

Exploring ChatGPT-based Augmentation Strategies for Contrastive Aspect-based Sentiment Analysis

Lingling Xu, Hong Kong Metropolitan University, Hong Kong, China

Haoran Xie, S. Joe Qin, Lingnan University, Hong Kong, China

Fu Lee Wang, Hong Kong Metropolitan University, Hong Kong, China

Xiaohui Tao, University of Southern Queensland, Queensland, Australia

Abstract—Aspect-based sentiment analysis (ABSA) involves identifying sentiment towards specific aspect terms in a sentence, and allows us to uncover people's nuanced perspectives and attitudes on particular aspects of a product, service, or topic. However, the scarcity of labeled data poses a significant challenge to training high-quality models. To address this issue, we explore the potential of data augmentation using ChatGPT, a well-performing large language model (LLM), to enhance the sentiment classification performance towards aspect terms. Specifically, we explore three data augmentation strategies based on ChatGPT: context-focused, aspect-focused, and context-aspect data augmentation techniques. Context-focused data augmentation focuses on changing the word expression of context words in the sentence while keeping aspect terms unchanged. In contrast, aspect-focused data augmentation aims to change aspect terms but keep context words unchanged. Context-Aspect data augmentation integrates the above two data augmentations to generate augmented samples. Furthermore, we incorporate contrastive learning into the ABSA tasks to improve performance. Extensive experiments show that all three data augmentation techniques lead to performance improvements, with the context-aspect data augmentation strategy performing best and surpassing the performance of the baseline models.

Aspect-based sentiment analysis (ABSA) is a fine-grained sentiment classification task that focuses on discerning the sentiment polarity of aspect terms in a given sentence. Take the following sentence as an example: “This restaurant’s decor is eye-catching, but the food is unsatisfactory.” The aspect term “decor” exhibits a positive sentiment polarity, whereas the aspect term “food” demonstrates a negative sentiment polarity. Recently, the success of contrastive learning has led to the emergence of various data augmentation methods [1], [2], [3], which have been applied to contrastive aspect-based sentiment classification tasks, further improving the performance of ABSA tasks. Moreover, the development of LLMs, particularly the introduction of ChatGPT, has revolutionized the field of natural language processing

(NLP). The advent of GPT-3.5 and GPT-4 has further expanded advanced applications across different domains, including ABSA tasks. However, directly using LLMs for fine-tuning with limited datasets poses challenges such as overfitting and expensive GPU costs. Therefore, a strategy to leverage the strong generative capabilities of LLMs is to use them for data augmentation. These augmented samples can then be employed in smaller models like BERT to undergo fine-tuning, thereby reducing computational demands and enhancing flexibility during fine-tuning.

Inspired by the work [4], [5], which leverages ChatGPT for data augmentation in contrastive sentence representation learning and few-shot named entity recognition, we further investigate the role of ChatGPT for data augmentation in ABSA tasks. We propose three data augmentation strategies: context-focused, aspect-focused, and context-aspect data augmentation methods. In context-focused data augmentation, we aim to replace context words with other words while

ensuring that the aspect terms and sentiment polarity remain unchanged. On the other hand, in aspect-focused data augmentation, we aim to replace aspect terms with other suitable aspect terms while preserving sentiment polarity. Finally, the context-aspect data augmentation combines both strategies mentioned above. Through comprehensive experiments, we observed that these three data augmentation techniques yielded improvements compared to the vanilla BERT, with context-aspect data augmentation demonstrating the best performance on both datasets. Furthermore, we found that employing data verification to ensure that generated aspect terms are completely different from the original ones does not consistently result in improved performance gains. Overall, the main contributions of this paper are as follows:

- We introduce how to prompt ChatGPT to perform context-focused and aspect-focused data augmentation for ABSA tasks.
- We explore the use of ChatGPT-based data augmentation and data verification strategies to boost the performance of ABSA tasks.
- We conduct exhaustive experiments to illustrate the effectiveness of ChatGPT for ABSA tasks in data augmentation.

Related Work

ABSA has gained a considerable amount of attention recently. A study from [6] employed attention mechanisms, Long Short-Term Memory (LSTM), and external commonsense knowledge to enhance ABSA performance. One study found in [7] introduced a hierarchical attention mechanism that combined context-focused and aspect-focused attention models. Another approach, Sentic GCN [8], utilized SenticNet¹ to enrich the dependency graphs of sentences using graph convolution networks. By enhancing the dependencies between contextual words and aspect terms, this method aimed to capture affective information about aspect terms. The latest version, SenticNet 8 [9], further improves this by integrating Emotion AI and Commonsense AI to enhance interpretability, trustworthiness, and explainability in affective computing, which is crucial for tasks like ABSA. Furthermore, contrastive learning [10], a self-supervised representation learning method, has received a great deal of attention and has been applied to various NLP tasks, including ABSA. This approach learns representations by bringing positive pairs (semantically similar sentences) closer to-

gether and pushing negative pairs (semantically different sentences) apart. For instance, some work [1], [3] proposed novel data augmentation strategies to increase the number of training datasets to improve ABSA through contrastive learning. More recently, the remarkable performance and powerful generative and rewriting capabilities of LLMs like ChatGPT have inspired researchers to explore their potential for data augmentation. In this regard, [4] leverages ChatGPT and word masking to generate new sentences and improve contrastive sentence representation learning. While [5] utilizes prompt and ChatGPT for data augmentation to improve the performance of few-shot named entity recognition.

Methodology

In this section, we begin by introducing the ABSA task. Next, we present three novel data augmentation methods based on ChatGPT for the ABSA datasets. Lastly, we outline the overall training objective of our proposed model framework.

Task Formulation

ABSA datasets are comprised of a group of triplet samples $\mathbf{D} = \{s_i, a_i, l_i\}_{i=1}^N$, in which s_i denotes review sentences, a_i means aspect term, and $l_i \in \{-1, 0, 1\}$ denotes the sentiment polarity towards aspect term (-1 denotes negative polarity, 0 denotes neutral polarity, 1 denotes positive polarity). Following the work [2], [11], we also choose the pretrained language model BERT [12] as the backbone model and construct sentence pair classification task to obtain the sentiment representations towards the aspect term in the given sentence. More specifically, the inputs are first transformed into “[CLS] s_i [SEP] a_i [SEP]” with [CLS] and [SEP] tokens. These transformed inputs are then fed into BERT to obtain the aspect-based sentiment representation:

$$h_i = \mathcal{F}([\text{CLS}]s_i[\text{SEP}]a_i[\text{SEP}]), \quad (1)$$

in which $\mathcal{F}(\cdot)$ denotes the BERT model.

Data Augmentation

We present three ChatGPT-based data augmentation methods for ABSA datasets with the assistance of GPT-3.5-turbo, which was developed by OpenAI². Moreover, some examples are presented in Table 1 to illustrate the distinctions among three data augmentation strategies.

¹<https://sentic.net/>

²<https://platform.openai.com/docs/models/gpt-3-5-turbo>

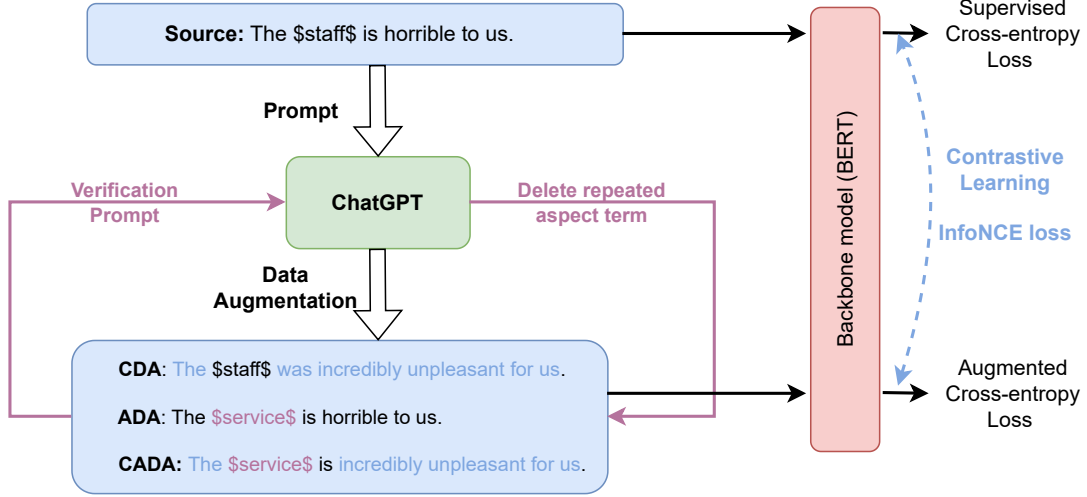


FIGURE 1. The overall framework of our proposed method.

TABLE 1. Comparing the generated samples with CDA, ADA, and CADA methods, the colored spans are changed part.

Method	Examples [dataset, aspect term, sentiment]
Source	The speed is incredible and I am more than satisfied. [Laptop, speed, positive]
CDA	The speed is extraordinary and I am more than content.
ADA	The performance is incredible and i am more than satisfied.
CADA	The performance is extraordinary and I am more than content.
Source	The palak paneer was standard, and I was not a fan of the malai kofta. [Restaurant, palak paneer, neutral]
CDA	The palak paneer was mediocre, and I did not enjoy the creamy vegetable balls.
ADA	The curry was standard, and I was not a fan of the malai kofta.
CADA	The curry was mediocre, and I did not enjoy the creamy vegetable balls.

Context-focused Data Augmentation (CDA) is a data augmentation technique that aims to change the contextual words in a sentence while keeping the aspect term and sentiment polarity of the aspect term unchanged. The goal of CDA is to increase the semantic richness and diversity of the training dataset while keeping the semantics of the sentences unchanged. Specifically, we adopt paraphrasing as the data augmentation strategy and utilize the excellent rewriting abilities and comprehensive world knowledge of ChatGPT for CDA to obtain augmented samples. However, constructing a suitable prompt for CDA is challenging, as it involves preserving aspect terms and their sentiment polarity unchanged, making it difficult to design a suitable prompt. Following the rule of principled guidance, the prompt for CDA is presented in Table 2. Notably, two examples are presented in the prompt to facilitate ChatGPT's comprehension of the intended task and generate the desired sentence. However, the selection of two examples varies to account for the

differences in domain within the training datasets.

TABLE 2. ChatGPT prompt for ADA. Place the source sentence to the "{sentence}" and place the aspect term in the source sentence to the "{aspect term}".

Given the sentence: "{sentence}", and given the aspect term "\${aspect term}\$" in above sentence.

Please generate one new sentence using paraphrasing. The new sentence should not paraphrase the aspect term "\${aspect term}\$" and should keep the aspect term "\${aspect term}\$", semantics of the sentence, and sentiment polarity towards the aspect term "\${aspect term}\$" unchanged.

Here are a few examples:

Source sentence: <source sentence> → New sentence: <CDA augmented sentence>

Please only output the New sentence.

Aspect-focused Data Augmentation (ADA) is a data augmentation method that focuses on replacing the original aspect term with a different semantically and logically suitable aspect term. At the same time, it preserves the context words in the sentence, and the sentiment polarity towards the new aspect term is unchanged. The purpose of ADA is to increase the diversity of aspect terms and improve the robustness of the model in detecting the opinion words associated with different aspect terms. Unlike existing approaches that generate new sentences based on given aspect terms in [1], our approach employs ChatGPT to randomly generate new aspect terms, thereby increasing the generality and diversity of aspect terms, especially for unseen aspect terms. To perform CDA for ABSA datasets, the ChatGPT prompt for CDA is presented in Table 3. Importantly, we have implemented a verification step to ensure that the newly generated aspect term differs from the original. In cases where the aspect term remains the same as the original, the source sentence is repeatedly fed back into ChatGPT with the same prompt until a distinct aspect term is generated.

TABLE 3. ChatGPT prompt for ADA. Place the source sentence to the “{sentence}” and place the aspect term in the source sentence to the “{aspect term}”.

Given the sentence: “{sentence}”, and given the aspect term “\${aspect term}\$” in above sentence.

Please replace the given aspect term in the given sentence with a new semantically and logically suitable aspect term and also keep the sentiment polarity towards the new aspect term unchanged.

Please only output the new aspect term.

Context-Aspect Data Augmentation (CADA) is a data augmentation method that combines the aforementioned two data augmentation strategies. Specifically, CADA first utilizes the CDA on the source sentence to generate a context-focused augmented sentence, and then leverages ADA on the same source sentence to obtain a new aspect term that aligns with the semantics and logic of the source sentence. The augmented samples using CADA are achieved by concatenating the context-focused augmented sentence with the newly generated aspect term. By integrating CDA and ADA in this manner, CADA not only helps to diversify the sentence structure and wording, but also the aspect terms.

Training Objective

In this section, we introduce the overall training objective of our proposed method. The overall framework is shown in Figure 1. The overall training objective is to perform aspect-based supervised sentiment classification on both source and augmented samples, and to conduct contrastive learning between source samples and augmented samples. Specifically, we perform supervised sentiment classification training (SSCT) on source and augmented sentences with the following objectives:

$$\mathcal{L}_{SSCT} = \frac{1}{N} \sum_{i=1}^N (L_{CE}(h_i W_s + b_s, l_i) + \alpha L_{CE}(h_i^+ W_s + b_s, l_i)), \quad (2)$$

in which h_i and h_i^+ are the aspect-based sentiment representations of source samples and augmented positive samples, and l_i is the ground-truth sentiment polarity label of source and augmented samples. α is the hyper-parameter to adjust the performance of the SSCT task.

To further enhance the performance robustness, we incorporate contrastive learning into our approach. In this method, the samples generated through the ChatGPT-based data augmentation strategy are considered positive pairs, while the rest of samples in the batch are regarded as negative pairs. The InfoNCE loss is used as our contrastive learning loss:

$$\mathcal{L}_{CL} = -\frac{1}{N} \sum_{i=1}^N \log \frac{\exp(\cos(h_i, h_i^+)/\tau)}{\sum_{j=1}^N \exp(\cos(h_i, h_j^+)/\tau)}. \quad (3)$$

Here, N is the number of samples in a batch, with τ being the temperature parameter to adjust the similarity between source samples and augmented samples.

Thus, the overall training objective of our proposed framework can be formulated as:

$$\mathcal{L} = \mathcal{L}_{SSCT} + \beta \mathcal{L}_{CL} \quad (4)$$

in which β is a controllable parameter.

Experiments

In this section, we begin by introducing training datasets and baseline models. Subsequently, we provide a comprehensive overview of our experimental settings and details. Additionally, we present and analyze the experimental results, and conduct sensitivity analysis to examine the impact of hyper-parameters on ABSA performance.

TABLE 4. Statistics of Laptop and Restaurant datasets.

Dataset	Positive		Neutral		Negative	
	Train	Test	Train	Test	Train	Test
Restaurant	2164	728	637	196	807	196
Laptop	994	341	464	169	870	128

Dataset

We conduct experiments on the two public ABSA datasets, Restaurant and Laptop, which are sourced from SemEval 2014 [13], comprising reviews in the domains of restaurants and laptops, respectively. The sample in the two ABSA datasets consists of a sentence, an associated aspect term, and the sentiment label towards the aspect term. The statistics and details of the two datasets are shown in Table 4.

Baselines

To evaluate the effectiveness of our proposed data augmentation approach, we compared it against four baseline models: 1) **BERT-base** directly uses the vanilla BERT model as the backbone and fine-tune BERT on the source training samples for aspect-based sentiment classification; 2) **C³DA** [1] proposes a contrastive cross-channel data augmentation method to obtain augmented samples for supervised aspect-based sentiment classification and contrastive learning; 3) **BERT-Scon** [2] leverages the characteristic of aspect-invariant and sentiment-invariant for supervised contrastive learning to enhance ABSA; 4) **BERT+SR** [3] utilizes synonym replacement for data augmentation to obtain augmented samples for contrastive aspect-based sentiment classification.

Implementation Details

We implemented our experiments using the pretrained language model BERT (bert-base-uncased) as the backbone model. The batch size was set to 32, and we employed Adam as the optimizer with a learning rate of 2e-5 and a dropout rate of 0.1. Additionally, we set the temperature parameter to 0.08 for contrastive learning. Regarding data augmentation, we utilized the GPT-3.5-turbo model as the ChatGPT, setting the temperature parameter to 0 for CDA and 1 for ADA, aiming to preserve the sentence semantics while diversifying the aspect terms. For the hyper-parameters α and β in the overall loss function, we adapted different values for each data augmentation strategy for better ABSA performance. Specific details can be found in Table 5. It was discovered that CADA achieves better performance when selecting a larger hyper-parameter. Furthermore, all models were trained for 50 epochs,

TABLE 5. The hyper-parameters setting for different data augmentation strategies. (veri) denotes to delete the repeated aspect terms

Dataset	α	β
CDA	0.2	0.2
ADA	0.6	0.5
ADA (veri)	0.1	0.2
CADA	0.2	0.4
CADA (veri)	0.4	0.6

employing an early-stop strategy that terminates training if there are no improvements within 10 epochs. All of our experiments were conducted with a single NVIDIA A800 (80G).

Main Results

As indicated in Table 6, ChatGPT-based augmentation techniques, including CDA, ADA, and CADA, consistently and significantly enhance performance across ABSA tasks. Notably, the BERT+CADA model achieved the highest performance in the Laptop and Restaurant datasets, surpassing all baseline models. The Laptop dataset demonstrated a remarkable improvement of 1.41% in accuracy and 3.06% in Macro F1 compared to the vanilla BERT model. Similarly, the Restaurant dataset showcased a notable enhancement of 1.16% in accuracy and 1.42% in macro F1 compared to the vanilla BERT model. Moreover, the incorporation of aspect term verification, utilizing GPT-3.5-turbo to eliminate repeated aspect terms, contributed to further accuracy improvements in the Restaurant dataset and macro F1 improvements in the Laptop dataset. We discovered that BERT+CADA (veri) outperformed BERT+CADA on both datasets. This may be due to the following reasons: first, repeated aspect terms are more in line with the semantics of the source sentence, and forcing the generation of different aspect terms is not semantically and logically appropriate for the sentence; second, certain repeated aspect terms in the CADA enhancement ensure that the semantic changes are not too large, leading to better performance. Additionally, BERT+CADA (veri) demonstrated superior performance compared to BERT+CDA and BERT+ADA in the Restaurant dataset, and better performance than BERT+ADA and BERT+ADA (veri) in the Laptop dataset. Unlike verification techniques used in LLMs to tackle hallucination and ensure accurate responses [14], our verification step concentrated on generating distinct aspect terms. Significantly, we discovered that generating unsuitable, yet distinct aspect terms may potentially impact the semantics of sentences and result in degraded per-

TABLE 6. Experimental results with different data augmentation strategies on the ABSA datasets.

Model	Restaurant		Laptop	
	Acc.	F1	Acc.	F1
BERT-base	85.98	79.84	79.47	74.65
BERT+SR [3]	85.80	79.70	80.56	76.11
BERT-Scon [2]	86.51	80.55	80.23	76.48
C ³ DA [1]	86.93	81.23	80.61	77.11
BERT+CDA	86.52	80.36	80.56	77.51
BERT+ADA	86.61	80.91	80.25	75.80
BERT+ADA (veri)	86.70	80.27	80.25	76.65
BERT+CADA	87.14	81.26	80.88	77.71
BERT+CADA (veri)	86.88	81.00	80.56	77.09

formance.

Sensitivity Analysis

As mentioned above, two hyper-parameters, α and β , were adopted in our proposed contrastive aspect-based sentiment classification framework. To investigate the influence of both hyper-parameters, a sensitivity analysis was conducted. We explored a range of values for α and β in $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$, and performed a grid search to identify suitable combinations that yielded decent performance on both the Laptop and Restaurant datasets. As depicted in Figure 2, we observed consistent trends in the Accuracy and F1 scores across different hyper-parameter values. The supervised aspect-based sentiment classification performance exhibited sensitivity to α , while the introduction of augmented cross-entropy loss yielded only marginal improvements, even a decrease, on the Laptop dataset across three data augmentation methods. However, notable improvements were observed in the Restaurant dataset. Additionally, as shown in Figure 3, we also investigated the impact of hyper-parameter β in the total loss function. We discovered that the overall performance of three different data augmentation strategies is sensitive to the hyper-parameter β . To ensure that our proposed framework achieves excellent performance in both the Laptop and Restaurant datasets across different methods, we employed distinct combinations of hyper-parameters for different data augmentation techniques, which are detailed in Table 5.

Conclusion

In this paper, we investigated three data augmentation methods based on ChatGPT, combined with contrastive learning, to enhance the performance of ABSA tasks. The experimental results demonstrate that all three data augmentation strategies yield notable im-

provements in ABSA performance, with CADA exhibiting the best performance. The proposed method, CADA, not only enriches the semantic diversity of context words, but also improves the diversity of aspect terms, resulting in significant enhancements in ABSA performance that surpass baseline models by a substantial margin. The experimental results provide strong evidence for the effectiveness of our proposed framework, underscoring ChatGPT as a powerful data augmentation strategy that generates high-quality data through carefully designed prompts.

Acknowledgements

Qin’s work was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (R1015-23). Xu, Xie and Wang’s work was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (UGC/FDS16/E17/23); and Xie’s work was also supported by the Faculty Research Grant (DB24C5) of Lingnan University, Hong Kong.

REFERENCES

1. B. Wang et al., “A Contrastive Cross-Channel Data Augmentation Framework for Aspect-Based Sentiment Analysis,” *Proceedings of the 29th International Conference on Computational Linguistics*, 2022, pp. 6691–6704.
2. B. Liang et al., “Enhancing aspect-based sentiment analysis with supervised contrastive learning,” *Proceedings of the 30th ACM international conference on information & knowledge management*, 2021, pp. 3242–3247.
3. L. Xu and W. Wang, “Improving aspect-based sentiment analysis with contrastive learning,” *Natural Language Processing Journal*, vol. 3, 2023, p. 100009.
4. Q. Cheng et al., “Improving Contrastive Learning of Sentence Embeddings from AI Feedback,” *Findings*

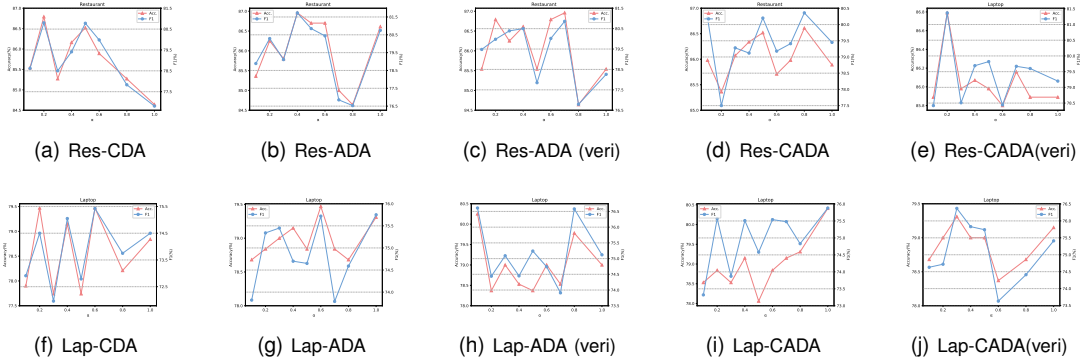


FIGURE 2. Sensitivity analysis of the augmented cross-entropy loss hyper-parameter α in supervised aspect-based sentiment classification under different data augmentation strategies.

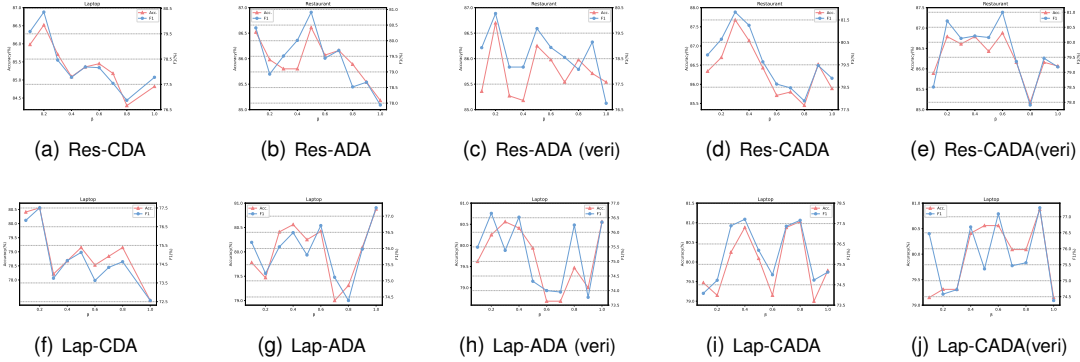


FIGURE 3. Sensitivity analysis of the contrastive learning hyper-parameter β in total training objective with three data augmentation strategies and data verification.

- of the Association for Computational Linguistics: ACL 2023, 2023, pp. 11122–11138.
5. J. Ye et al., “LLM-DA: Data Augmentation via Large Language Models for Few-Shot Named Entity Recognition,” *arXiv preprint arXiv:2402.14568*, 2024.
 6. Y. Ma et al., “Sentic LSTM: a hybrid network for targeted aspect-based sentiment analysis,” *Cognitive Computation*, vol. 10, 2018, pp. 639–650.
 7. Y. Ma, H. Peng, and E. Cambria, “Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM,” *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, 2018, p. 1.
 8. B. Liang et al., “Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks,” *Knowledge-Based Systems*, vol. 235, 2022, p. 107643.
 9. E. Cambria et al., “SenticNet 8: Fusing emotion AI and commonsense AI for interpretable, trustworthy, and explainable affective computing,” *International Conference on Human-Computer Interaction (HCI)*, 2024, pp. 1–20.
 10. T. Chen et al., “A simple framework for contrastive learning of visual representations,” *International conference on machine learning*, 2020, pp. 1597–1607.
 11. Z. Li et al., “Learning Implicit Sentiment in Aspect-based Sentiment Analysis with Supervised Contrastive Pre-Training,” *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 2021, pp. 246–256.
 12. J. Devlin et al., “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1*, 2019, pp. 4171–4186.
 13. M. Pontiki et al., “Semeval-2016 task 5: Aspect based sentiment analysis,” *ProWorkshop on Semantic Eval*

uation (*SemEval-2016*), 2016, pp. 19–30.

14. S. Dhuliawala et al., “Chain-of-verification reduces hallucination in large language models,” *arXiv preprint arXiv:2309.11495*, 2023.

Lingling Xu is currently pursuing a Ph.D. degree at the School of Science and Technology, Hong Kong Metropolitan University. Her research interests include aspect-based sentiment analysis, and contrastive sentence representation learning. Contact her at s1289650@live.hkmu.edu.hk.

Haoran Xie is currently the Acting Associate Dean and Professor at the School of Data Sciences, Lingnan University, Hong Kong. His research interests include artificial intelligence, big data, and educational technology. He has 414 research publications in international journals and conferences. Haoran Xie is the corresponding author of this article. Contact him at hrxie@ln.edu.hk.

S. Joe Qin is currently the Wai Kee Kau Chair Professor and President of Lingnan University, Hong Kong. His research interests include data science and analytics, machine learning, process monitoring, and model predictive control. He has over 470 publications in international journals and conferences. Contact him at joeqin@ln.edu.hk.

Fu Lee Wang is currently the Dean of the School of Science and Technology, Hong Kong Metropolitan University, Hong Kong. His research interests include educational technology, information retrieval, computer graphics, and bioinformatics. He has over 300 publications in international journals and conferences. Contact him at pwang@hkmu.edu.hk.

Xiaohui Tao is currently the School Head, Acting Dean, and Professor of School of Mathematics, Physics, and Computing, University of Southern Queensland, Australia. His research interests include artificial intelligence, knowledge engineering, and health informatics. He has more than 150 publications in international journals and conferences. Contact him at xtao@usq.edu.au.