

Aspect category sentiment analysis based on multiple attention mechanisms and pre-trained models

Xiaoan Zhan^{1a,*}, Chenxi Shi^{1b}, Lianwei Li², Kangming Xu³, Haotian Zheng⁴

^{1a}Electrical Engineering, New York University, NY, USA

^{1b}Telecommunication Systems Management, Northeastern University, MA, USA

²Computer Science, The University of Texas at Arlington, Arlington, USA

³Computer Science and Engineering, Santa Clara University, CA, USA

⁴Electrical & Computer Engineering, New York University, New York, NY, USA

*Corresponding author. E-mail: zgaq2502@gmail.com

Abstract. The rapid rise of e-commerce platforms has changed people's shopping habits, driving the popularity of online shopping. Users express their opinions on products and services by purchasing products on platforms and posting comments. These comment data contain rich user experiences, which are crucial for enterprises to understand user needs and improve product quality. Sentiment analysis of comment text is an important research direction in text mining, focusing on how to extract user evaluations of products from comment data to provide comprehensive, authentic, and accurate product feedback. This paper mainly investigates aspect-based fine-grained sentiment analysis. Because users express multiple sentiments towards different aspects in comments, coarse-grained sentiment analysis cannot accurately capture users' sentiment tendencies. This study utilizes artificial intelligence technology based on deep learning, firstly, it constructs auxiliary training samples to transform aspect sentiment tasks into machine reading comprehension or language inference tasks, using BERT model to extract text features and sentence features from comment data. To guide the model to focus on the features most relevant to the given aspect, a cross-attention mechanism is utilized to cross-focus the features of text with aspect category features. Finally, the sentiment polarity of given aspect category text can be predicted through a forward network. Experimental results on multiple datasets demonstrate that this method outperforms other deep learning models.

Keywords: Deep Learning, Pretrained BERT, Aspect-level Sentiment Analysis, Machine Learning.

1. Introduction

The proliferation of e-commerce platforms has profoundly altered consumer behavior, with a growing number of individuals turning to online shopping. As a result, user-generated comment data has become a valuable resource for both consumers and businesses alike. These comments offer rich insights into product experiences, serving as a crucial reference for users and a valuable source of feedback for companies. [1-2] Within the realm of text mining, sentiment analysis, particularly aspect-based sentiment analysis, has emerged as a pivotal research direction. Unlike traditional sentence-level sentiment analysis, aspect-based sentiment analysis delves deeper into user opinions by considering

multiple aspects within a single sentence, posing increased difficulty in prediction. While deep learning methods have been widely employed to tackle sentiment analysis tasks, shallow neural network models often struggle to effectively extract relevant features from comments. To address this limitation, our approach combines original comment texts with pre-trained language models to construct auxiliary training samples, [3] facilitating improved feature extraction. Additionally, we introduce an enhanced attention mechanism that focuses on important features pertinent to the current aspect category, enhancing the accuracy of polarity prediction. Experimental validation across multiple aspect sentiment datasets validates the efficacy of our proposed model, while further analysis sheds light on factors influencing its performance, providing valuable insights for future research endeavors and practical implementation in e-commerce settings.

2. Related Work

Aspect-based sentiment analysis has evolved significantly over time, progressing from rule-based methods relying on sentiment lexicons to machine learning techniques focusing on more effective text representation. Currently, deep learning neural networks are at the forefront of research in this field. For instance, the [4] Seq2Path model challenges the idea of fixed positional relationships between sentiment tuples and utilizes a seq2seq encoder-decoder architecture to generate sentiment paths resembling tree structures. On the other hand, the [5] PASTE model employs [6] Bi-LSTM and BERT to capture sentence features and employs a pointer network to accurately extract aspect words, sentiments, and opinions, enhancing triple extraction accuracy.

Another notable advancement is the BARTABSA [7] model, which addresses multiple subtasks within aspect-based sentiment analysis using a unified architecture based on BART. By transforming inputs into vectors and utilizing a pointer network for aspect category labeling, BARTABSA achieves comprehensive analysis within a single framework. Similarly, the [8] DOER model takes a multi-task approach, jointly modeling aspect sentiment analysis and aspect term extraction through sequence labeling. It utilizes shared framework units and cross-framework units to effectively extract and propagate information within the model, streamlining the complexity of the tasks.

Lastly, TAS-BERT represents a jointly trained model capable of identifying targets and sentiments simultaneously, addressing implicit aspects within sentences. By leveraging pre-trained language models and incorporating CRF or softmax decoders, TAS-BERT [9] achieves robust sentiment analysis while considering the nuanced relationships between targets and sentiments. Overall, these advancements underscore the importance of leveraging deep learning techniques and unified architectures for more accurate and comprehensive aspect-based sentiment analysis.

In this paper, we combine BERT fine-tuning tasks with a cross-attention mechanism for aspect sentiment polarity analysis. We construct auxiliary samples, fine-tune with pre-trained BERT models, extract features using a cross-attention network, conduct sentiment polarity classification, and achieve the best results to date.

3. Methodology

In this model as shown in Figure 1, the features of text and aspect markers are extracted by BERT. Auxiliary training samples are first constructed, transforming aspect sentiment tasks into reading comprehension or natural language inference tasks to extract text features and sentence features, Comments features are represented by $Vector - D$, while aspect tag features are represented by $Vector - Q$. Aspect tags can be single words or phrases. Applying document attention to $Vector - D$ yields the result represented by $Vector - S$. To ensure the model focuses on the most relevant features for a given aspect, we perform cross-attention feature extraction between comment features and aspect tag word features. We denote the result of this cross-attention as $Vector - R$, which is then concatenated with $Vector - S$. Finally, we can predict the sentiment polarity of the given aspect category text through a classification layer consisting of fully connected layers and softmax. Specifically, there are various methods for constructing auxiliary samples. In this paper, we use a mixture of methods to construct auxiliary samples. Sample question-answer method: Based on the provided sample, ask how

to evaluate a specific aspect, with the correct answer being the sentiment polarity of that aspect. Sample inference method: Provide a template, similar to a fill-in-the-blank question, "The user's evaluation of a certain aspect in the sentence is". Then, concatenate the auxiliary samples after the corresponding samples and input them into the [10-12] BERT model. For each word, its input representation is constructed by adding corresponding word, segment, and position embeddings bitwise. For classification tasks, the first token position of each sentence sequence is a special token ([CLS]). Use the first special token [CLS] of the final hidden state layer as the final vector representation of the input sequence.

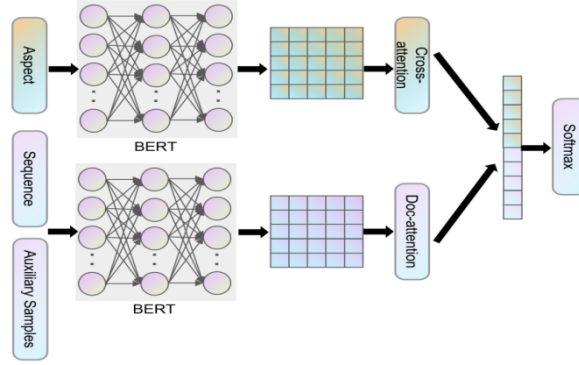


Figure 1. The architecture diagram of the model proposed in this paper.

The document attention can be represented as follows:

$$u_t = \tanh(d_t W_w + b_w) \quad (1)$$

$$\alpha_t = \frac{\exp(u_t u_w)}{\sum_{t=0}^n \exp(u_t u_w)} \quad (2)$$

$$s = \sum_{t=0}^n \alpha_t d_t \quad (3)$$

In this formula, we utilize a module composed of attention weights and parameters W_w and b_w , which constitute the bias terms, to extract output weights for various features, integrating them into a vector $s \in R^H$. For each different aspect category, we have a corresponding document attention layer, which, through a cross-attention mechanism, guides the network to focus on key parts of the comment text related to the aspect category. We can calculate the matching score between $D \in R^{(n+1) \times H}$ and $Q \in R^{(m+1) \times H}$ as follows:

$$M_{(i,j)} = d_i^T \cdot q_j \quad (4)$$

Where $M \in R^{(n+1) \times (m+1)}$. Then, by applying max pooling to M to obtain the maximum values along each row, the result is represented by $E \in R^{(n+1)}$. Since M represents the matching scores between text features and aspect labels, and each element in E can be interpreted as the relevance between a text word and a given aspect category [13]. By scaling E using softmax, we can obtain $\beta \in R^{(n+1)}$, and with this, weighted summation according to relevance of text features, we can obtain $r \in R^H$:

$$\beta = \text{softmax}(E) \quad (5)$$

$$r = \sum_{t=0}^n \beta_t d_t \quad (6)$$

Connecting the vectors $s \in R^H$ and $r \in R^H$ together, we can obtain $z \in R^{2H}$. Classification of sentiment polarity can be achieved through a classification layer, where p_j represents the sentiment polarity of the current aspect category.

$$p_j = \text{softmax}(\sum_{i=1}^n W_i z) \quad (7)$$

The loss function of the model is as follows:

$$loss_i = -\sum_j \sum_k weight_{ik} y_j^k \log \hat{y}_j^k \quad (8)$$

$$weight_{ik} = \log_{base} \frac{all_sample}{class_sample_{ik}} \quad (9)$$

$$loss_{total} = \sum_i loss_i \quad (10)$$

4. Experimental Results and Analysis

4.1. Datasets

Experimental work was conducted on three datasets (Table 1), the details of which are as follows:

- **ASAP [14]:** This dataset is for sentiment analysis and consists of restaurant review data. It includes real user review data from online food ordering platforms. Samples in this dataset contain users' star ratings for their dining experiences, as well as their evaluations of various aspects of the dining experience, such as food taste, environment, service, and 18 other predefined aspect categories. The dataset was randomly divided into training, validation, and test sets.
- **ACOS [15]:** This dataset is a quadruple-extraction dataset for aspect sentiment, namely aspect term - aspect category - opinion - sentiment quadruples. It contains text data from restaurant and laptop reviews. The laptop review data is sourced from the Amazon shopping website and is twice the size of past datasets. Samples in the dataset not only include explicitly expressed aspects and opinions but also implicit aspects and opinions, increasing the difficulty of learning from this dataset.
- **ASTE-Data-V25 [16]:** This dataset is a sentiment multi-tuple dataset involving aspect categories, sentiment polarities, and sentiment causality relationships. It is a highly challenging dataset in the aspect sentiment domain. Each sentence contains multiple aspect categories and corresponding sentiments. It includes explicit aspect categories when aspect terms appear in the sentence and implicit aspect categories when aspect terms do not appear in the sentence. For example, [('battery', 'durable', 'Positive').

Table 1. Datasets description

Dataset		Pos.	Neu.	Neg.
ASAP	Train	11712	10018	7455
	Dev	5185	2626	5898
	Test	6378	3055	5978
ACOS	Train	9870	3701	11677
	Dev	7001	4165	8698
	Test	5033	3311	8190
ASTE-Data-V25	Train	14262	10105	11070
	Dev	5547	4221	4694
	Test	13691	9389	11270

4.2. Training Protocol

To mitigate the risk of overfitting during model training, a data augmentation method involving random deletion of text content was employed. All methods incorporated a hyperparameter called Maximum Deletion Ratio (*MDR*), where *MDR* defined the maximum random deletion rate of input text during training, and the average deletion rate was set to *MDR* / 2. Pre-trained uncased BERT-base was used for fine-tuning. The model consisted of 12 Transformer blocks, with a hidden layer size of 768 and 12 self-attention modules, totaling 1.1 million parameters. During fine-tuning, a dropout probability of 0.1 was maintained, and the number of training epochs was set to 10. The initial learning rate was set to

$2e - 5$, with a batch size of 24. *Adam* optimizer was utilized for optimization. Each model was run 5 times, and the best results for all models were reported on the test dataset.

4.3. Evaluation Strategy

This paper utilizes *Accuracy* and *F1* scores for evaluation.

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \quad (11)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (12)$$

4.4. Results Analysis

Table 2. Experimental Result

Model	ASAP		ACOS		ASTE-Data-V25	
	ACC.	F1	ACC.	F1	ACC.	F1
DE-CNN	73.25%	70.17%	64.38%	63.96%	67.35%	61.61%
RNCRF	75.62%	75.35	66.08%	65.64	68.48%	61.48%
LCF-BERT	79.04%	77.46%	66.75%	64.08%	69.32%	63.37%
AOA	80.17%	77.79%	65.78%	64.23%	67.60%	64.39%
MemNet	81.68%	79.05%	68.58%	65.35%	68.98%	65.71%
BERT-PT	82.43%	81.17%	69.27%	68.76%	70.21%	67.11%
Our Model	85.94%	84.19%	71.16%	70.84%	70.69%	68.02%

Overall, as shown in Table 2, Our Model performs the best in terms of accuracy and F1 scores across all datasets. It is noteworthy that although Our Model's performance varies across different datasets, it remains the most stable model overall, indicating its strong generalization ability. It can be observed that all models exhibit similar performance on the [17] ACOS dataset compared to the ASTE-Data-V25 dataset, but a decrease in performance is noted when compared to the Data1 dataset. However, these models offer shorter training times and higher computational efficiency, making them suitable for lightweight applications. This highlights the advantage of simple model structures and efficient computational efficiency. Therefore, for simple tasks requiring efficient execution, shallow neural network-based models are recommended, while scenarios demanding high performance may benefit from models based on large language frameworks.

5. Conclusion

With the development of the Internet, people's shopping habits have undergone significant changes, from offline to online. E-commerce platforms have increasingly become the carriers of users' shopping needs. While purchasing goods, users express their opinions on products and services through comments. These comment data contain rich user experiences, which are crucial for companies to understand user sentiments. Sentiment analysis of comment text is a hot research direction in text mining. Compared to sentence-level sentiment analysis, aspect-based sentiment analysis can provide more in-depth and fine-grained information. Aspect sentiment analysis initially relied on rule-based methods, followed by machine learning methods that focused on finding more effective sentiment features at the text representation level. Now, the mainstream research method is based on deep learning approaches. This paper proposes an aspect category sentiment analysis model based on BERT, which can provide more accurate product feedback for platforms and merchants. The model extracts features using BERT and applies a cross-attention mechanism to improve model performance. Experimental results demonstrate that the model performs excellently on three datasets: ASAP, ACOS, and ASTE-Data-V25, validating the effectiveness of the model.

References

- [1] Schouten K, Frasincar F. Survey on aspect-level sentiment analysis[J]. IEEE transactions on knowledge and data engineering, 2015, 28(3): 813-830.
- [2] Do H H, Prasad P W C, Maag A, et al. Deep learning for aspect-based sentiment analysis: a comparative review[J]. Expert systems with applications, 2019, 118: 272-299.
- [3] Zhou J, Huang J X, Chen Q, et al. Deep learning for aspect-level sentiment classification: survey, vision, and challenges[J]. IEEE access, 2019, 7: 78454-78483.
- [4] Nazir A, Rao Y, Wu L, et al. Issues and challenges of aspect-based sentiment analysis: A comprehensive survey[J]. IEEE Transactions on Affective Computing, 2020, 13(2): 845-863.
- [5] Liu H, Chatterjee I, Zhou M C, et al. Aspect-based sentiment analysis: A survey of deep learning methods[J]. IEEE Transactions on Computational Social Systems, 2020, 7(6): 1358-1375.
- [6] Poria S, Hazarika D, Majumder N, et al. Beneath the tip of the iceberg: Current challenges and new directions in sentiment analysis research[J]. IEEE transactions on affective computing, 2020, 14(1): 108-132.
- [7] Mao Y, Shen Y, Yang J, et al. Seq2path: Generating sentiment tuples as paths of a tree[C]//Findings of the Association for Computational Linguistics: ACL 2022. 2022: 2215-2225.
- [8] Zhang W, Li X, Deng Y, et al. Towards generative aspect-based sentiment analysis[C]//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). 2021: 504-510.
- [9] Yan H, Dai J, Qiu X, et al. A unified generative framework for aspect-based sentiment analysis[J]. arXiv preprint arXiv:2106.04300, 2021.
- [10] Luo H, Li T, Liu B, et al. DOER: Dual cross-shared RNN for aspect term-polarity co-extraction[J]. arXiv preprint arXiv:1906.01794, 2019.
- [11] Wan H, Yang Y, Du J, et al. Target-aspect-sentiment joint detection for aspect-based sentiment analysis[C]//Proceedings of the AAAI conference on artificial intelligence. 2020, 34(05): 9122-9129.
- [12] Bu J, Ren L, Zheng S, et al. ASAP: A Chinese review dataset towards aspect category sentiment analysis and rating prediction[J]. arXiv preprint arXiv:2103.06605, 2021.
- [13] Cai H, Xia R, Yu J. Aspect-category-opinion-sentiment quadruple extraction with implicit aspects and opinions[C]//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). 2021: 340-350.
- [14] Xu L, Chia Y K, Bing L. Learning span-level interactions for aspect sentiment triplet extraction[J]. arXiv preprint arXiv:2107.12214, 2021.
- [15] Xu H, Liu B, Shu L, et al. Double embeddings and CNN-based sequence labeling for aspect extraction[J]. arXiv preprint arXiv:1805.04601, 2018.
- [16] Wang W, Pan S J, Dahlmeier D, et al. Recursive neural conditional random fields for aspect-based sentiment analysis[J]. arXiv preprint arXiv:1603.06679, 2016.
- [17] Zeng B, Yang H, Xu R, et al. Lcf: A local context focus mechanism for aspect-based sentiment classification[J]. Applied Sciences, 2019, 9(16): 3389.