

Table-Filling via Mean Teacher for Cross-domain Aspect Sentiment Triplet Extraction

Kun Peng
Institute of Information Engineering,
Chinese Academy of Sciences,
Beijing, China
School of Cyber Security, University
of Chinese Academy of Sciences,
Beijing, China
pengkun@iie.ac.cn

Lei Jiang*
Institute of Information Engineering,
Chinese Academy of Sciences,
Beijing, China
jianglei@iie.ac.cn

Qian Li
Beijing University of Posts and
Telecommunications,
Beijing, China
li.qian@bupt.edu.cn

Haoran Li
Hong Kong University of Science and
Technology,
HongKong, China
hlibt@connect.ust.hk

Xiaoyan Yu
Beijing Institute of Technology,
Beijing, China
xiaoyan.yu@bit.edu.cn

Li Sun
North China Electric Power
University,
Beijing, China
ccsunli@ncepu.edu.cn

Shuo Sun
Beihang University,
Beijing, China
sun.shuo@buaa.edu.cn

Yanxian Bi
China Academy of Electronic and
Information Technology,
Beijing, China
biyanxian@cetc.com.cn

Hao Peng
Beihang University,
Beijing, China
penghao@buaa.edu.cn

ABSTRACT

Cross-domain Aspect Sentiment Triplet Extraction (ASTE) aims to extract fine-grained sentiment elements from target domain sentences by leveraging the knowledge acquired from the source domain. Due to the absence of labeled data in the target domain, recent studies tend to rely on pre-trained language models to generate large amounts of synthetic data for training purposes. However, these approaches entail additional computational costs associated with the generation process. Different from them, we discover a striking resemblance between table-filling methods in ASTE and two-stage Object Detection (OD) in computer vision, which inspires us to revisit the cross-domain ASTE task and approach it from an OD standpoint. This allows the model to benefit from the OD extraction paradigm and region-level alignment. Building upon this premise, we propose a novel method named **Table-Filling via Mean Teacher (TFMT)**. Specifically, the table-filling methods encode the sentence into a 2D table to detect word relations, while TFMT treats the table as a feature map and utilizes a region consistency to enhance the quality of those generated pseudo labels. Additionally, considering the existence of the domain gap, a cross-domain consistency based on Maximum Mean Discrepancy is designed to

alleviate domain shift problems. Our method achieves state-of-the-art performance with minimal parameters and computational costs, making it a strong baseline for cross-domain ASTE.

CCS CONCEPTS

• **Computing methodologies** → **Information extraction**; • **Information systems** → **Sentiment analysis**.

KEYWORDS

Text Mining, Aspect Sentiment Triplet Extraction, Cross-domain

ACM Reference Format:

Kun Peng, Lei Jiang, Qian Li, Haoran Li, Xiaoyan Yu, Li Sun, Shuo Sun, Yanxian Bi, and Hao Peng. 2024. Table-Filling via Mean Teacher for Cross-domain Aspect Sentiment Triplet Extraction. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM '24)*, October 21–25, 2024, Boise, ID, USA. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3627673.3679543>

1 INTRODUCTION

Fine-grained sentiment analysis is an important research direction in Natural Language Processing (NLP). In recent years, there has been a constant stream of novel tasks being presented in this field, such as aspect-based sentiment analysis (ABSA) [24, 41], aspect opinion pair extraction (AOPE) [33], and aspect sentiment triplet extraction (ASTE) [4, 21, 33, 40]. These tasks vary in the elements they seek to extract, with ASTE representing the most intricate and challenging undertaking. As shown in the left half of Figure 1, ASTE involves the comprehensive extraction of triplets from a given sentence, encompassing aspect terms, opinion terms, and sentiment polarity (i.e., positive, neutral, and negative). Give the

*Corresponding author



This work is licensed under a Creative Commons Attribution International 4.0 License.

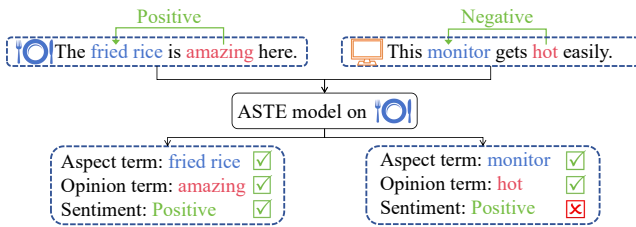


Figure 1: The ASTE model trained in the restaurant domain fails to produce accurate results in the laptop domain.

sentence “*The fried rice is amazing here.*”, the corresponding triplet is (*fried rice*, *amazing*, *Positive*). Although there have been many domain-specific ASTE works, they fail to meet the need to handle data from multiple domains in real-world scenarios. As shown in the right half of Figure 1, when we apply the ASTE model trained exclusively in the restaurant domain to predict data from the unseen laptop domain, the model generates incorrect predictions due to its inability to adapt to the new domain knowledge.

Cross-domain ASTE task [8] has newly been proposed with the objective of training models to perform ASTE tasks in the target domain, using only annotated data from the source domain and unlabeled data from the target domain. Previous works on cross-domain fine-grained sentiment analysis focused on learning cross-domain invariant pivotal features [15, 38]. However, this approach relies on task-specific design and may not fully capture the nuanced distinctions within each particular domain. Another more promising research direction is to leverage the power of pre-trained language models (PLMs) to generate a large amount of pseudo data that fits the distribution of the target domain [8, 13]. Although effective, compared to providing pseudo labels for unlabeled data, this method requires significant computational costs during the data generation process. Mean-teacher [1, 30] is a widely adopted method in unsupervised domain adaptation (UDA). It improves the performance by training a supervised student model and an unsupervised teacher model, utilizing their prediction differences on labeled and unlabeled target domain data. Although highly effective in computer vision (CV) tasks, mean-teacher is rarely utilized in ASTE, where generation paradigms are generally preferred. To leverage the embedded prior knowledge of PLMs, these methods [8, 13] choose to endure the high computational costs. This consensus hinders the application of the mean teacher in cross-domain research within the ASTE area. However, diverging from previous studies, we have discovered a striking similarity between the table-filling method used for ASTE and the two-stage Object Detection (OD) method [25] in terms of their workflow. Mapping sentiment triplets to regions on the figure allows the model to benefit from staged extraction (for ASTE) and region-level alignment (for cross-domain). This intriguing finding has inspired us to propose the Table-Filling via Mean Teacher (TFMT) approach.

Employing the table-filling method [10, 33] is an appealing direction for addressing the ASTE task. It encodes a sequence of sentences into a 2D table, where the rows and columns represent different entity categories (i.e., aspect terms and opinion terms). The relations between entities are marked in the off-diagonal grid of the

table, and supervised training is used to learn the relations between words. As shown in Figure 2, we observe that the region-level table-filling method shares a similar process logic with the two-stage OD, which involves four steps: 1) A given image is encoded into a feature map using a convolutional neural network (CNN). 2) A Region Proposal Network (RPN) is used to obtain all candidate regions. 3) the feature vectors within each candidate region are encoded into fixed-length representations using Region of Interest (RoI) pooling. 4) a Region Proposal Classifier (RPC) is used for region classification. In summary, the table-filling methods treat the target triplets as specific regions within the table. To detect these regions, they first identify candidate regions and then categorize them, mirroring the two-stage process of detecting and classifying objects in images. Many table-filling methods [4, 33, 40] adopt the two-stage formulation. Other work [19] employs a one-stage extraction framework in the relation extraction task. However, these works merely skim the surface and fail to thoroughly investigate the correlation between the two tasks or explore this direction further. We believe that addressing cross-domain ASTE tasks from a two-stage OD perspective will yield two key advantages for the model: 1) The staged extraction paradigm will better leverage the word relations within the regions; 2) assisting cross-domain models in focusing on crucial information within the regions.

Specifically, the proposed TFMT approach utilizes a mean-teacher architecture to guide domain adaptation. Mean-Teacher consists of two models with the same structures: a student model and a teacher model. The teacher model is pre-trained on the labeled source domain and then freezes its parameters in the training step, while the student model is randomly initialized. We reformat the ASTE task into an OD task by treating the encoded table as a feature map. The student network is supervisedly trained in each training epoch using source domain data. At the same time, it receives region-level consistency constraints from the teacher model’s pseudo labels generated using unlabeled target domain data. Because the triplets are labeled as regions, region-level consistency converts the extraction alignment from student and teacher into region alignment. After each training epoch, the teacher model parameters are updated via an exponential moving average (EMA) of the student model. Furthermore, since the model only possesses prior knowledge from the source domain, the pseudo-labels generated on the target domain inevitably contain noise. To alleviate this label shift phenomenon, we propose a multi-perspective consistency module based on Maximum Mean Discrepancy (MMD) [17].

Our main contributions can be summarised as follows:

- 1) We observe the similarity between table-filling methods in ASTE tasks and two-stage OD, which guides us to explore an effective solution for cross-domain ASTE.
- 2) The proposed multi-perspective MMD consistency can effectively mitigate the label shift caused by domain adaptation without increasing parameters.
- 3) The proposed TFMT framework avoids the paradigm of synthetic data generation, resulting in significantly fewer parameters and computational time costs compared to previous approaches.
- 4) Our framework is concise yet achieves state-of-the-art performance, serving as a strong baseline for follow-up research and inspiring the adoption of more advancing mean teacher methods in this field.

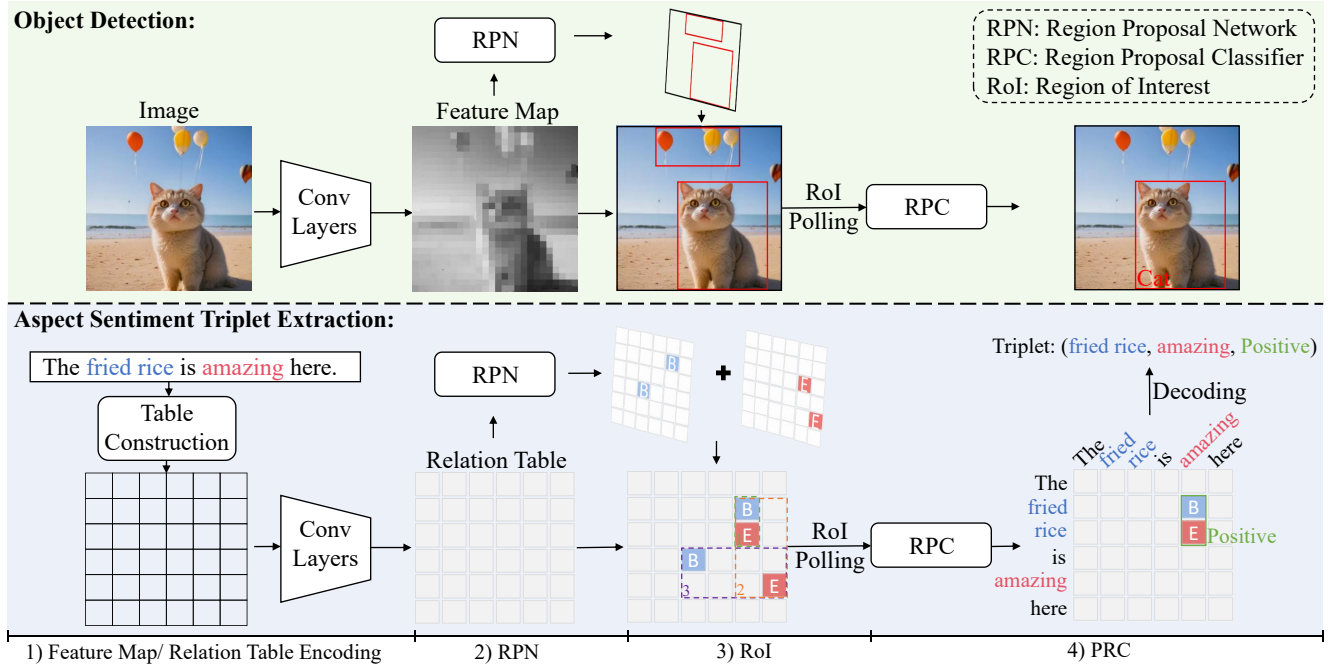


Figure 2: The general architecture of two-stage Object Detection (OD) and region-level table-filling method for Aspect Sentiment Triplet Extraction (ASTE) share similar underlying principles. *B* (Beginning) and *E* (Ending) represent the top-left and bottom-right corners of a candidate region, respectively.

2 RELATED WORK

2.1 Aspect Sentiment Triplet Extraction (ASTE)

ASTE is a recently emerged and challenging task. [20] introduced this task and employed a pipeline approach to extract aspect terms, opinion terms, and sentiment polarities in multiple stages. Recognizing the heavy reliance on word-level interactions in previous studies, SpanASTE [34] proposed a span-level method to consider span-to-span interactions explicitly. Some studies have treated ASTE as a sequence-to-sequence task and utilized generative models to address it. GAS [39] designed two normalized generative paradigms, annotation-style and extraction-style modeling. COM-MRC [37] formalized the ASTE task into a Machine Reading Comprehension (MRC) based framework and devised a context-enhancement strategy to identify sentiment triplets. RoBMRC [16] proposed a robustly optimized bidirectional machine reading comprehension method to tackle issues of query conflict and probability unilateral decrease encountered during generation. MvP [9] used human-like problem-solving processes to generate and summarize sentiment elements in different orders.

Another popular approach is the table-filling [18] method. GTS [33] was the first to apply the table-filling method to the ASTE task, they annotated sentiment triplets into a two-dimensional table for extraction. EMC-GCN [4] introduced a multi-channel GCN-based approach that further integrates syntactic and textual structural information into table-filling. On the other hand, BDTF [40] modified the original table-filling annotation scheme and proposed a boundary detection-based method.

2.2 UDA for Fine-grained Sentiment Analysis

Early UDA studies for fine-grained sentiment analysis primarily focused on learning domain-invariant features. The majority of these explorations introduced syntactic knowledge as a bridge to overcome domain gaps. [15] proposed a Selective Adversarial Learning (SAL) method to automatically capture the latent relations of words, thereby reducing reliance on external linguistic resources. [32] developed a novel RNN that can effectively mitigate word-level domain shifts through syntactic relations. EATN [38] designed two special tasks and a novel aspect-oriented multi-head attention mechanism to learn shared features from a well-labeled source domain and guide the classification performance in the target domain. Another line of research has placed emphasis on data. CDRG [36] and GCDDA [13] proposed a two-stage generative cross-domain data augmentation framework. Based on reviews labeled in the source domain, this framework first masks source-specific attributes and then converts the domain-independent review into a target-domain review, thereby generating target-domain reviews with fine-grained annotations. BGCA [8] further advanced this approach by proposing a bidirectional generative framework. This framework trains generative models in text-to-label and label-to-text directions to produce higher-quality target-domain data.

2.3 UDA for Object Detection (OD)

To learn domain-invariant visual features, some adversarial learning methods [6, 26] employed a domain discriminator and conducted adversarial training on the feature encoder and discriminator. Another

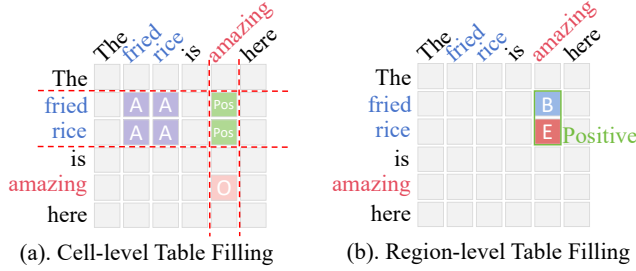


Figure 3: Examples of (a) cell-level and (b) region-level methods. As the latter closely resembles two-stage object detection, our TFMT is based on the region-level paradigm.

research line is based on data augmentation. These methods [3, 12] employed generative models (e.g., CycleGAN [42]) to produce interpolated samples between the source and target domains, thereby bridging the distribution gap between the domains. Mean-teacher [30] is a semi-supervised learning method, and recently [1] has been exploring its potential in UDA for OD. Further investigations [5, 7, 14] have been conducted to improve data augmentation or employ adversarial training techniques, with the goal of enhancing the quality of pseudo labels generated by the mean-teacher model. Contrastive Mean Teacher (CMT) [2] proposed an integrated framework to leverage the identified intrinsic alignment and synergy between contrastive learning and mean-teacher self-training.

3 METHOD

In this section, we first give the task formalization and introduce the table tagging scheme shown in Figure 3. Then, we present the proposed Table-Filling via Mean Teacher (TFMT) method as illustrated in Figure 4.

3.1 Task Formalization

Given a sentence s with length n , the ASTE task aims to extract all triplets $\tau = [(a, o, l)_1, (a, o, l)_2, \dots, (a, o, l)_{|\tau|}]$ from s , where a , o and l denote aspect term, opinion term and sentiment polarity in $\{Pos, Neu, Neg\}$, respectively. For cross-domain ASTE, given a set of source domain labeled sentences $D_s = \{(s_i^s, \tau_i^s)\}_{i=1}^{|D_s|}$ and a set of target domain unlabeled sentences $D_t = \{s_i^t\}_{i=1}^{|D_t|}$, where s_i^s and s_i^t are sentences in the source and target domains, respectively. τ_i^s is the triplets set of s_i^s . Our model aims to extract all triplets from the test set specific to the target domain.

Table Tagging Scheme. As shown in Figure 3, the relations between words in a sentence are represented in a square $n \times n$ table. The vertical and horizontal axes represent aspects and opinions, respectively. There exist two types of table-filling methods: cell-level and region-level. As shown in Figure 3 (a), the cell-level method annotates aspect terms (A) and opinion terms (O) on the diagonal of the table, while the cross-region between the two is used to annotate the corresponding sentiment polarity. It extracts the positions of aspect/opinion terms and then obtains the extraction results based on the category of the cells where they intersect.

Unlike the cell-level method, the region-level method [40] treats the target triplet as a specific region within the table and aims to

precisely identify its surrounding boundaries. As shown in Figure 3 (b), regions' boundaries are labeled using beginning (B) and ending (E), followed by labeling the sentiment of the annotated regions. Here, B and E are used to represent the upper left and lower right corners of a relation rectangle, respectively. While the rest of the numerous gray cells are marked as "None". This boundary-based tagging scheme converts word-level annotations in text sequences into region-level annotations on a feature map. Unlike cell-level methods that annotate all cells at once, region-level methods aligns closely with the two-stage object detection paradigm, hence adopted in our TFMT.

3.2 Table-Filling via Mean Teacher (TFMT)

The architecture of TFMT¹ is shown in Figure 4. Our TFMT framework consists of a student model and a teacher model, both sharing the same model architecture. The student model serves as the primary model, receiving supervised training on the source domain data while also undergoing unsupervised training using unlabeled target domain data.

The teacher model is utilized to guide the training of the student model. It is pretrained on the source domain and then frozen at each training step. During training step, the teacher model focuses on the target domain and generates pseudo labels for the student model. Specifically, in each training iteration, the target domain data D_t is augmented using a random word substitution to produce \tilde{D}_t . The teacher model then processes these augmented data and produces multiple region proposals along with their predicted softmax values \tilde{P}^r . Subsequently, we filter the region proposals and retain only those labels for which the confidence $\max_{i \in C} (\tilde{P}_i^r)$ (where C is the foreground label set) exceeds the threshold η . This confidence threshold η is applied to ensure that only high-confidence pseudo labels are assigned to the student model. After each training step, The teacher's parameters are updated through an Exponential Moving Average (EMA) [30] mechanism: $\Theta_t = \lambda \Theta_t + (1 - \lambda) \Theta_s$, where λ ($0 < \lambda < 1$) is a hyper-parameter. Θ_t and Θ_s denote the model parameters of teacher and student, respectively.

3.2.1 Supervised Training. To align with the two-stage OD paradigm for both teacher and student base models, we adopted Boundary-Driven Table-Filling (BDTF) [40] and restructured the framework accordingly. The architecture of the base model is shown in the lower half of Figure 2.

Table Encoding. Given the sentence $s = [x_1, x_2, \dots, x_n]$, we input it into a pre-trained language model and get the last hidden layer $[\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_n] \in \mathbb{R}^{n \times d}$ as the sentence representations, where n is the sentence length.

We donate the relation table as $T \in \mathbb{R}^{n \times n \times d}$, which maps the relations between any two words into separate cells. Specifically, for a cell $t_{ij} \in \mathbb{R}^d$ in T , it represents the relation between the i -th and j -th words in sentence s . To construct the relation table, BDTF calculates each t_{ij} as follows:

$$t_{ij} = \text{Linear}(\mathbf{h}_i \oplus \mathbf{h}_j \oplus \text{pooling}(\mathbf{h}_{i:j}) \oplus (\mathbf{h}_i)^T \mathbf{V} \mathbf{h}_j), \quad (1)$$

where Linear denotes a nonlinear projection with an activation function, \oplus denotes a concatenation operation. The third term

¹Our code and data are available at <https://github.com/KunPunCN/TFMT>

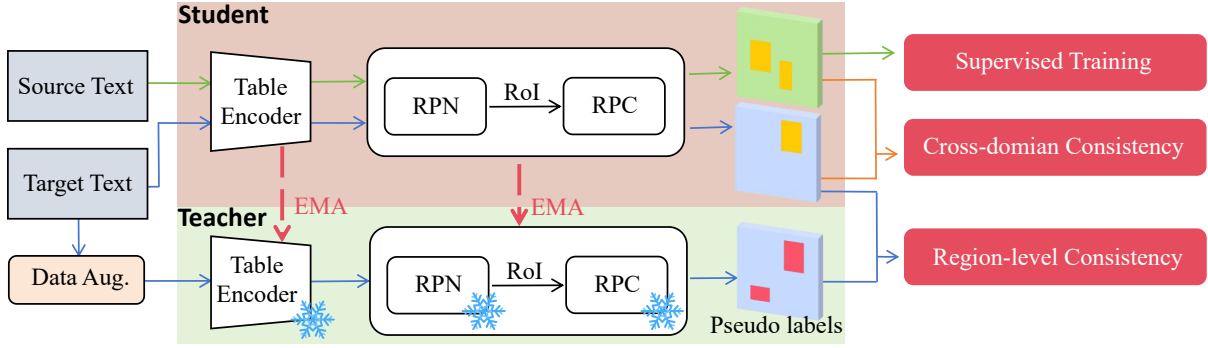


Figure 4: The architecture of TFMT. Our model consists of two identical architectures: 1) the teacher model, whose parameters are first pre-trained on the source domain and later frozen in each training step; 2) the student model, whose parameters are reinitialized. The student model is trained using source domain data and target domain pseudo-labels provided by the teacher model, and the teacher model parameters are updated using EMA after the training step.

$\text{pooling}(\mathbf{h}_{i:j})$ (assume $i < j$) is a max-pooling operation, which aims to capture essential contextual information between word pairs, thereby accentuating pertinent relational features. The fourth term $(\mathbf{h}_i)^T \mathbf{V} \mathbf{h}_j$ employs a tensor-based operation as described in [29], which is intended to harness the nuances of word-to-word interactions effectively. \mathbf{V} is a parameter matrix.

After obtaining the relational table T , we regard it as a feature map T^0 . To thoroughly learn and extract relational features, BDTF employs ResNet-style CNN [11] as an encoder. Specifically, given the $(l-1)$ -th CNN layer representation T^{l-1} , the l -th CNN layer representation $T^l \in \mathbb{R}^{n \times n \times d}$ is calculated as:

$$T^l = \text{Resnet}_l(T^{l-1}), \quad (2)$$

where Resnet_l denotes the l -th ResNet layer. The design of Resnet_l can be flexible, but it requires ensuring that the output dimensions of the convolutions are the same as the input dimensions $\mathbb{R}^{n \times n \times d}$.

Region Proposal Networks (RPN). The RPN aims to output a set of rectangular object proposals from the table, each with an objectness score. This process first requires categorizing each cell in the table as one of $\{B, E, \text{invalid}\}$. Given the feature map T^L output by the last convolutional layer (assume L layers in total), we calculate the label probabilities $P_{ij}^B, P_{ij}^E (i, j < n)$ for each cell as:

$$P_{ij}^B = \sigma(\text{Linear}_B(T^L)), \quad P_{ij}^E = \sigma(\text{Linear}_E(T^L)), \quad (3)$$

where Linear_B and Linear_E are two independent fully connected layers, σ is a sigmoid activation function. Instead of predicting the labels with a certain threshold, BDTF uses a top- k pruning strategy to retain the most promising candidates and mitigate the potential exposure bias [28]. First, RPN sorts the cells according to P_{ij}^B and P_{ij}^E , and then the top- k cells with the highest probabilities are selected to create candidate sets \mathcal{B} and \mathcal{E} , respectively, where k is a value determined based on sentence length.

After obtaining the candidate sets \mathcal{B} and \mathcal{E} , RPN constructs the candidate regions set based on their positional relations. For elements $B_{(a,b)}$ (at the (a, b) position) and $E_{(c,d)}$ (at the (c, d) position) in \mathcal{B} and \mathcal{E} respectively, if $B_{(a,b)}$ is in the upper-left corner of $E_{(c,d)}$

or they overlap (i.e., $a \leq c$ and $b \leq d$), then the corresponding rectangular region R_{abcd} is considered as a hit proposal. As illustrated in the RoI stage table in Figure 2, four labels B and E can form three corresponding candidate regions, which are respectively marked by green, orange, and purple boxes.

RoI Polling and Region Proposal Classifier (RPC). For a region proposal R_{abcd} , we define its representation $\mathbf{r}_{abcd} \in \mathbb{R}^{3d}$ as:

$$\mathbf{r}_{abcd} = \mathbf{t}_{ab}^L \oplus \mathbf{t}_{cd}^L \oplus \text{pooling}(\mathbf{t}_{a:c,b:d}^L), \quad (4)$$

where \mathbf{t}_{ab}^L and \mathbf{t}_{cd}^L are region vertices, $\mathbf{t}_{a:c,b:d}^L$ donate all the features encompassed within the region.

After obtaining the RoI representation, BDTF employs a classifier to predict its type in $\{\text{Pos}, \text{Neu}, \text{Neg}, \text{Invalid}\}$:

$$P_{abcd}^r = \text{softmax}(\text{Linear}_p(\mathbf{t}_{abcd})), \quad (5)$$

where Linear_p is a fully connected layer, softmax is a softmax function. $P_{abcd}^r \in \mathbb{R}^4$ is the sentiment label of the given region. During decoding, we drop those region proposals with $l_{abcd} = \text{Invalid}$ and generate the triplets from the remained proposals. For instance, if region r_{abcd} is predicted as POS , then the corresponding sentiment triplet is $(\text{aspect} = s[a : c], \text{opinion} = s[b : d], l = \text{POS})$.

Supervised Loss. Given the ground truth label $y_{ij}^B, y_{ij}^E \in \{0, 1\}$ for each cell in the relation table, the loss \mathcal{L}_{RPN} of RPN is calculated as the mean cross-entropy between P_{ij}^B/P_{ij}^E and y_{ij}^B/y_{ij}^E :

$$\mathcal{L}_{RPN} = \sum_{i=1}^n \sum_{j=1}^n L_{CE}(P_{ij}^B, y_{ij}^B) + L_{CE}(P_{ij}^E, y_{ij}^E), \quad (6)$$

where $0 \leq i, j \leq n$, L_{CE} is the cross entropy loss function. Given the ground truth label $y_i^r \in \{0, 1, 2, 3\}$ for each region in the candidate proposals, the loss \mathcal{L}_{RPC} of RPC is calculated as the mean cross-entropy between P_i^r and y_i^r :

$$\mathcal{L}_{RPC} = - \sum_{i=1}^m L_{CE}(P_i^r, y_i^r), \quad (7)$$

where $0 \leq i \leq m$. m is the number of region proposals. The overall supervised loss \mathcal{L}_{sup} is defined as the sum of \mathcal{L}_{RPN} and \mathcal{L}_{RPC} :

$$\mathcal{L}_{sup} = \mathcal{L}_{RPN} + \mathcal{L}_{RPC}. \quad (8)$$

3.2.2 Region-level Consistency. To compel alignment between teacher and student models in focusing on important information within a region, we adopt a region-level consistency as the unsupervised loss between the student and teacher models. We define the collection of pseudo-labeled regions detected by teacher RPN as \tilde{R} , in both the student and teacher models, the region predictions provided by the RPN are represented as P^r and \tilde{P}^r , respectively. The unsupervised loss is formulated as the Mean Squared Error between the predictions of the student and teacher:

$$\mathcal{L}_{uns} = \frac{1}{|\tilde{R}|} \sum_{r \in \tilde{R}} \|P_r^r - \tilde{P}_r^r\|_2^2. \quad (9)$$

Since our method labels all triplets as regions in the relation table, utilizing region-level consistency enables the alignment of predictions from student and teacher models and effectively reduces pseudo-label errors caused by unclear predicted word boundaries.

3.2.3 Cross-domain Consistency. Due to the fact that the knowledge learned by the teacher model mainly stems from training on the source domain labels, the use of pseudo labels generated by the teacher model in the target domain inevitably introduces noise, resulting in domain discrepancy. Different from using a domain discriminator to align the distributions of two domains [14], we utilize the Maximum Mean Discrepancy (MMD) to minimize the domain discrepancy, which does not require any additional model parameters. Specifically, given a source domain sample s^s , the student model generates a feature map T^s according to Eq. 1 to Eq. 2, from which we can extract the boundary features t_{ab}^s, t_{cd}^s , and the region feature r_{abcd}^s for the predicted regions by Eq. 4. Similarly, provided with a target domain sample s^t , we can extract the corresponding boundary features t_{ab}^t, t_{cd}^t and the region feature r_{abcd}^t . Then we define a boundary-level MMD and a region-level MMD:

$$\mathcal{L}_{boundary} = L_{MMD}(t_{ab}^s, t_{ab}^t) + L_{MMD}(t_{cd}^s, t_{cd}^t), \quad (10)$$

$$\mathcal{L}_{region} = L_{MMD}(r_{abcd}^s, r_{abcd}^t), \quad (11)$$

where L_{MMD} is the MMD function [17]. The cross-domain consistency \mathcal{L}_{mmd} is defined as the sum of the two terms:

$$\mathcal{L}_{mmd} = \mathcal{L}_{boundary} + \mathcal{L}_{region}. \quad (12)$$

3.2.4 Overall Loss. The overall loss of our TFMT is:

$$\mathcal{L} = \mathcal{L}_{sup} + \alpha \mathcal{L}_{uns} + \beta \mathcal{L}_{mmd}, \quad (13)$$

where α and β are the trade-off parameters, the loss terms are respectively defined in Eq. 8, Eq. 9 and Eq. 12.

4 EXPERIMENTS

4.1 Experimental Settings

4.1.1 Datasets. Following [8], we evaluate our model by employing a publicly available dataset sourced from the SemEval ABSA Challenges [22–24]. This dataset was released by [35] and consists of four subsets, namely R14, R15, R16, and L14, which are derived from two domains: restaurant reviews and laptop reviews. Table 1 summarizes the statistics of benchmark datasets.

Table 1: Statistics of datasets.

Dataset	Restaurant						laptop	
	R14		R15		R16		L14	
	#s	#t	#s	#t	#s	#t	#s	#t
train	1266	2338	605	1013	857	1394	906	1460
dev	310	577	148	249	210	339	219	346
test	492	994	322	485	326	514	328	543

4.1.2 Implementation. We opt for BERT [31] as our language encoder and a two-layer ResNet [11] as the feature encoder, with a dimension set at 768. The pruning threshold k and the confidence threshold η are set to 0.3 and 0.98, respectively. We directly use BERT-base-uncased² as the pretraining parameters to initialize both the student and teacher models. If not specified, the default setting for the rate smoothing coefficient λ of the EMA is 0.6. The loss weight factors $\alpha = 1$ and $\beta = 0.005$. For the data augmentation method, since the prerequisite for region-level consistency is to maintain the same encoded length of the augmented sentences, we randomly replaced 50% of the words in the sentence with words that have the same encoding lengths. We conduct model training on one RTX 3090 GPU for 10 epochs, utilizing a batch size of 4 and setting the learning rate to 3×10^{-5} . We assess the performance of the training model on the development set and save the best one. We use sentence-level F1 score as the evaluation metric. In this metric, a sentence is considered a true positive only when all sentiment triplets within it are correctly extracted. All reported results are averaged across five runs with different random seeds.

Cell-level TFMT (cTFMT). The cell-level method doesn't utilize a region proposal method like the two-stage OD paradigm but instead directly classifies each cell in the relation table. To adapt to the TFMT method, we treat each cell as an independent region (in \mathcal{L}_{uns} and \mathcal{L}_{mmd}) and keep the other TFMT modules (i.e., \mathcal{L}_{sup}) unchanged. Specifically, in Eq. 9, \tilde{R} represents all cells predicted as one of $\tilde{c} = \{A, O, POS, NEG, NEU\}$. In Eq. 12, \mathcal{L}_{mmd} is modified to the sum of MMD for all types of cells:

$$\mathcal{L}_{mmd} = \sum_{k \in \tilde{c}} L_{MMD}(t_k^s, t_k^t). \quad (14)$$

The cell-level TFMT (cTFMT) is a degradation of the original region-level TFMT.

4.1.3 Baselines. Although there are many baseline models for cross-domain sentiment analysis (e.g., EATN [38], CDRG [36], GCDDA [13] and BGCA [8]), these works mainly focus on sentence-level sentiment analysis or aspect-based sentiment analysis (ABSA). Among them, BGCA addresses both ASTE and AOPE tasks, while GCDDA specifically targets the AOPE task.

Due to the scarcity of work on UDA for ASTE currently, besides BGCA, we have also leveraged state-of-the-art models in a zero-shot manner within the in-domain setting. These models have been categorized into three groups:

1) **pipeline:** SpanASTE [34] captures the span-to-span interactions during the multi-step filtering process.

²<https://huggingface.co/bert-base-uncased>

Table 2: Results on cross-domain ASTE. The symbol † indicates that the results are obtained from [8].

Methods	R14→L14	R15→L14	R16→L14	Avg.	L14→R14	L14→R15	L14→R16	Avg.	Total Avg.
SpanASTE†	45.83	42.50	40.57	43.00	57.24	49.02	55.77	54.01	48.49
RoBMRC†	43.90	40.19	37.81	40.63	57.13	45.62	52.05	51.60	46.12
COM-MRC	44.87	40.55	38.23	41.22	57.42	47.68	53.11	52.74	46.98
MvP	48.42	41.19	43.78	44.46	63.55	54.30	61.63	59.83	52.15
GAS†	49.57	43.78	45.24	46.20	64.40	56.26	63.14	61.27	53.73
BGCA _{text→label} †	52.55	45.85	46.86	48.42	61.52	55.43	61.15	59.37	53.89
BGCA _{label→text} †	53.64	45.69	47.28	48.87	65.27	58.95	64.00	62.74	55.80
GTS-BERT	43.10±0.45	39.58±0.30	38.34±0.36	40.34±0.37	50.67±0.59	43.14±0.37	49.47±0.57	47.76±0.51	44.05±0.44
+ cTFMT	47.92±0.55	43.86±0.53	42.14±0.45	44.64±0.51	54.86±0.57	47.24±0.60	52.09±0.93	51.40±0.70	48.02±0.61
EMC-GCN	44.23±0.48	41.64±0.34	39.61±0.29	41.83±0.37	53.63±0.51	45.09±0.43	51.10±0.53	49.94±0.49	45.88±0.43
+ cTFMT	49.79±0.70	43.30±0.73	42.53±0.52	45.21±0.65	56.82±0.49	51.74±0.36	54.79±0.41	54.45±0.42	49.83±0.54
BDTF	45.63±0.67	42.37±0.44	40.59±0.63	42.86±0.58	56.25±0.76	49.57±0.63	54.08±0.77	53.50±0.72	48.08±0.65
+ TFMT	54.46±0.78	48.17±0.52	49.35±0.68	50.66±0.67	65.91±0.42	59.40±0.56	64.66±0.55	63.32±0.51	56.99±0.59
$\Delta(Abs.%)$	↑0.82 1.53%	↑2.32 3.97%	↑2.07 4.38%	↑1.79 3.66%	↑0.64 0.98%	↑0.45 0.76%	↑0.66 1.03%	↑0.58 0.92%	↑1.16 2.08%

Table 3: Results on cross-domain AOPE. The symbol † indicates that the results are obtained from [8].

Methods	R14→L14	R15→L14	R16→L14	Avg.	L14→R14	L14→R15	L14→R16	Avg.	Total Avg.
SpanASTE†	51.90	48.15	47.30	49.12	61.97	55.58	63.26	60.27	54.69
RoBMRC†	52.36	46.44	43.61	47.47	54.70	48.68	55.97	53.12	50.29
COM-MRC	53.36	45.79	44.14	47.76	54.58	49.27	56.63	53.49	50.63
GAS†	57.58	53.23	52.17	54.33	64.60	60.26	66.69	63.85	59.09
GCDDA	59.17	54.83	53.19	55.73	64.42	62.11	68.05	64.86	60.30
BGCA _{text→label} †	58.54	54.06	51.99	54.86	64.61	58.74	67.19	63.51	59.19
BGCA _{label→text} †	60.82	55.22	54.48	56.84	68.04	65.31	70.34	67.80	62.37
GTS-BERT	48.14±0.58	44.87±0.43	43.27±0.55	45.43±0.52	56.52±0.50	49.44±0.53	57.54±0.68	54.50±0.57	49.96±0.55
+ cTFMT	55.39±0.45	52.59±0.48	52.94±0.66	53.64±0.53	63.31±0.44	56.75±0.51	64.92±0.43	61.66±0.46	57.65±0.50
EMC-GCN	50.43±0.39	43.62±0.29	43.87±0.31	45.97±0.33	53.04±0.49	49.45±0.38	58.74±0.36	53.74±0.41	49.86±0.37
+ cTFMT	55.81±0.58	52.41±0.70	52.16±0.55	53.46±0.61	64.06±0.35	58.12±0.44	65.31±0.41	62.50±0.40	57.98±0.51
BDTF	50.77±0.48	47.31±0.50	45.38±0.61	47.82±0.53	61.29±0.31	53.92±0.24	61.83±0.41	59.01±0.32	53.42±0.43
+ TFMT	62.50±0.53	55.87±0.75	56.58±0.73	58.32±0.67	69.10±0.69	66.67±0.61	70.62±0.50	68.80±0.60	63.56±0.64
$\Delta(Abs.%)$	↑1.68 2.76%	↑0.65 1.18%	↑2.10 2.60%	↑1.48 1.56%	↑1.06 1.56%	↑1.36 2.08%	↑0.28 0.40%	↑0.90 1.33%	↑1.19 1.91%

2) **generative**: COM-MRC [37] and RoBMRC [16] address the ASTE task in a Machine Reading Comprehension (MRC) manner. GAS [39], GCDDA [13], MvP [9], and BGCA [8] views it as an index generation task.

3) **table-filling**: GTS-BERT [33] and EMC-GCN [4] follow the cell-level paradigm, while BDTF [40] follows the region-level paradigm. In the above methods, the pipeline and table-filling methods are based on BERT, while the generation methods are based on t5-base³.

4.2 Main Results

The main results are reported in Table 2, where the best results are highlighted in bold. According to the results, our findings can be summarized as follows: 1) Our top-performing model (BDTF-TFMT) outperforms the current SOTA baseline BGCA_{label→text} across all tasks, with an average improvement of 1.79% on the restaurant-to-laptop dataset, 0.58% on the laptop-to-restaurant dataset, and

an overall average improvement of 1.16%. This demonstrates the superiority of our BDTF-TFMT model. 2) The TFMT architecture brings about significant improvements to the table-filling model, with respective enhancements of 3.97%, 3.95%, and 8.83% observed for GTS-BERT, EMC-GCN, and BDTF models. This highlights the adaptability of the TFMT framework across different table-filling methods. Additionally, the most notable improvements are observed in the BDTF model, which aligns closely with the two-stage OD paradigm and thereby benefits the most from the region-level consistency component integrated within TFMT. We will further substantiate this finding through subsequent comparative experiments.

Result on the cross-domain AOPE task. In order to provide comprehensive evidence of the TFMT method’s enhancement in cross-domain fine-grained sentiment analysis, we conduct an extra experiment on the aspect opinion pair extraction (AOPE) task with the same dataset. Compared to ASTE, AOPE does not require sentiment polarity extraction; instead, its aim is to extract all (aspect term, opinion term) pairs from a given sentence. To align with the

³<https://huggingface.co/google-t5/t5-base>

AOPE extraction paradigm, we modify the classifier in the original RPC, which determines the sentiment polarity of regions, to a binary classifier that determines the region’s validity.

Table 3 shows the result of the cross-domain AOPE task, we have the following findings: 1) Our best-performing BDTF-TFMT model still comprehensively outperforms all the baselines. Compared to $BGCA_{label \rightarrow text}$, it achieves an average improvement of 1.48% on the restaurant-to-laptop dataset and 0.90% on the laptop-to-restaurant dataset, resulting in an overall average improvement of 1.19%. 2) TFMT also demonstrates superior performance on BDTF (average gain of 10.14%) compared to GTS-BERT (average gain of 7.69%) and EMC-GCN (average gain of 8.12%).

We observe that the findings from Table 3 and Table 2 are consistent, with similar conclusions drawn from the former clearly evident in the latter. These results demonstrate that TFMT provides benefits when tackling cross-domain fine-grained sentiment analysis tasks using OD formulas.

4.3 Computational Costs Analysis

We conduct a comparative experiment on computational costs. The average results are shown in Table 4. From the table, we have the following findings: 1) Comparing the results of the table-filling model before and after using TFMT, as TFMT is a pluggable module designed for guiding cross-domain learning, it is only utilized during the model training process, thus having no impact on the model’s inference speed. Furthermore, since TFMT does not introduce any additional parameters, the model’s parameter count does not change. 2) Compared to $BGCA_{label \rightarrow text}$, BDTF-TFMT exhibits a 19.1% reduction in average training time per sample due to the fact that BGCA spends more time on pseudo-data generation. 3) In comparison to $BGCA_{label \rightarrow text}$, BDTF-TFMT has 27.35% fewer parameters, yet achieves a 2.08% higher F1 score and the average inference time per sample decreases by 65.04%. This can be attributed to the fact that while BGCA operates on a generative backbone (encoder-decoder architecture, e.g., t5), the table-filling method relies on an auto-encoder backbone (encoder-only architecture, e.g., BERT). Consequently, the table-filling method typically entails lower parameter count and inference costs.

4.4 Comparative Study

To investigate the mechanism of the TFMT module, we conduct a comparative experiment, and the results are reported in Table 5. The designs for different domain adaptation components here in Table 5 are as follows:

- **TFMT/cTFMT**: The effect exhibited by TFMT and cTFMT.
- **s-t**: Replace TFMT with the self-training method. Similar to the mean-teacher approach, the model is first pre-trained on the source domain in this setup. Subsequently, the model generates pseudo-labels for the target domain data and incorporates them into the training of the next iteration. Since there are no paired teacher-student models, \mathcal{L}_{uns} and \mathcal{L}_{mmd} cannot be implemented.

Based on the data presented in the table, we can observe that both GTS-BERT and EMC-GCN exhibit lower TFMT performance compared to the self-training method (averaging 1.5% and 1.54%, respectively). In contrast, BDTF shows a significantly higher TFMT

Table 4: Comparison of computational costs for the cross-domain ASTE task. #Params. denotes the parameter count measured in millions (M). #Train and #Infer denote the time cost of training and inference, respectively. “/ms” denotes the average processing time per sample. “ $BGCA_{l \rightarrow t}$ ” denotes $BGCA_{label \rightarrow text}$. Abs.% represents the absolute gain of BDTF+TFMT compared to $BGCA_{l \rightarrow t}$.

Methods	backbone	Avg.F1	#Params.	#Train	#Infer
$BGCA_{l \rightarrow t}$	t5-base	55.80	223M	81.78 /ms	9.21 /ms
GTS-BERT	bert-base	44.05	162M	69.17 /ms	4.41 /ms
+ cTFMT		48.02		72.27 /ms	
EMC-GCN		45.88	121M	49.06 /ms	2.63 /ms
+ cTFMT		49.83		51.13 /ms	
BDTF	162M	53.42	64.24 /ms	3.22 /ms	
+ TFMT		56.99	66.16 /ms		
Abs.%	-	↑2.08%	↓27.35%	↓19.10%	↓65.04%

Table 5: Comparative study.

Method	ASTE		AOPE		Avg.	
	R→L	L→R	R→L	L→R		
GTS-BERT	+ cTFMT	44.64	51.40	53.64	61.66	52.84
	+ s-t	45.91	54.71	54.59	62.15	54.34
EMC-GCN	+ cTFMT	45.21	54.45	53.46	62.50	53.91
	+ s-t	46.96	57.35	52.37	65.13	55.45
BDTF	+ TFMT	50.66	63.32	58.32	68.80	60.28
	+ s-t	48.17	59.65	54.57	65.91	57.08
	+ cTFMT	43.16	52.23	51.49	60.65	51.88

performance than the self-training method (averaging 3.2%). We speculate that this discrepancy arises from the fact that GTS-BERT and EMC-GCN do not align with the two-stage OD paradigm, resulting in the degradation of the region-level consistency component in the original TFMT. To validate our hypothesis, we conduct additional ablation experiments on BDTF, as shown in the table under the “BDTF + cTFMT” row.

- **BDTF + cTFMT**: In this setting, we replace the region-level architecture with a cell-level architecture that is similar to GTS (as depicted in Figure 3(a)). Additionally, we keep the encoding component of the model unchanged. Since we treat each cell in the table as an independent region proposal, the original region-level consistency loss degrades to a cell-level consistency loss: $\mathcal{L}'_{uns} = \frac{1}{|C|} \sum_{k \in C} \|P_k - \tilde{P}_k\|_2^2$, where C denotes the set of cells with high-confidence pseudo labels.

As a consequence of degradation, we can observe that the “cTFMT” configuration showcases a decrease in performance of 8.4% compared to “TFMT” and a decrease of 5.2% compared to “self-training”.

4.5 Error Analysis of pseudo labels

As the quality of pseudo labels significantly influences cross-domain performance, we further investigate the effect of TFMT by examining the generation of pseudo labels in three settings: “TFMT”, “s-t”, and “cTFMT”. The results are presented in Figure 5. We categorize all types of errors into four types:

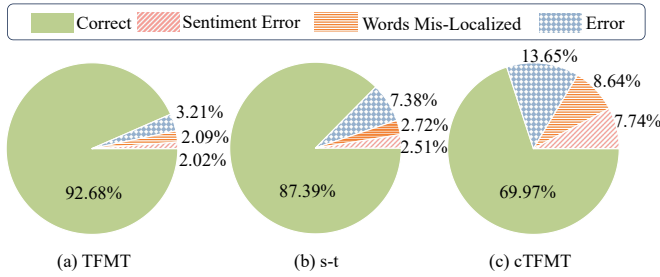


Figure 5: Error analysis of pseudo labels.

- **Correct:** Aspect term, opinion term, and sentiment polarity in the pseudo triplet are all accurate.
- **Sentiment Error:** Aspect term and opinion term are accurate, but sentiment polarity is incorrect.
- **Words Mis-Localized:** Sentiment polarity is correct, but the indexing for aspect term or opinion term is inaccurate.
- **Error:** Aspect term or opinion term is incorrect.

From the experimental results, we have the following findings: 1) The TFMT method effectively reduces the occurrence of incorrect pseudo-labels, resulting in a respective increase of 5.29% and 22.71% in the number of correctly labeled pseudo-labels compared to the s-t method and cTFMT method. 2) Compared to the s-t method, the TFMT method excels mainly in reducing the occurrence of error-type pseudo-labels. This can be attributed to the incorporation of Cross-Domain Consistency \mathcal{L}_{mmd} to constrain domain differences, thereby leading to a noticeable decrease in incorrectly labeled samples. 3) The cTFMT method exhibits more erroneous labels, attributed to its deviation from the object detection paradigm, where Region-level Consistency \mathcal{L}_{uns} and Cross-Domain Consistency \mathcal{L}_{mmd} fail to yield substantial effects.

4.6 More Analysis

Ablation Study. To delve into the module mechanisms within TFMT, we conduct a series of ablation experiments as shown in Table 6, and we have the following observations: 1) When the data augmentation module is removed, the model’s performance does not exhibit a significant decline. This is because the primary aim of data augmentation is to introduce perturbations to the teacher model’s inputs, thereby mitigating the risk of overfitting. 2) When we remove \mathcal{L}_{uns} or \mathcal{L}_{mmd} individually, there is a noticeable decline in model performance. Removing \mathcal{L}_{uns} led to performance drops of 0.93% and 1.49% on the cross-domain ASTE and APOE tasks, respectively. Similarly, removing \mathcal{L}_{mmd} results in performance decreases of 1.65% and 1.13% on the respective tasks. When \mathcal{L}_{uns} and \mathcal{L}_{mmd} both are removed simultaneously, the model’s performance severely degrades. The performance decreases of 3.14% and 3.69% on the respective tasks. This demonstrates the significant role played by both \mathcal{L}_{uns} and \mathcal{L}_{mmd} in TFMT.

Coefficients Study. To investigate the effect of weight coefficients α and β in Equation 13, we present the average performance curves in Figure 6. In Figure 6(a), fixing β at 0.005, we find the optimal performance when $\alpha = 1$. In Figure 6(b), setting α as 1, we observe that the best performance is attained when $\beta = 0.005$.

Table 6: Ablation study. “w/o Aug”, “w/o \mathcal{L}_{uns} ” and “w/o \mathcal{L}_{mmd} ” donate removing the data augmentation module, \mathcal{L}_{uns} and \mathcal{L}_{mmd} , respectively.

	Task	R→L	L→R	Avg.F1	Δ
BDTF-TFMT	ASTE	50.66	63.32	56.99	-
	AOPE	58.32	68.80	63.56	-
-w/o Aug.	ASTE	50.27	62.47	56.37	0.62
	AOPE	58.10	67.82	62.96	0.60
-w/o \mathcal{L}_{uns}	ASTE	49.51	62.61	56.06	0.93
	AOPE	57.34	66.80	62.07	1.49
-w/o \mathcal{L}_{mmd}	ASTE	49.17	61.89	55.53	1.65
	AOPE	57.85	67.01	62.43	1.13
-w/o \mathcal{L}_{uns} & \mathcal{L}_{mmd}	ASTE	48.67	59.02	53.85	3.14
	AOPE	53.69	66.05	59.87	3.69

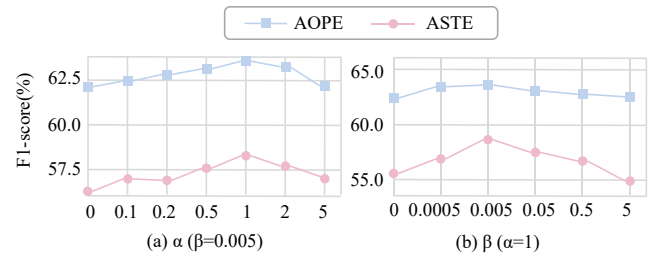


Figure 6: Effects of α and β .

Moreover, excessively large values of β can introduce an ambiguous decision boundary problem [27] and cause a decrease in the model’s classification performance.

5 CONCLUSION

In this paper, we find the inherent consistency between table-filling methods for ASTE and two-stage OD methods. This insight motivates our proposed TFMT method to perceive the ASTE task as an OD task and then leverages region-level consistency constraints for unsupervised training. Experimental results show that the proposed method significantly improves boundary-based paradigms, while the grid-based paradigm leads to a degradation in region-level consistency, further underscoring the alignment of the two tasks. To improve the quality of pseudo labels, we propose an MMD-based consistency constraint, which effectively mitigates the issue of domain discrepancy. Our method achieves state-of-the-art performance while operating under significantly reduced parameter counts and computational costs compared to previous approaches. This makes it an ideal and strong baseline for subsequent work in this field. In future work, we will continue to develop further advanced mean-teacher methods for cross-domain ASTE.

ACKNOWLEDGMENTS

This work was supported by Pilot Projects of Chinese Academy of Sciences (No.E3C0011101), NSFC through grant 62322202, Shijiazhuang Science and Technology Plan Project through grant 231130459A, and Guangdong Basic and Applied Basic Research Foundation through grant 2023B1515120020.

REFERENCES

- [1] Qi Cai, Yingwei Pan, Chong-Wah Ngo, Xinmei Tian, Lingyu Duan, and Ting Yao. 2019. Exploring Object Relation in Mean Teacher for Cross-Domain Detection. In *CVPR2019*.
- [2] Shengcao Cao, Dhiraj Joshi, Liang-Yan Gui, and Yu-Xiong Wang. 2023. Contrastive Mean Teacher for Domain Adaptive Object Detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 23839–23848.
- [3] Chaoqi Chen, Zebiao Zheng, Xinghao Ding, Yue Huang, and Qi Dou. 2020. Harmonizing transferability and discriminability for adapting object detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8869–8878.
- [4] Hao Chen, Zepeng Zhai, Fangxiang Feng, Ruifan Li, and Xiaojie Wang. 2022. Enhanced Multi-Channel Graph Convolutional Network for Aspect Sentiment Triplet Extraction. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2974–2985.
- [5] Meilin Chen, Weijie Chen, Shicai Yang, Jie Song, Xinchao Wang, Lei Zhang, Yunfeng Yan, Donglian Qi, Yueting Zhuang, Di Xie, et al. 2022. Learning domain adaptive object detection with probabilistic teacher. *arXiv preprint arXiv:2206.06293* (2022).
- [6] Yuhua Chen, Wen Li, Christos Sakaridis, Dengxin Dai, and Luc Van Gool. 2018. Domain adaptive faster r-cnn for object detection in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3339–3348.
- [7] Jinhong Deng, Wen Li, Yuhua Chen, and Lixin Duan. 2021. Unbiased mean teacher for cross-domain object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4091–4101.
- [8] Yue Deng, Wenxuan Zhang, Sinno Jialin Pan, and Lidong Bing. 2023. Bidirectional Generative Framework for Cross-domain Aspect-based Sentiment Analysis. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 12272–12285.
- [9] Zhibin Gou, Qingyan Guo, and Yujin Yang. 2023. MvP: Multi-view Prompting Improves Aspect Sentiment Tuple Prediction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (Eds.). Association for Computational Linguistics, Toronto, Canada, 4380–4397.
- [10] Pankaj Gupta, Hinrich Schütze, and Bernt Andrassy. 2016. Table Filling Multi-Task Recurrent Neural Network for Joint Entity and Relation Extraction. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. The COLING 2016 Organizing Committee, 2537–2547.
- [11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.
- [12] Han-Kai Hsu, Chun-Han Yao, Yi-Hsuan Tsai, Wei-Chih Hung, Hung-Yu Tseng, Maneesh Singh, and Ming-Hsuan Yang. 2020. Progressive domain adaptation for object detection. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. 749–757.
- [13] Junjie Li, Jianfei Yu, and Rui Xia. 2022. Generative Cross-Domain Data Augmentation for Aspect and Opinion Co-Extraction. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 4219–4229.
- [14] Yu-Jhe Li, Xiaoliang Dai, Chih-Yao Ma, Yen-Cheng Liu, Kan Chen, Bichen Wu, Zijian He, Kris Kitani, and Peter Vajda. 2022. Cross-domain adaptive teacher for object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7581–7590.
- [15] Zheng Li, Xin Li, Ying Wei, Lidong Bing, Yu Zhang, and Qiang Yang. 2019. Transferable End-to-End Aspect-based Sentiment Analysis with Selective Adversarial Learning. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 4590–4600.
- [16] Shu Liu, Kaiwen Li, and Zuhe Li. 2022. A Robustly Optimized BMRC for Aspect Sentiment Triplet Extraction. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 272–278.
- [17] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan. 2015. Learning Transferable Features with Deep Adaptation Networks. In *Proceedings of the 32nd International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 37)*. PMLR, 97–105.
- [18] Makoto Miwa and Yutaka Sasaki. 2014. Modeling Joint Entity and Relation Extraction with Table Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Alessandro Moschitti, Bo Pang, and Walter Daelemans (Eds.). 1858–1869.
- [19] Jinzhong Ning, Zhihao Yang, Yuanyuan Sun, Zhizheng Wang, and Hongfei Lin. 2023. OD-RTE: A One-Stage Object Detection Framework for Relational Triplet Extraction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Toronto, Canada, 11120–11135.
- [20] Haiyun Peng, Lu Xu, Lidong Bing, Fei Huang, Wei Lu, and Luo Si. 2020. Knowing What, How and Why: A Near Complete Solution for Aspect-Based Sentiment Analysis. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*. AAAI Press, 8600–8607.
- [21] Kun Peng, Lei Jiang, Hao Peng, Rui Liu, Zhengtao Yu, Jiaqian Ren, Zhifeng Hao, and Philip S Yu. 2024. Prompt Based Tri-Channel Graph Convolution Neural Network for Aspect Sentiment Triplet Extraction. In *Proceedings of the 2024 SIAM International Conference on Data Mining (SDM)*. SIAM, 145–153.
- [22] Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, and et al. 2016. SemEval-2016 Task 5: Aspect Based Sentiment Analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*. 19–30.
- [23] Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. 2015. SemEval-2015 Task 12: Aspect Based Sentiment Analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*. 486–495.
- [24] Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. SemEval-2014 Task 4: Aspect Based Sentiment Analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*. 27–35.
- [25] Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. 2015. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*. 91–99.
- [26] Kuniaki Saito, Yoshitaka Ushiku, Tatsuya Harada, and Kate Saenko. 2019. Strong-weak distribution alignment for adaptive object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6956–6965.
- [27] Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, and Tatsuya Harada. 2018. Maximum classifier discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3723–3732.
- [28] Florian Schmidt. 2019. Generalization in generation: A closer look at exposure bias. *arXiv preprint arXiv:1910.00292* (2019).
- [29] Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing*. 1631–1642.
- [30] Antti Tarvainen and Harri Valpola. 2017. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems* 30 (2017).
- [31] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).
- [32] Wenya Wang and Sinno Jialin Pan. 2018. Recursive Neural Structural Correspondence Network for Cross-domain Aspect and Opinion Co-Extraction. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2171–2181.
- [33] Zhen Wu, Chengcan Ying, Fei Zhao, Zhifang Fan, Xinyu Dai, and Rui Xia. 2020. Grid Tagging Scheme for Aspect-oriented Fine-grained Opinion Extraction. In *Findings of EMNLP 2020*. 2576–2585.
- [34] Lu Xu, Yew Ken Chia, and Lidong Bing. 2021. Learning Span-Level Interactions for Aspect Sentiment Triplet Extraction. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 4755–4766.
- [35] Lu Xu, Hao Li, Wei Lu, and Lidong Bing. 2020. Position-Aware Tagging for Aspect Sentiment Triplet Extraction. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2339–2349.
- [36] Jianfei Yu, Chengcong Gong, and Rui Xia. 2021. Cross-Domain Review Generation for Aspect-Based Sentiment Analysis. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli (Eds.). 4767–4777.
- [37] Zepeng Zhai, Hao Chen, Fangxiang Feng, Ruifan Li, and Xiaojie Wang. 2022. COM-MRC: A Context-Masked Machine Reading Comprehension Framework for Aspect Sentiment Triplet Extraction. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang (Eds.). Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 3230–3241.
- [38] Kai Zhang, Qi Liu, Hao Qian, Biao Xiang, Qing Cui, Jun Zhou, and Enhong Chen. 2023. EATN: An Efficient Adaptive Transfer Network for Aspect-Level Sentiment Analysis. *IEEE Transactions on Knowledge and Data Engineering* 35, 1 (2023), 377–389.
- [39] Wenxuan Zhang, Xin Li, Yang Deng, Lidong Bing, and Wai Lam. 2021. Towards Generative Aspect-Based Sentiment Analysis. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International*

- Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. 504–510.
- [40] Yice Zhang, Yifan Yang, Yihui Li, Bin Liang, Shiwei Chen, Yixue Dang, Min Yang, and Ruifeng Xu. 2022. Boundary-Driven Table-Filling for Aspect Sentiment Triplet Extraction. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. 6485–6498.
- [41] Xusheng Zhao, Hao Peng, Qiong Dai, Xu Bai, Huailiang Peng, Yanbing Liu, Qinglang Guo, and Philip S Yu. 2024. Rdgcn: Reinforced dependency graph convolutional network for aspect-based sentiment analysis. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*. 976–984.
- [42] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*. 2223–2232.