PRAF: A Part-of-speech-Ruled Adaptive Framework for Aspect Sentiment Triplet Extraction

Zhiyang Yu
Computer and Artificial Intelligence
Shandong Normal University
Jinan, China
Yzyang824@163.com

Guangjin Wang
Information Engineering
Qingdao Engineering Vocational College
Qingdao, China
guangjinwang24@163.com

Fuyong Xu
Computer and Artificial Intelligence
Shandong Normal University
Jinan, China
fyxu@outlook.com

Zhipeng Wang
Computer and Artificial Intelligence
Shandong Normal University
Jinan, China
2023028051@stu.sdnu.edu.cn

Ru Wang
Computer and Artificial Intelligence
Shandong Normal University
Jinan, China
ruwang0929@gmail.com

Peiyu Liu*
Computer and Artificial Intelligence
Shandong Normal University
Jinan, China
liupy@sdnu.edu.cn

***Abstract*—Aspect Sentiment Triplet Extraction (ASTE) is a fine-grained sentiment analysis task that aims to extract triplets consisting of aspect terms, opinion terms, and their corresponding sentiment polarities. Recent approaches have explored the integration of linguistic knowledge such as dependency syntax and semantic roles. However, these methods largely overlook the inherent part-of-speech (POS) rules exhibited by aspect and opinion terms, and fail to account for the interference caused by non-target POS categories. In this paper, we propose a Part-of-speech-Ruled Adaptive Framework (PRAF) that explicitly models and leverages POS rules for ASTE. Specifically, we design a POS-Powered Network (P-PN) to predict POS distributions, capturing the intrinsic POS regularities of target terms. To suppress noise from irrelevant words, we introduce a Rules-aware Attention Filtering Layer (RAFiL) that adaptively adjusts the contribution of POS features during attention computation. Experiments on four benchmark datasets demonstrate that PRAF achieves new state-of-the-art results, validating the effectiveness of leveraging POS rules for triplet extraction.**

***Index Terms*—ASTE, Natural Language Processing, Fine-grained Sentiment Analysis, Part-of-speech.**

I. Introduction

Aspect-Based Sentiment Analysis (ABSA) [1], as a critical task in Natural Language Processing (NLP), aims to provide fine-grained analysis of user reviews and identify the corresponding sentiment orientations. Three main subtasks have been extensively studied: Aspect Term Extraction (ATE) [2], [3], Opinion Term Extraction (OTE) [4], and Aspect-level Sentiment Classification (ASC) [5]. Building upon these subtasks, Aspect Sentiment Triplet Extraction (ASTE) [6] has emerged as a more comprehensive task, requiring the joint extraction of aspect term, opinion term, and their sentiment polarity as a triplet, as illustrated in Fig. 1(c).

Early work on ASTE adopted pipeline approaches [6] that extract aspects and opinions separately before pairing them, which suffers from error propagation. Subsequent joint learning methods [7] mitigated this issue by sharing parameters between subtasks. Recent advances have shifted toward end-to-end architectures, including table-filling methods [8], [9] that model word-pair interactions by filling 2D tables. These methods increasingly focus on the integration of syntactic relations and semantic roles [10]–[13].

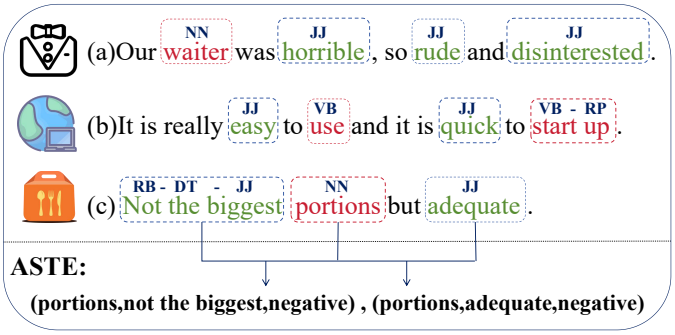

Fig. 1. Examples of ASTE task. Aspect terms are labeled in red and opinion terms in green, with the corresponding part-of-speech attached above. The triplets below show the extraction results for sentence (c).

Although contemporary researches have shown that incorporating dependency structures and semantic roles significantly improves the performance for ASTE, they still suffer from some weaknesses: (1) They do not fully consider inherent characteristics of part-of-speech rules: POS tags (e.g., JJ for adjectives, NN for nouns, RB for adverbs) reflect important linguistic rules embedded in reviews. (2) They lack comprehensive mining of specific information to explore the different part-of-speech patterns in aspect and opinion terms, while ignoring the potential influence of certain non-target part-of-speech. For example, opinion terms resulted from sentence (a) and (b) in Fig.1 have different part-of-speech while the part-of-speech of aspect terms are both “JJ”.

In this paper, we propose a novel Part-of-speech-Ruled Adaptive Framework (PRAF) to remedy the above problems.

*Corresponding author.

Unlike previous methods, PRAF systematically models and applies POS rules for ASTE. The framework comprises two key innovations: (1) POS-Powered Network (P-PN): A dedicated module that predicts the POS distribution for each word, learning which POS categories are indicative of aspect or opinion terms. This enables the model to capture task-relevant POS regularities beyond static tag embeddings. (2) Rules-aware Attention Filtering Layer (RAFiL): A dual-path attention mechanism that leverages the POS distributions from P-PN to adaptively adjust attention weights, suppressing the influence of non-target words and enhancing focus on candidate aspect and opinion terms. The results of extensive comparative experiments show that our method exhibits more convincing and superior performance compared with existing methods.

II. Methodology

A. Problem statement

For the ASTE task, the ultimate goal is to extract all identifiable sentiment triplets $T = (a_i, o_i, s_i)$ from a given sentence $W = w_1, w_2, w_3, ..., w_n$, where n is the length of the sentence, a and o represent the aspect term and opinion term in the review sentence respectively, s denotes the corresponding sentiment including three polarities: negative, positive, and neutral. For the datasets used, $i \geq 1$ indicates that each sentence contains at least one triplet.

B. Framework

The overall framework of our proposed module is shown in Fig.2. It consists of four main parts, details of each part are presented in subsequent subsections.

1) Embedding layer: For a given sentence, the pre-trained language model BERT [14] is used to obtain a context-hidden representation of the sentence. Moreover, we employ an alignment function, denoted as $\mathrm{align}(\cdot)$, to aggregate the subword representations from BERT into word-level representations. For a given word w_i, its representation h_i is obtained by averaging the vectors of its constituent subwords:

$$H_b = \text{BERT}([\text{CLS}], w_1, w_2, \ldots, w_n, [\text{SEP}]) \in \mathbb{R}^{T \times d_b}, \quad (1)$$

$$h_i = \text{align}(H_b, w_i) = \frac{1}{|S_i|} \sum_{t \in S_i} H_b[t]. \quad (2)$$

where T is the subword sequence length and d_b denotes the hidden state dimension of BERT, S_i is the index set of the subwords corresponding to word w_i in the subword sequence. After aligning all words, we obtain the word-level representation matrix, where n is the number of words in the sentence consistent with the original sentence length:

$$H = [h_1, h_2, \ldots, h_n] \in \mathbb{R}^{n \times d_b}, \quad (3)$$

2) Representation of part-of-speech information and POS-Powered Network: We use the context representation H_b obtained from BERT encoding and further integrate part-of-speech rules. First, let $\mathcal{P} = \{\text{tag}_1, \text{tag}_2, \ldots, \text{tag}_{N_{\text{pos}}}\}$ be the complete set of POS tags, where $N_{\text{pos}} = |\mathcal{P}|$ is the fixed total number of POS categories. And then, POS tags are obtained using the SpaCy [1] toolkit. We adopt the Universal POS tagset with an additional "OTHER" category for non-target words, resulting in $N_{\text{pos}} = 18$. Based on POS sequences, target words (aspect and opinion terms) receive specific POS labels, while non-target words are labeled as "OTHER"(18) to reduce noise interference. POS tags are mapped to dense vectors via a trainable embedding matrix:

$$h_i^{\text{pos}} = E^{\text{pos}}(p_i) \in \mathbb{R}^{d_{\text{pos}}}, \quad (4)$$

where $E^{\text{pos}} \in \mathbb{R}^{N_{\text{pos}} \times d_{\text{pos}}}$ is the POS embedding matrix, and $d_{\text{pos}} = 64$ is the POS embedding dimension. This dimension has been empirically validated to achieve the optimal balance between expressive power and computational efficiency.

Contextual and POS representations are concatenated to form enhanced representations:

$$h_i^{\text{enh}} = [h_i; h_i^{\text{pos}}] \in \mathbb{R}^{d_b + d_{\text{pos}}}. \quad (5)$$

P-PN takes the enhanced representation and predicts the POS distribution:

$$r_i^{\text{pos}} = \text{Linear}\left(W^{\text{pos}} h_i^{\text{enh}} + b^{\text{pos}}\right) \in \mathbb{R}^{N_{\text{pos}}}, \quad (6)$$

where $W^{\text{pos}} \in \mathbb{R}^{N_{\text{pos}} \times (d_b + d_{\text{pos}})}$ and $b^{\text{pos}} \in \mathbb{R}^{N_{\text{pos}}}$ are learnable parameters.

Let $l_i^{\text{pos}} \in \{1, 2, \ldots, N_{\text{pos}}\}$ denote the index of the true POS tag for word w_i in the tagset $\mathcal{P}$. The POS distribution loss is computed as:

$$\mathcal{L}_p = -\sum_{i=1}^{n} \log \frac{\exp\left(r_i^{\text{pos}}[l_i^{\text{pos}}]\right)}{\sum_{k=1}^{N_{\text{pos}}} \exp\left(r_i^{\text{pos}}[k]\right)} \quad (7)$$

3)Rules-aware Attention Filtering Layer: Learning POS distributions alone is insufficient to ensure effective application of rules. Standard attention mechanisms treat all POS features equally, failing to distinguish target words from non-target ones. To address this, we use RAFiL leverages the POS distributions predicted by P-PN to dynamically adjust attention weights.

For each word w_j, we compute its POS relevance score using the predicted distribution r_j^{pos} and a learnable target mask $m_{\text{target}} \in \mathbb{R}^{N_{\text{pos}}}$:

$$p_j = \sigma\left((r_j^{\text{pos}})^\top m_{\text{target}}\right),$$

where σ is the sigmoid function. This score $p_j \in [0, 1]$ indicates the likelihood that w_j belongs to a target-relevant POS category.

For a candidate aspect word w_i, the standard attention score is first computed as:

$$e_{ij}^a = \frac{\left(W_q^a h_i^{\text{enh}}\right)^\top \left(W_k^a h_j^{\text{enh}}\right)}{\sqrt{d_k}}.$$

where $W_q^a, W_k^a \in \mathbb{R}^{(d_b + d_{\text{pos}}) \times d_k}$ are learnable projection matrices. We then modulate it using the POS relevance score:

[1]https://spacy.io

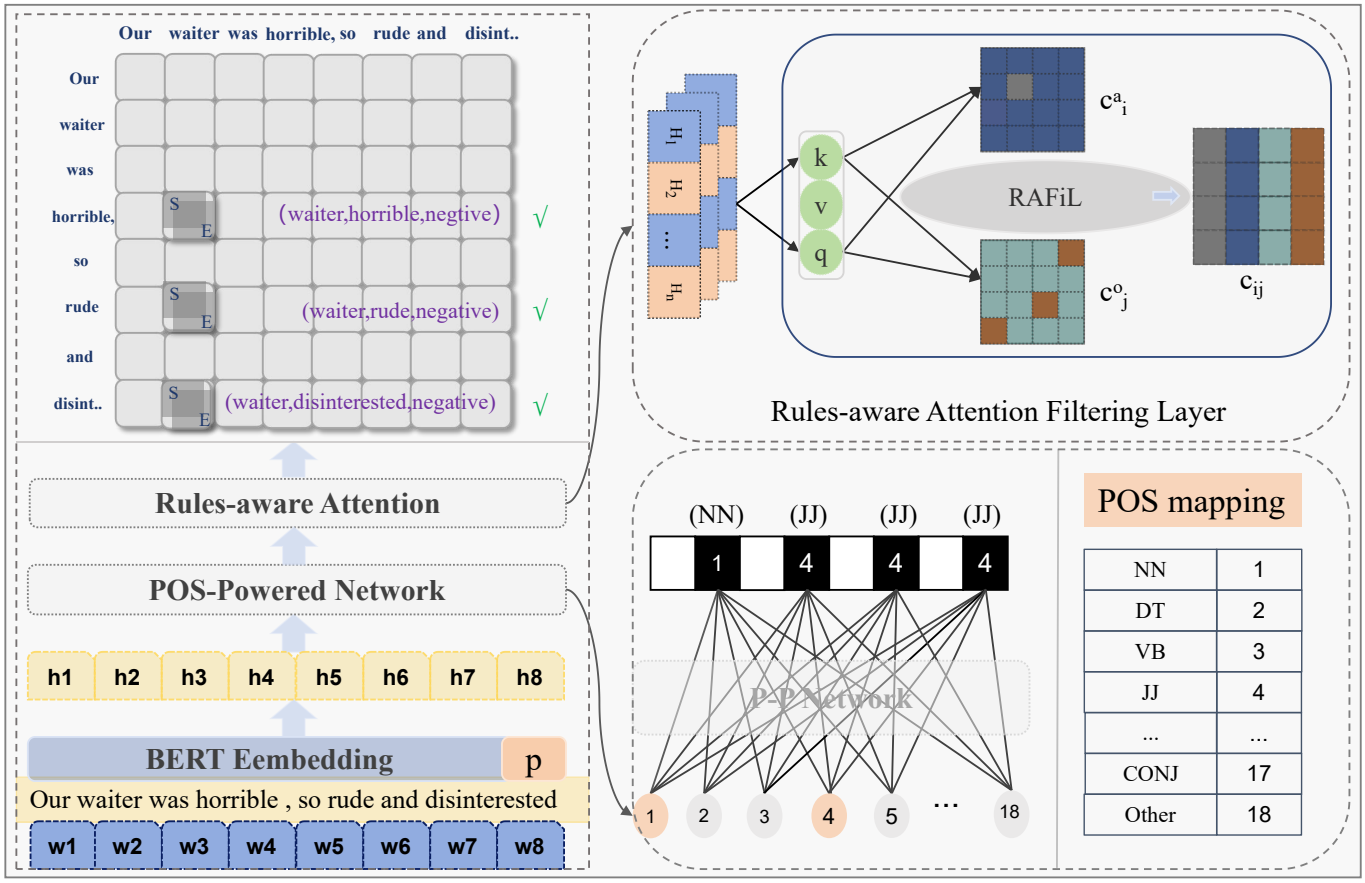

Fig. 2. The overall architecture of the model.

$$\tilde{e}^a_{ij} = e^a_{ij} + \gamma \cdot \log(p_j + \epsilon),$$

where γ is a learnable scalar controlling modulation strength. The attention weights are obtained via softmax:

$$\alpha^a_{ij} = \frac{\exp(\tilde{e}^a_{ij})}{\sum_{k=1}^{n} \exp(\tilde{e}^a_{ik})}.$$

The aspect-specific representation is:

$$c^a_i = \sum_{j=1}^{n} \alpha^a_{ij} \left(W^a_v h^{\text{enh}}_i\right) \in \mathbb{R}^{d_k}.$$

where $W^a_v \in \mathbb{R}^{(d_b+d_{\text{pos}})\times d_k}$ is a learnable projection matrix. For opinion terms, we apply the same mechanism with separate parameters W^o_q, W^o_k, W^o_v and a distinct target mask m^o_{target}, yielding c^o_j analogously.

This design ensures that POS distribution information directly and explicitly influences attention patterns, enabling the model to focus on linguistically plausible candidates while filtering non-target noise.

4)Extraction Layer: For each word w_i, RAFiL produces two distinct representations: c^a_i when w_i serves as a candidate aspect term, and c^o_i when it serves as a candidate opinion term. This dual representation design enables the model to evaluate every possible aspect-opinion pairing in the sentence. When assessing whether the word pair (w_i, w_j) constitutes a valid sentiment triplet, we concatenate the aspect representation of w_i and the opinion representation of w_j:

$$c_{ij} = [c^a_i; c^o_j] \in \mathbb{R}^{2d_k}. \tag{8}$$

This relation vector synthesizes "the contextual understanding of word w_i as a potential aspect term" and "the contextual understanding of word w_j as a potential opinion term," providing comprehensive evidence for triplet judgment.

Concatenating the original word representations with this relation vector, we obtain the final relation representation through a nonlinear transformation:

$$r_{ij} = \text{ReLU}\left(W_r[h_i; h_j; c_{ij}] + b_r\right) \in \mathbb{R}^{d_r}, \tag{9}$$

where $W_r \in \mathbb{R}^{d_r\times(2d_b+2d_k)}$, $d_r = 256$ is the relation representation dimension.

The region detection layer employs two binary classifiers to predict the boundaries(S and E) of aspect and opinion terms, and finally a three-way classifier predicts the sentiment polarity for each candidate pair.

5)Loss Function: The total loss is a weighted sum of four components:

$$\mathcal{L} = \mathcal{L}_a + \mathcal{L}_o + \mathcal{L}_s + \lambda\mathcal{L}_p, \tag{10}$$

TABLE I
DATA STATISTICS ON DATASET V2.

	Rest14				Lap14				Rest15				Rest16			
	#S	#A	#O	#T	#S	#A	#O	#T	#S	#A	#O	#T	#S	#A	#O	#T
Train	1266	2051	2061	2338	906	1280	1254	1460	605	862	935	1013	857	1198	1300	1394
Dev	310	500	497	577	219	295	302	346	148	213	236	249	210	296	319	339
Test	492	848	844	994	328	463	466	543	322	432	460	485	326	452	474	514

TABLE II
MAIN RESULTS ON ASTE FOR V2, WHERE THE STATE-OF-THE-ART F1 RESULTS IN BOLD AND THE SECOND-BEST RESULTS ARE UNDERLINED

Category	Methods	Lap14			Rest14			Rest15			Rest16		
		P.	R.	F1	P.	R.	F1	P.	R.	F1	P.	R.	F1
Pipeline	peng-two-stage	37.38	50.38	42.87	43.24	63.66	51.46	48.07	57.51	52.32	46.96	64.24	54.21
End-to-end	GTS-BERT	57.82	51.32	54.36	67.76	67.29	67.50	62.59	57.94	60.15	66.08	66.91	67.93
	EMC-GCN	61.70	56.26	58.81	71.21	72.39	71.78	61.54	62.47	61.93	65.62	71.30	68.33
	BDTF	68.94	55.97	61.74	75.53	73.24	74.35	68.76	63.71	66.12	71.44	73.13	72.27
	STAGE	71.98	53.86	61.58	78.58	69.58	73.76	73.63	57.90	64.79	77.67	68.44	72.75
	SimSTAR	66.46	58.23	62.07	76.23	71.63	73.84	71.71	59.59	65.09	72.02	74.12	73.06
	SA-Transformer	61.28	48.98	54.44	70.76	65.85	68.22	62.82	58.31	60.48	72.01	62.87	67.13
	MiniConGTS	66.82	60.68	<u>63.61</u>	76.10	75.08	75.59	66.50	63.86	65.15	75.52	74.14	<u>74.83</u>
	SAAG	62.63	56.88	59.62	76.03	70.75	73.30	61.51	62.26	61.38	69.26	71.51	70.37
	BTF-CCL	70.66	57.31	63.29	80.41	71.83	<u>75.88</u>	71.66	64.12	<u>67.68</u>	72.56	75.10	73.80
Ours	PRAF	73.52	58.49	**66.25**	78.60	75.34	**76.94**	73.45	64.68	**68.77**	78.34	74.77	**76.62**

where the boundary detection losses for aspect, opinion terms, and the sentiment classification loss are:

$$\mathcal{L}_a = -\sum_{i=1}^{n}\sum_{j=i}^{n} y_{ij}^{a} \log \hat{y}_{ij}^{a}, \tag{11}$$

$$\mathcal{L}_o = -\sum_{i=1}^{n}\sum_{j=i}^{n} y_{ij}^{o} \log \hat{y}_{ij}^{o}, \tag{12}$$

$$\mathcal{L}_s = -\sum_{(i,j)\in\mathcal{T}} \log P(\hat{s}_{ij} = s_{ij}). \tag{13}$$

Here, $\mathcal{T}$ is the set of valid word pairs, s_{ij} is the ground-truth sentiment label, and $P(\hat{s}_{ij} = s_{ij})$ is the predicted probability of the correct sentiment polarity. The hyperparameter $\lambda = 0.55$ balances the contribution of the POS loss. All loss components are optimized jointly during training.

III. EXPERIMENTS

A. Datasets and Metrics

We conduct experiments on ASTE-Data-V2 (V2) [7], which includes four public benchmark datasets on review data from the restaurant and laptop domains: REST16, REST15, REST14, and LAP14, sourced from SemEval 2016 Task 5 [15], SemEval 2015 Task 12 [16], and SemEval 2014 Task 4 [1]. The data statistics are shown in Table I.

B. Experimental Setting

The model uses BERT as the encoder, which has 110 million parameters, 12 attention heads, and consists of 12 transformer blocks with a hidden layer dimension size of 768. For the representation of part-of-speech labels, we use the SpaCy library to obtain part-of-speech tags. The model is trained for 10 epochs with a batch size of 8 and a learning rate of 2e-5. We run 5 times with different random seeds to select the best model parameters, and final report the mean score based on the F1 score.

C. Baselines

We compare our method with two categories of baselines: **Pipeline approaches:** Peng-two-stage [6] pioneeringly proposes and handles the ASTE task in a pipeline model with two stages to do the extraction and pairing tasks respectively; **End-to-end approaches:**The joint model provides a more diverse end-to-end approaches for ASTE task. Among them, SimSTAR [17] emphasizes the word-to-word relationship through tagging scheme; EMC-GCN [18], BDTF [9], GTS-BERT [8], and STAGE [19] combine the characteristic of table for filling and extracting the sentiment triples explicitly; MiniConGTS [20] minimizes the labeling scheme to improve and exploit the pre-trained representation; SA-Transformer [10] and SAAG [11] incorporate information such as syntax and semantics to enhance the word representation, which improves the accuracy of extraction. BTF-CCL [21] enhances sentence-level semantic consistency based on contrastive learning.

D. Main results

Table II respectively present the evaluation results compared to baselines on different datasets. The details of the analysis are as follows: (1)The superior performance of models such as SA-Transformer [10], and EMC-GCN [18] underscores the

importance of linguistic knowledge inherent in words. Our model further capitalizes on linguistic features by focusing on a more comprehensive exploration of effective part-of-speech rules, and thus exhibits a more leading performance to extract triplets. (2) PRAF leverages the strengths of table-filling and introduces an adaptive part-of-speech weight allocation mechanism to further improve the accuracy of locating aspect terms and opinion terms. The above two points are substantiated by the most comprehensive performance metric F1, where the F1 scores of V2 on LAP14, REST14, REST15, and REST16 increased by 2.64%, 1.06%, 1.09%, 1.79% respectively, demonstrating the effectiveness of the part-of-speech information captured by PRAF for the ASTE task.

TABLE III
RESULTS OF ABLATION EXPERIMENTS ON THE V2 FOR TWO COMPONENTS.

Model	Lap14		Rest14		Rest15		Rest16	
	F1	↓	F1	↓	F1	↓	F1	↓
Ours (Full Model)	66.25	-	76.94	-	68.77	-	76.62	-
w/o P-PN	63.57	2.68	74.02	2.92	66.73	2.04	74.58	2.04
w/o RAFiL	63.22	3.03	72.67	4.27	66.15	2.62	74.14	2.48

E. Ablation study

To further examine the contributions of different modules, such as the P-PN and the RAFiL module, we conduct ablation experiments on the four public datasets of V2. As indicated by the F1 experimental metrics shown in Table III, the ablation of different modules resulted in a certain degree of decline in model performance. Specifically, we remove the corresponding modules while ensuring the consistency of the remaining components. (1) For P-PN, without P-PN leads to a significant performance degradation on each dataset, especially on Rest14 which has the highest number of sentences under review. This suggests that the part-of-speech rules learned by our model are valid and effective to improve the accuracy of sentiment triplet extraction as the amount of data increases. (2) For the absence of RAFiL, the F1 scores on the four datasets are reduced by an average of 3.10%. The decrease in performance highlights the importance of RAFiL, suggesting that it helps the model to filter out irrelevant words and enhance overall performance.

F. Performance on ATE and OTE tasks

We conduct further evaluation of the PRAF on the Aspect Term Extraction (ATE) and Opinion Term Extraction (OTE) tasks. As shown in Figure3, compared to current methods such as DTF and PBLUN, PRAF demonstrates superior F1 metrics in both subtasks. Compared to the PBLUN method, the F1 scores on ATE and OTE improved by an average of 2.28% and 1.36%, while improving by an average of 2.50% and 1.09% over DTF. This result indicates that our method achieves a better part-of-speech exploration scope and emphasizes the more positive impact of the effective part-of-speech information mined on the recognition of aspect terms and opinion terms. Additionally, the performance improvement

Fig. 3. Comparative experimental results for the sub-tasks ATE and OTE tasks, where results of PBLUN [22] and DTF [23] are derived from the corresponding article data.

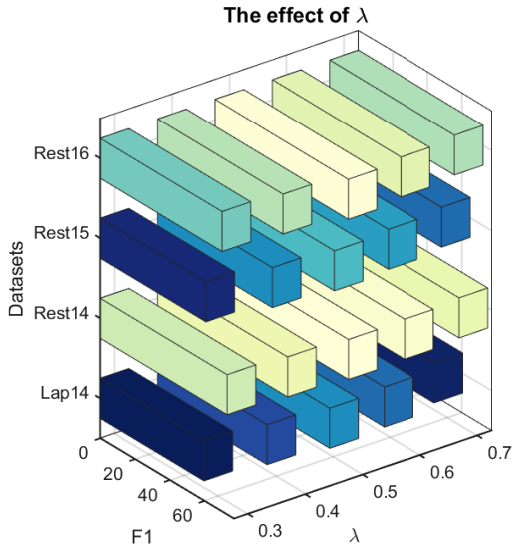

Fig. 4. The effect of λ, with peaks marked by squares.

of PRAF on the ATE and OTE subtasks is slightly lower than that on the ASTE task, which suggests that our method is influenced by the granularity of interaction between different tasks, achieving better interaction as task complexity increases.

G. More Analysis

1) Coefficients Study: As illustrated in Figure 4, we conduct a systematic parameter study to determine the optimal weight-

ing coefficient λ in Equation 10. The parametric analysis reveals a characteristic parabolic relationship between λ values and V2 metric performance, with the curve exhibiting a distinct maximum in the 0.3-0.7 parameter range. Through dense sampling with 0.25 intervals in the critical 0.45-0.55 subrange, we identified 0.55 as the empirically optimal value that maximizes V2 performance. This optimal zone represents a critical balance point-values below 0.3 induce signal undersaturation (12.3% V2 degradation at λ=0.2), while values exceeding 0.7 cause response oversaturation (9.7% performance drop at λ=0.8). This nonlinear response underscores the importance of precise λ calibration.

2) POS Category Analysis: To understand how POS information contributes to model performance, we analyze POS distributions, attention modulation, and error patterns. (1) We first examine the ground-truth POS distribution of aspect and opinion terms across datasets. Among aspect terms, NN(87.5%) and JJ(6.2%) have the highest POS frequency, while opinion terms are primarily JJ(72.2%) and RB(15.4%). This aligns with linguistic intuition and validates our choice of target POS masks. (2) We compute the average attention weight difference between words with high POS relevance and low relevance. Across all datasets, high-relevance words receive 23.7% higher attention on average, indicating that RAFiL effectively amplifies target-candidate words. (3) We further analyze misclassified examples, the most common error cases involve: ambiguous adjectives, like "interesting" can serve as both aspect (e.g., "the interesting part") and opinion ("it is interesting"), causing confusion. These findings suggest that future work could focus on handling POS ambiguity and improving compound term recognition.

IV. Conclusions

In this paper, we proposed PRAF, a Part-of-speech-Ruled Adaptive Framework for Aspect Sentiment Triplet Extraction. We identified that existing methods underutilize POS rules and introduced two key innovations: P-PN, which learns task-specific POS distributions to capture inherent patterns of aspect and opinion terms; RAFiL, which adaptively modulates attention weights using POS relevance scores to suppress non-target words. Experiments on four benchmark datasets demonstrate that PRAF achieves new state-of-the-art results, with average F1 improvements of 1.6% over the strongest baseline. Ablation studies confirm the contribution of each module, and POS analysis reveals that the model effectively amplifies attention on target POS categories. Future research could explore self-supervised part-of-speech tagging to reduce reliance on external data sources, as well as extend these methods to cross-lingual sentiment analysis.

References

[1] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "Semeval-2014 task 4: Aspect based sentiment analysis," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, 2014.

[2] J. Xu, J. Xie, Y. Cai, Z. Lin, H. Leung, Q. Li, and T. Chua, "Context-aware dynamic word embeddings for aspect term extraction," *IEEE Trans. Affect. Comput.*, vol. 15, pp. 144–156, 2023.

[3] Q. Wang, Z. Wen, Q. Zhao, M. Yang, and R. Xu, "Progressive self-training with discriminator for aspect term extraction," in *Proceedings of the Conference on EMNLP*, 2021, pp. 257–268.

[4] S. Wu, H. Fei, Y. Ren, B. Li, F. Li, and D. Ji, "High-order pair-wise aspect and opinion terms extraction with edge-enhanced syntactic graph convolution," *IEEE ACM Trans. Audio Speech Lang. Process.*, vol. 29, pp. 2396–2406, 2021.

[5] J. Zhou, Y. Lin, Q. Chen, Q. Zhang, X. Huang, and L. He, "Causalabsc: Causal inference for aspect debiasing in aspect-based sentiment classification," *IEEE ACM Trans. Audio Speech Lang. Process.*, vol. 32, pp. 830–840, 2023.

[6] H. Peng, L. Xu, L. Bing, F. Huang, W. Lu, and L. Si, "Knowing what, how and why: A near complete solution for aspect-based sentiment analysis," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, pp. 8600–8607.

[7] L. Xu, H. Li, W. Lu, and L. Bing, "Position-aware tagging for aspect sentiment triplet extraction," in *Proceedings of the Conference on EMNLP*, 2020, pp. 2339–2349.

[8] Z. Wu, C. Ying, F. Zhao, Z. Fan, X. Dai, and R. Xia, "Grid tagging scheme for aspect-oriented fine-grained opinion extraction," in *Findings of the Conference on EMNLP*, 2020, pp. 2576–2585.

[9] Y. Zhang, Y. Yang, Y. Li, B. Liang, S. Chen, Y. Dang, M. Yang, and R. Xu, "Boundary-driven table-filling for aspect sentiment triplet extraction," in *Proceedings of the Conference on EMNLP*, 2022, pp. 6485–6498.

[10] L. Yuan, J. Wang, L. Yu, and X. Zhang, "Encoding syntactic information into transformers for aspect-based sentiment triplet extraction," *IEEE Trans. Affect. Comput.*, vol. 15, no. 2, pp. 722–735, 2024.

[11] X. Sun, J. Qi, Z. Zhu, M. Li, H. Pei, and J. Meng, "Senticnet and abstract meaning representation driven attention-gate semantic framework for aspect sentiment triplet extraction," *EAAI*, p. 109625, 2025.

[12] G. Xu, Z. Yang, B. Xu, L. Luo, and H. Lin, "Span-based syntactic feature fusion for aspect sentiment triplet extraction," *Inf. Fusion*, vol. 120, p. 103078, 2025.

[13] G. Su, M. Wu, Z. Huang, Y. Zhang, T. Wang, Y. Hu, and Y. Sha, "Refine, align, and aggregate: Multi-view linguistic features enhancement for aspect sentiment triplet extraction," in *Findings ACL, 2024*, 2024, pp. 3212–3228.

[14] J. D. M.-W. C. Kenton and L. K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the Conference of NAACL-HLT*, 2019, pp. 4171–4186.

[15] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, A. Mohammad, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clercq *et al.*, "Semeval-2016 task 5: Aspect based sentiment analysis," *Proceedings of SemEval*, pp. 19–30, 2016.

[16] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, and I. Androutsopoulos, "Semeval-2015 task 12: Aspect based sentiment analysis," in *Proceedings of the SemEval@NAACL-HLT*, 2015, pp. 486–495.

[17] D. Li, Z. Yang, Y. Lan, Y. Zhang, H. Zhao, and G. Zhao, "Simple approach for aspect sentiment triplet extraction using span-based segment tagging and dua extractors," in *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*, 2023, pp. 2374–2378.

[18] H. Chen, Z. Zhai, F. Feng, R. Li, and X. Wang, "Enhanced multi-channel graph convolutional network for aspect sentiment triplet extraction," in *ACL (Volume 1: Long Papers)*, 2022, pp. 2974–2985.

[19] S. Liang, W. Wei, X. Mao, Y. Fu, R. Fang, and D. Chen, "STAGE: span tagging and greedy inference scheme for aspect sentiment triplet extraction," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023, pp. 13 174–13 182.

[20] Q. Sun, L. Yang, M. Ma, N. Ye, and Q. Gu, "Minicongts: A near ultimate minimalist contrastive grid tagging scheme for aspect sentiment triplet extraction," in *Proceedings of the Conference on EMNLP*, 2024, pp. 2817–2834.

[21] Q. Li, W. Wen, and J. Qin, "Boundary-driven table-filling with cross-granularity contrastive learning for aspect sentiment triplet extraction," in *Conference on ICASSP*, 2025.

[22] Y. Li, Q. He, and L. Yang, "Part-of-speech based label update network for aspect sentiment triplet extraction," *J. King Saud Univ. Comput. Inf. Sci.*, p. 101908, 2024.

[23] B. Wang, G. Wang, and P. Liu, "DTS: A decoupled task specificity approach for aspect sentiment triplet extraction," *Expert Syst. Appl.*, vol. 273, p. 126759, 2025.