

Research on Mini - Program for Chinese Language Style Transfer Based on Deep Learning and AI Enhancement

Yiming Chen

Shenzhen MSU-BIT University, Shenzhen, Guangdong, 518172, China

ABSTRACT

This paper presents a mini - program for Chinese language style transfer using deep learning and AI enhancement. It can transform modern texts into the classical style of "Dream of the Red Chamber" and includes a text analysis system. The system assesses classical word usage, sentence complexity, and literary quality in the input text. But the program has limitations like weak recognition of classical verbs and nouns, and obvious addition of classical connectives. This study analyzes its working principles, evaluates its performance, suggests improvements, and explores future research directions. Also, the mini - program shows potential in finishing unfinished literary works and classical literature research.

KEYWORDS

Chinese language style transfer; Deep learning; AI enhancement; Mini - program; Text analysis; Classical literature research

1 Introduction

1.1 Research background

In an era of globalization and digitization, cultural inheritance and language innovation are of great significance. Chinese, a language with a long history and rich resources, carries the history and culture of the Chinese nation through its diverse styles. Converting modern texts into classical literary styles helps preserve traditional culture and brings new vitality to literature, education, and communication.

Many outstanding literary works in history remained unfinished due to authors' deaths, illnesses, or accidents. These incomplete works, such as "Dream of the Red Chamber," have great cultural and artistic value but are hard to fully appreciate. With the development of natural language processing technology, it's possible to create a tool for finishing such works.

Studying classical literature also faces many challenges. The complex language of classical works and the changing times make understanding their meanings and cultural backgrounds difficult. So, using modern technology to assist classical literature research is of great practical significance.

1.2 Research purpose

This study aims to develop an efficient and accurate Chinese language style transfer mini - program to convert modern texts into the classical style of "Dream of the Red Chamber." It also aims to build a text analysis system to evaluate the converted texts and offer feedback and suggestions.

The mini - program aims to solve the problem of finishing unfinished literary works. By learning the unique language features and style patterns of the original author, it can generate texts in the same style, helping to complete the unfinished parts and protect cultural heritage. It also aims to play a role in classical literature research by assisting scholars in text comparison and analysis, digging up hidden information in literary works, promoting the digital study of classical literature, and contributing to its inheritance and development.

2 Related work

2.1 Research status of chinese language style transfer

With the rapid development of natural language processing technology, the field of Chinese language style transfer has made great progress in recent years. Some studies use rule - based methods, achieving style transfer through predefined vocabulary replacement and sentence pattern transformation rules. Others use machine learning and deep learning technologies, such as RNN, LSTM, and Transformer models, to learn text style features and perform transfers.

However, most existing studies focus on general style transfer tasks and don't specifically target the problem of finishing unfinished literary works. Also, there's limited research on using style transfer technology to support classical literature research. This study combines language style transfer technology with the goals of restoring incomplete

masterpieces and classical literature research, filling this gap.

2.2 Development of similar mini - programs

There are some Chinese language style transfer mini - programs on the market, but most have limited functions and unsatisfactory conversion effects. These mini - programs mainly focus on basic style transfer and lack the ability to accurately capture the unique styles of specific authors, which is crucial for finishing unfinished works and conducting in - depth classical literature research.

3 Functions and principles of the mini - program

3.1 Main functions of the mini - program

3.1.1 Text style transfer function

The core function of the mini - program is to convert input modern texts into the classical literary language style of "Dream of the Red Chamber." This function is not only suitable for general style transfer but also plays a key role in finishing unfinished literary works. By converting modern texts into the original author's style, the program can generate texts that seamlessly connect with the existing parts of the work.

In classical literature research, this function is also highly valuable. For example, when comparing different versions of classical works or their translations and adaptations, converting them into the original work's style can make comparisons more intuitive and accurate, helping scholars understand the essence of the original work in different versions.

3.1.2 Text analysis function

The mini - program also has a text analysis function that can comprehensively evaluate the input text, including the number of classical words used, sentence pattern complexity, and literary color scores. This function is crucial for understanding the original author's style features. By analyzing the existing text of unfinished works, the program can identify the author's unique language features and style patterns, which can guide the generation of new texts.

In classical literature research, the text analysis function can provide scholars with in - depth insights into the structure and style of classical works. It can help scholars identify key elements that contribute to a work's literary value, such as the use of specific vocabulary, sentence rhythm, and overall literary atmosphere. This information can be used to further explore the author's creative techniques and the cultural background of the work.

3.2 Implementation principles of the mini - program

3.2.1 Application of deep learning models

The mini - program uses deep learning models to learn the language feature patterns of "Dream of the Red Chamber." These models can capture the complex relationships between words, sentences, and styles in the text. By training on a large amount of "Dream of the Red Chamber" text, the model can learn the unique vocabulary, grammar, and rhetorical devices used by the author. This knowledge is then applied to generate new texts in the author's style.

In the process of classical literature research, these deep learning models can also be used to analyze the language features of other classical works. By training the models on different classical texts, scholars can gain a more comprehensive understanding of the language and style features of different periods and authors.

3.2.2 Use of AI enhancement technology

To improve the quality and accuracy of style transfer, the mini - program introduces AI enhancement technology. AI enhancement can further optimize the generated text by considering context, semantics, and style requirements. For example, it can adjust word selection, sentence structure, and rhetorical devices to make the text more natural and in line with the author's style. This technology is especially important for finishing unfinished works, as it ensures that the newly generated text perfectly matches the existing parts.

In classical literature research, AI enhancement technology can help improve analysis results. For example, when analyzing classical texts, AI can more accurately interpret obscure words and phrases and suggest possible connections between different parts of the text, greatly improving the efficiency and quality of research.

3.2.3 Construction of the text analysis system

The text analysis system is based on natural language processing technology. It analyzes the vocabulary, grammar, semantics, and other aspects of the text to evaluate its style features. The system can provide detailed information on the original author's style features, such as the frequency of certain word usage, the use of specific sentence patterns, and overall literary color. This information guides the style transfer process to ensure that the generated text is consistent with the author's style.

In the field of classical literature research, the text analysis system can be used to build a classical literature database. By analyzing a large number of classical texts, scholars can identify common patterns and unique features, which are important for studying literary history, cultural evolution, and author styles.

3.3 Core code and pseudocode

3.3.1 Pseudocode for style transfer

```
# Get input text and conversion intensity
input_text = get_input_text()
intensity = get_conversion_intensity()
# Replace modern words with classical words
for modern_word, classical_word in word_replacements:
    if random.random() < intensity and modern_word in input_text:
        input_text = input_text.replace(modern_word, classical_word)
# Add classical connectives
if random.random() < intensity * 0.6:
    connector = random.choice(classical_connectors)
    input_text = connector + input_text
# Replace "yes" with classical modifiers
if 'yes' in input_text and random.random() < intensity * 0.5:
    modifier = random.choice(classical_modifiers)
    input_text = input_text.replace('yes', modifier)
# Add classical ending words
if random.random() < intensity * 0.3:
    ending = random.choice(classical_endings)
    input_text = input_text.rstrip('。! ? ') + ending
# Output the converted text
output_text = input_text
```

3.3.2 Pseudocode for text analysis

```
# Get input text
input_text = get_input_text()
# Split sentences
sentences = split_sentences(input_text)
# Tokenize
tokens = tokenize(input_text)
# Calculate average sentence length
avg_sentence_length = sum(len(s) for s in sentences) / len(sentences) if sentences else 0
# Calculate total number of words
total_tokens = len(tokens)
# Calculate number of unique words
unique_tokens = len(set(tokens))
# Calculate lexical diversity
lexical_diversity = unique_tokens / total_tokens if total_tokens else 0
# Analyze part - of - speech distribution
pos_distribution = analyze_pos_distribution(tokens)
# Calculate proportion of classical words
classical_word_count = count_classical_words(tokens)
classical_word_ratio = classical_word_count / total_tokens if total_tokens else 0
```

```
# Output analysis results
analysis_result = {
    'avg_sentence_length': avg_sentence_length,
    'total_tokens': total_tokens,
    'unique_tokens': unique_tokens,
    'lexical_diversity': lexical_diversity,
    'pos_distribution': pos_distribution,
    'classical_word_ratio': classical_word_ratio
}
```

4 Problems of the mini - program and improvement measures

4.1 Existing problems

4.1.1 Weak recognition of classical verbs and nouns

The mini - program currently has weak recognition of classical verbs and nouns. This leads to inaccurate conversion of modern verbs and nouns into appropriate classical vocabulary during style transfer. This issue can seriously affect the quality of the generated text, especially when finishing unfinished works and conducting classical literature research. Inaccurate vocabulary replacement can cause style and semantic mismatches, making the newly generated text inconsistent with the existing parts of the work and possibly leading to misinterpretation of classical works.

4.1.2 Obvious addition of classical connectives

During the conversion process, the addition of classical connectives is too obvious, making the converted text seem stiff and unnatural. This also negatively affects the readability and authenticity of the generated text. When finishing unfinished works, the natural fluency of the language is crucial for maintaining the integrity of the work. In classical literature research, unnatural texts can distort the style of the original work, increasing the difficulty for scholars to accurately understand it and affecting the comparison and analysis of different texts.

4.2 Improvement measures

4.2.1 Improve recognition of classical verbs and nouns

Expand the vocabulary: Collect more classical verbs and nouns to build a richer vocabulary. With a wider range of classical vocabulary, the program will have more options for vocabulary replacement, reducing the possibility of inaccurate or missing replacements.

Optimize recognition algorithms: Adopt more advanced natural language processing algorithms to improve the accuracy of recognizing classical verbs and nouns. These algorithms can consider the context and semantics of the text to more accurately determine vocabulary replacement. In classical literature research, improved recognition can enhance the accuracy of text analysis and interpretation, helping scholars gain a deeper understanding of the original meaning of classical works.

4.2.2 Optimize the addition of classical connectives

Establish reasonable addition rules: Develop more reasonable rules for adding classical connectives based on the context and semantics of the text. These rules can more naturally and appropriately determine when and how to add classical connectives.

Introduce context - aware mechanisms: Use contextual information to determine whether to add connectives and what kind of connectives to add. This mechanism can make the addition of classical connectives more context - sensitive, improving the naturalness of the generated text. In classical literature research, more natural addition of classical connectives helps restore the original style of the text, facilitating in - depth research and comparison by scholars.

5 Experimental results and analysis

5.1 Experimental setup

To evaluate the performance of the mini - program, we conducted a series of experiments. The experimental data

included different types of modern texts, such as news reports, modern literary works, and everyday conversations. To specifically test the program's ability to finish unfinished works, we selected parts of "Dream of the Red Chamber" that are widely believed to be unfinished. Also, to assess the program's applicability in classical literature research, we included modern research papers on classical literature in the experimental data. By converting these papers into classical literary styles, we tested the program's performance in assisting research tasks.

5.2 Experimental results

5.2.1 Style transfer effect

Experiments showed that the mini - program could, to some extent, convert modern texts into the classical literary language style of "Dream of the Red Chamber." However, there are still some issues, such as weak recognition of classical verbs and nouns and obvious addition of classical connectives. When finishing the unfinished parts of "Dream of the Red Chamber," the program showed certain potential but also faced challenges. The generated text had some similarity in style to the existing parts but still needed improvements in some aspects.

In experiments related to classical literature research, the program also showed certain capabilities. When converting modern research papers into classical literary styles, the overall style had some similarity to the target style, but there were still issues in vocabulary replacement accuracy and text naturalness.

5.2.2 Text analysis accuracy

The text analysis system could fairly accurately evaluate the converted text, providing information on classical word usage, sentence pattern complexity, and literary color scores. This information helps understand the original author's style features and guides the style transfer process. However, in the context of finishing unfinished works and classical literature research, the accuracy of the text analysis system also needs to be further improved to ensure that the generated text better matches the author's style and provides more reliable evidence for research. For example, the analysis of classical word usage may not be detailed enough, and the evaluation of literary color may be somewhat subjective.

5.3 Results analysis

5.3.1 Advantages

Functional completeness: The mini - program has the basic functions of text style transfer and text analysis, meeting users' basic needs. These functions are also crucial for finishing unfinished works and conducting classical literature research, as they provide the necessary tools for style transfer and analysis.

AI enhancement effect: The AI enhancement function can effectively improve the quality of style transfer, making the converted text more natural and accurate. This is especially important for finishing unfinished works and classical literature research, as it helps narrow the gap between newly generated texts and existing parts of the work and improves the quality of research results.

5.3.2 Disadvantages

Classical verbs and nouns recognition: The ability to recognize classical verbs and nouns is insufficient, affecting the quality of the converted text and the accuracy of classical literature research. This issue needs to be resolved to ensure the generated text better matches the author's style and to ensure a more accurate understanding of classical works.

Addition of classical connectives: The addition of classical connectives is too obvious, reducing the fluency and naturalness of the text. This issue also needs to be resolved to improve the readability and authenticity of the generated text and to facilitate the conduct of classical literature research.

6 Conclusion and outlook

6.1 Summary of research results

This study successfully developed a Chinese language style transfer mini - program based on deep learning and AI enhancement. It can convert modern texts into the classical literary language style of "Dream of the Red Chamber" and has a text analysis system to evaluate the converted text. More importantly, the program shows potential in finishing unfinished literary works and classical literature research. By learning the original author's language features and style

patterns, it can generate texts in the author's style, providing a new way to restore incomplete masterpieces. At the same time, it can assist scholars in classical literature research tasks such as text comparison and hidden information mining.

6.2 Research limitations and directions for improvement

6.2.1 Research limitations

Vocabulary recognition and processing capable need improvement: The current ability to recognize classical verbs and nouns is weak and needs further improvement. This issue affects the quality of the generated text, especially when finishing unfinished works and conducting classical literature research.

The naturalness of connective addition needs optimization: The way classical connectives are added is not natural enough, affecting the quality of the text. This issue reduces the readability and authenticity of the generated text and is unfavorable for finishing unfinished works and classical literature research.

6.2.2 Directions for improvement

Enhance the application of deep learning models: Use more advanced deep learning models, such as Transformer, to improve the understanding and recognition of classical vocabulary. These models can better capture the complex relationships between words and styles, achieving more accurate vocabulary replacement.

Introduce more linguistic knowledge and rules: Combine linguistic knowledge and rules to optimize the addition of classical connectives, improving the fluency and naturalness of the text. By considering the context and semantics of the text, more natural and appropriate addition rules can be developed.

6.3 Future research outlook

Expand style types: In addition to the "Dream of the Red Chamber" style, other classical literary styles such as Tang poetry and Song lyrics can also be included. This expansion can provide users with more choices and help preserve and spread different classical literary styles.

Multilingual support: Consider supporting style transfer in multiple languages to expand the application scope of the mini - program. This can make the program useful in a wider cultural and linguistic context, facilitating cross - cultural literary research and exchange.

User interaction optimization: Further optimize the user interface and interaction experience of the mini - program to improve user satisfaction. A more user - friendly interface can make it easier for users, especially those not familiar with technical operations, to use the program.

Focus on finishing unfinished works: Continue to improve the program's ability to finish unfinished literary works. This may involve further refining style transfer and text analysis algorithms and collecting more data on unfinished works for model training. In this way, the program can become a more powerful tool for restoring incomplete masterpieces.

Deepen classical literature research: Further explore the application of the mini - program in classical literature research. This can include collaborating with scholars for in - depth research, improving the accuracy of text analysis, and developing more specialized functions for different types of classical works.

References

- [1] Brown, P. F., Pietra, S. D., Pietra, V. J., & Mercer, R. L. The Mathematics of Statistical Machine Translation: Parameter Estimation. *Computational Linguistics*, 1993, 19(2), 263 - 311.
- [2] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. Attention Is All You Need. *Advances in Neural Information Processing Systems*, 30. arXiv:1706.03762 [cs.CL]. 2017.
- [3] Shen, T., Lei, T., Barzilay, R., & Jaakkola, T. Style Transfer from Non - Parallel Text by Cross - Alignment. *Advances in Neural Information Processing Systems*, 30. arXiv:1705.09655 [cs.CL], 2017.
- [4] Fu, H., Johnson, M., & Hu, Y. Style Transfer in Text: Exploration and Evaluation. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. arXiv:1711.06861 [cs.CL]. 2018.
- [5] Gao, Y., & Huang, L. Chinese Classical Poem Generation with Pre - trained Language Models.[D]. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 2020.
- [6] Hu, Y., & Zhou, M. Deep Text Style Transfer. [D]. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. 2019.
- [7] Liu, B., & Zhang, L. Chinese Language Processing in the Age of Deep Learning. *Chinese Journal of Computational Linguistics*, 2020, 23(4), 321 - 335.
- [8] Zhang, X., & Zhao, J. Recent Advances in Chinese Text Generation. [D]. *Proceedings of the 2021 International Conference on Computational Linguistics*. 2021.