Big Data and
Information Analytics

BDIA, 10: 81–95.
DOI: 10.3934/bdia.2026005
Received: 08 December 2025
Accepted: 09 January 2026
Published: 18 March 2026

https://www.aimspress.com/journal/BDIA

Research article

Text data augmentation based on semantic roles

Yutong Wang[1], Guole Shang[1], Jianlin Wang[2,*] and Junfang Zhao[1]

[1] School of Science, China University of Geosciences, Beijing 100089, China
[2] Research Institute of Highway Ministry of Transport, Beijing 100088, China

*Correspondence: Email: Wangjl0072000@163.com.

Abstract: The performance of deep learning models in natural language processing (NLP) is heavily dependent on the volume and quality of training data. However, acquiring large amounts of high-quality labeled data is often costly and challenging. Text data augmentation presents a low-cost and efficient solution to expand datasets by generating diversified new samples. A critical limitation of traditional augmentation methods, such as synonym replacement and random manipulation, is the "non-core word replacement" problem. These methods often replace functionally unimportant words (e.g., prepositions, conjunctions), resulting in limited semantic variation and ineffective augmented data. To address this issue, this paper innovatively introduces semantic roles into the data augmentation process, proposing a novel semantic role-based data augmentation (SRDA) method. Our approach consists of three key components: (1) we constructed an advanced MacBERT-CRF model for semantic role labeling (SRL) that outperforms traditional architectures; (2) we developed a tailored masked prediction model with semantic role-level masking, an increased masking ratio, dynamic N-gram masking, and a novel label embedding mechanism; and (3) we demonstrated the method's effectiveness through rigorous evaluation on four distinct Chinese datasets. Experimental results demonstrated that our method consistently outperforms various traditional augmentation techniques across different data volumes and domains, significantly enhancing downstream text classification performance.

Keywords: text data augmentation; semantic role labeling; natural language processing; deep learning; MacBERT

1. Introduction

Deep learning has revolutionized natural language processing by automatically learning features from data. However, its success is profoundly dependent on large-scale, high-quality training datasets [1,2]. Acquiring such data is often impeded by factors like data privacy, access restrictions, and high annotation costs. Data augmentation (DA) emerges as a powerful technique to mitigate these issues by artificially expanding the dataset through various transformations, thereby improving model generalization and robustness [3].

In NLP, while DA is well-established in computer vision, text data augmentation poses unique challenges due to the discrete and compositional nature of language [4]. Common methods include synonym replacement, random insertion/deletion/swapping, back-translation, and more recently, model-based approaches using pre-trained language models like Bidirectional Encoder Representations from Transformers (BERT) [5–9]. A critical drawback of these methods is the random selection of words for manipulation, often leading to the replacement of non-core words (e.g., function words like prepositions and conjunctions) that contribute little to semantic change. Consequently, the generated sentences may lack diversity and fail to provide meaningful new information for model training.

Semantic role labeling (SRL) is a shallow semantic parsing technique that identifies the predicate-argument structure of a sentence, assigning roles such as Agent (Arg0), Patient (Arg1), Temporal (ArgM-TMP), and Locative (ArgM-LOC) to sentence constituents [10]. These semantic roles represent the core elements that convey the primary meaning of a sentence. Leveraging SRL allows for a more strategic approach to data augmentation by specifically targeting these semantically pivotal elements for replacement.

This paper proposes a novel data augmentation method grounded in semantic roles. Our approach first employs an advanced SRL model to identify core semantic constituents within a sentence. It then utilizes a specially designed masked prediction model to generate contextually appropriate and label-consistent words for replacing these identified roles. This ensures that the augmentation process alters the sentence meaning significantly at a semantic level while preserving its grammaticality and overall label, ultimately generating more diverse and higher-quality training samples.

The main contributions of this work are summarized as follows:

1) We propose a new MacBERT-CRF architecture for Chinese SRL, overcoming limitations of traditional models and achieving state-of-the-art performance on the Chinese Proposition Bank (CPB) dataset.

2) We design a novel masked prediction model with semantic role-aware masking strategies and a label embedding mechanism to control the generation process effectively.

3) We introduce a comprehensive semantic role-based augmentation pipeline and empirically demonstrate its superiority over various baseline methods across multiple datasets and data volumes.

4) We provide an in-depth analysis of key factors influencing augmentation performance, including replacement ratio, iteration count, and semantic role selection strategy, offering practical insights for future research and application.

2. Materials and methods

2.1. SRL model based on MacBERT-CRF

2.1.1 Model architecture

Traditional SRL models, often based on bi-directional long short-term memory-conditional random field (Bi-LSTM-CRF) [11–13], face challenges in feature extraction capacity and capturing long-range syntactic dependencies. To address this, we adopt MacBERT , a pre-trained language model optimized for Chinese [14], as our feature encoder. MacBERT's transformer-based architecture offers superior parallel processing, bidirectional context understanding, and efficient fine-tuning compared to recurrent neural network (RNN)-based models. The conditional random field (CRF) layer atop MacBERT handles the structured prediction task of sequence labeling [15], modeling dependencies between output labels. The overall architecture of our MacBERT-CRF SRL model is shown in Figure 1.

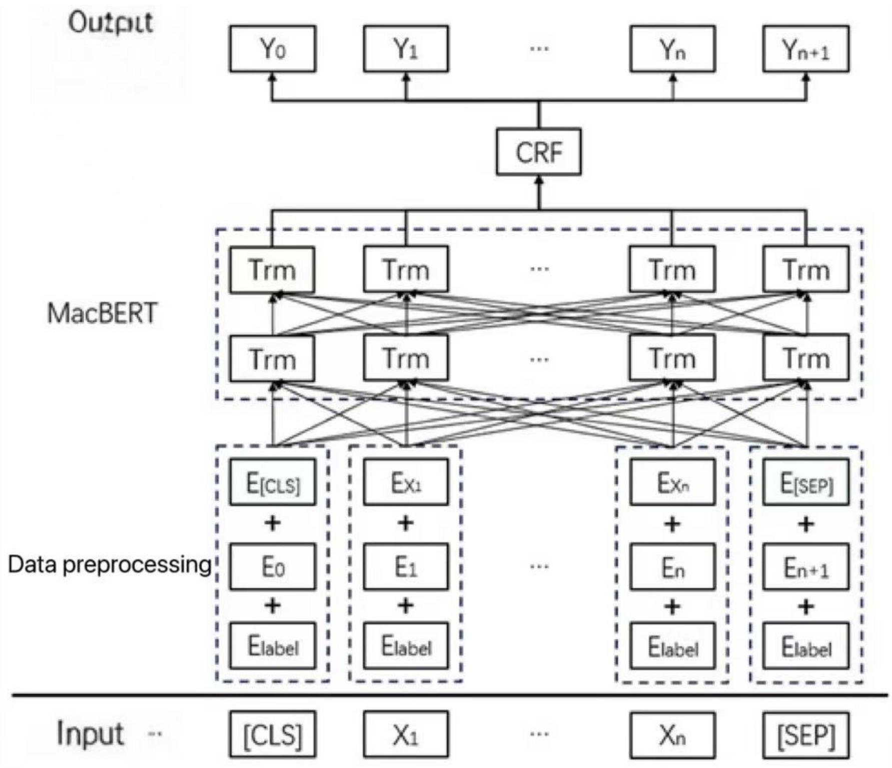

Figure 1. The architecture of the MacBERT-CRF semantic role labeling model.

2.1.2 Experimental setup and optimization

We trained and evaluated our model on the Chinese Proposition Bank (CPB) dataset [16], following the standard 80%/10%/10% split for training, validation, and testing. The dataset uses the Begin, Inside, Outside, End, Single (BIOES) annotation scheme and defines 20 semantic role types.

Hyperparameter tuning was crucial for optimal performance. We conducted experiments on learning rate, training epochs, and feature fusion strategies:

Learning rate: We tested learning rates in the range [0.1, 0.001]. A learning rate of 0.005 yielded the best F1 score (81.62%), balancing convergence speed and stability. Higher rates (0.1, 0.05) caused gradient oscillation, while lower rates (0.001) led to under-convergence within the training budget.

Training epochs: The model started overfitting after 9 epochs, where the validation accuracy peaked at 84.13%. Continuing training to 15 epochs increased training accuracy to 88.02% but decreased validation accuracy to 83.71%, indicating clear overfitting.

Feature fusion: We experimented with fusing different layers of the transformer encoder.

Concatenating the outputs from the last three layers (10th, 11th, 12th) provided the richest feature representation, achieving an F1 score of 81.15%, outperforming using only the final layer (79.73% F1). Fusing more than three layers led to performance degradation due to increased noise and redundancy.

2.1.3 Model comparison

As shown in Table 1, our optimized MacBERT-CRF model significantly outperformed traditional SRL models, including LSTM-CRF, Bi-LSTM-CRF, BERT-CRF, and BERT-BiLSTM-CRF, across accuracy, recall, and F1-score metrics, proving its effectiveness for Chinese SRL.

Table 1. Performance comparison of different semantic role labeling models.

Model structure	Accuracy	Recall rate	F1
LSTM-CRF	69.17%	62.58%	65.71%
BI-LSTM-CRF	74.03%	66.10%	69.84%
BERT-CRF	81.42%	72.94%	76.95%
BERT-Bi-LSTM-CRF	82.13%	74.13%	77.93%
MacBERT-CRF	**86.52%**	**76.40%**	**82.15%**

The 4.22% absolute improvement in F1 score over BERT-Bi-LSTM-CRF demonstrates MacBERT's superiority in Chinese language understanding and the effectiveness of our architectural choice and optimization strategy.

2.2. Masked prediction model based on MacBERT

2.2.1. Model adaptations for augmentation

The core task of the prediction model is to generate a replacement word that fits both the context and the original sentence's label. We chose MacBERT as our base model due to its superior Chinese language understanding.

We made three key adaptations to its pre-training task:

1) Semantic role masking: Instead of masking random tokens, we mask entire contiguous semantic role units identified by our SRL model. This ensures the model learns to generate words that fit specific semantic functions (e.g., generating a new "Agent" that makes sense for the predicate).

2) Increased masking ratio: The proportion of selected words for masking was raised from 15% to 50% to focus learning on semantic roles and compensate for their lower frequency compared to all words.

3) Dynamic N-gram masking: The probability of masking an N-gram semantic role is dynamically adjusted based on its length ($P(N = 1) = 0.4$, $P(N = 2) = 0.3$, $P(N = 3) = 0.2$, $P(N > 3) = 0.8/2$^N). This reflects the natural distribution of semantic role lengths in Chinese.

Furthermore, we replaced the standard segment embeddings in MacBERT's input with label embeddings. Each label in the dataset is assigned a trainable embedding vector, which is added to the token and position embeddings. This allows the model to incorporate label information directly during prediction, ensuring the generated words are consistent with the original sentence's class label. The input construction is illustrated in Figure 2.

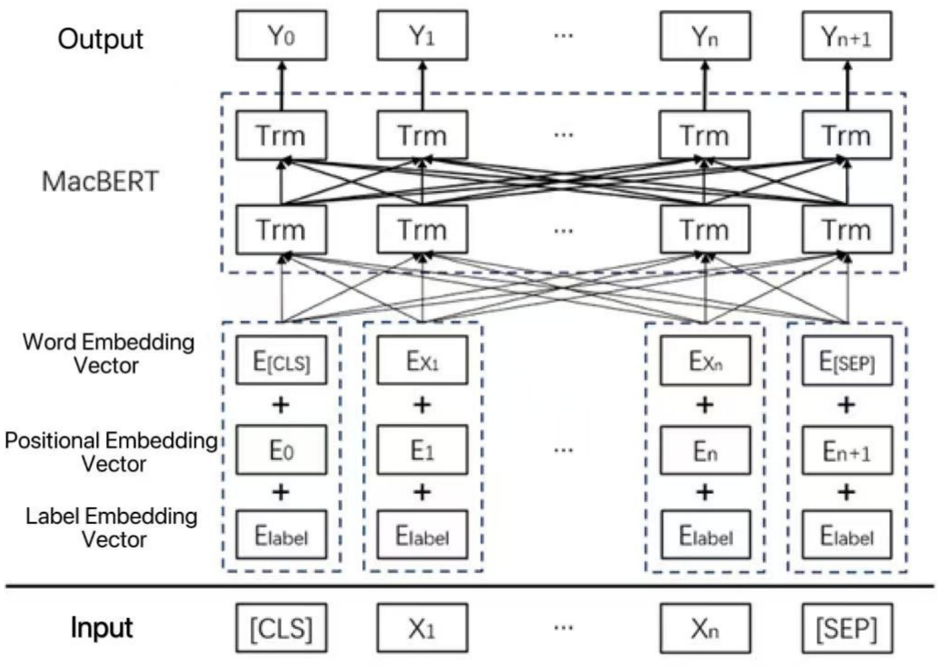

Figure 2. Input structure of the masked prediction model.

2.2.2. Evaluation of generated text

We evaluated the quality of sentences generated by our masked prediction model on four datasets using three metrics:

Jaccard similarity: measured lexical overlap (low values: 0.32–0.49, indicating sufficient surface-level difference).

BERT score: measured semantic similarity (high values: 0.86–0.93, indicating semantic preservation).

Perplexity (PPL): measured fluency using the BigScience Large Open-science Open-access Multilingual Language Model (BLOOM) model. Low absolute PPL and small changes (<13) indicated generated sentences remained coherent.

Results confirmed our model produces sentences that are lexically different, semantically equivalent, and fluent. The detailed results are presented in Table 2, showing the model's consistent performance across diverse datasets.

Table 2. Evaluation results of generated text across different datasets.

Dataset	Jaccard similarity	BERTScore	Sub-dataset PPL	Aug-dataset PPL	Δ Perplexity (ΔPPL)
Toutiao News	0.49	0.92	40.64	45.41	4.77
THUCNews (Cnews)	0.32	0.86	147.42	160.41	12.99
SMP2020 (Weibo)	0.43	0.89	50.21	54.16	3.95
Online Shopping Reviews	0.44	0.93	67.05	75.97	8.92

2.3. Semantic role-based data augmentation pipeline

The overall augmentation process, depicted in Figure 3, is straightforward:

1) SRL tagging: For a given sentence and its label, use the trained MacBERT-CRF model to identify all semantic roles.

2) Role selection: Randomly select a subset of these semantic roles for replacement (e.g., 50%).

3) Masked prediction and replacement: Mask the tokens within the selected roles. Use the trained masked prediction model to generate candidate words. Replace the original tokens with the most probable candidate that differs from the original.

4) Dataset expansion: Add the newly generated sentence with the original label to the training set.

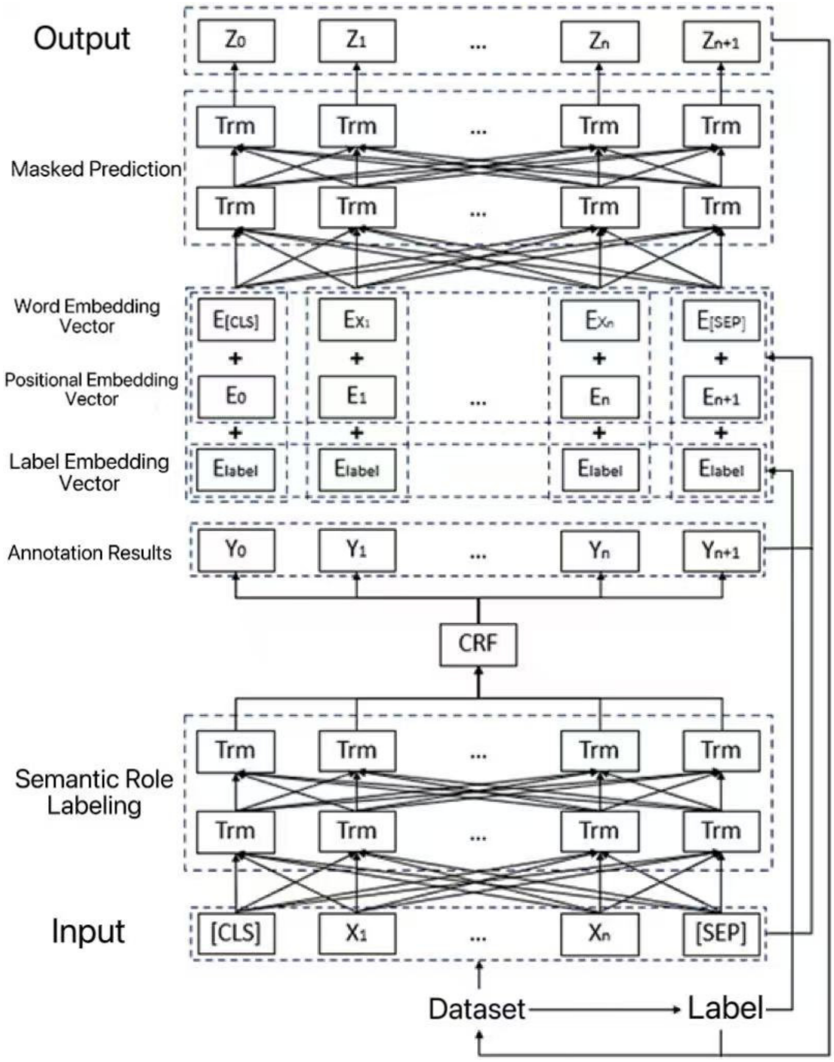

Figure 3. Workflow of the proposed semantic role-based data augmentation method.

2.4. Datasets

We used four Chinese datasets with varying characteristics to validate our method's generality:

Toutiao News: 382,688 short news headlines, 15 categories. (Avg. length: 23 chars)

THUCNews (Cnews): 74,000 long news articles, 14 categories. (Avg. length: 923 chars)

SMP2020 (Weibo): 7768 social media posts, 6 emotion categories. (Avg. length: 43 chars)

Online Shopping Reviews: 62,774 product reviews, 2 sentiment categories. (Avg. length: 58 chars)

This diversity in domain, length, and task type ensures our evaluation is comprehensive and robust. The statistical profiles of the four datasets utilized in this study are summarized in Table 3.

Table 3. Statistical information of experimental datasets.

Dataset	Classes	Samples	Avg. length
Toutiao News	15	382,688	23
THUCNews (Cnews)	14	65,000	923
SMP2020 (Weibo)	6	27,768	43
Online Shopping Reviews	2	62,774	58

3. Results

3.1. Ablation study on augmentation factors

We analyzed how different factors influence the augmentation effectiveness using LSTM classifiers. All experiments in this section start from a base subset of 500 samples per dataset.

3.1.1. Impact of SRL model quality

The accuracy of semantic role identification is fundamental to our method. We compared the downstream classification accuracy achieved when using different SRL models to drive the augmentation process. Table 4 clearly shows that using a more accurate SRL model (MacBERT-CRF > BERT-CRF > Bi-LSTM-CRF) to select words for replacement consistently led to better performance of the classifier trained on the augmented data. This confirms that precise semantic role identification is crucial, as it ensures that the most semantically meaningful units are targeted for replacement, leading to higher-quality augmented samples.

Table 4. Impact of SRL model quality on final classification accuracy.

Semantic role annotation model	Toutiao News	THUCNews (Cnews)	SMP2020 (Weibo)	Online Shopping Reviews
No semantic role annotation model	51.24%	81.13%	60.45%	70.78%
BiLSTM-CRF	51.97%	81.56%	60.89%	71.55%
BERT-CRF	52.43%	81.89%	61.09%	71.93%
BERT-BiLSTM-CRF	52.52%	81.98%	61.26%	72.21%
MacBERT-CRF	**52.71%**	**82.34%**	**61.87%**	**72.41%**

3.1.2. Impact of semantic role replacement ratio

The replacement ratio (percentage of semantic roles replaced in a sentence) is a critical hyperparameter. As shown in Table 5, its effect is nonlinear and consistent across datasets. A ratio of 40–60% proved optimal. Lower ratios (e.g., 20%) produced insufficient diversity, limiting the performance gain. Higher ratios (80–100%) caused excessive semantic distortion, altering the core meaning of the sentence and effectively introducing label noise, which harmed classifier performance. This finding underscores the importance of striking a balance between diversity and semantic fidelity.

Table 5. Impact of the semantic role replacement ratio on classification accuracy.

Semantic role substitution ratio	Toutiao News	THUCNews (Cnews)	SMP2020 (Weibo)	Online Shopping Reviews
20%	51.41%	81.15%	60.72%	71.41%
40%	52.70%	82.46%	**62.33%**	72.45%
60%	**52.93%**	**82.51%**	61.09%	**72.76%**
80%	51.74%	81.90%	60.72%	72.17%
100%	51.35%	81.47%	60.32%	71.80%

3.1.3. Impact of augmentation iterations

Iterative augmentation (applying augmentation multiple times to the growing dataset) can further expand data but risks noise accumulation. Table 6 shows the performance trend across multiple datasets. Performance peaks around 3 iterations and then declines. Early iterations add beneficial diversity. However, later iterations compound small semantic shifts and errors from previous rounds, leading to semantic drift and a degradation in data quality. This effect is more pronounced in longer-text datasets like THUCNews, where more roles are replaced per sentence, accelerating the drift.

Table 6. Impact of augmentation iterations on model accuracy.

Number of data enhancement iterations	Toutiao News	THUCNews (Cnews)	SMP2020 (Weibo)	Online Shopping Reviews
1	52.70%	82.20%	61.94%	72.35%
2	52.91%	**82.44%**	62.40%	72.44%
3	52.93%	82.39%	**62.56%**	**72.69%**
4	53.11%	81.96%	62.38%	72.46%
5	**53.20%**	81.45%	61.81%	71.71%
6	52.96%	81.10%	61.63%	71.30%

3.1.4. Impact of the semantic role selection strategy

Not all semantic roles contribute equally to a sentence’s meaning. We investigated whether selectively replacing only the most frequent and impactful roles (Arg0, Arg1, ArgM-ADV, ArgM-TMP, REL) could yield better results than replacing all roles. Table 7 shows that this selective strategy yielded comparable or even better performance than replacing all roles, while being more computationally efficient. Replacing only Arg1 and REL, the two core arguments, provided a solid baseline, but adding adjunct roles (ArgM-ADV, ArgM-TMP) captured more nuanced semantic variations, leading to the best performance. This strategy helps avoid replacing rare or peripheral roles that might introduce noise rather than meaningful variation.

Table 7. Impact of the semantic role selection strategy on classification accuracy.

Semantic role substitution combinations	Toutiao News	THUCNews (Cnews)	SMP2020 (Weibo)	Online Shopping Reviews
All semantic roles	55.72%	83.56%	63.64%	77.10%
Arg0+Arg1+ArgM-ADV+ArgM-TMP+REL	55.79%	**83.81%**	**63.81%**	**77.58%**
Arg0+Arg1+ArgM-ADV+REL	**55.81%**	83.74%	63.70%	75.41%
Arg0+Arg1+REL	54.61%	83.40%	63.21%	74.87%
Arg1+REL	53.87%	82.81%	62.87%	73.94%

3.2. Comparison with traditional data augmentation methods

We compared our method (SRDA) against several strong baselines: synonym replacement (SR), random insertion (RI), random swap (RS), random deletion (RD), EDA (combining SR, RI, RS, RD) [7], TF-IDF replacement [8], back translation (BT) [5], and CBERT [9]. We conducted two types of experiments:

3.2.1. Fixed data volume

The effectiveness of data augmentation techniques exhibits a significant correlation with dataset scale. To systematically evaluate the performance of different augmentation methods, it is essential to maintain identical dataset sizes. In this experiment, while preserving the original class distribution, a subset of 500 samples was extracted from each dataset. Data augmentation methods were subsequently applied to these subsets. The augmented datasets were then used to train both RNN and LSTM models as classifiers. The performance of these models serves as the metric for assessing the comparative effectiveness of the augmentation strategies. The specific experimental results are presented below:

Table 8. Comparative results on the Toutiao News subset.

Data augmentation method	RNN	LSTM
No augmentation	48.25%	49.62%
Synonym replacement	50.34%	51.62%
Random insertion	49.54%	51.25%
Random swap	49.70%	50.96%
Random deletion	49.31%	51.01%
TF-IDF replacement	49.47%	50.17%
Back translation	51.51%	53.45%
EDA	52.57%	54.78%
CBERT	52.69%	55.22%
Semantic role-based augmentation	**53.21%**	**55.79%**

Table 9. Comparative experimental results on the Cnews dataset subset.

Data augmentation method	RNN	LSTM
No augmentation	71.21%	78.12%
Synonym replacement	76.98%	81.90%
Random insertion	76.65%	81.42%
Random swap	76.77%	80.95%
Random deletion	75.53%	81.65%
TF-IDF replacement	76.14%	82.87%
Back translation	77.23%	83.51%
EDA	79.51%	83.87%
CBERT	79.53%	84.41%
Semantic role-based augmentation	**80.21%**	**84.87%**

Table 10. Comparative experimental results on the SMP2020 dataset subset.

Data augmentation method	RNN	LSTM
No augmentation	57.41%	60.25%
Synonym replacement	62.01%	62.87%
Random insertion	61.51%	62.97%
Random swap	61.39%	62.15%
Random deletion	61.11%	62.09%
TF-IDF replacement	60.82%	62.14%
Back translation	61.90%	63.07%
EDA	62.57%	63.34%
CBERT	62.61%	63.41%
Semantic role-based augmentation	**62.79%**	**63.63%**

Table 11. Comparative experimental results on the Online Shopping Reviews dataset subset.

Data augmentation method	RNN	LSTM
No augmentation	67.25%	71.15%
Synonym replacement	71.56%	73.95%
Random insertion	71.11%	74.12%
Random swap	70.23%	73.65%
Random deletion	70.56%	73.05%
TF-IDF replacement	69.80%	73.26%
Back translation	72.14%	74.02%
EDA	73.30%	74.58%
CBERT	72.98%	74.89%
Semantic role-based augmentation	**74.21%**	**75.40%**

3.2.2. Variable data volume (5% to 100% of original data)

The efficacy of data augmentation methods is influenced by the scale of the dataset, which inherently varies across different datasets. To investigate the sensitivity of augmentation techniques to dataset size, this study employed a stratified sampling strategy: while maintaining the original class distribution, subsets were generated by sampling 5%, 10%, 20%, 40%, 60%, 80%, and 100% of the data from each source. Following the application of data augmentation to each subset, LSTM models were trained and evaluated. The performance of these classifiers was then compared to analyze the impact of varying data volumes on the effectiveness of the augmentation methods. The experimental results are presented in the figures below.

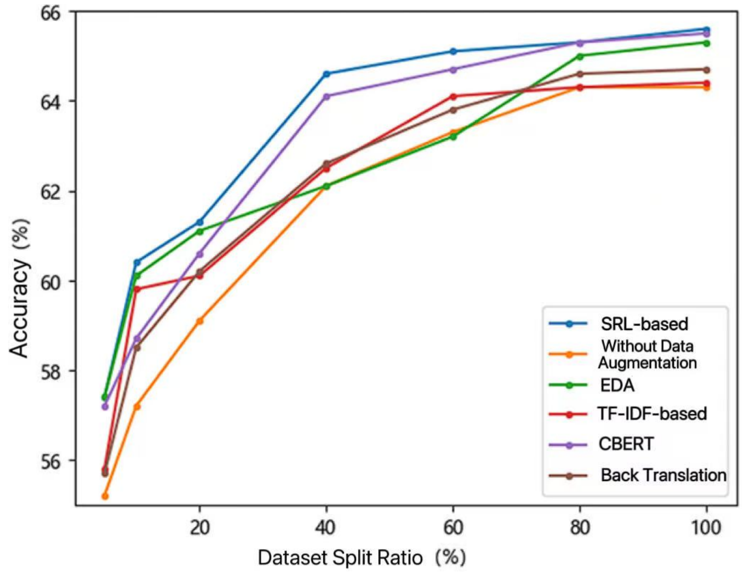

Figure 4. Accuracy trends on the Toutiao dataset under different training data volumes using LSTM.

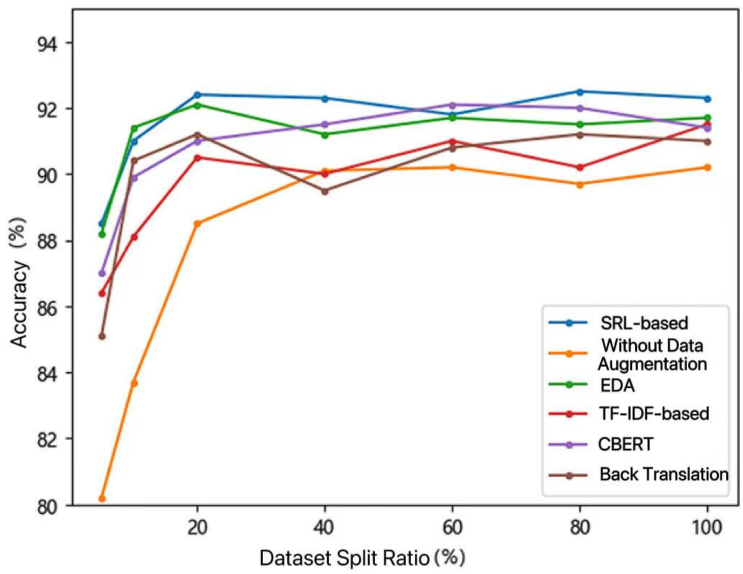

Figure 5. Accuracy trends on the Cnews dataset under different training data volumes using LSTM.

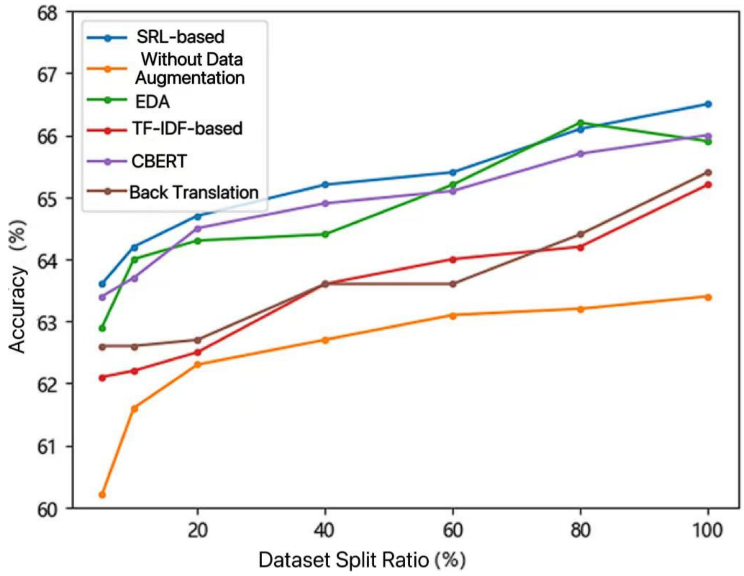

Figure 6. Accuracy trends on the SMP2020 dataset under different training data volumes using LSTM.

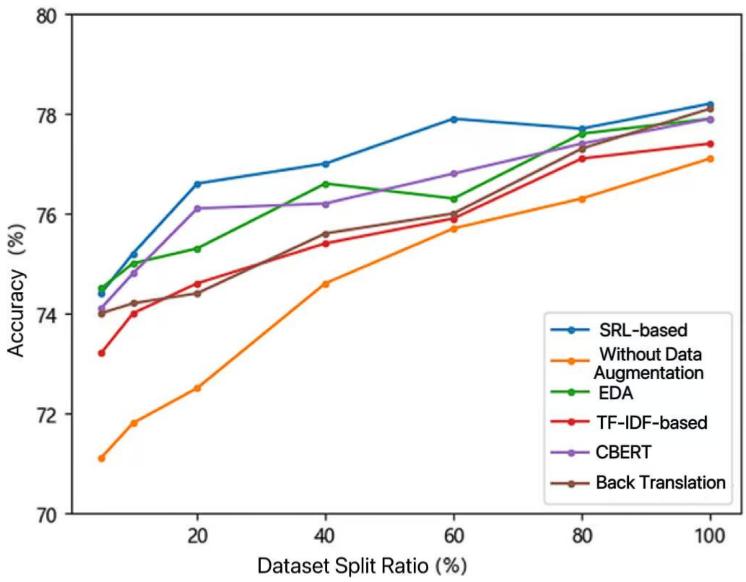

Figure 7. Accuracy trends on the SMP2020 dataset under different training data volumes using LSTM.

Experimental results demonstrate a clear correlation between the efficacy of data augmentation techniques and the scale of the training set: when the amount of available annotated data is limited, all data augmentation methods significantly improve the accuracy of the classification model. As the volume of training samples continues to increase, although data augmentation methods still effectively enhance model performance, the magnitude of improvement gradually diminishes.

It is noteworthy that the semantic role-based data augmentation method proposed in this study consistently demonstrates stable and superior performance compared to traditional augmentation approaches in most experimental scenarios. This indicates that semantic role analysis can effectively capture deep semantic features, and that incorporating semantic roles into data augmentation significantly enhances model generalization and robustness.

4. Discussion

Our experimental results comprehensively validate the proposed semantic role-based data augmentation method. The ablation studies reveal several key insights:

First, the performance of the SRL model is paramount. It acts as the precision guidance system for our method. An inaccurate SRL model would target the wrong words for replacement, rendering the subsequent generation step ineffective. The significant performance gap between different SRL models underscores that advancement in SRL technology directly translates to improvement in augmentation quality.

Second, the existence of an optimal replacement ratio (40–60%) confirms a fundamental trade-off in data augmentation: diversity versus fidelity. Too little change fails to provide new learning signals, while too much change corrupts the original meaning and introduces noise. This optimal range provides a practical guideline for applying our method.

Third, the danger of iterative augmentation beyond a certain point (3 iterations in our case) reveals a limitation of sequential augmentation methods. The errors and semantic shifts compound, leading to data degradation. This suggests that for large-scale augmentation, generating all new samples from the original pristine data, rather than iteratively from previously augmented data, might be preferable.

The consistent and significant outperformance of our method over all traditional techniques is the strongest evidence for our hypothesis. By leveraging linguistic structure (semantic roles), we move beyond naive random manipulation. We are not merely altering the surface form of the text but are making strategic, meaningful changes to its semantic core. This results in augmented data that is not only diverse but also semantically coherent and valuable for model learning.

The success of selective role replacement further indicates that not all semantic changes are equally valuable. Focusing on the core roles (agents, patients, predicates) and key modifiers (time, manner) yields the best return on investment, maximizing impact while minimizing the risk of introducing nonsense.

5. Conclusions

This paper presented a novel text data augmentation method based on semantic roles. To overcome the limitations of existing methods, we developed an advanced MacBERT-CRF model for accurate semantic role labeling in Chinese and a tailored MacBERT-based masked prediction model capable of generating contextually and label-appropriate words. By integrating these components, our method selectively replaces core semantic elements in sentences, effectively addressing the "non-core word replacement" problem pervasive in traditional augmentation techniques.

Extensive experiments on four Chinese datasets demonstrate that our method significantly outperforms various strong baselines. It consistently improves the performance of downstream text classification models, especially in low-data scenarios, by generating higher-quality and more diverse training samples. Our detailed analysis of key factors provides valuable insights for practitioners to apply the method effectively.

Future work will focus on extending this method to other languages and NLP tasks beyond classification, such as question answering and summarization. We also plan to explore more sophisticated generation control mechanisms to further enhance the quality and diversity of the generated samples.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

The authors declare that they have no conflicts of interest.

References

1. Munappy A, Bosch J, Olsson HH, Arpteg A, Brinne B, (2019) Data management challenges for deep learning, In: *2019 45th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, 140–147. https://doi.org/10.1109/SEAA.2019.00030
2. Li B, Hou Y, Che W, (2022) Data augmentation approaches in natural language processing: A survey. *Ai Open* 3: 71–90. https://doi.org/10.1016/j.aiopen.2022.03.001
3. Feng SY, Gangal V, Wei J, Chandar S, Vosoughi S, Mitamura T, et al. (2021) A survey of data augmentation approaches for NLP, In: *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, 968–988. https://aclanthology.org/2021.findings-acl.84
4. Sennrich R, Haddow B, Birch A, (2016) Improving neural machine translation models with Monolingual data, In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 86–96.
5. Kobayashi S, (2018) Contextual augmentation: Data augmentation by words with paradigmatic relations, In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 452–457. https://doi.org/10.18653/v1/N18-2072
6. Wei J, Zou K, (2019) 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 (EMNLP-IJCNLP)*, 6382–6388. https://doi.org/10.18653/v1/D19-1670
7. Wu X, Lv S, Zang L, Han J, Hu S, (2019) Conditional bert contextual augmentation, In: *International Conference on Computational Science*, 84–95. https://doi.org/10.1007/978-3-030-22747-0_7
8. Xie Q, Dai Z, Hovy E, Luong T, Le Q, (2020) Unsupervised data augmentation for consistency training. *Adv Neural Inf Process Syst* 33: 6256–6268.
9. Cui Y, Che W, Liu T, Qin B, Wang S, Hu G, (2020) Revisiting pre-trained models for Chinese natural language processing, In: *Findings of the Association for Computational Linguistics: EMNLP 2020*, 657–668. https://doi.org/10.18653/v1/2020.findings-emnlp.58
10. Gildea D, Jurafsky D, (2002) Automatic labeling of semantic roles. *Comput Linguist* 28: 245–288. https://doi.org/10.1162/089120102760275983
11. Zhang H, (2018) *Research on Semantic Role Labeling Based on Bi-LSTM-CRF*, Master's thesis, Beijing Institute of Technology.

12. Wang Y, Wan F, Ma N, (2020) Chinese semantic role labeling integrating multi-level features. *CAAI Trans Intell Syst* 15: 107–113.
13. Wang XH, Li R, Wang ZQ, Chai QH, Han XQ, (2022) Syntax-aware Chinese frame semantic role labeling based on self-attention. *J Chin Inf Process* 36: 38–44.
14. Devlin J, Chang MW, Lee K, Toutanova K, (2019) BERT: Pre-training of deep bidirectional transformers for language understanding, In: *Proceedings of NAACL-HLT 2019*, 4171–4186. https://doi.org/10.18653/v1/N19-1423
15. Lafferty J, McCallum A, Pereira FCN, (2001) Conditional random fields: Probabilistic models for segmenting and labeling sequence data, In: *Proceedings of ICML 2001*, 282–289.
16. Palmer M, Xue N, (2009) Adding semantic roles to the Chinese Treebank. *Natural Lang Eng* 15: 143–172. https://doi.org/10.1017/S1351324908004902