Postprint: Comparative Sentence Processing in Attention-Based Sentiment Analysis

Authors: Zhang Rong, Liu Yuan, Li Yang

Date: 2022-05-10T11:22:55Z

Abstract

Aspect-level sentiment analysis aims to determine the sentiment polarity toward specific aspects in reviews; however, few studies have investigated the impact of complex sentences on sentiment classification. Based on this, we propose an aspect-level sentiment analysis model based on BERT and a self-attention network with relative position. First, a dynamic weighted sampling method is employed to address the scarcity of contrastive sentences, enabling the model to learn more feature information from such sentences. Second, a dual-head self-attention network is utilized to extract feature representations with relative position, which are jointly trained with the absolute position feature representations obtained from the pre-trained model. Finally, label balancing techniques are applied to regularize the model, stabilizing its identification of neutral samples. The model is evaluated on SemEval 2014 Task 4 Sub Task 2, demonstrating improvements in both Accuracy and Macro-f1 metrics on two datasets. Experimental results show that the model is effective in contrastive sentence classification and also outperforms other baseline models on the entire test set.

Full Text

Preamble

Handling Contrastive Sentences in Sentiment Analysis with Attention Networks

Zhang Rong[1],[2]†, Liu Yuan[2], Li Yang[1] 1. School of Internet of Things Engineering, Jiangsu Vocational College of Information Technology, Wuxi, Jiangsu 214000, China 2. School of Artificial Intelligence & Computer, Jiangnan University, Wuxi, Jiangsu 214000, China

Abstract: Aspect-level sentiment analysis aims to determine the sentiment polarity towards specific aspects in reviews. However, little research has investigated the influence of complex sentences on sentiment classification. To address

this gap, we propose an aspect-level sentiment analysis model based on BERT and a self-attention network with relative position encoding. First, we employ a dynamic weighted sampling method to balance the scarcity of contrastive sentences, enabling the model to learn more features from these critical samples. Second, we utilize a dual-head self-attention network to extract feature representations with relative position information, which are jointly trained with absolute position representations obtained from the pre-trained model. Finally, we apply label balancing techniques for model regularization to stabilize identification of neutral samples. Our model is evaluated on SemEval 2014 Task 4 Sub Task 2, demonstrating improvements in both Accuracy and Macro-F1 metrics across two datasets. Experimental results confirm the model' s effectiveness for contrastive sentence classification and its superior performance over baseline models on the entire test set.

Keywords: aspect-level sentiment analysis; contrastive sentences; attention network; BERT model; relative position encoding

0 Introduction

Text sentiment analysis enables enterprises to accurately analyze user evaluations of product features, providing effective references for detailed product improvement strategies. Fine-grained sentiment analysis methods have attracted significant attention from both academia and industry due to their wide application in dialogue systems, online reviews, and social networks [**?**]. Aspect-level sentiment classification (ASC) [**?**] represents a fine-grained sentiment analysis task that aims to determine the sentiment polarity (positive, negative, or neutral) of aspect terms within a sentence. A sentence may contain multiple aspect terms, each potentially bearing different sentiment polarities, necessitating specification of target aspects for accurate polarity determination.

In recent years, deep learning has demonstrated strong performance in sentiment analysis by automatically constructing neural networks for feature extraction [**?**]. Pre-trained models (PTMs) such as BERT [**?**] represent state-of-the-art approaches, achieving leading performance on GLUE benchmarks including text classification. BERT is a language model pre-trained on large Wikipedia corpora, with its specialized architecture enabling fine-tuning for supervised downstream tasks like ASC. While PTMs acquire general linguistic knowledge from massive corpora, effectively adapting this knowledge to downstream tasks remains a critical challenge [**?**]. Additionally, since Wikipedia articles are predominantly objective statements rather than subjective reviews with sentiment, BERT learns insufficient emotional content. This limitation, combined with the small training samples typically available for ASC, poses severe challenges for this already complex task [**?**].

Moreover, ASC tasks suffer from both limited labeled data and complex sentence structures, including contrastive sentiment sentences, implicit sentiment

sentences, and misleading neutral reviews. Table 1 illustrates these challenges: the review "air has higher resolution but the fonts are small" contains two target aspects ("resolution" and "fonts") with opposing polarities ("higher" as positive, "small" as negative). In "The waiter poured water on my hand and walked away," no explicit sentiment words appear, yet the target aspect "waiter" clearly conveys negative sentiment. The review "The service was typical short-order, dinner type" expresses neutrality toward "service" through very subtle language. Such complex sentences exceed the learning capacity of existing models [**?**].

This paper investigates BERT' s limitations as a pre-trained model and analyzes the distribution and characteristics of complex sentences in ASC datasets. Building upon BERT-DK [**?**], we optimize fine-tuning techniques related to complex sentence features in sampling and feature extraction. Our main contributions include: (1) empirical analysis of error samples on the validation set to systematically summarize characteristics of hard-to-classify instances; (2) a novel aspect-level sentiment analysis framework that improves classification performance for both complex sentences and overall test samples; (3) joint training of relative position features extracted via attention modules with absolute position features from BERT-DK to enhance positional information capture; and (4) the first application of weighted random sampling to aspect-level sentiment analysis.

1 Related Work

1.1 Aspect-Level Sentiment Analysis

ASC tasks can be trained independently or jointly with Aspect Extraction (AE) [**?**, **?**]. ASC requires attention to subtle opinions about specific aspects, making it more complex than document-level or sentence-level classification [**?**]. The small-sample problem in ASC has received considerable research attention, typically addressed through two approaches: model optimization to better capture syntactic and semantic features, and augmentation with external sentiment lexicons or in-domain corpora.

For model optimization, Sun Chi et al. [**?**] transformed ASC from single-sentence classification to sentence-pair classification by constructing auxiliary sentences for aspects, similar to machine reading comprehension and natural language inference tasks, achieving better performance through BERT fine-tuning. Karimi et al. [**?**] proposed BERT Adversarial Training (BAT), using adversarial processes to generate data similar to real examples in embedding space for joint adversarial training of AE and ASC tasks. For external knowledge integration, He Ruidan et al. [**?**] proposed the PRET+MULT framework, transferring sentiment knowledge from document-level classification trained on Amazon review datasets to ASC through shared shallow embeddings and LSTM layers. Xu et al. [**?**] introduced a post-training approach using additional domain-specific data to adapt BERT from its source domain to the ASC domain and task.

However, recent neural network methods have paid limited attention to com-

plex sentences in ASC datasets. Xu Hu et al. [**?**] demonstrated through concrete data and experiments that contrastive sentences (sentences with multiple aspects bearing different polarities) are extremely rare in ASC datasets, causing existing ASC classifiers to inadequately learn these patterns and "degrade" to sentence-level classifiers. They proposed an Adaptive Re-weighting (ARW) scheme that assigns each training sample a weight representing training importance, dynamically guiding the model to emphasize contrastive sentence training and effectively improving their classification. Li Zhengyan et al. [**?**] studied implicit sentiment sentences specifically, partitioning ASC datasets into explicit and implicit sentiment expression slices, finding approximately 30% of reviews contain implicit sentiment expressions. They introduced external sentiment knowledge through supervised contrastive pre-training on large-scale sentiment-annotated corpora, aligning implicit expressions with same-label explicit expressions, and employed aspect-aware fine-tuning to improve aspect-based sentiment recognition.

1.2 Adjusting Sample Weights

Most machine learning algorithms for classification assume balanced classes, yet appropriately balanced data is uncommon in real-world scenarios. Natural language processing tasks frequently exhibit class imbalance, most classically in sequence labeling where categories are severely imbalanced [**?**]. In named entity recognition, entities are obviously far fewer than non-entities, creating severe imbalance. A fundamental approach to mitigating class imbalance is adjusting sample weights [**?**], assigning higher weights to minority class samples so their error losses contribute more to network weight updates. Sample weight adjustment has been applied in domain adaptation [**?**] and sentiment analysis [**?**], though with completely different weighting purposes and methods. This paper improves the influence of rare but critical samples in training by adjusting contrastive sentence sampling weights.

1.3 Self-Attention Mechanism

Since its introduction, the self-attention mechanism [**?**] has rapidly advanced natural language processing. For tasks like text classification and recommendation where input is a sequence but output is not, attention can learn relationships between each token and relevant tokens in the same input sequence. Attention weights aim to capture how two words in the same sequence are related, where the relevance concept depends on the primary task [**?**].

For a given word embedding output sequence, three vectors (Q-query, K-key, V-value) are created for each sequence position. Attention is then implemented using Q, K, V for each position, yielding output containing information about the position and its relationships with all other positions. These Q, K, V vectors are generated using feed-forward layers. The self-attention calculation formula is:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

For a given word vector dimension d_{model}, multi-head self-attention performs attention h times on projection matrices (Q, K, V) of dimension d_{model}. For each head, self-attention (Q, K, V) is uniquely projected to dimension d_{model}/h, with output dimension also d_{model}/h. Each head' s output is concatenated and linearly projected again to obtain output with the same dimension as performing self-attention once on the original (Q, K, V) matrices. The entire process is described by:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W^O$$
$$\text{where head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$$

Projection weight matrices are $W_i^Q \in \mathbb{R}^{d_{\text{model}} \times d_k}$, $W_i^K \in \mathbb{R}^{d_{\text{model}} \times d_k}$, $W_i^V \in \mathbb{R}^{d_{\text{model}} \times d_v}$, and $W^O \in \mathbb{R}^{hd_v \times d_{\text{model}}}$.

When designing ASC "pre-training + fine-tuning" architectures, self-attention mechanisms are frequently applied in downstream task fine-tuning [**?**]. BERT and similar pre-trained models are fundamentally Transformers with two main components: self-attention and position-wise feed-forward layers, both permutation-equivariant and insensitive to input token order. To make models position-aware, absolute position encoding is added to word embeddings at each character position through self-attention mechanisms. However, this absolute position encoding approach causes loss of some fragment position information [**?**].

Inspired by Shaw et al. [**?**], this paper proposes DWS+RpSAN to address BERT's position information loss and complex sentence challenges in ASC tasks. The main differences from previous work are: (1) treating contrastive sentence scarcity as a simple class imbalance problem and dynamically adjusting training sample sampling weights to increase contrastive sentence sampling frequency; (2) combining absolute position features extracted by BERT pre-training with relative position features extracted by self-attention modules through parallel training, compensating for pre-trained models' deficiencies in position information extraction.

2 Dataset Analysis

We evaluate our algorithm on the popular ABSA benchmark dataset SemEval 2014 Task 4 Sub Task 2 [**?**], which covers two domains: restaurants (Rest) and laptops (Lap). The dataset includes three sentiment labels: positive, negative, and neutral. Each review contains zero, one, or multiple target aspects with irregular lexical units and syntactic patterns, making the data noisy, sparse,

and high-dimensional. For fair comparison with prior work and ensuring experimental consistency, we adopt the data processing from Xu Hu et al. [**?**], which removes conflicting sentences from the original dataset, adds "contra" labels to each review, and extracts contrastive sentences from the test set to create an independent dataset for performance comparison. Detailed statistics are shown in Tables 2 and 3.

Pontiki et al. [**?**] observed during annotation that laptop reviews often treat the product as a whole, and when commenting on specific aspects, users typically use adjectives to implicitly refer to aspects (e.g., "expensive," "heavy") rather than explicit target terms (e.g., "price,""weight"). Consequently, the Rest dataset contains more target aspects. As shown in Table 2, aspect-containing sentences comprise 75% of the Rest training set but only 47.75% of the Lap training set. Additionally, laptop reviews frequently mention functional descriptions without expressing sentiment (e.g., "Has a 5-6 hour battery life"), resulting in more neutral samples in the Lap dataset.

Contrastive sentiment sentences, which best demonstrate and evaluate fine-grained sentiment classification performance, are rare in both domains' training and test sets. Contrastive sentences constitute approximately 16% of the Rest training set and only 11% of the Lap training set—fewer even than annotation errors (which can be considered noise). Machine learning models trained on such contrastive-scarce datasets tend to degrade into coarse-grained (sentence-level) sentiment classifiers. For example, the review "The screen is good and also the battery" contains two target aspects ("screen" and "battery") with positive polarity for both, making the entire sentence positive. The model processes such reviews essentially as sentence-level classification tasks. Since most samples dominate training, rare but important samples are easily overlooked and may even be treated as noise, representing a common and widespread problem for machine learning models that can be viewed as imbalanced data issues.

3 Methodology

3.1 Problem Definition

Given a context sequence $X = \{x_1, x_2, \ldots, x_n\}$ where aspect term $a = \{x_i, \ldots, x_j\}$ is a subsequence of X, the aspect-level sentiment analysis task aims to predict the sentiment orientation of target aspect a in sentence X. Figure 1 illustrates the overall architecture of the proposed Relative Position Self-Attention Encoder Network (DWS+RPSAN), which primarily consists of a domain-aware BERT-DK embedding layer, a relative position self-attention encoder layer, and an output layer.

3.2 Dynamic Weighted Random Sampling

Given that contrastive sentences are critical yet rare for fine-grained sentiment analysis, we must consider how to enable machine learning models to learn

from these scarce samples. Assuming the dataset is divided into contrastive and non-contrastive classes $\{c_{\text{contra}}, c_{\text{nocontra}}\}$, with uniform random sampling probabilities $p(x \in c_i) = \frac{\#\{c_i\}}{\#\{\text{train}\}}$, we observe that the Rest training set has $c_{\text{nocontra}} : c_{\text{contra}} \approx 5 : 1$ while the Lap training set shows $9 : 1$ imbalance. Training on such datasets exposes models to non-contrastive sentences far more frequently than contrastive ones, making it difficult for deep learning models to learn from rare samples.

Gao et al. [**?**] found that during early training stages, most samples' losses dominate the total loss and determine model parameter update directions. In later iterations, although rare samples may dominate the total loss, they might not contribute sufficiently. In the worst case, when the optimizer begins overfitting minor details in majority samples, it may only then consider losses from rare examples, meaning validation might stop training before rare samples are well-optimized.

To address this issue where rare but important samples are easily overlooked, we must solve two problems: (1) increase contrastive sentence samples during early training stages, and (2) increase (or rebalance) sampling opportunities for poorly-optimized samples before the validation process finds the optimal model. A natural solution is balancing the training set through majority-class oversampling or minority-class oversampling. Since data is extremely sparse, undersampling the majority class is suboptimal as it may lose meaningful samples. Therefore, oversampling the minority class is preferable [**?**]. Deep learning models typically train on a batch-by-batch basis, making per-class weight adjustment natural at the end of each epoch when every sample has participated in learning once, allowing the model to focus on poorly-handled (misclassified) samples.

Based on this analysis, our goal is to design a dynamic adaptive scheme that continuously adjusts sampling weights for known contrastive sentences in the training set. Since probability values cannot explicitly indicate whether the model errs on a sample, we use accuracy-based weighting, assigning greater sampling weights to contrastive sentences. To prevent over-adaptation to the minority class, after each epoch we identify misclassified samples and their classes, dynamically updating weights based on validation set error rates for both classes. Let $w_i^{(n)}$ denote the sampling weight for class i in epoch n, with initial weights set as $w_{\text{contra}}^{\text{init}} = \frac{\text{total\_samples}}{\text{total\_samples} - N_{\text{contra}}}$ and $w_{\text{nocontra}}^{\text{init}} = \frac{\text{total\_samples}}{N_{\text{contra}}}$. The weight update formula is:

$$w_i^{(n)} = w_i^{(n-1)} + \varepsilon \times \text{error\_rate}_i^{(n-1)}$$

where $\text{error\_rate}_i^{(n-1)}$ is the validation set classification error rate for class i in epoch $n-1$, and ε is an update factor regulating the influence of misclassified sample categories on weighted sampling. Larger ε values increase the impact of misclassified categories on 下一轮迭代 weighted sampling. We experimented

with $\varepsilon \in \{0.0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3\}$. When $\varepsilon = 0.0$, sampling weights remain at initialization, already providing some performance improvement on contrastive sentence classification. Performance initially increases then decreases with larger ε, peaking at $\varepsilon = 0.1$, indicating that excessive ε overemphasizes classification results' influence on sampling weights.

3.3 BERT-DK Embedding Layer

The word embedding layer uses the pre-trained BERT-DK model [**?**] to generate sequence word vectors. This model addresses BERT's insufficient learning of sentiment content and poor applicability to review classification, particularly fine-grained sentiment analysis. It first performs masked language modeling, then uses unsupervised domain-specific (restaurant or laptop) review datasets for sentence prediction on pre-trained BERT weights, enhancing domain awareness before fine-tuning with supervised ASC data. Word vectors processed through the BERT-DK layer thus possess domain awareness to some extent.

3.4 Relative Position Self-Attention Layer

Attention mechanisms are non-recurrent models that cannot capture element order in input sequences, requiring explicit position encoding. Currently, three embedding methods are common: sinusoidal position encoding, learned position encoding, and relative position representation. The BERT-DK embedding module already uses sinusoidal signals to embed absolute position information for sequence processing, but relative position information is lost during self-attention computation, causing deviation from actual fine-tuning data. To incorporate this missing relative position information, we adopt relative position embeddings proposed by Shaw et al. [**?**]. Rather than using fixed embeddings per position, relative position embeddings generate different learned embeddings based on offsets between "key" and "query" in the self-attention mechanism.

Building upon original self-attention, we introduce two position-related vectors:

$$a_{ij}^K, a_{ij}^V \in \mathbb{R}^{d_{\text{model}}}$$

These learn relative position information between every two sequence positions using a set of trainable embedding vectors to represent each word' s position encoding in the input sentence. When computing attention features for target word x_i on x_j, these two vectors are additionally considered. We also introduce a tunable parameter k to limit the maximum distance between two sequence positions. Experiments tried $k \in \{1, 2, ..., 12\}$, finding no improvement when $k > 8$, indicating attention is relatively sensitive to relative positions within an 8-token window. Beyond this window, relative positions need not be distinguished, and we replace them with average pooling of embeddings within the window (gram=8). Defining these as trainable vectors essentially learns relative position information between every two sequence positions.

3.5 Classification Model

Existing ASC models exhibit unstable performance on neutral sentiment classification, tending to misclassify neutral samples as positive/negative or vice versa. The BERT-DK+DWS+RPSAN model introduces a Label Smoothing Regularization (LSR) term [**?**] into the original cross-entropy loss function to penalize low-entropy output distributions and suppress model confidence, stabilizing neutral class identification. For training sample x, assuming its actual probability distribution is $q(k|x)$, the smoothed distribution is:

$$q'(k|x) = (1-\lambda)q(k|x) + \lambda u(k)$$

where $u(k)$ is the prior probability distribution over labels and λ is the smoothing parameter. In experiments, the prior label distribution is uniformly set as $u(k) = 1/C$.

LSR is equivalent to the KL divergence between prior label distribution $u(k)$ and model prediction distribution $p_\theta(k|x)$. The LSR definition is:

$$L_{\text{lsr}} = \text{KL}(u(k) \| p_\theta(k|x))$$

Thus, LSR adds a cross-entropy loss to the original cross-entropy loss. The entire model' s objective function (loss function) to be optimized is:

$$L = L_{\text{Absolute}} + L_{\text{Relative}} + L_{\text{lsr}}$$

where L_{Absolute} and L_{Relative} represent cross-entropy losses for absolute and relative position feature representations, respectively.

The parameter optimization process for BERT-DK+DWS+RPSAN is shown in Algorithm 1. The algorithm comprises three stages: dynamic weighted random sampling, BERT-DK word embedding preprocessing, and joint fine-tuning with relative position self-attention mechanism.

4 Experiments

4.1 Experimental Environment and Hyperparameter Settings

All experiments and benchmarks run on a single GPU (GTX 1080 Ti) with an Intel Core i7-8700K@4.7 GHz CPU and 16GB RAM.

For fine-tuning, hyperparameters generally align with referenced experiments, with adjustments for new model characteristics. The self-attention module' s multi-head count matches Shaw et al. [**?**], with 2 heads yielding optimal results. For dropout rate, unlike the high 0.7 rate in prior work, we prefer a lower 0.1 rate, consistent with typical BERT sentiment classification settings. Initial learning rates of 2e-5 and 3e-5 have been validated as effective for Rest and

Lap datasets, respectively. Through extensive ablation experiments, we found setting Rest' s initial learning rate to 2e-5 and Lap' s to 3e-5 is optimal, likely because Rest contains more aspect-level sentences than Lap. Batch size is set to 32, consistent with BERT-DK, with each batch constructed through weighted random sampling from the training set. Training uses 20 epochs, saving the model with maximum accuracy. The relative position self-attention model' s word vector dimension d_{model} matches BERT-DK' s 300-dimensional output. Adam optimizer updates all parameters with label smoothing parameter $\lambda = 0.2$. All results are averaged over 10 runs.

4.2 Baseline Models

We select four classifiers as baselines and conduct ablation studies on our proposed modules, demonstrating that all components contribute to final performance. Dynamic weighted random sampling provides the greatest contribution on the Rest dataset, while the relative position self-attention layer contributes most on the Lap dataset.

AOA [**?**]: Introduces an attention-over-attention neural network that jointly models target aspects and sentences, explicitly capturing interactions between aspects and sentence contexts.

MGAN [**?**]: Proposes a multi-grained attention network framework that also uses target aspect alignment loss to describe aspect-level interactions between target aspects with the same context.

BERT-DK [**?**]: Builds upon BERT by first performing masked language modeling (MLM) and next sentence prediction (NSP) using domain-specific (laptop or restaurant) reviews on pre-trained BERT weights, then fine-tuning with supervised ASC data.

4.3 Experimental Results and Discussion

Model performance is evaluated using Accuracy and Macro-F1 metrics. Table 4 summarizes all experimental results.

Dynamic weighted sampling improves contrastive test set performance by approximately 8.4% on Rest and 11% on Lap compared to BERT-DK, and by about 2% compared to BERT-DK alone. After weighted sampling, full dataset performance improves on Rest but slightly decreases on Lap, possibly because weighted sampling is unsuitable for learning when Lap' s noisy samples (annotation errors) far exceed contrastive sentences, causing the model to learn more annotation errors and degrade overall performance.

BERT-DK+RPSAN improves contrastive test set performance by approximately 6.4% on Rest and 15.5% on Lap, achieving optimal performance on Lap' s contrastive test set. Compared to BERT-DK, it shows slight improvement on Rest' s contrastive test set and approximately 4.5% improvement on Lap' s.

Full dataset performance improves over both BERT-DK and BERT-DK+ARW on both domains.

BERT-DK+DWS+RPSAN shows performance improvements across all metrics except slightly lower performance on Lap' s contrastive test set compared to BERT-DK+RPSAN without weighted sampling. The improvement is particularly significant on Rest' s contrastive test set, proving our approach effective, especially for enhancing overall performance on contrastive sentences and truly testing fine-grained sentiment classification capability.

Figure 2 shows the distribution of non-contrastive and contrastive sentence classes sampled from the last 10 batches of the Rest test set using random sampling versus dynamic weighted random sampling. Weighted random sampling effectively balances the severe scarcity of critical contrastive sentences.

Table 5 analyzes hard samples on the validation set for BERT-DK+DWS, defining "Hard Samples" as those misclassified more than 5 times across 10 runs. Analysis reveals the model is most error-prone on neutral classification (over 70% of hard samples in both datasets), tending to predict neutral labels as other polarities or vice versa. This likely stems from neutral sentiment being inherently ambiguous and unreliable human annotation. Additionally, sentences containing comparative opinions (including contrastive sentences) show high error rates (over 58% in both datasets: 20 in Rest, 18 in Lap), with aspect term position being critically important.

Two non-contrastive hard samples illustrate neutral classification instability and how relative distance between aspect terms and sentiment words affects classification difficulty when comparative opinions exist in the sentence (Figures 3 and 4). After adding the relative position self-attention module, HardSample1' s aspect term [leather carrying case] errors decrease from 7 to 4 in 10 predictions, no longer qualifying as a hard sample. HardSample2' s aspect term [application] errors decrease from 10 to 8 but remain hard. This demonstrates that self-attention modules improve neutral class discrimination and model generalization, while also showing that neutral instability and comparative opinion sentences remain bottlenecks in ASC tasks—our future research focus. Label Smoothing Regularization (LSR) improves neutral sample prediction accuracy by approximately 0.12-0.2% across three ablation models, though Table 4 does not detail these increments.

5 Conclusion

This study addresses BERT' s position information loss in sentiment classification tasks through empirical analysis of hard samples on the validation set, confirming the importance of position information for classification. Further investigation of complex sentences in ASC datasets reveals that improving ASC classifier performance requires not only solving contrastive sentence scarcity but also addressing neutral class instability and challenges posed by comparative opinions in sentences for accurate aspect-specific classification. We balance con-

trastive and non-contrastive training samples through dynamic weighted sampling, jointly train relative position features from self-attention networks with absolute position features from pre-trained models, and apply label smoothing regularization. Experimental results demonstrate our model achieves new breakthroughs in handling contrastive sentences crucial for ASC tasks while maintaining excellent classification performance on the entire test set.

References

[1] Zhang Yan, Li Tianrui. Review of comment-oriented aspect-based sentiment analysis [J]. Computer Science, 2020, 47(6): 200-206.

[2] Thet T T, Na J C, Khoo C S. Aspect-based sentiment analysis of movie reviews on discussion boards [J]. Journal of Information Science, 2010, 36(6): 823-848.

[3] Zhang Lei, Wang Shuai, Liu Bing. Deep learning for sentiment analysis: A survey [J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2018, 8(4): e1253.

[4] Devlin J, Chang Mingwei, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding [C]// Proc of the Conference of the North American Chapter of the Association for Computational Linguistics: Language Technologies. 2019: 4171-4186.

[5] Wang A, Singh A, Michael J, et al. GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding [C]// Proc of the EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP. 2018: 353-355.

[6] Qiu Xipeng, Sun Tianxiang, Xu Yige, et al. Pre-trained models for natural language processing: A survey [J]. Science China Technological Sciences, 2020: 1-26.

[7] Xu Hu, Liu Bing, Shu Lei, et al. BERT Post-Training for Review Reading Comprehension and Aspect-based Sentiment Analysis [C]// Proc of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2019: 2324-2335.

[8] Wang Kai, Shen Weizhou, Yang Yunyi, et al. Relational Graph Attention Network for Aspect-based Sentiment Analysis [C]// Proc of the 58th Annual Meeting of the Association for Computational Linguistics. 2020: 3229-3238.

[9] Zeng Yifu, Lan Tian, Wu Zufeng, et al. Bi-memory based attention model for aspect level sentiment classification [J]. Chinese Journal of Computers, 2019, 42(8): 1845-1857.

[10] Yang Heng, Zeng Biqing, Yang Jianhao, et al. A multi-task learning model for chinese-oriented aspect polarity classification and aspect term extraction [J]. Neurocomputing, 2021, 419: 344-356.

[11] Ambartsoumian A, Popowich F. Self-attention: A better building block for sentiment analysis neural network classifiers [C]// Proc of the 9th Workshop on Computational Approaches to Subjectivity: Sentiment and Social Media Analysis. 2018: 130-139.

[12] Sun Chi, Huang Luyao, Qiu Xipeng. Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence [C]// Proc of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2019: 380-385.

[13] Karimi A, Rossi L, Prati A. Adversarial training for aspect-based sentiment analysis with bert [C]// the 25th International Conference on Pattern Recognition. IEEE, 2021: 8797-8803.

[14] He Ruidan, Lee W S, Ng H T, et al. Exploiting document knowledge for aspect-level sentiment classification [C]// Proc of the 56th Annual Meeting of the Association for Computational Linguistics. 2018: 579-585.

[15] Xu Hu, Liu Bing, Shu Lei, et al. A failure of aspect sentiment classifiers and an adaptive re-weighting solution [J]. arXiv preprint arXiv:1911.01460, 2019.

[16] Li Zhengyan, Zou Yicheng, Zhang Chong, et al. Learning Implicit Sentiment in Aspect-based Sentiment Analysis with Supervised Contrastive Pre-Training [C]// Proce of the Conference on Empirical Methods in Natural Language Processing. 2021: 246-256.

[17] Akkasi A, Varoğlu E, Dimililer N. Balanced undersampling: a novel sentence-based undersampling method to improve recognition of named entities in chemical and biomedical text [J]. Applied Intelligence. 2018, 48(8): 1965-1978.

[18] Guo X, Yin Y, Dong C, et al. On the class imbalance problem [C]// The 4th international conference on natural computation. IEEE, 2008, 4: 192-197.

[19] Wang Rui, Utiyama M, Liu Lemao, et al. Instance weighting for neural machine translation domain adaptation [C]// Proc of the Conference on Empirical Methods in Natural Language Processing. 2017: 1482-1488.

[20] Pappas N, Popescu-Belis A. Explaining the stars: Weighted multiple-instance learning for aspect-based sentiment analysis [C]// Proc of the Conference on Empirical Methods In Natural Language Processing. 2014: 455-466.

[21] Yang Zichao, Yang Diyi, Dyer C, et al. Hierarchical attention networks for document classification [C]// Proc of the conference of the North American chapter of the association for computational linguistics: human language technologies. 2016: 1480-1489.

[22] Chaudhari S, Mithal V, Polatkan G, et al. An attentive survey of attention models [J]. ACM Transactions on Intelligent Systems and Technology, 2021, 12(5): 1-32.

[23] Yang Heng, Zeng Biqing, Yang Jianhao, et al. A multi-task learning model for chinese-oriented aspect polarity classification and aspect term extraction [J]. Neurocomputing, 2021, 419: 344-356.

[24] Yang Zhilin, Dai Zihang, Yang Yiming, et al. Xlnet: Generalized autoregressive pretraining for language understanding [J]. Advances in neural information processing systems, 2019, 32.

[25] Shaw P, Uszkoreit J, Vaswani A. Self-attention with relative position representations [C]// Proc of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2018: 464-468.

[26] Pontiki M, Galanis D, Papageorgiou H, et al. Semeval-2014 task 4: Aspect based sentiment analysis [C]// International workshop on semantic evaluation. 2014: 19-30.

[27] Gao T, Jojic V. Sample importance in training deep neural networks [J]. Workshop on Computational Approaches to Subjectivity: Sentiment and Social Media Analysis, 2018.

[28] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision [C]// Proc of the IEEE conference on computer vision and pattern recognition. 2016: 2818-2826.

[29] Huang Binxuan, Ou Yanglan, Carley K M. Aspect level sentiment classification with attention-over-attention neural networks [C]// International Conference on Social Computing: Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation. Springer, Cham, 2018: 197-206.

[30] Li Zheng, Wei Ying, Zhang Yu, et al. Exploiting coarse-to-fine task transfer for aspect-level sentiment classification [C]// Proc of the AAAI Conference on Artificial Intelligence. 2019, 33(01): 4253-4260.

Note: Figure translations are in progress. See original paper for figures.