A Data Augmentation Approach to Sentiment Analysis of MOOC Reviews

Guangmin Li[1], Long Zhou[2], Qiang Tong[3], Yi Ding[4], Xiaolin Qi[5], Hang Liu[6]
School of Computer and Information Engineering, Hubei Normal University, Huangshi, China[1,2,3,4]
Academic Affairs Office, Wuhan Technology and Business University, Wuhan, China[5]
College of Physical Science and Technology, Central China Normal University, Wuhan, China[6]

***Abstract*—To address the lack of Chinese online course review corpora for aspect-based sentiment analysis, we propose Semantic Token Augmentation and Replacement (STAR), a semantic-relative distance-based data augmentation method. STAR leverages natural language processing techniques such as word embedding and semantic similarity to extract high-frequency words near aspect terms, learns their word vectors to obtain synonyms and replaces these words to enhance sentence diversity while maintaining semantic consistency. Experiments on a Chinese MOOC dataset show STAR improves Macro-F1 scores by 3.39%-8.18% for LCFS-BERT and 1.66%-8.37% for LCF-BERT compared to baselines. These results demonstrate STAR's effectiveness in improving the generalization ability of deep learning models for Chinese MOOC sentiment analysis.**

***Keywords*—Data augmentation; sentiment analysis; MOOC; natural language processing; deep learning**

I. Introduction

The rise of information technology has facilitated the sharing of experiences via online platforms, leading to significant growth in User Generated Content (UGC). Researchers have utilized Natural Language Processing (NLP) and machine learning to extract valuable insights from UGC on topics such as product attribute extraction [1], [2], [3], consumer preference patterns [4], [5], [6], and public sentiment monitoring [7], [8]. These studies enhance text mining applications and support data-driven consumer behavior analysis, product improvement, and market strategies.

Massive Open Online Courses (MOOCs) have generated extensive online course reviews, providing insights into learner preferences and teaching effectiveness. Analyzing these reviews aids educators in improving courses. Automated sentiment analysis, using machine learning algorithms, efficiently processes large volumes of review data, offering insights for student feedback [9], course evaluation [10], and teaching quality assessment [11]. These techniques facilitate data-driven decision-making in education.

Despite the growing interest in sentiment analysis for Chinese MOOC review data, challenges still exist including a scarcity of annotated corpora, leading to overfitting and class imbalance issues. Data augmentation (DA) techniques can expand training datasets while preserving labels. Although DA techniques have been applied in various NLP tasks, including natural language inference [12], [13] and sentiment analysis [14], [15], [16], there remains a research gap in DA strategies for Chinese educational review data.

This paper proposes a novel approach, Semantic Token Augmentation and Replacement (STAR), to address Chinese text augmentation challenges in the educational domain. STAR uses semantic relative distance calculation to augment training datasets and balance sentiment polarity in Chinese MOOC reviews.

The main contributions of this work are as follows:

1) We propose STAR, a novel augmentation method that utilizes external knowledge bases and semantic relative distance calculation to rapidly augment small sample datasets. STAR maintains the semantic accuracy of original review sentences while balancing sentiment distribution in augmented samples. We demonstrate its effectiveness across three BERT model variants (LCFS-BERT, LCF-BERT, and BERT-SPC) for Chinese MOOC sentiment analysis.
2) We conduct a comparative analysis of Word2Vec and BERT Encoding for word vector generation in STAR, employing uniform sampling for synonym replacement. This comparison provides insights into the optimal word embedding approach for data augmentation in aspect-based sentiment analysis of Chinese MOOC reviews.
3) To the best of our knowledge, we first present the synonym replacement-based data augmentation to Chinese review datasets, demonstrating its effectiveness in improving aspect-based sentiment analysis performance in this specific domain.

The rest of this paper is organized as follows: Section II presents related work. Section III and Section IV describe the proposed method in detail. The experimental results and analysis are presented in Section V and Section VI. Section VII concludes our study and points out the research work in the future.

II. Related Work

Data augmentation aims to increase the diversity of training samples by synthesizing new data from existing datasets [17], [18]. It has been widely applied in various NLP tasks, including named entity recognition[19], [20], natural language inference [21], and sentiment analysis [22], [23]. Wei et al. [24] introduced an Easy Data Augmentation (EDA) that employs methods like synonym replacement, random insertion, deletion, and swapping. Their experiments improved significantly performance on text classification tasks, especially for smaller datasets. To address the limitations of random insertion and swapping approaches, Karimi et al. [25] introduced an Easier Data Augmentation (AEDA), which randomly inserts punctuations into the original text while maintaining word order. This method offers easier implementation and preserves all input information, leading to enhanced generalization performance. For sentence-level named entity extraction, Dai et al. [18] improved experimental results through label-oriented token and synonym replacement, demonstrating their approach using Transformer models on biomedical and materials science datasets.

Synonym Replacement (SR) stands out as a simple yet effective data augmentation method. It replaces selected words with synonyms from WordNet or similar words from word embeddings, maintaining the original sentence semantics [26], [27]. Claude et al. [28]implemented synonym replacement and spell checking via Cloud API, achieving 4.3%-21.6% accuracy improvements in deep networks like LSTM and BiLSTM. For Aspect-Based Sentiment Analysis (ABSA), Zhang Rong et al. [29] developed Multi-Level Data Augmentation (MLDA), improving Accuracy and Macro-F1 by 1.2% 3% on the Rest dataset compared to LSA-BERT.

However, most studies have focused on English text, with limited exploration of Chinese text augmentation, particularly for educational review data. To address this gap, we propose Semantic Token Augmentation and Replacement (STAR), a novel approach leveraging semantic relative distance calculation for Chinese MOOC course reviews. STAR utilizes external corpus knowledge to augment training data and balance sentiment polarity distribution. We evaluate STAR using three BERT variants (LCFS-BERT, LCF-BERT, and BERT-SPC), demonstrating significant improvements in model performance and generalization capability for educational data mining tasks.

III. METHODOLOGY

Let $S = \{t_1, t_2, \ldots, t_n\}$ denote a sentence of n tokens, where t_n represents the nth token. We define the aspect term set $A = \{t_{m+1}, \ldots, t_{m+k}\}$ and the opinion term set $O = \{t_{n-k+1}, \ldots, t_n\}$, where $A, O \subseteq S$. The sentiment polarity set is given by $P = \{\text{positive, negative}\}$. The sentiment classification model computes the sentiment polarity of various aspect terms within a review. It computes a function $f : A \times O \rightarrow P$, which maps aspect terms and their associated opinion terms to sentiment polarities based on local context analysis.

For instance, in the sentence, as shown in Fig. 1, *The course is really good, but the picture is blurred.* we have aspect terms A = ”course”, ”picture” and opinion terms O = ”good”, ”blurred”. The model would classify f(”course”, ”good”) as positive and f(”picture”, ”blurred”) as negative, demonstrating its ability to discern different sentiment polarities for various aspects within a review.

The lack of annotated corpus resources presents a significant challenge in training deep learning models, especially in specialized domains requiring expert evaluation of annotation quality. To address this limitation and enhance the generalization capability of neural network models, researchers have employed various data augmentation techniques, including word transformation, sentence order alteration, and back-translation

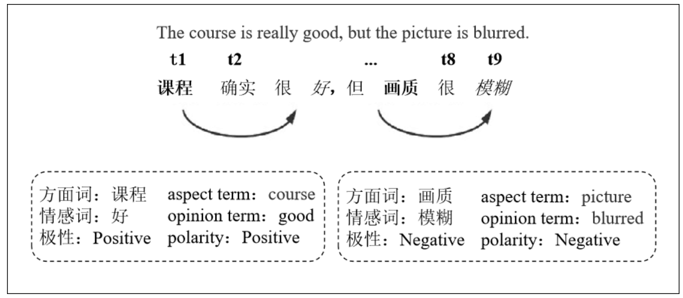

Fig. 1. Course aspect sentiment analysis example.

[16], [24], [30], [31]. These techniques aim to generate diverse training sentences while preserving semantic integrity.

To address these challenges, we propose three data augmentation approaches tailored specifically to Chinese MOOC course review data. These approaches aim to expand the training sample size while preserving the original labels, thereby reducing the loss function value during the deep learning model training process, as shown in Eq. 1. This approach seeks to enhance the extraction and identification of course-related content and its associated sentiment polarities.

$$\arg\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} L(y_i, f(x_i, \theta)) \tag{1}$$

Where $L(,)$ denotes the cross-entropy loss function, θ represents the optimization parameter set, $f(x_i, \theta)$ is the decision function, and N is the number of training samples.

Due to the imbalance in the distribution of sentiment polarity in the training data, we employ Eq. 2 to keep the number of positive and negative samples the same size, thereby addressing the overfitting issue during the training process.

$$N' = (1 + N) \times rate \tag{2}$$

Here, N represents the original dataset sample size, $rate$ denotes the augmentation rate, and N' indicates the sample size in the new dataset after augmentation, accounting for the augmentation rate and sentiment polarity distribution ratio.

IV. PROPOSED DATA AUGMENTATION METHODS

We propose and implement three augmentation methods, namely, Semantic Token Augmentation and Replacement (STAR), Aspect Replacement (AR), and Token Replacement TF-IDF (TRT). We first focus on the STAR method, followed by AR and TRT. The STAR method is the core augmentation approach in this study, leveraging semantic relative distance calculations to enhance the training dataset. The AR method focuses on aspect word replacement, while the TRT method utilizes TF-IDF values to identify replacement candidates. These methods aim to expand the training dataset while preserving original labels, thereby improving model performance in sentiment analysis tasks.

A. Semantic Token Augmentation and Replacement (STAR)

This approach employs token replacement based on semantic relative distance, utilizing Word2Vec to identify similar words and generate a new training dataset. To preserve semantic integrity, the augmentation process avoids replacing words near core words, as these typically better express sentiment or opinion. Our analysis revealed that 75% of sample sentences exceed 18 words in length. Consequently, we set the Semantic-Relative Distance (SRD) threshold to 5. When a non-stop word's semantic relative distance from the core word surpasses this threshold, it is replaced with a similar word, generating a new sentence as shown in Fig. 2.

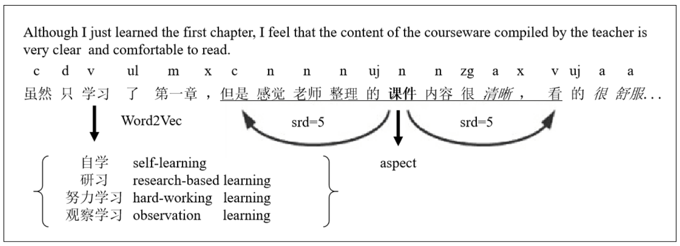

Fig. 2. STAR augmentation approach.

To maintain the semantic relationships of sentiment polarity words in the original sentence, adjective and adverb tokens outside the aspect word area remain unchanged during augmentation.

The SRD calculation process involves the following steps:

1) Initialize the SRD threshold to 5 and set the augmented quantity *augNumI* to 0. While *augNumI* is less than the specified augmentation quantity *augNum*, iterate through the original data and calculate the quantity to be augmented.
2) Designate aspect words from the specified area as stop words. Use the Jieba word segmentation tool to annotate parts of speech and add adjectives to the stop word set. The remaining words become the *randomWords* for augmentation. During candidate word traversal, replace words using similar words from the augmentation dictionary via uniform sampling.
3) Regulate the sample augmentation quantity to ensure the training set's augmentation meets requirements. Save results in *augData* and return the augmented dataset.

Algorithm **??** presents the STAR algorithm's pseudocode, with input parameters including *index* (augmentation dictionary), *dataframe* (data to be augmented), and *augNum* (augmentation quantity). The algorithm outputs the augmented dataset.

B. Aspect Replacement (AR)

This approach leverages BERT for aspect word semantic similarity matching. It randomly selects aspect words from a synonym dictionary using uniform sampling, replaces original aspect words in the sentence with these selections, and updates the corresponding aspect words in the training data.

Algorithm 1 Semantic Token Augmentation and Replacement (STAR) algorithm pseudo-code

```
Input: index, dataframe, augNum;
Output: Augment dataset augData
srd = 5
augNumI = 0
while augNumI ¡ augNum do
    for each row_i in dataframe do
        augProportion = maxNum of augment per sentence
        left = aspect_index - srd
        right = aspect_index + srd
        srdWords = words[left : right + 1]
        adjWords = Filter tags starting with A by jieba
        stopWords = stopWords ∪ srdWords ∪ adjWords
        randomWords = Wordlist ∉ stopWords
        rowAug = 0
        for each word_i in randomWords do
            if word not in index or similar words is NULL then
                continue
            end if
            synonymsList = synonym for word_i in index
            for each synonym_s in synonymsList do
                sentence_n = Replace word in sentence with synonym_s
                augData = sentence ∪ augData
                augNumI = augNumI + 1
                rowAug = rowAug + 1
                if augNumI = augNum then
                    return augData
                end if
                if rowAug ¿ augProportion then
                    break
                end if
            end for
        end for
        if rowAug ¿ augProportion then
            break
        end if
    end for
end while
```

C. Token Replacement TF-IDF (TRT)

Token Replacement TF-IDF (TRT) follows these steps:

1) Calculate the TF-IDF value for each token.
2) Identify replacement candidates: non-stop words, non-aspect words, with a distance greater than 1 from aspect words.
3) For each candidate, compute its weight as $w_i = \text{tfidf}_{\text{max}} - \text{tfidf}$.
4) Determine each candidate's position in the sentence.
5) Using the synonym dictionary in the augmentation index, uniformly sample one token from 10 candidate synonyms for each replacement candidate. If no synonyms are available, retain the original token.
6) Save the augmented dataset and return the results.

These data augmentation methods expand the training dataset while preserving original labels, thereby enhancing model performance in downstream tasks.

V. Experimental Settings

We conducted experiments on a real-world dataset of Chinese MOOC course comments to evaluate the effectiveness of our proposed data augmentation methods. Our experiments focused on aspect-level sentiment classification, comparing the performance of LCFS-BERT, LCF-BERT, and BERT-SPC models across different training set sizes. We analyzed the impact of our data augmentation methods on model performance and evaluated the effectiveness of these methods in enhancing model generalization capability.

A. Data Source

The data used in this study is sourced from course reviews on the China University MOOC website. After data cleaning and annotation, we obtained 1,971 valid data points, which were classified into positive and negative categories based on sentiment polarity. Among them, there are 1,550 positive reviews and 421 negative reviews, with a positive-to-negative sample ratio of approximately 3.7:1. Due to this imbalance in the number of positive and negative samples, there is a certain degree of imbalance in the performance of the training samples across these categories.

We analyzed the distribution of positive and negative polarities across various course aspects, as detailed in Table I. For sentences containing multiple aspect words, we split them into multiple samples to ensure that each sample contains only one aspect word.

B. Experimental Procedure

In the experiment, we first randomly shuffle the original data and then split it into training and testing sets at an 8:2 ratio. To ensure that the testing set and training set are mutually exclusive, meaning that testing samples do not appear during training, we implemented strict data partitioning measures.

Subsequently, we extract 50, 150, 300, and 500 samples from the training set to form new training subsets, ensuring that the sentiment polarity ratio in the new training subsets remains consistent with the original training set. For each new training subset, we use three data augmentation methods to expand it, ensuring that the augmented training data remains balanced in terms of sentiment polarity.

To compare the effectiveness of different data augmentation methods, we kept the sentiment polarity ratio unchanged in the testing set. For example, STAR_150_3 indicates that 150 samples are extracted according to the sentiment polarity ratio in the original training set. Using the STAR augmentation method, we expand it to 600 samples, including the original samples. After replacing aspect terms with T, these data are input into the deep learning model for training.

C. Parameter Configuration

To balance the training efficiency and generalization ability of each model, ensuring good performance on MOOC data, training stops when the loss function value on the validation set is minimal for the LCFS-BERT, LCF-BERT, and BERT-SPC models.

In terms of model training, the Chinese BERT [32] is used as the pre-trained language model, with the Adam optimizer updating the model weights. The learning rate is set to 0.00002 to control the rate at which the model weights are updated. The training dataset is iterated for 10 epochs, with a batch size of 16 for each iteration. To reduce the risk of overfitting and improve the model's generalization ability, a dropout rate of 0.5 is used. Additionally, to control sequence length, reduce the model's computational complexity, and avoid negative impacts on the loss function from overly long sequences, the maximum sequence length is set to 80.

D. Experimental Results

In the experiment, aspect-based sentiment classification models based on BERT were selected: LCFS-BERT [33], LCF-BERT [34], and BERT-SPC [35]. We compared the performance of our three proposed augmentation methods against the baseline results from training on the original dataset. Furthermore, we analyzed the performance of these models across different training set sizes, using Macro-F1 as the evaluation metric for the model's effectiveness.

As shown in Table II, we can see the performance of the three augmentation methods across different sample sizes and models. Bold underlined values indicate the highest score in the current training batch.

With the increase in training data, different augmentation methods showed the varying performance improvement on each model. For instance, after applying the STAR method, the Macro-F1 score of the LCFS-BERT model improved from 71.9% to 92.27%. Models trained with augmentation methods outperformed those trained with original data alone. For the LCFS-BERT model, after using STAR, AR, and TRT methods, the Macro-F1 scores were 92.27%, 78.92%, and 78.2%, respectively. However, for the BERT-SPC model, the original dataset's Macro-F1 score was better than the results with data augmentation methods at 150 training samples. It was mainly due to the introduction of excessive noise text in the augmentation experiment, weakening the semantic correlation between aspect words and sentiment words.

To comprehensively evaluate and compare the effectiveness of the STAR augmentation method across different models and dataset sizes, the model performance curve shown in Fig. 3 visually demonstrates the improvement effect of the STAR augmentation method on model performance.

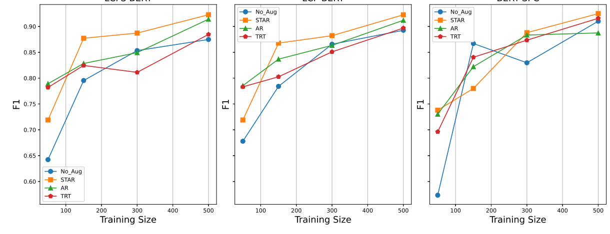

Fig. 3. Model performance comparison chart.

As shown in Fig. 3, the STAR method consistently outperforms the other two augmentation techniques, particularly in enhancing model generalization on small datasets. Compared

TABLE I. STATISTICS FOR SENTIMENT POLARITY OF COURSE ASPECTS

Sentiment Polarity	**Teaching Evaluation**	**Teaching Content**	**Teaching Cost**	**Platform Health**	**Teaching Interaction**	**Teacher**	**Course Structure**	**Course Video**
Positive (count)	62	571	16	11	639	76	117	58
Negative (count)	74	133	3	38	65	23	33	52
Total	136	704	19	49	704	99	150	110
Positive Ratio	45.59%	81.11%	84.21%	22.45%	90.77%	76.77%	78.00%	52.73%

TABLE II. COMPARISON OF MODELS' PERFORMANCE(%)

Model	Method	50	150	300	500
LCFS-BERT	**STAR**	71.90	**87.72**	**88.72**	**92.27**
	AR	**78.92**	82.84	84.93	91.41
	TRT	78.20	82.46	81.12	88.46
	No_Aug	64.23	79.54	85.33	87.48
LCF-BERT	**STAR**	71.90	**86.78**	**88.22**	**92.26**
	AR	**78.54**	83.68	86.34	91.20
	TRT	78.32	80.26	85.09	89.69
	No_Aug	67.79	78.41	86.56	89.26
BERT-SPC	**STAR**	**73.80**	78.00	**88.80**	**92.49**
	AR	73.03	82.22	88.34	88.75
	TRT	69.62	84.04	87.35	91.58
	No_Aug	57.34	**86.70**	82.97	91.02

to training results without augmentation, all three augmentation methods generally improved Macro-F1 scores, with STAR showing outstanding comprehensive performance across all three model types.

VI. DISCUSSION AND ANALYSIS

This paper explores the application of augmentation methods in sentiment analysis tasks for text data. The analysis of experimental results indicates that augmentation methods positively impact model performance. The following analysis is conducted from two aspects: the implementation ideas of the augmentation methods and the performance of the models.

- From the perspective of the implementation ideas of the augmentation methods:
    - The AR augmentation method differs from the other two augmentation methods. It replaces aspect words in sentences and uses BERT Encoding for synonym generation, which can reduce the impact of replacement words on the context. According to experimental results, AR improves the model's classification performance compared to the baseline model, but the improvement is not as significant as that of STAR. The main limitation of AR is that it only replaces aspect terms without altering other tokens in the sentence. Consequently, AR generates sentences with a uniform structure, offering less potential for improving the model's generalization capability compared to the other augmentation methods.
    - For the TRT augmentation method, replacing words within the local context of a term can introduce textual noise. This can impact the model's ability to extract syntactic features, thereby affecting its classification performance.
- From the perspective of model performance:
    - The overall enhancement effect of the LCF-BERT model is slightly inferior to that of the LCFS-BERT model. Considering that the LCF-BERT model calculates semantic relative distance based on word distance, both AR and TRT methods perform word replacements close to aspect words, so the actual word distance does not change. However, for the LCFS-BERT model, which calculates semantic relative distance based on syntax trees, the model can better understand the local context related to aspect words.
    - Both AR and TRT methods perform word replacements within the local context of aspect words, introducing some noise in the model's recognition of local context. Therefore, their effectiveness is weaker than that of the STAR method.

According to experimental results, the STAR augmentation approach shows better enhancement effects on small sample training sets compared to AR and TRT methods and can also quickly augment according to the specified multiplier. Currently, the experimental results only provide a statistical analysis of the classification performance of each model when the original dataset is augmented by a single multiplier. However, for high-multiplier augmentation methods, further experimental verification is needed to compare the performance of the three augmentation methods.

VII. Conclusion and Future Work

In this paper, we propose a novel semantic relative distance-based augmentation method (STAR) to address the lack of labeled data corpus for Chinese course reviews. This method combines augmentation techniques with the BERT pre-trained model to mine the sentiment tendencies in the text. To improve the generalization ability of the deep learning model and discover more sentiment tendencies related to course teaching in the text, the STAR method integrates external knowledge and semantic relative distance calculation and then conducts experiments.

The experimental results show that after using the STAR method, the Macro-F1 values of the LCFS-BERT model and the LCF-BERT model increased by 3.39% to 8.18% and 1.66% to 8.37%, respectively, indicating the effectiveness of this augmentation method. Our experiments are based on aspect-level sentiment classification models using a local context attention mechanism.

In the future, it is necessary to further explore and validate the effectiveness of high-ratio data augmentation and the applicability of the STAR augmentation method under other language models. Additionally, it is essential to deeply explore and optimize data augmentation techniques to adapt to the characteristics of Chinese education review data, thus providing more accurate and effective support for data analysis and decision-making in the education field.

Acknowledgment

This research was supported by Postgraduate Education and Teaching Reform Research Project of Hubei Province (2023392), Funds For Philosophy and Social Science of Hubei Province (21Y145), Huangshi Innovation and Development Joint Fund Project (2024AFD002) and Special Fund for Scientific Planning of Private Higher Education (GA202003).

References

[1] A. Brinkmann, R. Shraga, and C. Bizer, "Product attribute value extraction using large language models," *arXiv preprint arXiv:2310.12537*, 2023.

[2] K. Roy, P. Goyal, and M. Pandey, "Attribute value generation from product title using language models," in *Proceedings of The 4th Workshop on e-Commerce and NLP*, 2021, pp. 13–17.

[3] K. Kumar and A. Saladi, "Pave: Lazy-mdp based ensemble to improve recall of product attribute extraction models," in *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 2022, pp. 3233–3242.

[4] S. M. Babu, P. P. Kumar, S. Devi, K. P. Reddy, M. Satish *et al.*, "Predicting consumer behaviour with artificial intelligence," in *2023 IEEE 5th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA)*. IEEE, 2023, pp. 698–703.

[5] R. Sleiman, K.-P. Tran, and S. Thomassey, "Natural language processing for fashion trends detection," in *2022 International Conference on Electrical, Computer and Energy Technologies (ICECET)*. IEEE, 2022, pp. 1–6.

[6] L. Ren, B. Zhu, and Z. Xu, "Robust consumer preference analysis with a social network," *Information Sciences*, vol. 566, pp. 379–400, 2021.

[7] H. Liang, U. Ganeshbabu, and T. Thorne, "A dynamic bayesian network approach for analysing topic-sentiment evolution," *IEEE Access*, vol. 8, pp. 54 164–54 174, 2020.

[8] R. Koonchanok, Y. Pan, and H. Jang, "Tracking public attitudes toward chatgpt on twitter using sentiment analysis and topic modeling," *arXiv preprint arXiv:2306.12951*, 2023.

[9] X. Chen, F. L. Wang, G. Cheng, M.-K. Chow, and H. Xie, "Understanding learners' perception of moocs based on review data analysis using deep learning and sentiment analysis," *Future Internet*, vol. 14, no. 8, p. 218, 2022.

[10] B. Du, "Research on the factors influencing the learner satisfaction of moocs," *Education and Information Technologies*, vol. 28, no. 2, pp. 1935–1955, 2023.

[11] K. F. Hew, X. Hu, C. Qiao, and Y. Tang, "What predicts student satisfaction with moocs: A gradient boosting trees supervised machine learning and sentiment analysis approach," *Computers & Education*, vol. 145, p. 103724, 2020.

[12] M. Sadat and C. Caragea, "Learning to infer from unlabeled data: A semi-supervised learning approach for robust natural language inference," *arXiv preprint arXiv:2211.02971*, 2022.

[13] J. Li and Y. Ning, "Anti-asian hate speech detection via data augmented semantic relation inference," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 16, 2022, pp. 607–617.

[14] Z. Feng, H. Zhou, Z. Zhu, and K. Mao, "Tailored text augmentation for sentiment analysis," *Expert Systems with Applications*, vol. 205, p. 117605, 2022.

[15] B. Wang, L. Ding, Q. Zhong, X. Li, and D. Tao, "A contrastive cross-channel data augmentation framework for aspect-based sentiment analysis," *arXiv preprint arXiv:2204.07832*, 2022.

[16] G. Li, H. Wang, Y. Ding, K. Zhou, and X. Yan, "Data augmentation for aspect-based sentiment analysis," *International Journal of Machine Learning and Cybernetics*, vol. 14, no. 1, pp. 125–133, 2023.

[17] S. Y. Feng, V. Gangal, J. Wei, S. Chandar, S. Vosoughi, T. Mitamura, and E. Hovy, "A survey of data augmentation approaches for nlp," *arXiv preprint arXiv:2105.03075*, 2021.

[18] X. Dai and H. Adel, "An analysis of simple data augmentation for named entity recognition," *arXiv preprint arXiv:2010.11683*, 2020.

[19] S. Chen, G. Aguilar, L. Neves, and T. Solorio, "Data augmentation for cross-domain named entity recognition," *arXiv preprint arXiv:2109.01758*, 2021.

[20] T. Kang, A. Perotte, Y. Tang, C. Ta, and C. Weng, "Umls-based data augmentation for natural language processing of clinical research literature," *Journal of the American Medical Informatics Association*, vol. 28, no. 4, pp. 812–823, 2021.

[21] J. Singh, B. McCann, N. S. Keskar, C. Xiong, and R. Socher, "Xlda: Cross-lingual data augmentation for natural language inference and question answering," *arXiv preprint arXiv:1905.11471*, 2019.

[22] G. Li, H. Wang, Y. Ding, K. Zhou, and X. Yan, "Data augmentation for aspect-based sentiment analysis," *International Journal of Machine Learning and Cybernetics*, vol. 14, no. 1, pp. 125–133, 2023.

[23] H. Q. Abonizio, E. C. Paraiso, and S. Barbon, "Toward text data augmentation for sentiment analysis," *IEEE Transactions on Artificial Intelligence*, vol. 3, no. 5, pp. 657–668, 2021.

[24] J. Wei and K. Zou, "Eda: Easy data augmentation techniques for boosting performance on text classification tasks," *arXiv preprint arXiv:1901.11196*, 2019.

[25] A. Karimi, L. Rossi, and A. Prati, "Aeda: an easier data augmentation technique for text classification," *arXiv preprint arXiv:2108.13230*, 2021.

[26] X. Zhang, J. Zhao, and Y. LeCun, "Character-level convolutional networks for text classification," *Advances in neural information processing systems*, vol. 28, 2015.

[27] S. Y. Feng, V. Gangal, D. Kang, T. Mitamura, and E. Hovy, "Genaug: Data augmentation for finetuning text generators," *arXiv preprint arXiv:2010.01794*, 2020.

[28] C. Coulombe, "Text data augmentation made simple by leveraging nlp cloud apis," *arXiv preprint arXiv:1812.04718*, 2018.

[29] L. Y. ZHANG Rong, "Multi-level data augmentation method for aspect-based sentiment analysis," *Frontiers of Data and Computing*, vol. 5, pp. 140–153, 2023.

[30] H. Shi, K. Livescu, and K. Gimpel, "Substructure substitution: Structured data augmentation for nlp," *arXiv preprint arXiv:2101.00411*, 2021.

[31] C. Shorten, T. M. Khoshgoftaar, and B. Furht, "Text data augmentation for deep learning," *Journal of big Data*, vol. 8, no. 1, p. 101, 2021.

[32] Y. Cui, W. Che, T. Liu, B. Qin, and Z. Yang, "Pre-training with whole word masking for chinese bert," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 3504–3514, 2021.

[33] M. H. Phan and P. O. Ogunbona, "Modelling context and syntactical features for aspect-based sentiment analysis," in *Proceedings of the 58th annual meeting of the association for computational linguistics*, 2020, pp. 3211–3220.

[34] B. Zeng, H. Yang, R. Xu, W. Zhou, and X. Han, "Lcf: A local context focus mechanism for aspect-based sentiment classification," *Applied Sciences*, vol. 9, no. 16, p. 3389, 2019.

[35] Y. Song, J. Wang, T. Jiang, Z. Liu, and Y. Rao, "Attentional encoder network for targeted sentiment classification," *arXiv preprint arXiv:1902.09314*, 2019.