CL-XABSA: Contrastive Learning for Cross-Lingual Aspect-Based Sentiment Analysis

Nankai Lin, Yingwen Fu, Xiaotian Lin, Dong Zhou, Aimin Yang, and Shengyi Jiang

Abstract—Aspect-based sentiment analysis (ABSA), an extensively researched area in the field of natural language processing (NLP), predicts the sentiment expressed in a text relative to the corresponding aspect. Unfortunately, most languages lack sufficient annotation resources; thus, an increasing number of recent researchers have focused on cross-lingual aspect-based sentiment analysis (XABSA). However, most recent studies focus only on cross-lingual data alignment instead of model alignment. Therefore, we propose a novel framework, CL-XABSA: contrastive learning for cross-lingual aspect-based sentiment analysis. Based on contrastive learning, we close the distance between samples with the same label in different semantic spaces, achieving convergence of semantic spaces of different languages. Specifically, we design two contrastive objectives, token-level contrastive learning of token embeddings (TL-CTE) and sentiment-level contrastive learning of token embeddings (SL-CTE), to unify the semantic space of source and target languages. Since CL-XABSA can receive datasets in multiple languages during training, it can be further extended to multilingual aspect-based sentiment analysis (MABSA). To further improve the model performance, we perform knowledge distillation with target-language unlabeled data. In the distillation XABSA task, we further explore the effectiveness of different data (source dataset, translated dataset, and code-switched dataset). The results demonstrate that the proposed method has a certain improvement in the three XABSA tasks, distillation XABSA and MABSA.

Index Terms—Contrastive Learning, cross-lingual aspect-based sentiment analysis, knowledge distillation, multilingual aspect-based sentiment analysis.

I. Introduction

A

SPECT-BASED sentiment analysis (ABSA) is the task of predicting the sentiment expressed in a text relative to the corresponding aspect, which has recently gained much attention [8], [25], [30], [34]. Research on the ABSA mainly focuses on high-resource languages, and there is limited research on low-resource languages. However, in real-world scenarios such as e-commerce websites, users’ opinions are usually expressed in multiple different languages, including some low-resource languages [18], [28].

To address this issue, there has recently been increasing interest in cross-lingual aspect-based sentiment analysis (XABSA). In XABSA, only the source-language labeled dataset is used for model training, and then the model is used to predict the target-language unlabeled dataset. An example is shown in Fig. 1, where we regard English as the source language and French as the target language. After training on the English labeled dataset, the model can directly identify the French sentence “Bonnes glaces.” mentioned the aspect term “glaces” with the corresponding sentiment of “positive”.

Since there are few publicly available resources for low-resource languages, scholars have gradually considered zero-shot transfer for cross-lingual aspect-based sentiment analysis tasks [16], [32]. However, existing methods mainly implement the zero-shot method from the cross-lingual data alignment perspective. The data alignment method enables the model to indirectly learn the target-language knowledge by means of data generation when there is no target-language labeled dataset. Li et al. [24] generated pseudo data in the target language by translating the source-language labeled data. Zhang et al. [39] generated code-switched data for XABSA tasks in a more advanced way. Their methods alleviate the XABSA data-hungry problem to a certain extent, but the implemented models are both BERT-based token classification models, which do not consider the zero-shot problem from the perspective of model alignment. The data alignment method simply merges the generated data with the source-language data for training, while the model alignment encourages the model to learn the “alignment pattern” between languages.

In addition, the contrastive learning technique has significantly advanced progress in self-supervised representation learning [3], [13], [27], [38]. Its notation is to shorten the distance between anchor points and “positive” samples in the embedding space and widen the distance between anchor points and many “negative” samples. Among them, positive examples are usually
data-augmented samples generated from anchor points, and negative examples are the other samples in the same mini-batch. Khosla et al. [19] first extended contrastive learning to fully supervised settings, which is called supervised contrastive learning. It uses label information to distinguish positive and negative anchor point samples. Contrastive learning has not only been shown to have superior performance in monolingual tasks [10] but is also applicable to cross-lingual tasks [5], [11]. In cross-lingual settings, contrastive learning has shown its potential to learn the sample representations with different labels by pulling the sample representations with the same label from different languages in semantic space. Thus, the bilingual vector space can be implicitly brought closer. Therefore, we propose a novel framework CL-XABSA for XABSA based on contrastive learning. Specifically, there are two modules in CL-XABSA: the token classification module and the contrastive learning module. CL-XABSA uses different datasets for training, including a source-language labeled dataset, a target-language translated dataset from the source language, and code-switched datasets. Code-switched datasets are constructed by switching the aspect terms between the source and translated sentences to construct two bilingual sentences. The token classification module is a BERT-based sequence labeling model. The contrastive learning module contains two alternative objectives: token-level contrastive learning of token embeddings (TL-CTE) and sentiment-level contrastive learning of token embeddings (SL-CTE). The contrastive learning module fuses and compares different datasets to learn richer language-independent knowledge and cross-lingual ability. Furthermore, we extend CL-XABSA to multilingual ABSA. To further improve the model performance, we perform knowledge distillation technology with the target-language unlabeled data. In summary, the contributions of this article are as follows:

1) To the best of our knowledge, we are the first to apply contrastive learning to the XABSA task. Our method can make the semantic space of different languages more consistent.
2) We propose TL-CTE and SL-CTE to shorten the distance between tokens with the same label or the same sentiment in the source, translated, and code-switched datasets.
3) By comparing different datasets, we further explore contrastive learning effectiveness in multi-teacher model distillation.
4) Furthermore, we also apply contrastive learning to MABSA and distillation MABSA.
5) The results show that the proposed method has a certain improvement in the three XABSA tasks, multi-teacher distillation ABSA and multilingual distillation ABSA.

II. RELATED WORK

A. XABSA

Existing works on cross-lingual ABSA mainly focus on its subtasks, including cross-lingual aspect term extraction and aspect sentiment classification. The related works for XABSA are mainly divided into two groups: data alignment and embedding learning.

Early methods for XABSA tasks mainly employ word alignment algorithms [9], [32]. In addition, Zhou et al. [40] proposed adopting a translate-then-align strategy to obtain pseudo-labeled data of the target language through machine translation [22]. Zhang et al. [39] cleverly generated code-switched data for XABSA tasks. However, these methods heavily depend on the bilingual dictionary quality and the translation system performance.

In addition to data alignment, there are also some works investigating semantic representation construction to improve the XABSA performance. Ruder et al. [31] used cross-lingual word embeddings trained on a large parallel bilingual corpus for the XABSA task. By switching the word embeddings between different languages, the model can be used in a language-agnostic manner [1], [2]. Jebbara and Cimiano incorporated multilingual word embeddings, which are based on a common shared vector space across various languages, into a convolutional neural network architecture for opinion target extraction [16].

B. Contrastive Learning

Recently, contrastive learning has shown significant improvements for various NLP tasks, such as ABSA [17] and text classification [33]. Gao et al. [10] proposed a simple contrastive learning framework for semantic textual similarity (STS) tasks, which is adaptable to both unsupervised and supervised tasks. In addition, contrastive learning has received attention and applications in cross-lingual NLP tasks. Mohtarami et al. [26] introduced a novel memory network-based contrastive language adaptation approach to align the stances in the source and target languages. Guo et al. proposed a contrastive learning method to obtain effective representations for cross-lingual text classification based on the BERT model [11]. Choudhary et al. proposed learning the resource-poor and resource-rich language in a common emoji space based on contrastive learning [4].

III. METHODOLOGY

In this section, we introduce the overall architecture of the proposed CL-XABSA. We regard the ABSA task as a sequence labeling task [12], [23]. Given an input sentence $S$ that is composed of a token sequence $\{w_1, w_2, w_3, \ldots, w_n\}$ and a label sequence $\{l_1, l_2, l_3, \ldots, l_n\}$, where $n$ is the length of the sentence, the proposed model infers the label $l_i$ for each token $w_i$ and outputs a label sequence, where $l_i \in Y = \{B, I, E, S\} - \{POS, NEU, NEG\} \cup \{O\}$. $B$, $I$, $E$ represent the beginning, middle, and end of an aspect item, respectively, and $S$ indicates that the aspect item is only one word. $POS$, $NEU$ and $NEG$ indicate that the sentiment of the corresponding aspect item is positive, neutral, and negative, respectively. For example, $l_i = B - POS$ means $w_i$ is the beginning of a positive aspect term.

As shown in Fig. 2, CL-XABSA is composed of two modules: 1) the token classification module and 2) the contrastive learning module. In the token classification module, the model regards ABSA as a sequence labeling task and chooses the label with the highest probability for each token. In the contrastive learning
module, we design two contrastive objectives, TL-CTE and SL-CTE.

Contrastive learning has been widely used in related research on natural language processing. Contrastive learning focuses on improving the model’s ability to distinguish a given data point from “positive” examples (points that share the same label) and “negative” examples (different labels). The definition of “positive” examples determines which samples should be narrowed by the contrastive learning method. By defining the set of positive samples and the set of negative samples, a contrastive loss brings the latent representations of samples belonging to the same class closer together. Based on contrastive learning, we close the distance between samples with the same label in different semantic spaces, thus achieving a convergence of semantic spaces of different languages.

We propose TL-CTE and SL-CTE, which use contrastive loss for XABSA tasks by defining positive examples from different levels. TL-CTE treats tokens with the same label as positive examples. In contrast, SL-CTE performs contrastive learning from the sentiment level and treats tokens with the same sentiment as positive samples. The two objectives narrow the distance between tokens from different levels. TL-CTE narrows the distance between tokens of the same label, and SL-CTE compares tokens of the same sentiment. Therefore, TL-CTE is more fine-grained than SL-CTE. Through contrastive learning objectives at different levels, we can integrate the semantic spaces of different languages from different perspectives. Li et al. [21] proposed the use of probability contrastive learning to replace feature contrastive learning so that each module of the model can be updated instead of only the representation layer parameters. Therefore, in the contrastive learning module, we choose the probability contrastive learning method.

In addition to leveraging the labeled dataset in the source language $D_S$, we translate the source language $D_S$ to the target language and construct a translated dataset $D_T$. Moreover, we utilize the code-switched datasets proposed by Zhang et al. [39], which are called $D_{S_C}$ and $D_{T_S}$. $D_{S_C}$ is constructed by replacing the aspect terms appearing in the source-language dataset $D_S$ with those appearing in the target-language dataset $D_T$. In contrast, $D_{T_S}$ is generated by replacing the aspect terms appearing in the target-language dataset $D_T$ with those appearing in the source dataset $D_S$. Fig. 3 shows an example of the source language, translated and code-switched datasets.

### A. Token Classification

We build the token classification model based on the BERT model. Given a text sequence $S = \{s_1, s_2, s_3, \ldots, s_n\}$ with $n$ words, the BERT model encodes $S$ into context-aware feature representations $h = \{h_1, h_2, h_3, \ldots, h_n\}$. Specifically, $h_i$ is represented as follows:

$$h_i = m_{BERT} (s_i)$$

Then, a multi-layer perceptron (MLP) classifier with the softmax function is leveraged to calculate the label probability distribution of $w_i$:

$$g_i = \text{softmax} (W \cdot h_i + b)$$

where $W$ and $b$ are learnable parameters, and $g_i$ is the predicted label distribution of the $i$-th context-aware feature representation. Then, the training objective $L_{CE}$ is computed as the cross-entropy loss:

$$L_{CE} = \frac{1}{N} \sum_{i=0}^{N} \left[ -\frac{1}{n} \sum_{i=0}^{n} y_i \log (g_i) \right]$$

where $N$ refers to the number of sentences in the training set, $n$ is the number of tokens for each sentence, and $y_i$ represents the label of the $i$-th token.

### B. Contrastive Learning

#### TL-CTE

To narrow the distance between tokens of the same label, we propose TL-CTE to fuse and compare different datasets to learn richer language-independent knowledge and cross-lingual ability. In Fig. 4, we can see how TL-CTE applies contrastive learning on the XABSA task. Specifically, we denote a batch of sample and label pairs as $\{x_i, y_i\}_{i \in I}$, where $I = \{1, \ldots, K\}$ is the indices of the samples and $K$ is the batch size. Given a set of samples $P$ containing the same labels as the anchor, its positive set is defined by $P = \{p : p \in I, y_p = y_i \land p \neq i\}$, with size $|P|$, where $y_i \in \text{Y}_{token} = \{B, I, E, S\} - \{POS, NEU, NEG\} \cup \{O\}$ in TL-CTE. Our contrastive loss function for each entry $i$ across the batch is:

$$L_{TL_i} = -\frac{1}{|P|} \sum_{p \in P} \log \frac{\exp (\text{sim} (g_i, g_p) / \tau)}{\sum_{k \in I \setminus \{i\}} \exp (\text{sim} (g_i, g_k) / \tau)}$$

where $\text{sim}(\cdot)$ indicates the cosine similarity function. The contrastive loss for the TL-CTE module is:

$$L_{TL-CTE} = \sum_{i=1}^{K} L_{TL_i}$$

where $\tau$ is the temperature hyper-parameter. Larger values of $\tau$ scale down the dot-products, creating more difficult comparisons. After obtaining the contrastive loss through the TL-CTE module from 5, we further weighted the contrastive loss $L_{TL-CTE}$ and the cross-entropy loss $L_{CE}$ to obtain the final model loss $L$:

$$L = \alpha \cdot L_{TL-CTE} + (1 - \alpha) \cdot L_{CE}$$
SL-CTE. Different from TL-CTE, SL-CTE narrows the distance of samples at the sentiment level, which is shown in Fig. 5. The positive set is given by $P = \{ p : p \in I, y_p = y_i \land p \neq i \}$, with size $|P|$, where $y_i \in Y_{sen} = \{ POS, NEU, NEG \} \cup \{ O \}$ in SL-CTE. In addition, our contrastive loss function for each entry $i$ across the batch is

$$L_{SL-i} = -\frac{1}{|P|} \sum_{p \in P} \log \frac{\exp(\frac{\text{sim}(g_i, g_p)}{\tau})}{\sum_{k \in I \setminus \{i\}} \exp(\frac{\text{sim}(g_i, g_k)}{\tau})}$$

(7)

Similar to the TL-CTE module, the contrastive loss for the SL-CTE module is:

$$L_{SL-CTE} = \sum_{i=1}^{K} L_{SL-i}$$

(8)

The final loss function $L$ also consists of two parts, cross-entropy loss $L_{CE}$ and contrastive loss $L_{SL-CTE}$:

$$L = \alpha \cdot L_{SL-CTE} + (1 - \alpha) \cdot L_{CE}$$

(9)

Although the computational process is similar to that of TL-CTE, SL-CTE performs contrastive learning from a coarser-grained sentence level, so its fitting goal is easier to achieve.
C. Distillation XABSA

Zhang et al. [39] verified the remarkable performance of knowledge distillation on the XABSA task, especially multi-teacher distillation, which achieved the best performance on XABSA; therefore, we further applied our CL-XABSA framework to the multi-teacher distillation task.

Based on contrastive learning, we use the translated dataset and code-switched datasets to train different teacher models, as shown in Fig. 6. To fully exploit the knowledge in different datasets, we design a multi-teacher contrastive distillation model with different teacher models separately trained on three dataset combinations: \( D_T \cup D_S \), \( D_T \cup D_{S_T} \), and \( D_T \cup D_{S_T} \). Each teacher model contains the translated dataset \( D_T \) to involve some language-specific knowledge and one of the remaining datasets to share the same sentence semantics (\( D_S \)), the same context sentence (\( D_{S_T} \)), and the same aspect term (\( D_{S_T} \)). Assuming that the prediction probability of the k-th teacher model is \( g_{Tk} \), the label prediction probability obtained after fusing the three teacher models is

\[
p_k = \sum_{k=1}^{3} \omega_k g_{Tk}
\]

where \( \omega_k \) is the weight for each teacher model. With the combined soft label \( g_s \), a student model can be trained similarly by using only the token classification module. Since the student model receives more informative soft labels than hard labels [14], the student model replaces the original cross-entropy loss with the mean squared error loss \( L_{KD} \) during training:

\[
L_{KD} = \frac{1}{|D_U|} \sum_{x \in D_U} \left[ \frac{1}{n} \sum_{j=0}^{n} MSE(g_{Tj}, g_{sj}) \right]
\]

where \( n \) is the number of tokens contained in a sentence and \( j \) represents the \( j \)-th token in the sentence. In addition, \( D_U \) denotes the unlabeled dataset in the target language, and \( g_{sj} \) is the \( j \)-th token’s predicted probability of the student model. We use the mean squared error loss \( (MSE(\cdot)) \) to measure the difference between the two probability distributions.

D. Multilingual XABSA

We further adapt the CL-XABSA framework to make it suitable for the MABSA task. As shown in Fig. 7, during the training phase, we utilize the source dataset \( D_S \) for the English language and multiple datasets for the other four languages to train the teacher model. Specifically, considering the zero-shot setting, the multiple datasets contain target-language translated datasets \( D_T \) and code-switched datasets \( D_{S_T} \) and \( D_{T_S} \). Moreover, we perform knowledge distillation technology leveraging data from the unlabeled target language. The unlabeled data come from the target language training set provided by the original evaluation task, and then we discard its labels. In multilingual mode, whether it is a teacher model or a student model, the received data are datasets containing multiple languages. Similar to the multi-teacher distillation task, the student model uses only the token classification module for learning and utilizes mean squared error loss as the loss function.

IV. EXPERIMENTS

A. Dataset

We conduct experiments on the SemEval-2016 dataset [29], which includes real user reviews in English (EN), French (FR), Spanish (ES), Dutch (NL), Russian (RU), and Turkish (TR). However, due to the limitation in the Turkish data size, many studies excluded this language during the experiment; similarly, we also only evaluated the other five languages. We use the data processed by Zhang et al. [39]. For each language, they split the data into training, validation, and test sets, as well as constructing code-switched datasets, where \( D_{S_T} \) and the source-language dataset \( D_S \) have the same sentence context but aspect terms in
different languages, while $D_{T_S}$ and the source-language dataset $D_S$ have different sentence contexts but the same aspect term in the target language. The dataset details are shown in Table I.

### B. Baselines

We adopt the following approaches for comparisons: ZERO-SHOT, a method utilizing labeled source data to fine-tune the model and directly conduct inference on the target data, which has been shown to be a strong baseline for cross-lingual adaptation [6], [37]. To compare with the previous translation-based method, we adopt the baseline that utilizes the pseudo-labeled data with the translate-then-align paradigm [20], [24] (TRANSLATION-TA) and the combination of the source data with such translated data (BILINGUAL-TA). In addition, we compare our method with the state-of-the-art (SOTA) method ACS [39].

### C. Experimental Settings

We regard English as the source language and other languages as target languages. To simulate an unsupervised setting, following previous works [15], [16], [39], we use the English validation set to select the best model. For the original training set of each target language, we discard all labels and treat the set as unlabeled data in the distillation stage [35], [36], [39]. We use only the training set with labels in Section V-C to verify the generalization of our MABSA model.

We conduct experiments based on the cased multilingual BERT (mBERT) [7]. Similar to [39], we train the model up to 2000 steps and select the best model in the last 500 steps. For distilling the XABSA task, we initialize the student model with the TRANSLATION model, which is trained on the target translated dataset $D_T$. Then, we continue the training up to 1000 steps and select the best model on the last 500 steps. When training the baseline methods ACS, TRANSLATION, MTL-AF, MTL-ACS, and MTL-ACS-D, following [39], we set the learning rate and batch size as 5e-5 and 16, respectively. For TL-CTE and SL-CTE, we train the model with a larger batch size because a larger batch size is favorable to the training of contrastive learning. We select the best training hyper-parameters by conducting a grid search on batch size. The range of batch size is {16, 32, 64}. We use batch sizes of 32 and 64 for TL-CTE and SL-CTE, respectively. Note that training the student model required more graphics memory, and we could only set the batch size as 32 for SL-CTE. The results with different batch sizes are shown in Table II. We set hyperparameter $\tau$ to 0.07 consistently. For multi-teacher distillation, we treat each teacher equally, which means $\omega_i = 1/3$ in (10). The loss weighting factor $\alpha$ in (6) and (9) is set as 0.5.

Micro-F1 is used as the evaluation metric where a prediction is judged as correct only if both its boundary and sentiment polarity are correct. For all experiments, we report the average F1 scores over 5 runs with different random seeds.

### V. The Results and Analysis

#### A. Cross-Lingual ABSA Results

As shown in Table II, the proposed two contrastive objectives outperform the existing SOTA models, which indicates that they could achieve a convergence of semantic spaces of different languages and make the semantic space more consistent. We can also observe that the TL-CTE method performs better on French and Spanish, while on Dutch and Russian, the SL-CTE method performs better. On average, the SL-based method outperforms the TL-based method on the XABSA task, with an average Micro-F1 value reaching 53.10%. The performance of the contrastive learning model is significantly affected by the batch size [3]. Therefore, we explore the optimal batch size for our two objectives, and the results are shown in Table III. We can see that for the SL-CTE strategy, the larger the batch size, the better the performance the model can obtain, while for the TL-CTE strategy, the optimal batch size value is 32. More negative samples lead to better performance [19]. The negative sample number of each sample in CL-XABSA(TL) is higher than that in CL-XABSA(SL) with the same batch size. This demonstrates that CL-XABSA(SL) needs a larger batch size to achieve a large negative example number for optimal performance.

---

1[Online]. Available: https://huggingface.co/bert-base-multilingual-cased

---

| Dataset | Type | EN | FR | ES | NL | RU |
|---------|------|----|----|----|----|----|
| Train   | $D_S$ | 1600 | 1664 | 1656 | 1378 | 2924 |
|         | $D_T$ | - | 1573 | 1571 | 1586 | 1577 |
|         | $D_{T_S}$ | - | 1600 | 1600 | 1600 | 1600 |
|         | $D_{S_T}$ | - | 1572 | 1571 | 1586 | 1577 |

| Dev     | - | 400 | 332 | 414 | 344 | 731 |
| Test    | - | 676 | 668 | 881 | 575 | 1209 |

| Method | Batch Size | FR | ES | NL | RU | Avg. |
|--------|-------------|----|----|----|----|------|
| ZERO-SHOT | 16 | 49.33 | 60.12 | 49.61 | 51.51 | 52.64 |
| TL-CTE | 32 | 48.53 | 60.64 | 50.96 | 50.77 | 52.73 |
| SL-CTE | 64 | 49.02 | 59.85 | 50.62 | 50.24 | 52.43 |

| Method | Batch Size | FR | ES | NL | RU | Avg. |
|--------|-------------|----|----|----|----|------|
| TRANSLATION | 16 | 47.12 | 59.41 | 49.43 | 52.98 | 52.01 |
| ACS | 32 | 49.08 | 60.25 | 50.13 | 50.62 | 52.52 |
| SL-CTE | 64 | 49.50 | 61.62 | 50.64 | 50.65 | 53.10 |
Looking at the performance across languages, we can see that the performance of any model on Spanish is superior to that of the other three languages. We further analyze the proportion of labels in different languages to explore the reason. We counted the label distribution of the test set of each language, and the results are shown in Fig. 8. Generally, the “POS” class accounts for the highest part of all test sets, which indicates that the “POS” class is potentially the easiest to correctly identify among the three categories. The Spanish test set contains the highest proportion of the “POS” class. Therefore, from the performance comparison of the four languages, the Spanish model performs best.

To conduct more in-depth analyses on the specific performance of our model on each sentiment class, we take Spanish as an example and further display the F1 value of each class, and the results are shown in Table IV. All three models perform poorly on neutral sentiment. In the three sentiment classes, our two methods have brought obvious improvements to the ACS method, among which the improvement for the neutral class is the most obvious. This demonstrates that our model can alleviate the class-imbalance problem.

Furthermore, to explore the reason why the model performs poorly on the neutral sentiment class, we further conducted a case study, which is shown in Table V. When performing recognition on shorter sentences that contain only one aspect, the three models can basically recognize the sentiment correctly. When dealing with more complex sentences, the three models make different errors. ACS recognizes non-aspect tokens as aspects and recognizes tokens with positive sentiment as neutral sentiment. ACS also has a boundary recognition error. CL-XABSA(SL) mainly misses identifying aspect tokens or misidentifies non-aspect tokens. The main error case of CL-XABSA(TL) is to identify the aspect of neutral emotion as the aspect of positive emotion. Although our model performs better than ACS, there are still many problems that need to be further addressed and researched in the future.

B. Multi-Teacher Distillation Results

To further improve the model performance, we introduce distillation technology. It is apparent from the results in Table VI that multi-teacher distillation can significantly enhance XABSA performance. During multi-teacher distillation, although the result of the CL-XABSA(TL) model is slightly lower than that of the ACS model, our framework still shows a certain improvement in that the performance of the CL-XABSA(SL) model is 0.56 higher than that of the ACS model. We believe that due to the limitation in the data size, it is difficult for the initialized CL-XABSA student and teacher models to obtain better model performance. Therefore, we further analyze the performance of the initialized student model and different teacher models. The results are shown in Table VII. We can see in the table that when using $D_T$ to train the model, TL-CTE and SL-CTE aim only to shorten the distance between token and sentiment in the monolingual vector space, which is similar to the goal of the monolingual sentiment classification model, so the ACS model outperforms the contrastive learning models. For the model trained by $D_S \cup D_T$ and $D_{ST} \cup D_T$, the correlation between the source and target languages is relatively limited compared with the model trained by $D_{TS} \cup D_T$. Specifically, aspect terms are language-independent in $D_S$ and $D_T$. In contrast, $D_{ST}$ and $D_T$ are associated with only aspect term information, which means that their semantic representations are not in a uniform space. However, as mentioned above, since the TL-CTE model needs to regularize more token spaces to be more uniform, its performance can only be better improved on datasets that contain more information. For the models trained on $D_{ST} \cup D_T$ and $D_S \cup D_T$, the cross-language information provided by these two corpora is not enough to support the TL-CTE model to achieve good results, while the SL-CTE model that requires less learning information can achieve a better effect. Considering the dataset of $D_{TS} \cup D_T$, aspect terms for different languages appear only in the semantic representation of the target language. Although the semantic space of the source language is lacking, the label information provided by the aspect term can better improve the XABSA task, which indicates that TL-CTE can perform better when there is enough information for learning.

C. Multilingual Distillation Results

In addition to XABSA, we further report on our work on MABSA. As shown in Table VIII, after knowledge distillation, CL-XABSA based on TL-CTE achieves the best results, with an average Micro-F1 reaching 55.93, which is 1.94 higher than the state-of-the-art model MTL-ACS-D. Furthermore, although the CL-XABSA model based on SL-CTE does not exceed the CL-XABSA based on TL-CTE, it still has a significant improvement, with the average Micro-F1 value being 1.67 higher than the MMTL-ACS-D model. In contrast, for the model without knowledge distillation, the teacher model of our method does not outperform the existing models (MTL-ACS and MTL-AF), while the student model achieves...
Fig. 9. 2D visualization of different models’ semantic space. Subfigures (a. *) are the MTL-ACS. Subfigures (b. *) are the MTL-CL-XABSA(SL). Subfigures (c. *) are the MTL-CL-XABSA(TL).
better performance. To explore the reasons for this result, we further compared the performance of our model with the MTL-ACS model on the target language training set provided by the SemEval workshop (i.e., distillation data). The results are shown in Table IX. On distillation data, our model outperforms other models, and our (teacher) model performs better in terms of the average results of training and test sets, meaning that the training and test set distribution provided by the SemEval workshop are inconsistent and our model has strong generalization performance. Overall, our model outperforms existing models on datasets in different distribution spaces.

Taking a closer look at the performance of each language, we can see that the CL-XABSA results perform better on the French and Russian languages and relatively poorly on the Spanish and

---

**TABLE V**

**CASE STUDY**

| Text                                                | Answer                  |
|-----------------------------------------------------|-------------------------|
| En cuanto a la comida, bien pero nada exclusivo ni sorprendente. | (comida, NEU)           |
| ACS                                                 | (comida, NEU)           |
| CL-XABSA(TL)                                        | (comida, NEU)           |
| CL-XABSA(SL)                                        | (comida, NEU)           |

| Text                                                | Answer                  |
|-----------------------------------------------------|-------------------------|
| Un buen restaurante relación calidad precio muy buena, local con decoración cuidada, luz baja pero suficiente y un hilo musical muy bien seleccionado. | (restaurante, POS), (local, POS), (luz, NEU), (hilo musical, POS) |
| ACS                                                 | (restaurante, POS), (local, NEU), (decoración, NEU), (luz, NEU), (musical, POS) |
| CL-XABSA(TL)                                        | (restaurante, POS), (decoración, POS), (luz, POS), (hilo musical, POS) |
| CL-XABSA(SL)                                        | (restaurante, POS), (decoración, NEU), (luz, NEU) |

**TABLE VI**

**THE RESULTS OF MULTI-TEACHER DISTILLATION EXPERIMENTS**

| Method                                      | FR  | ES   | NL   | RU   | Avg  |
|---------------------------------------------|-----|------|------|------|------|
| ACS + Multi-teacher Distillation            | 50.13 | 63.82 | 53.51 | 54.18 | 55.41 |
| CL-XABSA(TL) + Multi-teacher Distillation  | 51.57 | 63.51 | 52.87 | 53.08 | 55.26 |
| CL-XABSA(SL) + Multi-teacher Distillation  | 51.69 | 63.54 | 54.28 | 53.98 | 55.97 |
and English are parts of the large Indo-European language family. However, they have distinct differences because they belong to different branches. English and Dutch are from the Germanic branch, Russian is from the Slavonic branch, and Spanish is from the Romance branch. The superior performance of CL-XABSA on French and Russian indicates that CL-XABSA performs better on language pairs with larger language discrepancies. The two contrastive learning objectives proposed in CL-XABSA effectively facilitate the construction of multilingual semantic space, thereby making knowledge transferred across languages more language-independent. This gives CL-XABSA better generalization performance, and it has the potential to be extended to more language pairs with large language discrepancies.

One main concern is that although Spanish and English are from different language branches, the performance of CL-XABSA on Spanish is inferior to that of ACS. One possible reason could be the label distribution of different sets of ESs. Let’s take a closer look at the test results of each sentiment for MABSA task, which are shown in Table X. Overall, CL-XABSA tends to achieve significant improvement on the NEU and NEG classes, demonstrating that CL-XABSA can alleviate the class-imbalance problem. One specific observation is that CL-XABSA works poorly at the POS class of Spanish, and the POS class takes a large account of the Spanish test set, which in turn causes CL-XABSA’s overall performance on Spanish to be inferior to that of ACS.

Dutch languages, especially on Spanish. Some in-depth analyses are as follows:

- From the perspective of the language family of the source and target languages, French, Russian, Spanish, Dutch and English are parts of the large Indo-European language family. However, they have distinct differences because they belong to different branches. English and Dutch are from the Germanic branch, Russian is from the Slavonic branch, and Spanish is from the Romance branch. The superior performance of CL-XABSA on French and Russian indicates that CL-XABSA performs better on language pairs with larger language discrepancies. The two contrastive learning objectives proposed in CL-XABSA effectively facilitate the construction of multilingual semantic space, thereby making knowledge transferred across languages more language-independent. This gives CL-XABSA better generalization performance, and it has the potential to be extended to more language pairs with large language discrepancies.

One main concern is that although Spanish and English are from different language branches, the performance of CL-XABSA on Spanish is inferior to that of ACS. One possible reason could be the label distribution of different sets of ESs. Let’s take a closer look at the test results of each sentiment for MABSA task, which are shown in Table X. Overall, CL-XABSA tends to achieve significant improvement on the NEU and NEG classes, demonstrating that CL-XABSA can alleviate the class-imbalance problem. One specific observation is that CL-XABSA works poorly at the POS class of Spanish, and the POS class takes a large account of the Spanish test set, which in turn causes CL-XABSA’s overall performance on Spanish to be inferior to that of ACS.

- From the perspective of the language family of the source and target languages, French, Russian, Spanish, Dutch

**TABLE VII**
The Results of the Initialized Student Model and Different Teacher Models

| Data   | Method               | FR   | ES   | NL   | RU   | Avg  |
|--------|----------------------|------|------|------|------|------|
| $D_T$  | ACS                  | 47.29| 60.43| 49.84| 50.11| 51.92|
|        | CL-XABSA(TL)        | 46.96| 60.20| 49.42| 49.39| 51.49|
|        | CL-XABSA(SL)        | 46.34| 59.85| 49.38| 48.96| 51.13|
| $D_S \cup D_T$ | ACS                  | 47.34| 60.06| 50.61| 51.56| 52.39|
|        | CL-XABSA(TL)        | 48.87| 60.50| 51.20| 50.72| 52.82|
|        | CL-XABSA(SL)        | 49.84| 61.29| 51.22| 50.39| 53.19|
| $D_S \cup D_T$ | ACS                  | 46.93| 59.48| 48.77| 50.24| 51.36|
|        | CL-XABSA(TL)        | 47.56| 59.12| 48.68| 50.14| 51.38|
|        | CL-XABSA(SL)        | 46.58| 60.78| 50.55| 49.87| 51.95|
| $D_T \cup D_T$ | ACS                  | 46.87| 59.15| 50.43| 49.52| 51.49|
|        | CL-XABSA(TL)        | 48.43| 59.97| 50.87| 48.97| 52.06|
|        | CL-XABSA(SL)        | 47.15| 59.20| 49.41| 49.58| 51.34|

**TABLE VIII**
The Results of Multilingual Distillation Experiments. + Denotes Results From Li et al. [24]

| Method               | FR   | ES   | NL   | RU   | Avg  |
|----------------------|------|------|------|------|------|
| MTL-DA              | 40.72| 54.14| 49.06| 43.89| 46.95|
| MTL-WS              | 46.93| 58.18| 49.87| 44.88| 49.96|
| MTL-AF              | 48.73| 59.54| 52.15| 50.99| 52.85|
| MTL-AC              | 48.60| 60.22| 51.43| 50.36| 52.65|
| MTL-CL-XABSA(TL)    | 50.01| 59.05| 51.22| 50.59| 52.72|
| MTL-CL-XABSA(SL)    | 49.92| 58.54| 49.70| 49.76| 51.98|
| MTL-ACS-D           | 49.70| 61.65| 52.41| 52.11| 53.97|
| MTL-CL-XABSA(TL)-D  | 53.03| 62.01| 54.04| 54.63| 55.93|
| MTL-CL-XABSA(SL)-D  | 52.23| 62.19| 54.25| 53.87| 55.64|

**TABLE IX**
The Results of Multilingual Distillation Experiments in Training Datasets

| Method               | Dataset | FR   | ES   | NL   | RU   | Avg  |
|----------------------|---------|------|------|------|------|------|
| MTL-AC              | Test    | 48.60| 60.22| 51.43| 50.36| 53.13|
|                     | Train   | 53.57| 58.70| 49.23| 52.90| 53.80|
| CL-XABSA(TL)        | Test    | 50.01| 59.05| 51.22| 50.59| 54.32|
|                     | Train   | 55.84| 58.34| 50.99| 54.32| 53.43|
| CL-XABSA(SL)        | Test    | 49.92| 58.54| 49.70| 49.76| 53.43|
|                     | Train   | 55.84| 58.34| 50.99| 54.32| 53.43|

**TABLE X**
The Test Results of Each Sentiment for the MABSA Task

| POS   | NEU   | NEG   |
|-------|-------|-------|
| ACS   | CL-XABSA(SL) | CL-XABSA(TL) |
| FR    | 54.35 | 10.93 | 45.77 |
| CL-XABSA(SL) | 54.39 | 18.00 | 47.77 |
| CL-XABSA(TL) | **55.00** | **18.09** | **48.00** |
| ES    | CL-XABSA(SL) | CL-XABSA(TL) |
| ACS   | **68.70** | 13.33 | 37.60 |
| CL-XABSA(SL) | 67.25 | 16.41 | 36.27 |
| CL-XABSA(TL) | **67.15** | **20.12** | **37.97** |
| RU    | CL-XABSA(SL) | CL-XABSA(TL) |
| ACS   | 58.15 | 8.64  | **38.91** |
| CL-XABSA(SL) | 57.91 | 11.76 | 37.12 |
| CL-XABSA(TL) | **58.41** | 11.29 | 38.02 |
| NL    | CL-XABSA(SL) | CL-XABSA(TL) |
| ACS   | 58.17 | 5.91  | **43.69** |
| CL-XABSA(SL) | 56.56 | 2.71  | 41.69 |
| CL-XABSA(TL) | **58.37** | **12.87** | **41.67** |
TABLE XI
THE CALINSKI–HARABASZ INDEX OF DIFFERENT MODELS

| Model                  | Calinski–Harabasz index |
|------------------------|-------------------------|
| mBERT                  | 264.10                  |
| MTL-ACS                | 57.40                   |
| MTL-CL-XABSA(TL)       | 22.16                   |
| MTL-CL-XABSA(SL)       | 23.09                   |

D. Semantic Space Verification

To demonstrate the ability of our approach to achieving a convergence of semantic spaces of different languages, we plot 2D visualization of the hidden sentence representations of different models using PCA. We visualized the results of the MTL-ACS, MTL-CL-XABSA(TL), MTL-CL-XABSA(TL), and mBERT models without fine-tuning. In addition to the other three methods, except for mBERT, we visualize the results of five experiments. The results are shown in Fig. 9. We treat the samples with the same language as the same cluster. In the semantic space of mBERT, the semantic space of each language is relatively scattered, while the other three methods are closer. In addition, although the other three methods close the semantic space of different languages, the semantic space of MTL-ACS is more scattered than the other two methods we proposed, which means that MTL-CL-XABSA (TL) and MTL-CL-XABSA(TL) better achieve convergence of semantic spaces of different languages. (the horizontal and vertical coordinates of the subgraphs (a. *) are relatively large).

To more intuitively measure the consistency of the semantic space, we use the classical clustering evaluation metric—the Calinski–Harabasz metric—for evaluation. When the Calinski–Harabasz metric is larger, the boundaries between different clusters are more obvious; when the Calinski–Harabasz metric is smaller, the degree of coincidence between different clusters is higher; that is, the semantic space representation is more consistent. As shown in Table XI, the mBERT model without fine-tuning has a large Calinski–Harabasz value. Although MTL-ACS can narrow the semantic space by data alignment to a certain extent, its Calinski–Harabasz value is higher than that of our two methods. From the Calinski–Harabasz metric point of view, MTL-CL-XABSA(TL) works best with the goal of achieving semantic space consistency.

VI. CONCLUSION

To the best of our knowledge, we are the first to apply contrastive learning to XABSA and MABSA tasks. The results show the superior performance and generalization of our framework CL-XABSA. We also verify the effectiveness of knowledge distillation on XABSA and MABSA tasks. Our framework can be applied to a variety of multilingual pre-trained models. Among the mBERT-based models, our model achieves the best performance, and our method can be regarded as a new state-of-the-art for XABSA and MABSA tasks. In the future, we will further explore the following problems: 1) the possibility of contrastive learning on XABSA tasks, 2) the possibility of using contrastive learning methods to better learn cross-lingual alignment patterns, and 3) the impact of language-specific knowledge on ABSA and MABSA tasks.

REFERENCES

[1] M. S. Akhtar, P. Sawant, S. Sen, A. Ekbkal, and P. Bhattacharyya, “Solving data sparsity for aspect based sentiment analysis using cross-linguality and multi-linguality,” in Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Hum. Lang. Technol., 2018, vol. 1, pp. 572–582, doi: 10.18653/v1/N18-1053.
[2] J. B. Barnes, P. Lambert, and T. Badia, “Exploring distributional represen-
tations and machine translation for Aspect-based Cross-lingual sentiment classification,” in Proc. 26th Int. Conf. Comput. Linguistics: Tech. Papers, 2016, pp. 1613–1623, [Online]. Available: https://aclanthology.org/C16-1152.
[3] T. Chen, S. Kornbluth, M. Norouzi, and G. Hinton, “A simple framework for contrastive learning of visual representations,” in Proc. 37th Int. Conf. Mach. Learn., 2020, pp. 1597–1607.
[4] N. Choudhary, R. Singh, I. Bindlish, and M. Shrivastava, “Contrastive learning of Emojis-based representations for resource-poor languages,” in Proc. Int. Conf. Computat. Linguistics Intell. Process., 2018, pp. 129–141, doi: 10.1007/978-3-030-23804-8_11.
[5] N. Choudhary, R. Singh, and M. Shrivastava, “Cross-Linguistic task-specific representation learning for text classification in resource poor languages,” 2018, arXiv: 1806.03590.
[6] A. Conneau et al., “Unsupervised cross-lingual representation learning at scale,” in Proc. 58th Annua. Meetting Assoc. Comput. Linguistics, 2020, pp. 8440–8451, doi: 10.18653/v1/2020.acl-main.747.
[7] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Hum. Lang. Technol, 2019, pp. 4171–4186.
[8] X. Dong and G. de Melo, “Cross-lingual propaganda for deep senti-
ment analysis,” in Proc. 32nd AAAI Conf. Artif. Intell., 30th Innov. Appl. Artif. Intell. Conf., 8th AAAI Symp. Educ. Adv. Artif. Intell., 2018, pp. 5771–5778.
[9] C. Dyer, V. Chalaneau, and N. A. Smith, “A simple, fast, and effective reparameterization of IBM model 2,” in Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Hum. Lang. Technol., 2013, pp. 644–648, [Online]. Available: https://aclanthology.org/N13-1073.
[10] T. Gao, X. Yao, and D. Chen, “SimCSE: Simple contrastive learning of sentence embeddings,” in Proc. Conf. Empirical Methods Natural Lang. Process., 2021, pp. 6894–6910, doi: 10.18653/v1/2021.emnlp-main.552.
[11] Z. Guo, Z. Liu, Z. Ling, S. Wang, L. Jin, and Y. Li, “Text classification by contrastive learning and Cross-lingual data augmentation for Alzheimer’s disease detection,” in Proc. 28th Int. Conf. Computat. Linguistics, 2020, pp. 6161–6171, doi: 10.18653/v1/2020.coling-main.542.
[12] B. He et al., “An interactive multi-task learning network for End-to-End aspect-based sentiment analysis,” in Proc. 57th Annua. Meet. Assoc. Comput. Linguistics, 2019, pp. 504–515.
[13] O. Henaff, “Data-efficient image recognition with contrastive predictive coding,” in Proc. Int. Conf. Mach. Learn., 2020, pp. 4182–4192.
[14] G. Hinton, O. Vinyals, and J. Dean, “Distilling the knowledge in a neural network,” 2015, arXiv:1503.02531.
[15] J. Hu, S. Ruder, A. Siddhant, G. Neubig, O. Firat, and M. Johnson, “Xreme: A massively multilingual multi-task benchmark for evaluating cross-lingual generalisation,” in Proc. Int. Conf. Mach. Learn., 2020, pp. 4411–4421.
[16] S. Jebbara and P. Cimiano, “Zero-shot cross-lingual opinion target extrac-
tion,” in Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Hum. Lang. Technol., 2019, pp. 2486–2495, doi: 10.18653/v1/N19-1257.
[17] Z. Ke, B. Liu, H. Xu, and L. Shu, “CLASSIC: Continual and con-
trastive learning of aspect sentiment classification tasks,” in Proc. Conf. Empirical Methods Natural Lang. Process., 2021, pp. 6871–6883, doi: 10.18653/v1/2021.emnlp-main.550.
[18] P. Keung, Y. Lu, G. Szarvas, and N. A. Smith, “The multilingual amazon reviews corpus,” in Proc. Conf. Empirical Methods Natural Lang. Process., 2020, pp. 4563–4568, doi: 10.18653/v1/2020.emnlp-main.369.
[19] P. Khosla et al., “Supervised contrastive learning,” in Proc. Adv. Neural Inf. Process. Syst., 2020, vol. 33, pp. 18661–18673, [Online]. Available: https://proceedings.neurips.cc/paper/2020/file/1d89a66c7e80a29b1bbdb0f2e1a94a8-Paper.pdf.
[20] R. Klinger and P. Cimiano, “Instance selection improves cross-lingual model training for fine-grained sentiment analysis,” in Proc. 19th Conf. Comput. Natural Lang. Learn., 2015, pp. 153–163, doi: 10.18653/v1/K15-1016.

[21] J. Li, Y. Zhang, Z. Wang, and K. Tu, “Probabilistic contrastive learning for domain adaptation,” 2021, arXiv: 2111.06021.

[22] L. Li, T. Tayir, Y. Han, X. Tao, and J. D. Velásquez, “Multimodality information fusion for automated machine translation,” Inf. Fusion, vol. 91, pp. 352–363, 2023.

[23] X. Li, L. Bing, W. Zhang, and W. Lam, “Exploiting BERT for End-to-End Aspect-based sentiment analysis,” in Proc. 5th Workshop Noisy User-Generated Text, 2019, pp. 34–41, doi: 10.1145/3340531.3411868.

[24] X. Li, L. Bing, W. Zhang, Z. Li, and W. Lam, “Unsupervised Cross-lingual adaptation for sequence tagging and beyond,” 2020, arXiv: 2010.12405.

[25] B. Liang, R. Yin, L. Gui, J. Du, Y. He, and R. Xu, “Aspect-invariant sentiment features learning: Adversarial multi-task learning for aspect-based sentiment analysis,” in Proc. 29th ACM Int. Conf. Inf. Knowl. Manage., 2020, pp. 825–834, doi: 10.1145/3340531.3411868.

[26] M. Mohtarami, J. Glass, and P. Nakov, “Contrastive language adaptation for cross-lingual stance detection,” in Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process., 2019, pp. 4442–4452, doi: 10.18653/v1/D19-1452.

[27] A. V. D. Oord, Y. Li, and O. Vinyals, “Representation learning with contrastive predictive coding,” 2018, arXiv:1807.03748.

[28] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, “SemEval-2014 task 4: Aspect based sentiment analysis,” in Proc. 8th Int. Workshop Semantic Eval., 2014, pp. 27–35, doi: 10.3119/v1/IS14-2004.

[29] M. Pontiki et al., “SemEval-2016 task 5: Aspect based sentiment analysis,” in Proc. 10th Int. Workshop Semantic Eval., 2016, pp. 19–30, doi: 10.18653/v1/S16-1002.

[30] N. Reddy, P. Singh, and M. M. Srivastava, “Does BERT understand sentiment? Leveraging comparisons between contextual and non-contextual embeddings to improve aspect-based sentiment models,” 2020, arXiv: 2011.11673.

[31] S. Ruder, I. Vulić, and A. Søgaard, “A survey of cross-lingual word embedding models,” J. Artif. Intell. Res., vol. 65, no. 1, pp. 569–630, 2019, doi: 10.1613/jair.11640.

[32] K. Sattar, Q. Umer, D. G. Vashieva, S. Chung, Z. Latif, and C. Lee, “A multi-layer network for aspect-based cross-lingual sentiment classification,” IEEE Access, vol. 9, pp. 133961–133973, 2021, doi: 10.1109/ACCESS.2021.3116053.

[33] V. Suresh and D. Ong, “Not all negatives are equal: Label-aware contrastive loss for Fine-grained text classification,” in Proc. Conf. Empirical Methods Natural Lang. Process., 2021, pp. 4381–4394, doi: 10.18653/v1/2020.emnlp-main.359.

[34] O. Wallaart and F. Frasincar, “A hybrid approach for aspect-based sentiment analysis using a lexicalized domain ontology and attentional neural models,” in Proc. Eur. Semantic Web Conf., 2019, pp. 363–378.

[35] W. Wang and S. J. Pan, “Transition-based adversarial network for Cross-lingual aspect extraction,” in Proc. 27th Int. Joint Conf. Artif. Intell., 2018, pp. 4475–4481, doi: 10.24963/ijcai.2018/622.

[36] O. Wallaart and F. Frasincar, “A hybrid approach for aspect-based sentiment analysis using a lexicalized domain ontology and attentional neural models,” in Proc. Eur. Semantic Web Conf., 2019, pp. 363–378.

[37] S. Wu and M. Dredze, “Beto, bentz, becas: The surprising cross-lingual effectiveness of BERT,” in Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process., 2019, pp. 833–844, doi: 10.18653/v1/D19-1077.

[38] Z. Wu, Y. Xiong, S. X. Yu, and D. Lin, “Unsupervised feature learning via non-parametric instance discrimination,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 3733–3742.

[39] W. Zhang, R. He, H. Peng, L. Bing, and W. Lam, “Cross-lingual aspect-based sentiment analysis with aspect term code-switching,” in Proc. Conf. Empirical Methods Natural Lang. Process., 2021, pp. 9220–9230, [Online]. Available: https://aclanthology.org/2021.emnlp-main.727.

[40] X. Zhou, X. Wan, and J. Xiao, “CLOpinionMiner: Opinion target extraction in a cross-language scenario,” IEEE/ACM Trans. Audio Speech Lang. Process., vol. 23, no. 4, pp. 619–630, Apr. 2015.