related Work

Stance detection has been well studied on English datasets (Mohammad et al., 2016; Sobhani et al., 2017; Conforti et al., 2020; Allaway and Mckeown, 2020; Glandt et al., 2021). Previous methods for stance detection mainly focus on monolingual setting to learn a supervised classification model with labeled data. The mainstream research centers on stance detection for pre-defined targets (Du et al., 2017; Zhou et al., 2017; Wei et al., 2018; Sun et al., 2018; Li and Caragea, 2019), cross-target stance detection (Xu et al., 2018; Wei and Mao, 2019; Zhang et al., 2020)and few/zero-shot stance detection (Allaway et al., 2021; Liang et al., 2022).

Compared to the abundant resources in English, there are much fewer data resources in other languages (Xu et al., 2016; Lozhnikov et al., 2018; Baly et al., 2018; Khouja, 2020; Cignarella et al., 2020). To promote stance detection in low-resource languages, some researchers make efforts to construct multilingual stance datasets (Taulé et al., 2017; Vamvas and Sennrich, 2020; Lai et al., 2020), while other research develops the methods for cross-lingual stance detection (Mohtarami et al., 2019; Hardalov et al., 2022). Hardalov et al. (2022) conducts a comprehensive empirical study on pretraining the language model with additional corpora in the source language, to acquire and transfer the knowledge to the target language through prompt-tuning.

Mohtarami et al. (2019) proposes a contrastive language adaptation method to align the representations across languages, which encourages samples with the same label in different languages to be closer in the embedding space. However, their method only considers the contrastive adaptation at the language level, ignoring the fine-grained modeling of target relations that is essential to the compensation for target inconsistency and can facilitate cross-lingual stance detection in general. Another drawback of the previous method

(Mohtarami et al., 2019) is that it only considers cross-language contrastive alignment and ignores the in-language target relation learning and contrastive alignments. Therefore, in our work, we consider both target-level modeling and language-level alignments, and develop our computational method with in-language and cross-language solutions to tackle the target inconsistency issue for cross-lingual stance detection.

3 Proposed Method

Figure 1 illustrates the overall structure of our proposed method TaRA. We first encode the input target and text with the *Encoder Module* and get the textual representations. Then, we construct the *Target Relation Graph* to learn both in-language and cross-language associations between targets and get target representations with aggregated information. After that, we concatenate the textual representations and target representations for classification. Meanwhile, we align the representations with related targets within and across languages using the *Target Relation Alignment Strategies*.

3.1 Problem Statement

We denote the src language data as $D_{src} = \{(t_i^s, c_i^s), y_i^s\}_{i=1}^{N_s}$, where t_i^s is the i-th target, c_i^s is the i-th text, and y_i^s is the stance label of text c_i^s towards target t_i^s, N_s is the number of samples in D_{src}. We also denote the target set of D_{src} as $\mathcal{T}_s = \{T_i^s\}_{i=1}^{n_s}$, where n_s is the number of the targets in D_{src}. Similarly, we denote the tgt[1] language data as $D_{tgt} = \{(t_i^t, c_i^t), y_i^t\}_{i=1}^{N_t}$ with the target set $\mathcal{T}_t = \{T_i^t\}_{i=1}^{n_t}$. Typically, $N_t \ll N_s$ in cross-lingual stance detection. We use both D_{src} and D_{tgt} to train the model and predict the stance label of each sample in the tgt language test set.

3.2 Encoder Module

To map words in different languages into the same embedding space, we use a language model mBERT (Devlin et al., 2019) as the encoder module, which is pre-trained on a large-scale multilingual corpus. Specifically, given a pair of target t and text c, we encode them with the encoder module and obtain a textual representation $\boldsymbol{h} \in \mathbb{R}^d$:

$$\boldsymbol{h} = \text{mBERT}([CLS]t[SEP]c[SEP])_{[CLS]} \quad (1)$$

where t and c are sequences of words in the target and text respectively.

[1] In this paper, we abbreviate target (language) as "tgt" and source (language) as "src" to avoid any confusion.

3.3 In-Language Target Relation Graphs

To reduce the impact of the language gap on relation learning across languages, we first learn the target relations within the language to provide preliminary knowledge for cross-lingual target relation modeling. Specifically, we construct src and tgt target relation graphs $\mathcal{G}^* = \langle \mathcal{V}^*, \mathcal{A}^* \rangle$, $* \in \{s, t\}$, where $\mathcal{V}^*$ represents the node features and $\mathcal{A}^*$ is the adjacent matrix.

Graph Construction Each target is treated as a node in the graph, and the correlations of nodes reflect target relations. Intuitively, the relationship between targets is characterized by the relationship between their corresponding text sets. Hence, for each target T_i^*, we also use mBERT to derive all the textual representations with target T_i^* and calculate the mean vector as the feature $\boldsymbol{v}_i^* \in \mathbb{R}^d$ of the i-th node:

$$\boldsymbol{v}_i^* = \text{Average}\left([\boldsymbol{h}_{i,j}^*]_{j=1}^{N_i^*}\right) \quad (2)$$

$$\boldsymbol{h}_{i,j}^* = \text{mBERT}([CLS]T_i^*[SEP]c_j^*[SEP])_{[CLS]} \quad (3)$$

where $\boldsymbol{h}_{i,j}^* \in \mathbb{R}^d$ is the j-th textual representation with target T_i^* and N_i^* is the number of samples with target T_i^*.

After obtaining node features $\mathcal{V}^* = \{\boldsymbol{v}_i^*\}_{i=1}^{n_*}$ for $\mathcal{G}^*$, we construct the adjacent matrix $\mathcal{A}^* \in \{0,1\}^{n_* \times n_*}$. We use the semantic similarities between targets as the start point, which can be viewed as an approximation of semantic associations of targets and used as a basis for subsequent target relation learning. Specifically, we calculate the cosine similarity score $score_{i,j}^*$ between $\boldsymbol{v}_i^*$ and $\boldsymbol{v}_j^*$ and filtering them with threshold θ_0:

$$score_{i,j}^* = f(\boldsymbol{v}_i^*, \boldsymbol{v}_j^*) = \frac{\boldsymbol{v}_i^* \cdot \boldsymbol{v}_j^*}{\|\boldsymbol{v}_i^*\| \|\boldsymbol{v}_j^*\|} \quad (4)$$

$$\mathcal{A}_{i,j}^* = \begin{cases} 1 \text{ if } score_{i,j}^* > \theta_0 \\ 0 \text{ otherwise} \end{cases} \quad (5)$$

where $f(\cdot)$ is the cosine similarity function.

Target Relation Calculation To dynamically model the associations between targets, we adopt Graph Attention Network (GAT) (Veličković et al., 2018) to learn the weights between nodes and obtain high-level target representations with aggregated information. Specifically, we feed the node features $\mathcal{V}^*$ and the adjacent matrix $\mathcal{A}^*$ into GAT, and derive the target representations $\mathcal{U}^* = \{\boldsymbol{u}_i^*\}_{i=1}^{n_*}$ ($\boldsymbol{u}_i^* \in \mathbb{R}^d$) and the attention weight matrix $\mathcal{W}^* \in \mathbb{R}^{n_* \times n_*}$.

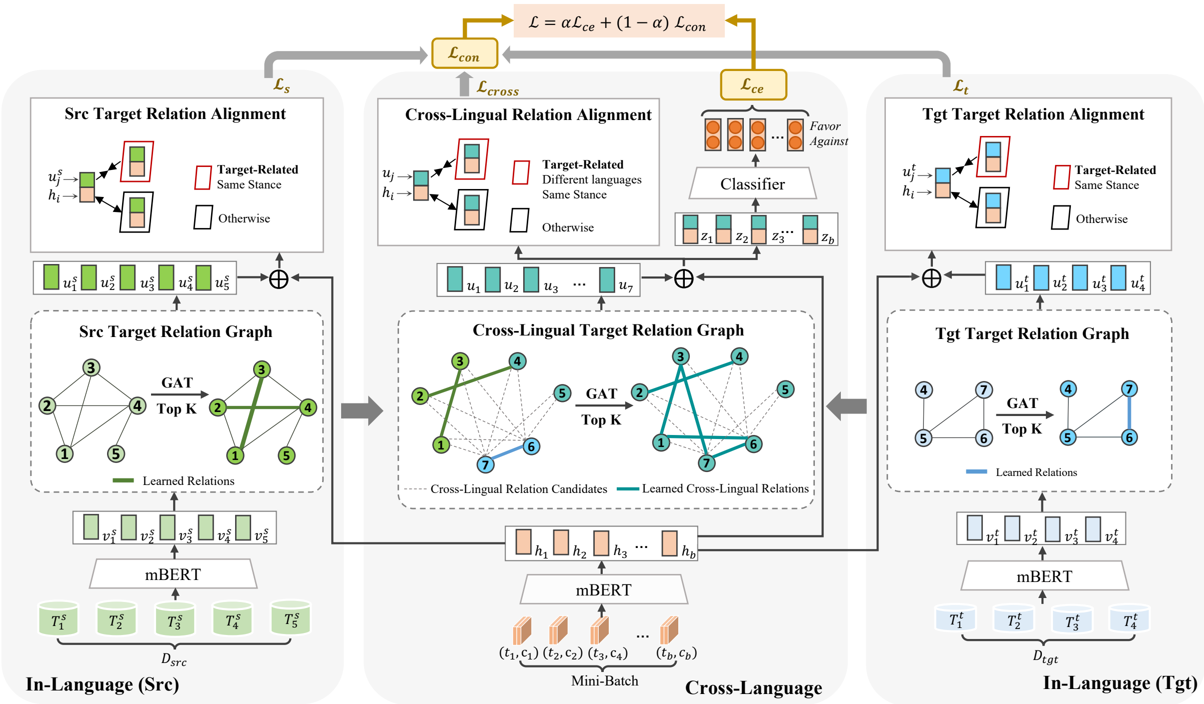


Figure 1: The overall architecture of our proposed method TaRA.

We adopt Top K to convert the weight matrix $\mathcal{W}^*$ learned from $\mathcal{G}^*$ into the target relation matrix $\mathcal{R}^* \in \{0,1\}^{n_* \times n_*}$, $* \in \{s,t\}$. Specifically, for the i-th target, we treat the targets with the first K highest weights as its *related* targets:

$$\text{Index}^*(i) = \text{Top K}(\mathcal{W}^*[i,:], k_*) \tag{6}$$

$$\mathcal{R}^*_{i,j} = \begin{cases} 1 & \text{if } j \in \text{Index}^*(i) \text{ and } i \in \text{Index}^*(j) \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where $\text{Index}^*(i)$ is the set of selected indices of targets with the Top K operation and k_* is a hyperparameter denoting the value of K in the corresponding language.

To utilize the high-level aggregated target information, we concatenate the learned target representation $\boldsymbol{u}_i^*$ and the textual representations with target T_i^* to obtain the target-enhanced representations within the language for in-language relation alignment:

$$\boldsymbol{z}^*_{i,j} = \boldsymbol{h}^*_{i,j} \oplus \boldsymbol{u}^*_i \tag{8}$$

where $\boldsymbol{z}^*_{i,j} \in \mathbb{R}^{2d}$ is the target-enhanced representation of the j-th sample with target T_i^* and $\oplus$ denotes the concatenation operation.

3.4 Cross-Lingual Target Relation Graph

To explore the relationships between targets across languages, we further construct the cross-lingual target relation graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{A} \rangle$ with all targets from the two languages $\mathcal{T} = \{T_k\}_{k=1}^n$, where n is the total number of targets in two languages. The learned in-language target associations and representations are utilized as the start point, for the purpose of reducing the impact of the language gap and providing reliable prior target information.

Graph Construction We calculate the node features $\mathcal{V} = \{\boldsymbol{v}_k\}_{k=1}^n$ with target representations $\mathcal{U}^s$ and $\mathcal{U}^t$. Especially, for the targets shared across languages, we initialize them with the mean vectors of target representations in the two languages:

$$\boldsymbol{v}_k = \begin{cases} \boldsymbol{u}^s_{k_s} & \text{if } T_k \in \complement_{\mathcal{T}}\mathcal{T}_t \\ \frac{1}{2}\left(\boldsymbol{u}^s_{k_s} + \boldsymbol{u}^t_{k_t}\right) & \text{if } T_k \in \mathcal{T}_s \cap \mathcal{T}_t \\ \boldsymbol{u}^t_{k_t} & \text{if } T_k \in \complement_{\mathcal{T}}\mathcal{T}_s \end{cases} \tag{9}$$

where k_s and k_t are the corresponding indices of target T_k in $\mathcal{T}_s$ and $\mathcal{T}_t$ respectively, and $\complement_{\mathcal{T}}\mathcal{T}_s$ denotes the complementary set of $\mathcal{T}_s$ (i.e., the set of targets only in $\mathcal{T}_t$), and $\complement_{\mathcal{T}}\mathcal{T}_t$ denotes the complementary set of $\mathcal{T}_t$.

Then, we calculate the adjacency matrix $\mathcal{A} \in \{0,1\}^{n \times n}$ with target relation matrices $\mathcal{R}^s$ and $\mathcal{R}^t$. To compensate for the target inconsistency between languages, we also establish the connections between cross-language targets, forcing the model to pay attention to the cross-language target relations.

Specifically, for those targets only in the tgt language, we connect them with each remaining target by setting $\mathcal{A}_{i,j} = \mathcal{A}_{j,i} = 1$:

$$\mathcal{A}_{i,j} = \begin{cases} \mathcal{R}^s_{i_s,j_s} & \text{if } T_i, T_j \in \mathcal{T}_s \\ 1 & \text{if } T_i \in \complement_{\mathcal{T}}\mathcal{T}_s \text{ and } T_j \in \mathcal{T}_s \\ 1 & \text{if } T_i \in \mathcal{T}_s \text{ and } T_j \in \complement_{\mathcal{T}}\mathcal{T}_s \\ \mathcal{R}^t_{i_t,j_t} & \text{if } T_i, T_j \in \complement_{\mathcal{T}}\mathcal{T}_s \end{cases} \tag{10}$$

where i_s and j_s are the corresponding indices of targets T_i and T_j in $\mathcal{T}_s$, i_t and j_t are the corresponding indices of targets T_i and T_j in $\mathcal{T}_t$.

Cross-Lingual Target Relation Calculation We also adopt GAT to learn the cross-lingual target representations $\mathcal{U} = \{\boldsymbol{u}_k\}_{k=1}^n$ ($\boldsymbol{u}_k \in \mathbb{R}^d$) and attention weight matrix $\mathcal{W} \in \mathbb{R}^{n\times n}$. The learned weight matrix $\mathcal{W}$ is transformed into the cross-lingual relation matrix $\mathcal{R} \in \{0,1\}^{n\times n}$ via the Top K operation to acquire the target relations across languages:

$$\text{Index}(i) = \text{Top K}(\mathcal{W}[i,:], k) \tag{11}$$

$$\mathcal{R}_{i,j} = \begin{cases} 1 & \text{if } j \in \text{Index}(i) \text{ and } i \in \text{Index}(j) \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

where $\text{Index}(i)$ is the set of selected indices with the Top K operation and k is a hyperparameter denoting the value of K in Top K operation.

Similarly, we concatenate the learned cross-lingual target representation $\boldsymbol{u}_k$ and the textual representations with target T_k to get the target-enhanced representations for cross-lingual relation alignment and classification:

$$\boldsymbol{z}_{k,j} = \boldsymbol{h}_{k,j} \oplus \boldsymbol{u}_k \tag{13}$$

where $\boldsymbol{z}_{k,j} \in \mathbb{R}^{2d}$ is the target-enhanced representation of the j-th sample with target T_k.

3.5 Relation Alignment Strategies

We devise target relation alignment strategies to align representations between highly correlated targets so that semantic associations like shared background knowledge can be transferred across languages. Inspired by Mohtarami et al. (2019) and Lin et al. (2022), we further take the target relations into consideration based on contrastive learning and devise the relation alignment strategies within the language and across languages.

3.5.1 In-Language Relation Alignment

We first devise the in-language alignment strategies for the src and tgt languages to realize the in-language target relation alignment and optimize relationships of targets within the language. For each anchor $\hat{\boldsymbol{z}}^*_i$ in the mini-batch, we select positive samples within the language which are target-related to $\hat{\boldsymbol{z}}^*_i$ and have the same stance label with it, and treat other samples within the language as negative samples. We use the following loss function to pull the positive pairs closer and push the negative pairs away:

$$\mathcal{L}_* = -\frac{1}{b_*}\sum_{i=1}^{b_*}\frac{1}{b'_*}\sum_{j=1}^{b'_*}\Psi^*(i,j)\,\mathbb{I}(y^*_i = y^*_j)\cdot \log\frac{\mathbb{I}(i\neq j)\exp(f(\hat{\boldsymbol{z}}^*_i,\hat{\boldsymbol{z}}^*_j)/\tau)}{\sum_{k=1}^{b_*}\mathbb{I}(i\neq k)\exp(f(\hat{\boldsymbol{z}}^*_i,\hat{\boldsymbol{z}}^*_k)/\tau)} \tag{14}$$

$$\Psi^*(i,j) = \begin{cases} 1 & \text{if } \mathcal{R}^*_{t^*_i,t^*_j} = 1 \text{ or } t^*_i = t^*_j \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

where $\hat{\boldsymbol{z}}^*_i$ is the i-th target-enhanced representation within the language in the mini-batch, $\Psi^*(i,j)$ calculates the related targets through the target relation matrix $\mathcal{R}^*$, $\mathbb{I}(\cdot)$ is an indicator function, $f(\cdot)$ denotes the cosine similarity function, τ is the parameter of temperature, and b_* and b'_* are the numbers of samples and positive samples within the language.

3.5.2 Cross-Lingual Relation Alignment

To align representations across languages, we design a cross-lingual relation alignment strategy, which transfers the knowledge between semantically correlated targets across languages. The cross-lingual relation alignment enables us to make up for the lack of the tgt language data using src language data with the most relevant targets. For each anchor $\hat{\boldsymbol{z}}_i$ in the mini-batch, we select positive samples that are target-related, with the same stance label and from different languages, and take others as negative samples:

$$\mathcal{L}_{cross} = -\frac{1}{b}\sum_{i=1}^{b}\frac{1}{b'}\sum_{j=1}^{b'}\Psi(i,j)\mathbb{I}(y_i = y_j, l_i \neq l_j)\cdot \log\frac{\mathbb{I}(i\neq j)\exp(f(\hat{\boldsymbol{z}}_i,\hat{\boldsymbol{z}}_j)/\tau)}{\sum_{k=1}^{b}\mathbb{I}(i\neq k)\exp(f(\hat{\boldsymbol{z}}_i,\hat{\boldsymbol{z}}_k)/\tau)} \tag{16}$$

$$\Psi(i,j) = \begin{cases} 1 & \text{if } \mathcal{R}_{t_i,t_j} = 1 \text{ or } t_i = t_j \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

Algorithm 1: Training Process of TaRA.

Input: $D = D_{src} \cup D_{tgt}, \mathcal{V}^s, \mathcal{A}^s, \mathcal{V}^t, \mathcal{A}^t$
Output: The Overall Loss $\mathcal{L}$
for *each mini-batch* $\mathcal{X} = \{(\boldsymbol{t}_i, \boldsymbol{c}_i), \boldsymbol{y}_i\}_{i=1}^{b}$ *in* D **do**
  $\boldsymbol{h}_i = \mathrm{mBERT}(\boldsymbol{t}_i, \boldsymbol{c}_i)$;
  `/* In-language target relation */`
  $\mathcal{U}^*, \mathcal{W}^* \leftarrow \mathrm{GAT}(\mathcal{V}^*, \mathcal{A}^*), * \in \{s, t\}$;
  $\mathcal{R}^* \leftarrow \mathrm{Top\ K}(\mathcal{W}^*)$;
  $\boldsymbol{z}_i^* = \boldsymbol{h}_i \oplus \boldsymbol{u}^*$;
  `/* Cross-lingual target relation */`
  Construct Graph $\mathcal{G}$ with $\mathcal{U}^*, \mathcal{R}^*$;
  $\mathcal{U}, \mathcal{W} \leftarrow \mathrm{GAT}(\mathcal{V}, \mathcal{A})$;
  $\mathcal{R} \leftarrow \mathrm{Top\ K}(\mathcal{W})$;
  $\boldsymbol{z}_i = \boldsymbol{h}_i \oplus \boldsymbol{u}$;
  `/* Target relation alignment */`
  Align target-related samples within the language and calculate $\mathcal{L}_* = F_*(\{\boldsymbol{z}_i^*\}, \mathcal{R}^*)$;
  Align target-related samples across languages and calculate $\mathcal{L}_{cross} = F_c(\{\boldsymbol{z}_i\}, \mathcal{R})$;
  $\mathcal{L}_{in} = \gamma\mathcal{L}_s + (1-\gamma)\mathcal{L}_t$;
  $\mathcal{L}_{con} = \beta\mathcal{L}_{cross} + (1-\beta)\mathcal{L}_{in}$;
  `/* Classification */`
  $\hat{\boldsymbol{y}}_i = \mathrm{FFN}(\boldsymbol{z}_i)$;
  Calculate $\mathcal{L}_{ce}$ with $\hat{\boldsymbol{y}}_i$ and $\boldsymbol{y}_i$;
  $\mathcal{L} = \alpha\mathcal{L}_{ce} + (1-\alpha)\mathcal{L}_{con}$;
end
return $\mathcal{L}$

where $\hat{\boldsymbol{z}}_i$ is the i-th target-enhanced representation in the mini-batch, $\Psi(i, j)$ calculates the related targets through $\mathcal{R}$, l_i is the language of $\hat{\boldsymbol{z}}_i$, and b and b' are the numbers of samples and positive samples in the mini-batch.

3.6 Stance Classifier

The cross-lingual target-enhanced representations are fed into a two-layer feed-forward network with a softmax function for classification. We adopt a cross-entropy loss $\mathcal{L}_{ce}$ to optimize the classifier:

$$\hat{\boldsymbol{y}}_i = \mathrm{Softmax}(\mathrm{FFN}(\hat{\boldsymbol{z}}_i)) \tag{18}$$

$$\mathcal{L}_{ce} = -\frac{1}{b}\sum_{i=1}^{b} \boldsymbol{y}_i^\top \log \hat{\boldsymbol{y}}_i \tag{19}$$

where $\hat{\boldsymbol{y}}_i$ is the predicted label, $\boldsymbol{y}_i$ is the ground truth, and FFN is a two-layer feed-forward network with the activation function $\mathrm{ReLU}(\cdot)$.

3.7 Model Training

Algorithm 1 presents the training process of our method. We optimize the whole target-oriented relation alignment method by minimizing the overall loss function $\mathcal{L}$, consisting of the cross-entropy loss $\mathcal{L}_{ce}$ and the combined contrastive alignment loss $\mathcal{L}_{con}$. Formally, $\mathcal{L}_{con}$ is defined as follows:

$$\mathcal{L} = \alpha\mathcal{L}_{ce} + (1-\alpha)\mathcal{L}_{con} \tag{20}$$

$$\mathcal{L}_{con} = \beta\mathcal{L}_{cross} + (1-\beta)\underbrace{(\gamma\mathcal{L}_s + (1-\gamma)\mathcal{L}_t)}_{\text{in-language alignment}} \tag{21}$$

where α, β, γ are trade-off hyperparameters for balancing different losses.

4 Experiments

4.1 Datasets

X-Stance (Vamvas and Sennrich, 2020) is a multilingual stance dataset in German, French and Italian (with no training data), in which German is used as the src language and French as the tgt language. We focus on two political domains "*Foreign Policy*" and "*Immigration*" in X-Stance, with 31 targets in total. Based on them, we construct two datasets with different target settings for our experiments. **X-Stance-all** contains 5926 and 2582 texts in the src and tgt languages with the complete overlap of all the 31 targets. **X-Stance-partial** contains 3406 and 1806 texts in the src and tgt languages with partial overlap of targets. More details on datasets and targets are provided in Appendix A.

Multilingual Political Dataset (Lai et al., 2020) is comprised of 4 datasets, including two election datasets and two other datasets that contain only one target. We use the two election datasets for our experiments, and English and French are as the src and tgt languages respectively. **Election-none** contains 1691 and 1116 texts in the src and tgt languages. The targets in src include *Hillary Clinton* and *Donald Trump*, and *Emmanuel Macron* and *Marine Le Pen* are the targets for tgt.

4.2 Experimental Settings

In our experiments, we use mBERT to extract 768-dimensional textual representations. The threshold θ_0 for the adjacent matrix is set to 0.4. For Top K in target relation calculation, k, k_s, k_t are set to 10, 10, 4 for X-Stance-all; 10, 8, 5 for X-Stance-partial; 2, 1, 1 for Election-none. τ is set to 0.3 in the target relation alignment loss. For the trade-off hyperparameters, α, β and γ are set to 0.7, 0.6 and 0.7, respectively. All parameters are optimized by Adam (Kingma and Ba, 2015) with a learning rate of 2e-5 and a batch size of 64 for X-Stance-all and 32 for X-Stance-partial and Election-none. We train the model for 15 epochs with early stopping.

Methods	X-Stance-all de→fr		X-Stance-partial de→fr		Election-none en→fr	
	Acc (%)	$\boldsymbol{F_1}$ (%)	***Acc*** (%)	$\boldsymbol{F_1}$ (%)	***Acc*** (%)	$\boldsymbol{F_1}$ (%)
BiCond	69.6 ± 2.5	68.9 ± 2.4	68.9 ± 1.1	68.5 ± 1.1	70.9 ± 2.4	56.7 ± 3.7
TAN	67.1 ± 1.9	66.8 ± 1.6	65.2 ± 3.6	64.8 ± 3.4	70.2 ± 5.6	52.9 ± 2.8
TGMN	73.5 ± 0.9	73.0 ± 1.3	69.0 ± 2.4	68.6 ± 2.1	69.3 ± 1.8	54.7 ± 1.5
ADAN	61.4 ± 2.0	61.1 ± 1.9	57.0 ± 1.6	57.0 ± 1.6	55.9 ± 7.5	47.2 ± 3.3
CLA	77.1 ± 1.8	76.9 ± 1.7	76.2 ± 1.3	76.1 ± 1.3	74.6 ± 1.1	57.8 ± 2.4
ACLR	77.2 ± 2.7	77.0 ± 2.8	76.2 ± 1.5	76.0 ± 1.5	75.5 ± 2.0	58.1 ± 3.9
mBERT-FT	77.6 ± 1.6	77.5 ± 1.6	75.6 ± 2.0	75.5 ± 1.9	74.0 ± 3.5	57.8 ± 4.9
TaRA	$\mathbf{79.3 \pm 1.4}^\dagger$	$\mathbf{79.0 \pm 1.4}^\dagger$	$\mathbf{78.1 \pm 1.2}^\dagger$	$\mathbf{78.0 \pm 1.1}^\dagger$	$\mathbf{75.8 \pm 2.9}$	$\mathbf{62.3 \pm 3.0}^\dagger$

Table 2: Experimental results of the comparative methods and TaRA on the three datasets. For each method, we report the average scores and standard deviations of 5 runs. The best performance of each evaluation metric is marked in bold, and † means that our proposed TaRA is statistically significantly better than the baselines ($p < 0.05$).

The whole method is implemented with PyTorch on NVIDIA GeForce RTX 3090. The mBERT is Multilingual Cased BERT-Base model, which is 12-layer, 768-hidden, and 12-head, with about 110M parameters, and is implemented in the Transformers framework. For the three datasets, the running time is around 1 GPU hour.

4.3 Comparison Methods

We select the following methods for cross-lingual tasks as the comparative methods. (1) **ADAN** (Chen et al., 2018) is an adversarial based method for cross-lingual sentiment classification; (2) **CLA** (Mohtarami et al., 2019) aligns the representations in the two languages with contrastive language adaptation; (3) **ACLR** (Lin et al., 2022) improves the alignment method of CLA by devising two different alignments for the src and tgt languages respectively for cross-lingual rumor detection; (4) **mBERT-FT** (Devlin et al., 2019) fine-tunes the language model mBERT with the training data.

In addition, we also choose the following monolingual methods for stance detection and adapt them to the cross-lingual stance detection task by replacing the original word embeddings with the hidden vectors of mBERT. (1) **BiCond** (Augenstein et al., 2016) incorporates target representations into text representations with bidirectional conditional LSTMs; (2) **TAN** (Du et al., 2017) learns the target-specific representations with attention mechanism; (3) **TGMN** (Wei et al., 2018) utilizes a multi-hop memory network to obtain the implicit clues for stance detection.

4.4 Main Results

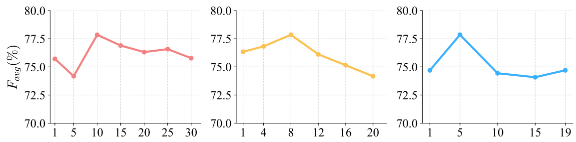
We use accuracy and the average F1 score of "*Favor*" and "*Against*" as the evaluation metric. Table 2 gives the experimental results of the comparative methods and our proposed TaRA on the three datasets. It can be seen from the table that our method outperforms all the baseline methods on the three datasets. In general, cross-lingual methods perform better than monolingual methods, indicating the importance of knowledge transfer across languages for cross-lingual tasks. As for the cross-lingual methods, we can see that the performance of fine-tuning mBERT is relatively good on the three datasets, which benefits from the superiority of the pre-trained language model. ACLR and CLA perform better among the cross-lingual methods, demonstrating the advantage of cross-lingual alignment with contrastive learning. More importantly, the results in Table 2 reveal that with the decrease of the number of topics shared between languages, the improvements of our method compared to the suboptimal methods become greater. Specifically, TaRA achieves **1.5%**, **1.9%** and **4.2%** performance gains on F1 scores compared to the suboptimal results on X-Stance-all, X-Stance-partial and Election-none, respectively. The experimental results verify the effectiveness of our relation alignment method for dealing with target inconsistency.

4.5 Ablation Study

Table 3 gives the ablation results of all the variants of our proposed TaRA on the three datasets. It can be seen from the table that removing the in-language target relation graph $\mathcal{G}_s$ and $\mathcal{G}_t$ decreases the performance, showing the necessity of incorporating in-language target relations to provide the preliminary information for cross-lingual relation graph. It can also be seen from the table that removing the cross-lingual relation graph $\mathcal{G}$ causes larger drops in performance, indicating that cross-

Variants	X-Stance-all				X-Stance-partial				Election-none			
	Acc	F_1	Δ*Acc*	ΔF_1	*Acc*	F_1	Δ*Acc*	ΔF_1	*Acc*	F_1	Δ*Acc*	ΔF_1
TaRA (Ours)	**79.3**	**79.0**	-	-	**78.1**	**78.0**	-	-	**75.8**	**62.3**	-	-
- $\mathcal{G}^s$, $\mathcal{G}^t$	78.1	78.0	-1.2	-1.1	76.9	76.8	-1.2	-1.2	75.6	61.0	-0.2	-1.3
- $\mathcal{G}$	77.4	77.4	-1.9	-1.7	76.5	76.4	-1.6	-1.6	74.2	59.8	-1.6	-2.5
- $\mathcal{G}^s$, $\mathcal{G}^t$, $\mathcal{G}$	77.1	77.0	-2.2	-2.0	76.5	76.4	-1.6	-1.6	74.8	59.7	-1.0	-2.6
- $\mathcal{L}_s$, $\mathcal{L}_t$	78.3	78.1	-1.0	-0.9	76.3	76.1	-1.8	-1.9	75.6	59.1	-0.2	-3.2
- $\mathcal{L}_{cross}$	77.5	77.3	-1.8	-1.7	75.8	75.6	-2.3	-2.4	74.0	58.8	-1.8	-3.5
- $\mathcal{L}_s$, $\mathcal{L}_t$, $\mathcal{L}_{cross}$	76.6	76.4	-2.8	-2.6	75.3	75.1	-2.8	-2.9	75.6	57.4	-0.2	-4.9
- *Target Relation*	77.7	77.6	-1.6	-1.4	75.3	75.1	-2.8	-2.9	73.6	58.4	-2.2	-3.9
- *Language*	77.5	77.4	-1.8	-1.6	75.8	75.7	-2.3	-2.3	75.0	59.7	-0.8	-2.6
- *Both*	76.6	76.4	-2.7	-2.6	74.7	74.6	-3.4	-3.4	74.2	57.2	-1.6	-5.1

Table 3: Ablation results of all the variants of our proposed TaRA on the three datasets.



(a) k (b) k_s (c) k_t

Figure 2: Impact of hyperparameters k in Top K Calculation on X-Stance-partial.

lingual relation graph is vital for addressing the target inconsistency issue.

Regarding learning objectives, it can be seen from the table that the F1 scores decline without target relation alignment within the language $\mathcal{L}_s$, $\mathcal{L}_t$ or across languages $\mathcal{L}_{cross}$, demonstrating the effectiveness of our proposed relation alignment strategies and the validity of model optimization.

Furthermore, the table also shows the results of the variants of relation alignment strategies. Excluding the "target relation" has a greater influence in target inconsistency cases (X-Stance-partial, Election-none) than excluding the "language", whereas removing the "language" has a greater impact in target consistency case (X-Stance-all) than removing the "target relation". This indicates that the alignment between semantically correlated targets across languages is more effective than the sole alignment of language for target inconsistency.

4.6 Analysis of Hyperparameters

Impact of Top K in Target Relation Graph The value of k in the Top K operation determines the number of related targets. We conduct experiments on X-Stance-partial. As shown in Figure 2, the performance is low in the beginning. When k (or k_s, k_t) is small, the related targets are rather few, resulting in that most samples in the mini-batch have no target-related positive samples. As the value of k (or k_s, k_t) increases, the performance gradually improves until it reaches a peak value. After that, increasing the value of k (or k_s, k_t), leads to too many positive samples with low correlations, resulting in a decrease in contrastive ability.

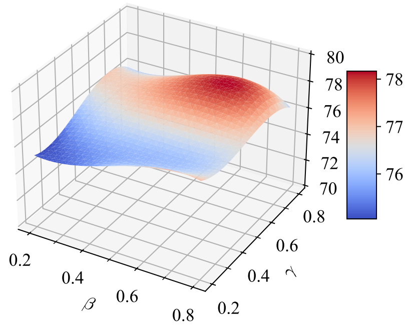


Figure 3: Performances of our method with different trade-off hyperparameters β and γ when $\alpha = 0.7$.

Analysis of Trade-off Hyperparameters We use three hyperparameters α, β and γ to balance different losses in our method. We set $\alpha = 0.7$ empirically. As shown in Figure 3, we conduct experiments to analyze the influence of different values of β and γ on X-Stance-partial. It can be seen that our method performs best when β and γ are around $0.6 \sim 0.8$. When β is greater than 0.5, the model gains higher performance because it pays more attention to the cross-lingual relation alignment, so that the shared knowledge between correlated targets can be transferred across languages. However, when β is too large, the drop in performance indicates that in-language target relations are the basis for cross-lingual relation learning.

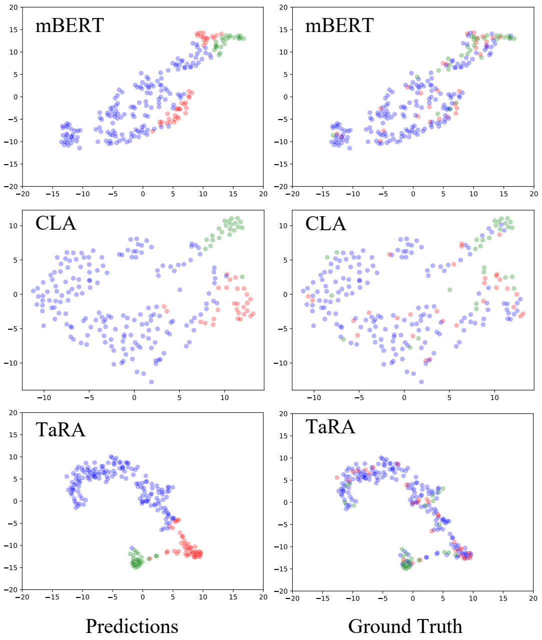


Figure 4: Visualization of representations. Red, blue and green points denote the samples of *Favor*, *Against* and *None* classes, respectively.

4.7 Visualization of Representations

We compare the representations learned by our method TaRA and the baseline methods mBERT-FT and CLA on the test set of Election-none. We use t-SNE to visualize them on a two-dimensional plane, as shown in Figure 4. It can be seen from the left column that the representations of mBERT and CLA overlap in *Favor* and *Against* samples. Our method clearly separates the three classes of samples, with higher in-class compactness. Comparing the predicted results visualized in the left column and the ground truth in the right column, it can be seen that our method gets fairly consistent results. This shows that our method can better handle target inconsistency with fine-grained alignment strategies at the both target and language levels.

5 Conclusion

We first identify the issue of target inconsistency in cross-lingual stance detection and propose a target-oriented relation alignment method TaRA, which considers relation associations and language alignments at both target and language levels. Our method explores the in-language and cross-language associations of targets via target relation graphs and aligns samples between highly correlated targets within the language and across languages through the fine-grained relation alignment strategies. Experimental results demonstrate the effectiveness of our method for cross-lingual stance detection.

6 Limitations

For the Top K operation in target relation calculation, we set K's three hyperparameters (i.e., k_s, k_t and k) to determine the number of the related targets. To explore the influence of the selection of K on model performance, a grid search on these three hyperparameters needs to be conducted to iterate each combination. However, due to the time and resource limits, we explore the impact of one hyperparameter in K by controlling the other two hyperparameters. Based on the empirical findings from this, we then set the value of K so as to achieve an appropriate performance.

7 Acknowledgments

This work is supported in part by the Ministry of Science and Technology of China under Grants #2020AAA0108401 and #2021ZD0111200, and National Natural Science Foundation of China under Grants #72293575, #62206287 and #62206282.

References

Abeer AlDayel and Walid Magdy. 2021. Stance detection on social media: State of the art and trends. *Information Processing & Management*, 58(4):102597.

Emily Allaway and Kathleen Mckeown. 2020. Zero-shot stance detection: A dataset and model using generalized topic representations. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 8913–8931.

Emily Allaway, Malavika Srikanth, and Kathleen Mckeown. 2021. Adversarial learning for zero-shot stance detection on social media. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4756–4767.

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

Ramy Baly, Mitra Mohtarami, James Glass, Lluís Màrquez, Alessandro Moschitti, and Preslav Nakov. 2018. Integrating stance detection and fact checking in a unified corpus. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 21–27.

Xilun Chen, Yu Sun, Ben Athiwaratkun, Claire Cardie, and Kilian Weinberger. 2018. Adversarial deep averaging networks for cross-lingual sentiment classification. *Transactions of the Association for Computational Linguistics*, 6:557–570.

Alessandra Teresa Cignarella, Mirko Lai, Cristina Bosco, Viviana Patti, Rosso Paolo, et al. 2020. Sardistance@ evalita2020: Overview of the task on stance detection in italian tweets. *EVALITA Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian*, pages 1–10.

Costanza Conforti, Jakob Berndt, Mohammad Taher Pilehvar, Chryssi Giannitsarou, Flavio Toxvaerd, and Nigel Collier. 2020. Will-they-won't-they: A very large dataset for stance detection on twitter. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 1715–1724.

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 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4171–4186.

Jiachen Du, Ruifeng Xu, Yulan He, and Lin Gui. 2017. Stance classification with target-specific neural attention networks. In *Proceedings of the International Joint Conferences on Artificial Intelligence*, pages 3988–3994.

Kyle Glandt, Sarthak Khanal, Yingjie Li, Doina Caragea, and Cornelia Caragea. 2021. Stance detection in covid-19 tweets. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing*, pages 1596–1611.

Momchil Hardalov, Arnav Arora, Preslav Nakov, and Isabelle Augenstein. 2021. Cross-domain label-adaptive stance detection. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 9011–9028.

Momchil Hardalov, Arnav Arora, Preslav Nakov, and Isabelle Augenstein. 2022. Few-shot cross-lingual stance detection with sentiment-based pre-training. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 10729–10737.

Jude Khouja. 2020. Stance prediction and claim verification: An arabic perspective. In *Proceedings of the Third Workshop on Fact Extraction and VERification*, pages 8–17.

Diederik P Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *Proceedings of the International Conference on Learning Representations*.

Dilek Küçük and Fazli Can. 2020. Stance detection: A survey. *ACM Computing Surveys*, 53(1):1–37.

Mirko Lai, Alessandra Teresa Cignarella, Delia Irazú Hernández Farías, Cristina Bosco, Viviana Patti, and Paolo Rosso. 2020. Multilingual stance detection in social media political debates. *Computer Speech & Language*, 63:1–27.

Yingjie Li and Cornelia Caragea. 2019. Multi-task stance detection with sentiment and stance lexicons. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing and the International Joint Conference on Natural Language Processing*, pages 6299–6305.

Bin Liang, Qinglin Zhu, Xiang Li, Min Yang, Lin Gui, Yulan He, and Ruifeng Xu. 2022. Jointcl: A joint contrastive learning framework for zero-shot stance detection. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 81–91.

Hongzhan Lin, Jing Ma, Liangliang Chen, Zhiwei Yang, Mingfei Cheng, and Chen Guang. 2022. Detect rumors in microblog posts for low-resource domains via adversarial contrastive learning. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 2543–2556.

Nikita Lozhnikov, Leon Derczynski, and Manuel Mazzara. 2018. Stance prediction for russian: data and analysis. In *International Conference in Software Engineering for Defence Applications*, pages 176–186.

Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. Semeval-2016 task 6: Detecting stance in tweets. In *Proceedings of the International Workshop on Semantic Evaluation (SemEval-2016)*, pages 31–41.

Mitra Mohtarami, James Glass, and Preslav Nakov. 2019. Contrastive language adaptation for cross-lingual stance detection. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing and the International Joint Conference on Natural Language Processing*, pages 4442–4452.

Parinaz Sobhani, Diana Inkpen, and Xiaodan Zhu. 2017. A dataset for multi-target stance detection. In *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics*, pages 551–557.

Qingying Sun, Zhongqing Wang, Qiaoming Zhu, and Guodong Zhou. 2018. Stance detection with hierarchical attention network. In *Proceedings of the International Conference on Computational Linguistics*, pages 2399–2409.

Mariona Taulé, M Antonia Martí, Francisco M Rangel, Paolo Rosso, Cristina Bosco, Viviana Patti, et al. 2017. Overview of the task on stance and gender detection in tweets on catalan independence at ibereval 2017. In *Workshop on Evaluation of Human Language Technologies for Iberian Languages*, volume 1881, pages 157–177.

Jannis Vamvas and Rico Sennrich. 2020. X-stance: A multilingual multi-target dataset for stance detection. In *Proceedings of the SwissText & KONVENS Joint Conference*, pages 1–9.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph attention networks. In *International Conference on Learning Representations*.

Penghui Wei and Wenji Mao. 2019. Modeling transferable topics for cross-target stance detection. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1173–1176.

Penghui Wei, Wenji Mao, and Daniel Zeng. 2018. A target-guided neural memory model for stance detection in twitter. In *Proceedings of the International Joint Conference on Neural Networks*, pages 1–8.

Chang Xu, Cecile Paris, Surya Nepal, and Ross Sparks. 2018. Cross-target stance classification with self-attention networks. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 778–783.

Ruifeng Xu, Yu Zhou, Dongyin Wu, Lin Gui, Jiachen Du, and Yun Xue. 2016. Overview of nlpcc shared task 4: Stance detection in chinese microblogs. In *Natural Language Understanding and Intelligent Applications*, pages 907–916.

Bowen Zhang, Min Yang, Xutao Li, Yunming Ye, Xiaofei Xu, and Kuai Dai. 2020. Enhancing cross-target stance detection with transferable semantic-emotion knowledge. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 3188–3197.

Yiwei Zhou, Alexandra I Cristea, and Lei Shi. 2017. Connecting targets to tweets: Semantic attention-based model for target-specific stance detection. In *International Conference on Web Information Systems Engineering*, pages 18–32.

A More Details on Datasets

We use two political domains in X-Stance (Vamvas and Sennrich, 2020), "*Foreign Policy*" and "*Immigration*", which have a similar target-text ratio to that of the original X-Stance dataset (with 10 domains in total). Table 4 shows the complete list of the 31 targets and their descriptions (translated into English). The X-Stance-all and X-Stance-partial datasets in our experiments are constructed as follows: In **X-Stance-all**, all the 31 targets in Table 4 are adopted in the tgt language, and the targets in the src language are exactly the same as those in the tgt language. In **X-Stance-partial**, we use the first 20 targets in Table 4 in the src language and the last 19 targets in Table 4 in the tgt language, with 8 targets overlapping.

	Domain	Target (English)
1	Immigration	Are you in favour of legalizing the status of sans papiers immigrants (i.e. immigrants who have no official paperwork) through a one-off, collective granting of residency permits?
2	Immigration	Would you support foreigners who have lived for at least ten years in Switzerland being given voting and electoral rights at municipal level throughout Switzerland?
3	Immigration	Should the state provide more funding for the integration of foreigners?
4	Immigration	Should access to "facilitated naturalization" via the Federation be made more difficult?
5	Immigration	The United Nations High Commissioner for Refugees (UNHCR) is seeking host countries for groups of refugees known as "quota refugees". Should Switzerland accept more of these groups?
6	Immigration	A popular initiative has been launched that wants to regulate immigration and thus limit migration-related population growth to 0.2% annually. Do you support this idea?
7	Foreign Policy	Would you support the introduction of the automatic exchange of bank client data between Switzerland and foreign tax authorities?
8	Foreign Policy	Should Switzerland embark on negotiations in the next four years to join the EU?
9	Foreign Policy	Should Switzerland conclude an agricultural free trade agreement with the EU?
10	Immigration	Do you support the existing agreement with the EU on the free movement of peoples?
11	Foreign Policy	Today, the Swiss Army can take part in UN or OSCE peace-keeping missions abroad, armed for self-defence purposes. Do you approve?
12	Foreign Policy	For a number of years, Switzerland has pursued a more active and open foreign policy that is less geared to strict neutrality. Do you welcome this change?
13	Foreign Policy	Should compliance with human rights play a greater role when deciding whether to enter into economic agreements with other countries (e.g. free trade agreements)?
14	Immigration	Would you support that foreigners who have lived for at least ten years in Switzerland being given voting and electoral rights at municipal level throughout Switzerland?
15	Immigration	Are you in favour of legalizing the status of sans papiers immigrants (i.e. immigrants who have no official paperwork) through a one-off, collective granting of residency permits?
16	Immigration	Do you think Switzerland should accept an increased number of refugees directly from crisis regions for which the United Nations High Commissioner for Refugees (UNHCR) needs host countries (what is called quota refugees)?
17	Foreign Policy	Should Switzerland embark on negotiations in the next four years to join the EU?
18	Foreign Policy	Should Switzerland start negotiations with the USA on a free trade agreement?
19	Foreign Policy	Should liability regulations for companies operating from Switzerland be tightened with regard to the compliance with human rights and environmental standards?
20	Foreign Policy	Do you think that Swiss foreign policy should increasingly be oriented to a strict interpretation of neutrality?
21	Foreign Policy	Should Switzerland terminate the Schengen Agreement with the EU and reintroduce increased identity checks directly on the border?
22	Immigration	Should the federal government provide more support for the integration of foreigners?
23	Immigration	Should foreigners who have lived in Switzerland for at least ten years be given the right to vote and be elected at the municipal level?
24	Immigration	Is limiting immigration more important to you than maintaining the bilateral treaties with the EU?
25	Immigration	Should sans-papiers be able to obtain a regularized residence status more easily?
26	Immigration	Are you in favor of further tightening the asylum law?
27	Immigration	Should the requirements for naturalization be increased?
28	Foreign Policy	Should Switzerland start membership negotiations with the EU?
29	Foreign Policy	Should Switzerland strive for a free trade agreement with the USA?
30	Foreign Policy	An initiative calls for liability rules for Swiss companies with regard to compliance with human rights and environmental standards abroad to be tightened. Do you support this proposal?
31	Foreign Policy	Are you in favour of Switzerland's candidacy for a seat on the UN Security Council?

Table 4: Targets in X-Stance-all.

ACL 2023 Responsible NLP Checklist

A For every submission:

☑ A1. Did you describe the limitations of your work?
Section 6 Limitations.

☒ A2. Did you discuss any potential risks of your work?
We conduct experiments on publicly available datasets.

☑ A3. Do the abstract and introduction summarize the paper's main claims?
Abstract; Section 1 Introduction.

☒ A4. Have you used AI writing assistants when working on this paper?
Left blank.

B ☑ Did you use or create scientific artifacts?

Section 3 Proposed Method; Section 4.1 Datasets.

☑ B1. Did you cite the creators of artifacts you used?
Section 3.2 Encoder Module; Section 4.1 Datasets.

☑ B2. Did you discuss the license or terms for use and / or distribution of any artifacts?
Section 3.2 Encoder Module; Section 4.1 Datasets; Section 4.2 Experimental Settings.

☑ B3. Did you discuss if your use of existing artifact(s) was consistent with their intended use, provided that it was specified? For the artifacts you create, do you specify intended use and whether that is compatible with the original access conditions (in particular, derivatives of data accessed for research purposes should not be used outside of research contexts)?
Section 3.2 Encoder Module; Section 4.1 Datasets; Section 4.2 Experimental Settings; Appendix A More Details on Datasets.

☐ B4. Did you discuss the steps taken to check whether the data that was collected / used contains any information that names or uniquely identifies individual people or offensive content, and the steps taken to protect / anonymize it?
Not applicable. Left blank.

☑ B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?
Section 3.2 Encoder Module; Section 4.1 Datasets; Section 4.2 Experimental Settings; Appendix A More Details on Datasets.

☑ B6. Did you report relevant statistics like the number of examples, details of train / test / dev splits, etc. for the data that you used / created? Even for commonly-used benchmark datasets, include the number of examples in train / validation / test splits, as these provide necessary context for a reader to understand experimental results. For example, small differences in accuracy on large test sets may be significant, while on small test sets they may not be.
Section 4.1 Datasets; Appendix A More Details on Datasets.

C ☑ Did you run computational experiments?

Section 4 Experiments.

☑ C1. Did you report the number of parameters in the models used, the total computational budget (e.g., GPU hours), and computing infrastructure used?
Section 4.2 Experimental Settings.

The Responsible NLP Checklist used at ACL 2023 is adopted from NAACL 2022, with the addition of a question on AI writing assistance.

☑ C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
Section 4.2 Experimental Settings; Section 4.6 Analysis of Hyperparameters; Section 6 Limitations.

☑ C3. Did you report descriptive statistics about your results (e.g., error bars around results, summary statistics from sets of experiments), and is it transparent whether you are reporting the max, mean, etc. or just a single run?
Section 4.4 Main Results.

☑ C4. If you used existing packages (e.g., for preprocessing, for normalization, or for evaluation), did you report the implementation, model, and parameter settings used (e.g., NLTK, Spacy, ROUGE, etc.)?
Section 4.2 Experimental Settings.

D ☒ Did you use human annotators (e.g., crowdworkers) or research with human participants?

Left blank.

☐ D1. Did you report the full text of instructions given to participants, including e.g., screenshots, disclaimers of any risks to participants or annotators, etc.?
No response.

☐ D2. Did you report information about how you recruited (e.g., crowdsourcing platform, students) and paid participants, and discuss if such payment is adequate given the participants' demographic (e.g., country of residence)?
No response.

☐ D3. Did you discuss whether and how consent was obtained from people whose data you're using/curating? For example, if you collected data via crowdsourcing, did your instructions to crowdworkers explain how the data would be used?
No response.

☐ D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
No response.

☐ D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?
No response.