LLM-as-an-Augmentor: Improving the Data Augmentation for Aspect-Based Sentiment Analysis with Large Language Models

Mengyang Xu, Qihuang Zhong (✉), and Juhua Liu

School of Computer Science, Wuhan University, Wuhan 430072, China
{xumengyang,zhongqihuang,liujuhua}@whu.edu.cn

Abstract. Aspect-Based Sentiment Analysis (ABSA) is a vital fine-grained sentiment analysis task that aims to determine the sentiment polarity towards an aspect in a sentence. Due to the expensive and limited amounts of labeled data, data augmentation (DA) methods have become the de-facto standard for ABSA. However, current DA methods usually suffer from 1) poor fluency and coherence; 2) lack of the diversity of generated data. To this end, we propose a novel simple-yet-effective DA method for ABSA, namely **LLM-as-an-Augmentor**, which leverages the powerful capability of third-party large language models (LLMs) to improve the quality of generated data. Specifically, we introduce several text reconstruction strategies and use them to guide the LLMs for automatic data generation via a carefully-designed prompting method. Extensive experiments on 5 baseline methods and 3 widely-used benchmarks show that our **LLM-as-an-Augmentor** can bring consistent and significant performance gains among all settings. More encouragingly, given only 15% labeled data, our method can achieve comparable performance to that of full labeled data. To the best of our knowledge, our work is one of the rare works to leverage LLMs to generate fine-grained training data for the ABSA task. We hope our work could promote more research in related fields.

Keywords: Aspect-based Sentiment Analysis · Large Language Model· Data Augmentation · Prompt Engineering.

1 Introduction

Aspect-based sentiment analysis (ABSA) is an important fine-grained sentiment analysis task, which focuses on identifying and analyzing the sentiments expressed toward specific aspects in a sentence [1]. Although currently popular large language models (LLMs) (e.g., GPT-4 [2], LLaMA2 [3]) have achieved widespread success in a wide of downstream tasks, they might fall short in dealing with the fine-grained classification tasks against the discriminative language models (e.g., BERT [4]), which has been proven by prior works [5]. Hence, employing BERT-style models is still a viable option in the field of ABSA. However, due to the limited amounts of labeled training data, fine-tuning the models usually results in sub-optimal performance.

Data augmentation (DA) is a common technology to enrich the quantity of training data by changing the original data [6] or generating more data [7,8,9]. Specifically, DA methods can be generally divided into two categories: word-level and sentence-level

DA [10]. Word-level DA methods involve replacing or inserting words into sentences, leveraging techniques such as word synonym dictionaries or contextual word embeddings [6]. These methods aim to introduce linguistic variations that maintain the sentiment orientation and aspect context of the original sentence. While sentence-level DA methods focus on generating new sentences that preserve the sentiment and aspect associations of the original text but rephrasing them using paraphrasing methods [9], generative models [8,9], or machine translation [7] techniques. This allows the ABSA models to learn from different sentence structures and lexical variations.

Although these DA methods have achieved remarkable performance, they usually suffer from some issues: 1) **Poor fluency and coherence**, as the word-level DA methods distort the sentence meaning or structures, and existing sentence-level DA methods struggle to generate fluent and coherent sentences. 2) **Lack of the diversity of generated data**, as most of the prior DA methods do not reconstruct the structure of original sentence, limiting the diversity of generated sentences. Intuitively, while LLMs fall short in determining the fine-grained information for ABSA, they have the great potential to deal with the above issues of DA methods, as they can generate fluent and high-quality text [11]. As stated by Wei et al. [12], instruction tuning enables the LLMs to well follow human instructions. By carefully designing some prompting methods, LLMs can be used for various applications, such as text completion, text rewriting, etc.

Motivated by this, we propose a simple-yet-effective DA method for ABSA, namely **LLM-as-an-Augmentor**, which leverages the powerful in-context learning ability of LLMs to generate more high-quality and diverse training data. In particular, the proposed method contains three-stage processes. First, we introduce three text reconstruction strategies to reconstruct the sampled training data, and obtain the "original-reconstructed" text pairs. Then, by using the "original-reconstructed" text pairs as demonstrations, we design a prompting method to guide the LLMs for automatic data generation in a few-shot manner. Lastly, we filter incorrect data and form a new dataset. In general, instead of directly prompting LLMs to predict the fine-grained sentiment polarity, we convert it to an auxiliary text generation task and carefully design some novel prompting methods to guide the generation of LLMs. By doing so, we can take full advantage of LLMs to generate more fluent and diverse training data for the ABSA task.

Extensive experiments on three widely-used ABSA benchmarks, Restaurant14, Laptop14 [13] and Resuaurant15 [14], show that: 1) our proposed DA method can bring consistent and significant performance gains among 5 baseline ABSA models; 2) given only 15% labeled data, our method can achieve comparable performance to that of full labeled data; 3) our method outperforms the other DA counterparts by a clear margin. To summarize, our contributions are two-fold: (1) We propose a simple-yet-effective DA method for ABSA by leveraging the powerful in-context learning ability of LLMs. (2) Extensive results on 3 widely-used ABSA benchmarks show the effectiveness and superiority of our proposed method.

The remainder of this paper is designed as follows. We review the related work in Section 2. In Section 3, we introduce our proposed method in detail. In Section 4, we present the experimental results. Conclusions are described in Section 5.

2 Related Work

2.1 Aspect-based Sentiment Analysis

Aspect-based Sentiment Analysis (ABSA) can be roughly regarded as a fine-grained sentiment analysis task that makes a judgment of the sentiment polarity towards an aspect [15]. To solve this task, several neural network-based models were proposed, which we categorize into two groups: context- and syntax-based methods. Context-based models first utilize CNN to obtain aspect features from context [16]. Leveraging LSTM's significant advantages in processing sequence data tasks, Tang et al. [17] propose target-dependent LSTM (TD-LSTM) to capture aspect information. On the other hand, syntax-based methods generally utilize dependency information with graph-based networks. Zhang et al. [18] use graph convolutional networks (GCN) to learn node representations from dependency trees. On this basis, Wang et al. [19] propose a novel aspect-oriented dependency tree structure.

With the advent of Transformer architecture [20], pre-trained language models (PLMs) based on BERT [4] massively influence the situation of ABSA. After Li et al. [21] first introduce BERT into end-to-end ABSA assignment, lots of BERT-based models are further proposed. Rietzler et al. [22] fine-tune the BERT language model and conduct cross-domain research. The LCF model proposed by Zeng et al. [23] captures both local context features and global context features via self-attention mechanisms.

2.2 Data Augmentation

Data augmentation (DA) generates new data by changing the original data through various methods, which enlarges the training dataset to alleviate the issue of data scarcity. EDA [6] is a simple text data enhancement technique containing four operations: synonym substitution, random insertion, random exchange, and random deletion. Back-Translation [7] translates data to a chosen pivot language and then back to the original language, which generates more diverse sentences without changing their meanings [24]. CBERT [8] integrates label information into the masked language modeling task through segmentation embedding to realize the prediction of replacement words, considering not only context but also label information. Mixup [9] is a DA method that modifies both inputs and labels by mixing up inputs of samples and their labels, where labels are commonly represented with one-hot encoding.

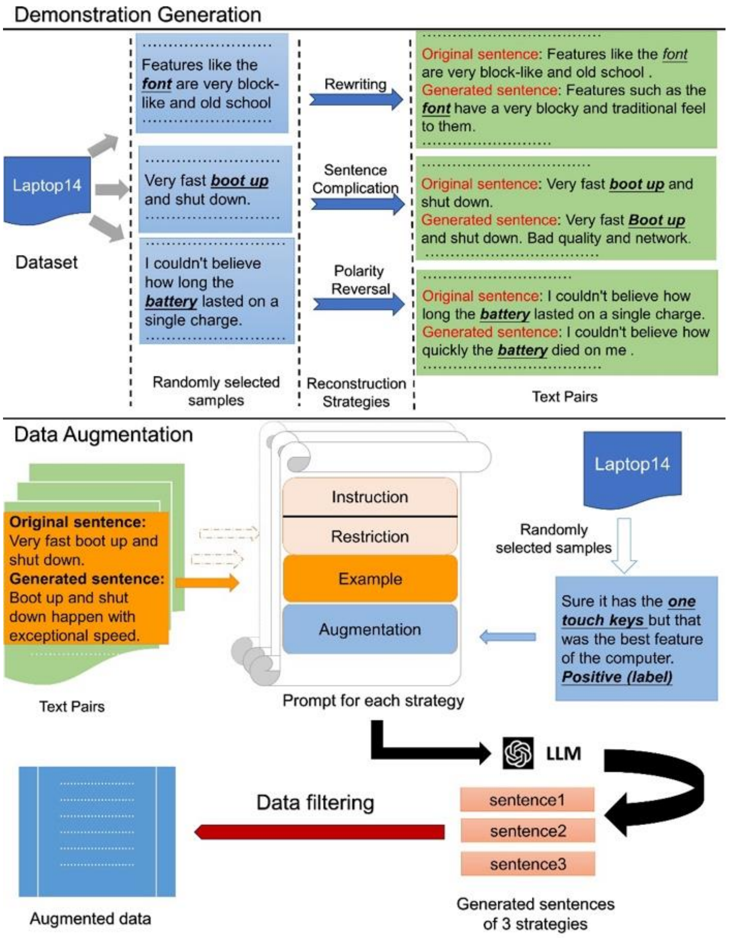
2.3 Large Language Model

Recently, large language models (LLMs) like LLaMA2 [3] and GPT-4 [2] have achieved great success in the NLP community. However, Zhong et al. [5] show that for the fine-grained tasks, LLMs might perform worse than the BERT-style models, as they might fall short in extracting the fine-grained information. Hence, for the ABSA task, the currently popular methods are almost based on the BERT-based models. Despite this, LLMs excel in generating fluent and high-quality contexts and show powerful instruction-following and in-context learning capabilities. Inspired by it, we attempt to take advantage of LLMs and enforce them to generate more high-quality data for boosting the performance of ABSA models.

3 Method

This section explains details about our proposed **LLM-as-an-Augmentor** method for ABSA tasks, which contains three parts: demonstration generation, data augmentation, and data filtering. The framework is shown in Fig. 1.

Fig. 1. Illustration of our proposed **LLM-as-an-Augmentor** method. In the demonstration generate stage, we introduce 3 text reconstruction strategies. Then, we design several prompting methods to guide the LLMs for automatic data augmentation. Finally, we conduct data filtering to remove incorrect generated data.


Demonstration Generation
Laptop14
Dataset
Features like the font are very block-like and old school
Very fast boot up and shut down.
I couldn't believe how long the battery lasted on a single charge.
Randomly selected samples
Rewriting
Sentence Complication
Polarity Reversal
Reconstruction Strategies
Original sentence: Features like the font are very block-like and old school .
Generated sentence: Features such as the font have a very blocky and traditional feel to them.
Original sentence: Very fast boot up and shut down.
Generated sentence: Very fast Boot up and shut down. Bad quality and network.
Original sentence: I couldn't believe how long the battery lasted on a single charge.
Generated sentence: I couldn't believe how quickly the battery died on me .
Text Pairs
Data Augmentation
Original sentence:
Very fast boot up and shut down.
Generated sentence:
Boot up and shut down happen with exceptional speed.
Text Pairs
Instruction
Restriction
Example
Augmentation
Prompt for each strategy
Laptop14
Randomly selected samples
Sure it has the one touch keys but that was the best feature of the computer.
Positive (label)
LLM
sentence1
sentence2
sentence3
Generated sentences of 3 strategies
Data filtering
Augmented data

3.1 Demonstration Generation

The core of our method is to introduce three text reconstruction strategies, which are then used to guide the data augmentation of LLMs. Specifically, we first randomly sample some instances from the original training datasets. For each sampled instance, we perform the text reconstruction processes, respectively, and obtain the "original-reconstructed" pairs. These pairs will be used as the demonstrations in the latter DA process, thus we denote this stage as "Demonstration Generation". In particular, the introduced text reconstruction strategies are as follows:

1) **Rewriting:** To improve the diversity of training datasets, we are inspired by Fan et al. [25] and attempt to rewrite the sentence without changing its meaning. Specifically, to ensure fluency and coherence, we enforce third-party LLMs (e.g., ChatGPT [26]) to rewrite the sentence with the following prompts: "*please rewrite the given sentence with no more than 30 words. Note that the word [aspect given by dataset label] must appear in the new sentence without any change and the new sentence should have a similar tone as the given sentence.*"
2) **Sentence Complication:** Training on the samples with multi-aspects in a sentence can boost the robustness of ABSA models [27]. Motivated by this, we also try to construct the multi-aspect training data by prompting the ChatGPT to complicate the original sentence with a new sub-sentence, which is similar in structure but has a different aspect and polarity.
3) **Polarity Reversal:** Generally, the diversity of sentiment polarity has a great impact on the performance of the model [28]. However, the number of samples with different emotional polarities varied slightly. Thus, we let LLMs generate sentences with opposite sentiments, using the following prompts: "*please generate a sentence using the same sentence pattern and tone by imitating the example above. Note that [aspect given by dataset label] must appear in the new sentence with no change. the sentiment of [aspect given by dataset label] should be [sentiment set].*"

3.2 Data Augmentation

Here, we perform the data augmentation by prompting the LLMs in a few-shot setting. Considering the powerful LLMs' capability of in-context learning, we randomly select several "original-reconstructed" text pairs as the demonstrations and then force the LLMs to generate new training data following the style of demonstrations. Specifically, we first set the **instruction** of this task as "*we need you to complete a sentence generation task with following requirements*". Then, we set some **restrictions** to improve the successful rate of generating needed sentences. For instance, "*the new sentence should have the similar sentence structure and tone as original sentence*". Subsequently, we provide some **examples** for each strategy to guide the generation of LLMs. Ultimately, the original sentence is presented to the LLMs for **augmentation**.

3.3 Data filtering

Although the LLMs excel in following the instructions, they might generate incorrect sentences that do not satisfy the given requirements. Hence, to improve the quality of generated data, we filter the noisy data by designing some rules, e.g., deduplication, removing the modified aspects and etc. More specifically, we find that there are three main error types, i.e., aspect missing (the given aspect is missing in the generated sentence), aspect changing (the given aspect is not appear in the generated sentence) and ambiguity (there are two identical aspects in the sentence). Thus, we manually design the rules to filter these errors.

4 Experiments

4.1 Dataset and Settings

We conduct our experiments on three public and widely-used datasets, i.e., Laptop14, Restaurant14 [13], and Restaurant15 [14]. For our data augmentation pipeline, we use the widely-used GPT-3.5-Turbo-1106 as the LLM. In practice, we set the temperature as 0.7 to balance the diversity and coherence of generated sentences, and set the max length of generated sentences as 256. The statistics of original and generated datasets are listed in Table 1. We train the models on the augmented training datasets and report the test performance on the original test sets. For evaluation, we use the Accuracy ("Acc") and Macro-F1 ("F1") metrics to measure the performance.

To investigate the effectiveness of our methods, we employ our augmented data to various ABSA baseline models, which are based on three backbones, i.e., GloVe [29] with 300 dimensions, BERT [4] and RoBERTa [30]. These models are listed below:

- **ATAE-LSTM** [31]: A LSTM-based model for aspect-level sentiment classification using aspect embedding and attention mechanism.
- **ASGCN** [32]: The first ABSA model to represent sentences with dependency trees and use GCN to explore the syntactical information.
- **Vanilla-BERT** [33]: A variant of BERT that can truncate long text sequences into fixed-length segments without dependency.
- **RGAT** [19]: A model uses an aspect-oriented dependency tree structure to reshape and prune ordinary dependency parse trees to better model syntax information.
- **KGAN** [34]: A knowledge graph augmented network, where different information is encoded as multi-view representations to augment the semantic features.

We run most of the models in their default settings given by corresponding papers. For the others, we set learning rates as 1e-3 for Glove-based models, 3e-5 for BERT-based and RoBERTa-based models. The batch size is set as 32 for all models and we apply dropout on the word embeddings with a drop rate of 0.1 for Glove and 0.3 for BERT and RoBERTa. We use Adam [35] as the optimizer to achieve optimization and training. Every experiment is conducted three times with random seed, and we choose the best performance as our result.

Table 1. Statistics of original and generated datasets.

Datasets	Division	Positive	Neutral	Negative
Laptop14	Train (Original)	994	870	464
	Train (Generated)	368	392	288
	Test	134	128	169
Restaurant14	Train (Original)	2164	807	637
	Train (Generated)	745	363	516
	Test	728	196	196
Restaurant15	Train (Original)	912	36	256
	Train (Generated)	312	16	214
	Test	326	34	128

4.2 Main Result

We randomly extract 15%[1] labeled samples to apply data augmentation and provide one text pair as demonstration in LLM prompt. Table 2 lists the results of different models using augmented data.

Table 2. Results of the proposed DA method on various baseline methods. Notably, the original results are collected from [18], [19], [34], while those with "*" are our re-produced results. "+only-gen" denotes that we only train the models on the generated data, while "+merged" denotes that we train the models on the mix of original and generated training data. The best results are in **bold**.

Embedding	Models	Laptop14		Restaurant14		Restaurant15	
		Acc	F1	Acc	F1	Acc	F1
GloVe	**ATAE-LSTM**	68.88	63.94	78.06	67.02	78.48	62.84
	+only-gen	65.54	60.96	74.18	63.15	76.12	59.30
	+merged	**70.03**	**64.52**	**78.29**	**67.95**	**79.05**	**64.01**
	ASGCN	75.55	71.05	80.86	78.19	79.34	60.78
	+only-gen	72.72	67.08	77.41	75.64	76.95	57.45
	+merged	**76.50**	**71.76**	**81.40**	**78.77**	**80.23**	**61.78**
	RGAT	77.42	73.76	83.30	76.08	75.09*	61.76*
	+only-gen	74.32	70.00	79.03	72.59	77.68	59.45
	+merged	**78.57**	**74.68**	**84.01**	**76.84**	79.73	63.25
	KGAN	78.91	75.21	84.46	77.47	83.09	67.90
	+only-gen	75.35	72.87	81.80	73.56	80.96	63.84
	+merged	**80.13**	**76.68**	**85.23**	**78.32**	**84.10**	**68.55**
BERT	**Vanilla-BERT**	77.58	72.38	85.62	78.28	83.40	65.28
	+only-gen	74.04	68.35	82.11	73.96	80.20	60.97
	+merged	**79.38**	**73.56**	**86.87**	**81.09**	**77.81**	**75.22**
	RGAT-BERT	78.21	74.40	86.60	81.35	83.22	69.73
	+only-gen	75.29	70.41	84.29	76.99	81.75	65.54
	+merged	**79.22**	**75.41**	**87.54**	**82.42**	**85.18**	**72.02**
	KGAN-BERT	82.66	78.98	87.15	82.05	86.21	74.20
	+only-gen	78.95	74.18	83.48	78.17	81.11	70.03
	+merged	**84.19**	**80.43**	**89.01**	**83.54**	**88.00**	**75.42**
RoBERTa	**Vanilla-RoBERTa**	83.78	80.73	87.37	80.96	84.56	70.16
	+only-gen	80.45	76.76	83.81	76.08	81.98	66.43
	+merged	**85.01**	**82.32**	**88.31**	**82.10**	**87.20**	**71.94**
	RGAT-RoBERTa	83.33	79.95	87.52	81.29	84.65*	70.30*
	+only-gen	79.43	75.21	84.92	76.85	81.01	66.56
	+merged	**85.02**	**81.69**	**87.65**	**82.03**	**85.98**	**71.99**
	KGAN-RoBERTa	83.28	80.14	87.78	83.05	88.60	74.36
	+only-gen	78.02	75.31	83.49	79.22	84.90	70.49
	+merged	**83.92**	**81.08**	**88.84**	**84.29**	**89.98**	**75.95**

[1] The analysis on the ratio of sampled labeled data can be found in Section 4.3.

Firstly, when using merged data (contains original data and generated data), most of the models outperform the baseline, indicating that our method brings consistent performance gains in all settings. Moreover, when it comes to using generated data only, we can also find that it could achieve (nearly) comparable performance to that of original training data. These results can prove the effectiveness of our proposed method.

Secondly, our methods perform better when using BERT and RoBERTa models than those of Glove-based models. We attribute it to the powerful capability of pretrained language models, as they can learn more informative context embeddings than static GloVe embeddings.

Lastly, we compare our performance with other typical data augmentation methods, i.e., EDA [6], Back-translation [7], CBERT [8] and C3DA [27]. Specifically, taking the RGAT and RGAT-BERT as examples, we report the contrastive results in Table 3. As seen, our method outperforms the other counterparts by a clear margin, indicating its superiority.

Table 3. Comparison of different DA methods.

Method	Laptop14		Restaurant14	
	Acc	F1	Acc	F1
RGAT	77.48	73.76	83.30	76.08
+EDA	78.09	74.03	83.61	76.34
+Back-translation	78.02	74.12	83.72	76.70
+CBERT	78.30	74.57	83.70	76.74
+C3DA	78.33	74.66	83.82	**76.99**
+OURS	**78.57**	**74.68**	**84.01**	76.84
RGAT-BERT	78.21	74.40	86.60	81.35
+EDA	78.59	74.82	86.52	81.47
+Back-translation	79.70	75.01	86.85	81.02
+CBERT	78.62	74.96	87.01	82.19
+C3DA	79.16	75.40	87.22	82.69
+OURS	**79.22**	**75.41**	**87.54**	**82.72**

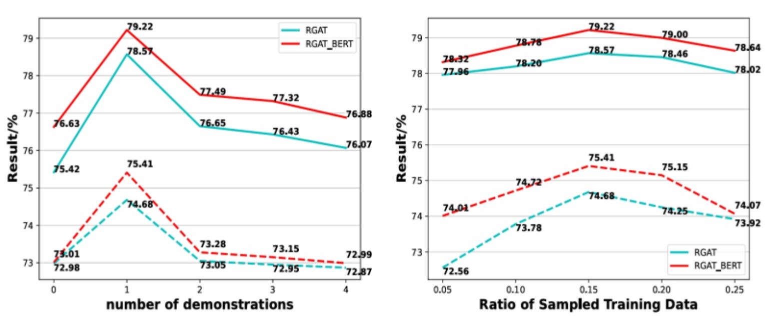
4.3 Analysis and Discussion

Effect of Different Numbers of Demonstrations. Fig. 2 (a) shows the result of different numbers of demonstrations (used in in-context learning), which range from [0,1,2,3,4]. As seen, zero-shot generation performs worst as there is not enough information to guide the generation of LLMs. However, if there are too many demonstrations in the context, the LLMs might misunderstand the instructions, thus leading to sub-optimal performance. When using 1 demonstration, our method performs best, thus leaving as the default setting.

Effect of Different Ratios of Sampled Training Data. As mentioned in Section 3, we randomly sample some instances from the original training data for DA Here, we investigate the effect of difference sampling data ratio and illustrate the results in Fig. 2

(b). As seen, when the sampling ratio is increased, the model performance first increases and then decreases. One possible reason is that the sampling data are relatively similar, i.e., the randomly-sampled data might lack representativeness, thus hindering the effectiveness of our method. This also indicates that a more sophisticated sampling method is required, which is in our future work. When the ratio of sampled data is 15%, our method performs best, and we thus use it as the default setting.

Fig. 2. Ablation study on different numbers of demonstrations(a) and different ratios of sampled data(b). We report the results of RGAT and RGAT BERT on the Laptop14 benchmark.



(a)

(b)

Effect of Different In-Context Learning Strategies. To investigate the effectiveness of our text reconstruction strategies, we remove them and analyze the performance degradation, respectively. The results are listed in Table 4. As seen, compared to the full three strategies, removing any strategy will cause performance degradation, which proves the effectiveness of these strategies. More specifically, when removing the polarity reversal, the performance degradation is the most significant, indicating that the polarity-aware DA takes a more important role in our method.

Table 4. Ablation study on different In-Context Learning strategies. Here, we report the results of RGAT and RGAT-BERT on the Laptop14 benchmark.

Method(onLaptop14)	only-gen		merged	
	Acc	F1	Acc	F1
RGAT+Ours	**74.32**	**70.00**	**78.57**	**74.68**
-w/o rewriting	73.84	69.12	78.12	74.26
-w/o sentence complication	73.69	68.98	78.01	74.32
-w/o polarity reversal	70.47	65.57	78.09	74.15
RGAT-BERT+Ours	**75.29**	**70.41**	**79.22**	**75.41**
-w/o rewriting	74.72	69.85	79.02	75.05
-w/o sentence complication	74.56	69.73	78.65	74.37
-w/o polarity reversal	71.68	66.66	78.84	74.83

Whether Our Method Works Well in other Scenarios. To verify whether our DA method can be used in other tasks, we apply it to the other NLP task, i.e., SST-2 [36], which is a widely-used sentence-level sentiment analysis task. In practice, we use the GPT-3.5-Turbo to obtain the generated training data, and merge them with the original training data. The mixed training data is used to train various baseline models, i.e. BERT-base [4], XLNet-base [37], MT-DNN- base [38] and RoBERTa-base [30]. The contrastive results are listed in Table 5. As seen, our DA method brings consistent and significant performance gains among all baseline models. These results prove that our method can also work well in other sentiment analysis tasks.

Table 5. Results of various Base-size baseline models in the SST-2 [36] task. Notably, "Original training data" denotes that we train the baseline models using the original training data of SST-2, and "+Our generated data" means training on the mix of original and generated training data.

Method	BERT	XLNet	MT-DNN	RoBERTa
Original training data	92.78	93.35	94.30	94.40
+Our generated data	**93.24**	**94.08**	**95.02**	**95.15**

5 Conclusion

We propose a simple-yet-effective DA method for ABSA, namely LLM-as-an-Augmentor, which leverages the powerful capability of LLMs to generate fluent and diverse training data. Three text reconstruction strategies are introduced to guide the data generation of LLMs. Extensive experiments show that our method brings consistent and significant performance gains among 5 baseline models and 3 benchmarks. More analyses also prove that our method outperforms the other DA counterparts by a clear margin. However, we still notice several limitations in our work. For example, our sampled training data for augmentation may not be representative enough, thus influencing the diversity of generated data. Future work could attach more attention to data clustering to obtain more informative demonstrations for LLMs.

References

1. Schouten, K., Frasincar F.: Survey on Aspect-Level Sentiment Analysis. IEEE Transactions on Knowledge and Data Engineering **28**(3), 813-830 (2016)
2. OpenAI, et al.: GPT-4 Technical Report. arXiv preprint arXiv:2303.08774 (2023)
3. Touvron, H., et al.: Llama 2: Open Foundation and Fine-Tuned Chat Models. arXiv preprint arXiv:2307.09288 (2023)
4. Devlin, J., et al.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1, pp. 4171 –4186. Minnesota, ACL, (2019)
5. Zhong, Q.H., et al.: Can chatgpt understand too? a comparative study on chatgpt and finetuned bert. arXiv preprint arXiv:2302.10198 (2023)

6. Wei, J., Zou K.: EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, pp. 6382–6388. ACL, Hong Kong (2019)
7. Fan A, et al.: Beyond English-centric multilingual machine translation. The Journal of Machine Learning Research **22**(1), 4839–4886 (2021)
8. Wu, Xing et al.: Conditional BERT Contextual Augmentation. In: International Conference on Conceptual Structures, pp. 84–95. Springer, Berlin, (2018)
9. Guo, H.Y., Mao, Y.Y., Zhang, R.C.: Augmenting Data with Mixup for Sentence Classification: An Empirical Study. arXiv preprint arXiv:1905.08941 (2019)
10. Tavor, A., et al.: Not enough data? deep learning to the rescue! In: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34. New York (2020)
11. Long, O.Y., et al.: Training language models to follow instructions with human feedback. In: Proceedings of the 36th International Conference on Neural Information Processing Systems, pp. 27730-27744. LA (2022)
12. Wei, J., et al.: Finetuned Language Models Are Zero-Shot Learners. arXiv preprint arXiv: 2109.01652 (2021)
13. Pontiki, M., et al.: SemEval-2014 Task 4: Aspect Based Sentiment Analysis. In: Proceedings of the 8th International Workshop on Semantic Evaluation, pp. 27–35. ACL, Dublin (2014)
14. Pontiki, M., et al.: SemEval-2015 Task 12: Aspect Based Sentiment Analysis. In: Proceedings of the 9th International Workshop on Semantic Evaluation, pp, 486–495. ACL, Denver (2015)
15. Zhang, W., et al.: A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges. IEEE Transactions on Knowledge and Data Engineering **35**(11), 11019-11038 (2022)
16. Li, X., Bing, L., Lam, W., Shi, B.: Transformation networks for target-oriented sentiment classification. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, pp. 946–956. ACL, Melbourne (2018)
17. Tang, D., et al.: Effective lstms for target dependent sentiment classification. In: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pp. 3298–3307. COLING, Osaka (2016)
18. Zhang, C., Li, Q., Song, D.: Aspect-based sentiment classification with aspect-specific graph convolutional networks. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, pp. 4568–4578. ACL, Hong Kong (2019)
19. Wang, K., et al.: Relational graph attention network for aspect-based sentiment analysis. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 3229–3238. Online (2020)
20. Vaswani, A., et al.: Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing, pp. 6000–6010. California (2017)
21. Li, X., et al.: Exploiting BERT for end-to-end aspect-based sentiment analysis. In: Proceedings of the 5th Workshop on Noisy User-generated Text, pp. 34–41. ACL, Hong Kong (2019)
22. Rietzler, A., et al.: Adapt or Get Left Behind: Domain Adaptation through BERT Language Model Finetuning for Aspect-Target Sentiment Classification. In: Proceedings of the Twelfth Language Resources and Evaluation Conference, pp. 4933–4941. ELRA, Marseille (2020)
23. Zeng, B., Yang, H., Xu, R., Zhou, W., Han, X.: LCF: A Local Context Focus Mechanism for Aspect-Based Sentiment Classification. Applied Sciences **9**(16), 3389 (2019)
24. Yu, A.W., et al.: Fast and Accurate Reading Comprehension by Combining Self-Attention and Convolution. In: 6th International Conference on Learning Representations. Vancouver (2018)

25. Fan, L.J., et al.: Improving CLIP Training with Language Rewrites. In: the 37 Annual Conference on Neural Information Processing Systems, vol.36, pp.35544–35575. New Orleans (2023)
26. ChatGPT. https://chat.openai.com/
27. Wang, B., et al.: A Contrastive Cross-Channel Data Augmentation Framework for Aspect-Based Sentiment Analysis. In: Proceedings of the 29th International Conference on Computational Linguistics, pp. 6691–6704. ICCL, Gyeongju (2022)
28. Wu, D.M., et al.: A Novel Counterfactual Data Augmentation Method for Aspect-Based Sentiment Analysis. In: Proceedings of the 15th Asian Conference on Machine Learning, vol. 222, pp. 1479-1493. PMLR (2023)
29. Pennington, J., Socher, R., Manning, C.: GloVe: Global Vectors for Word Representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, pp. 1532–1543. ACL, Doha (2014)
30. Liu, Y.H., et al.: Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692 (2019)
31. Wang, Y., et al.: Attention-based LSTM for aspect-level sentiment classification. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp. 606–615. ACL, Austin (2016)
32. Zhang, C., Li, Q.C., Song, D.: Aspect-based sentiment classification with aspect-specific graph convolutional networks. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, pp. 4568–4578. ACL, Hong Kong (2019)
33. Devlin, J., et al.: Bert: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, pp. 4171– 4186. ACL, Minneapolis (2019)
34. Zhong, Q., et al.: Knowledge Graph Augmented Network Towards Multiview Representation Learning for Aspect-Based Sentiment Analysis. IEEE Transactions on Knowledge and Data Engineering **35**(10), 10098-10111 (2023)
35. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: 3rd International Conference on Learning Representations, Poster. CA (2015)
36. Socher, R., et al.: Recursive deep models for semantic compositionality over a sentiment treebank. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 1631–1642. ACL, Washington (2013)
37. Yang, Z.L., et al.: XLNet: generalized autoregressive pretraining for language understanding. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems, pp. 5753–5763. New York (2019)
38. Liu, X.D., et al.: Multi-Task Deep Neural Networks for Natural Language Understanding. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 4487–4496. ACL, Florence (2019)