Chinese Couplet Generator With Attention-Based Transformer Mechanism

Jianing Zhang jz5212@nyu.edu
Haotong Wu hw2933@nyu.edu
Xinwei Xie xx2185@nyu.edu

CSCI-UA 0469: Natural Language Processing

Abstract

Chinese couplets, with their rich cultural significance and intricate linguistic structure, present a challenging yet fascinating domain for Natural Language Processing. This project introduces a Chinese Couplet generator with a Transformer architecture that can automate the generation of Chinese couplets. We employ tokenization and encoding techniques to maintain the semantic integrity of the couplets. The model is trained on a large and curated dataset, where it learns to mirror the syntactic harmony and thematic coherence emblematic of traditional couplets. To evaluate our model performance, we have two parts, the quantitative analysis and human evaluation, which we use loss function and questionnaires respectively. The results indicate that our model not only achieves a high level of linguistic precision but also captures the aesthetic essence of Chinese couplets, demonstrating potential applications in cultural preservation, and computational creativity in the realm of AI-generated literature.

1 Introduction and Motivation

During the Spring Festival, one of the most cherished traditions is the exchange of Chinese couplets, known as *duilián*. These artistic expressions are more than mere decorations; they are profound cultural symbols steeped in the hopes and aspirations for the new year. Typically crafted on vibrant red paper, couplets are displayed prominently in homes and public places as part of the New Year ' s adornments. Each couplet consists of two lines—the Upper Line and the Lower Line—that are harmoniously balanced in meter, tone, and meaning. While these lines can sometimes form short sentences, they are predominantly phrases or poetic lines that convey wishes for prosperity, happiness, and good fortune.

The art of crafting a couplet is deeply rooted in ancient Chinese literary tradition, requiring not only a mastery of language but also a deep understanding of cultural nuances and historical contexts. This dual-line structure, with its inherent symmetry and balance, reflects the fundamental Taoist principles of harmony and duality. The crafting of each line demands meticulous attention to detail, ensuring that both syntactic and semantic elements align perfectly to evoke the desired sentiment.

Inspired by the unique structure and the symmetrical beauty of couplets, we aim to develop a Chinese Couplet Generator. This technological endeavor would automate the creative process by generating the Lower Line in response to a given Upper Line, preserving the essential characteristics of traditional couplets. By leveraging advanced computational techniques and deep learning models, this generator would not only honor this ancient art form but also make it accessible to a broader audience, enabling more people to partake in this expressive cultural practice during the Spring Festival and beyond. This fusion of technology and tradition offers a fascinating glimpse into how cultural heritage can be preserved and enhanced through innovation.

2 Related Works

The development of Chinese text generators has historically faced two primary challenges: tokenization of Chinese characters and generating contextually appropriate sentences. Unlike alphabetic languages where spaces naturally delineate words, Chinese script consists of logograms without explicit word boundaries, complicating the tokenization process. Effective tokenization is crucial as it directly influences the quality of generated text by ensuring that the semantic units are correctly identified and processed by the model.

Furthermore, generating contextually coherent

sentences in Chinese requires the model to not only consider the immediate linguistic environment but also maintain thematic and stylistic consistency across a larger text body. This involves understanding and integrating complex aspects of the language such as idiomatic expressions, cultural nuances, and syntactic structures.
This section reviews models commonly employed in Chinese text generation, emphasizing their effectiveness in learning contextual dependencies.

2.1 Long Short-term Memory Networks

Long Short-term Memory Networks (LSTMs) are an advanced type of recurrent neural networks specifically designed to address the vanishing gradient problem commonly encountered in traditional RNNs (Hochreiter and Schmidhuber, 1997). By integrating memory cells that can maintain information for long periods, LSTMs are particularly adept at handling tasks that require understanding long-term dependencies in text data. This capability makes them well-suited for applications like generating Chinese couplets, where maintaining thematic and semantic coherence between lines is crucial.
Despite their strengths, LSTMs in a traditional Encoder-Decoder setup sometimes struggle to fully capture the deeper, more nuanced relationships between lines of text. This limitation can lead to the generation of couplets that lack meaningful semantic connections, undermining the quality and artistic integrity of the output. Recent advancements in neural network architectures and techniques may offer improvements in these areas, enhancing the LSTM's ability to generate more contextually and thematically consistent couplets (Hochreiter and Schmidhuber, 1997).

2.2 Sequence to Sequence Models

The Sequence to Sequence (Seq2Seq) framework, introduced by Sutskever et al. (Sutskever et al., 2014), employs an encoder-decoder architecture. In this architecture, the encoder, an RNN, first transforms the source sequence into a fixed-length vector, encoding the input sequence's information. Subsequently, the decoder, another RNN, utilizes this vector to generate the target sequence. This model excels in tasks requiring the transformation of one sequence into another, such as generating couplets. However, due to the inherent limitations of traditional RNNs, like the vanishing gradient problem, the Seq2Seq model may encounter challenges when processing longer sequences.
The selection of the model was a strategic decision based on its ability to comprehend and generate human-like texts, a critical capability for our project. The essence of couplets lies not only in their structural alignment but also in the intricate interplay between the lines. Each character must correspond in sound, part of speech, and meanings within the poetic lines, making the choice of model pivotal. In the end, the multi-head transformer stand out for being more coherent and more accurate in Chinese poetic generation. The Transformer's unique structure allows it to excel in capturing latent relationships between sentences, leading to more effective generation of couplets.
Learning from the featured models, this project introduces a well-developed combination of advanced methods in data processing, model architecture, and evaluation to enhance the generation of Chinese couplets.

3 Architecture and Methodology

3.1 Token Embedding

The ‘TokenEmbedding‘ module in our model employs PyTorch's ‘nn.Embedding‘ to transform input tokens into high-dimensional vectors. This transformation scales each embedding by the square root of the embedding size, enhancing the model ’ s ability to process and learn from the input effectively.

3.2 Position Encoding

The ‘PositionalEncoding‘ module adds unique position information to the embedding, enhancing the model's ability to distinguish the sequence order of inputs. Implemented using PyTorch, this module generates position-specific signals using sinusoidal functions, which are then added to the token embedding. It also includes dropout for regularization. This addition of positional information is crucial for models like the Transformer that do not inherently capture sequence order.

3.3 Transformer

In the architecture of our Chinese Couplets Generator, we incorporate the Transformer model, which operates within an Encoder-Decoder framework. The Transformer exclusively employs various attention mechanisms, including self-attention, to construct its neural network architecture (Vaswani

et al., 2017).
The attention mechanism, initially proposed by (Vaswani et al., 2017), has revolutionized the field of neural networks by enabling models to focus on different parts of the input sequence, enhancing their ability to understand complex relationships in data. This mechanism is foundational to the development of transformer models, which are now prevalent in a variety of natural language processing tasks and it is adept at focusing on specific parts of the text that are crucial for the task at hand. This capability significantly enhances the performance of language representation models in specialized applications such as song lyrics generation (Wang et al., 2016) and ancient Chinese text generation (Tian et al., 2020) .
Within the Transformer, the multi-head attention module runs the attention mechanism in parallel across multiple 'heads.' This arrangement allows the model to attend to information from different representational subspaces and positions simultaneously. Each 'head' computes the scaled dot-product attention using three vectors—queries (Q), keys (K), and values (V)—which are derived either from the input embeddings or from the output of the previous layer.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

where d_k is the dimension of the queries and keys.

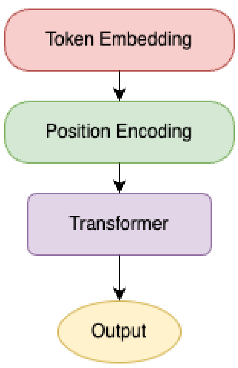

Figure 1: Our model architecture.

The encoder and decoder layers in our model leverage self-attention mechanisms and feedforward neural networks to process the sequence effectively.
Encoder: This component processes the upper line of the couplet. Each character or token is first embedded and then processed through multiple layers of multi-head attention and position-wise feedforward networks. The encoder transforms the input sequence into a comprehensive set of feature representations driven by attention, capturing the contextual nuances of the upper line.
Decoder: The decoder is responsible for generating the lower line of the couplet, utilizing the feature representations provided by the encoder. It incorporates multi-head attention layers, similar to the encoder, but adds a cross-attention mechanism that focuses on the encoder ' s outputs. This structure ensures that the semantic and stylistic context of the first line is effectively reflected in the generated second line.

4 Experiments and Evaluation

4.1 Dataset

Our dataset (Dat, 2019) includes 74,000 Chinese couplets, carefully curated to exclude any improper language, including insults and other inappropriate content. This is crucial, especially for couplets used during the Spring Festival, which should be positive and uplifting. Consequently, the couplets with inappropriate terms were completely removed by cross-referencing with another inappropriate word list. The dataset we applied is well-recognized and frequently utilized by researchers focused on generating Chinese couplets. Initially, we loaded the couplets from two distinct files, stripping any extraneous whitespace to maintain consistency. We divided the dataset into training and validation subsets using an 80-20 split.
The next step in our model's data preparation phase involves constructing a comprehensive vocabulary from our dataset. A 'Vocab' class is initiated to build a vocabulary from the tokenized dataset. This class counts the frequency of each token across all lines in the dataset, considering tokens from both parts of each couplet. We tokenize the vocabulary into following cateogories:

- self.word2idx is a dictionary that maps words with special tokens that serve different functions in the processing of sequences:
  - Vocab.UNK represents words that are not found in the vocabulary, i.e. out-of-vocabulary (OOV) words.
  - Vocab.PAD is used to fill in sequences to ensure that all input sequences have the same length when batched together for model training.

- Vocab.BOS is used to indicate the start of a sequence, which helps the model recognize sequence initiation.
- Vocab.EOS is used to signify the end of a sequence, allowing the model to determine when a generated sequence should terminate.

- self.idx2word is the reverse mapping of self.word2idx and converts integer indices back to their respective word representations. This reverse mapping is crucial for the decoding phase when the model's output needs to be translated from indices back to words to form a human-readable sequence.

4.2 Experimental Process

We adjusted different batch sizes as well as different learning rates and tested the results at 10, 20, 30, and 40 epochs to get the best model.

5 Evaluation

5.1 Quantitative analysis

In evaluating the performance of our transformer model, we employ two principal metrics: cross-entropy loss and perplexity.

Cross-entropy loss

In our transformer model, we employ cross-entropy loss as the loss function for backpropagation. This method is crucial for updating the model's weights effectively, steering them toward an optimal solution. Specifically, cross-entropy loss quantifies the disparity between the predicted probability distribution and the actual distribution of the target data. By minimizing this loss, we enhance the model's ability to align its predictions more accurately with the expected results.

Perplexity

Perplexity is another critical metric used to assess the model's ability to predict the next token in a sequence, providing a measure of the model's overall predictive performance. Lower perplexity values indicate a better performing model, as they suggest a lower level of uncertainty in predictions. It is computed using the formula:

$$PPL(X) = \exp\left\{-\frac{1}{t}\sum_{i}\log p_\theta(x_i|x_{<i})\right\} \quad (2)$$

where t is the length of the sequence, x_i is the ith element of the sequence, and $p_\theta(x_i|x_{<i})$ is the probability of x_i given the preceding elements $x_{<i}$, as predicted by a model with parameters θ.

5.2 Human Evaluation

As a Chinese Couplet generator need to generate something that can be easily understood by human beings, except quantitative analysis, we also adopt human evaluation to evaluate the couplet generator. We did the evaluation twice with two methods during the process of training and after training:

Comparing

We conducted the human assessment twice with 10 Chinese native speakers after training for either 10 or 40 epochs. The participants were asked to evaluate a set of 60 couplets that were generated during our testing phase. The evaluation utilized a double-blind trial methodology, where participants were presented with the first line of each couplet followed by two potential second lines: one was the original, and the other was generated by our model. Participants chose the second line that they believed best complemented the first, without knowing some of the lines were generated by the model. This approach helped to ensure that the selections were unbiased by perceptions of AI-generated content. The number of times the AI-generated lines were chosen was recorded, providing direct feedback on the model ' s performance in creating couplets that were indistinguishable from those written by humans.

Ranking

Evaluation Criteria	The Specification	Score
Syntactic Relation	Antithetical or not	1-5
Semantics	Fluency or not	1-5
Semantics	Content Consistent or not	1-5
Overall Impression	Overall	1-5

Table 1: Human Evaluation Criteria.

In this evaluation phase, three Chinese native speakers participated in rating a set of 60 couplets generated during our testing phase. The evaluation criteria were based on syntactic relation, semantics, and overall impression, with each aspect scored on a scale of 1-5.

For syntactic relations, we assessed whether the antecedent and subsequent clauses were of equal length and if the positional words conformed to the antithesis relation. Regarding semantics, we evaluated whether the antecedent and subsequent clauses were semantically meaningful and coherent.

The ratings were conducted separately for couplets generated after training the model for 10 and 40 epochs, allowing us to analyze the model's performance progression. Given that all group members are Chinese native speakers, we were able to make informed subjective judgments on the quality of the generated outputs. This approach helped in providing nuanced insights into how well the couplets adhered to traditional Chinese couplet standards and their overall coherence and appeal.

5.3 Result

5.3.1 Quantitative Results

# of Epoch	Perplexity	Cross-entropy Loss
10	27.4422	3.386854
30	22.9492	3.087754
40	22.2854	3.023477

Table 2: Performance of Transformer

The model shows consistent improvement in both perplexity and cross-entropy loss as the number of training epochs increases. Perplexity decreases from 27.4422 at 10 epochs to 22.2854 at 40 epochs, indicating that the model is becoming more efficient at predicting the next token in the sequence, suggesting better handling of the language model's uncertainty. Similarly, the cross-entropy loss, which measures the difference between the predicted and actual outputs, decreases from 3.386854 to 3.023477, showing that the model's predictions are becoming more accurate over time.

5.3.2 Comparing

10 Epoch Training:

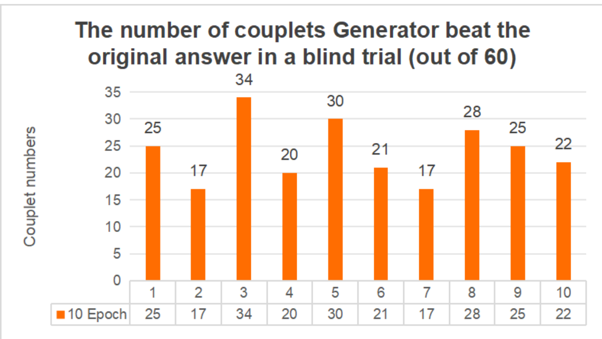

Figure 2: The number of preferences for the couplet generator after training for 10 epochs in a blind trial involving 60 couplets. It shows the number of results where the couplets generated by the model were preferred over the original answers.

- The generator's couplets were preferred over the original answers in a blind trial, with an average of 23.9 selections out of 60.
- The results exhibited variability, as indicated by a standard deviation of 5.05.
- Preference for the AI-generated couplets ranged from a minimum of 17 to a maximum of 36 selections in individual trials.

40 Epoch Training:

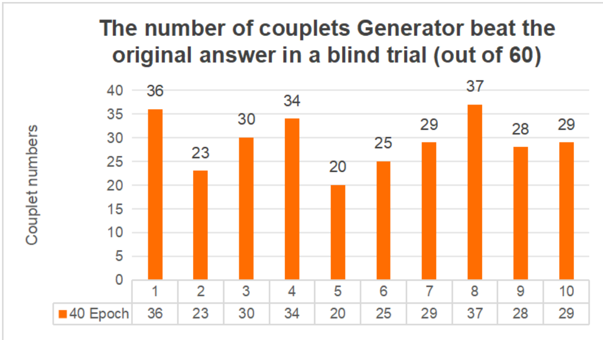

Figure 3: The number of preferences for the couplet generator after training for 40 epochs in a blind trial involving 60 couplets. It shows the number of results where the couplets generated by the model were preferred over the original answers.

- Following extended training, the generator's couplets were chosen 29.1 times on average out of 60.
- This phase showed slightly lower variability with a standard deviation of 4.98.
- The number of couplets preferred ranged from 20 to 37 selections per trial.

5.3.3 Ranking

# of Epoch	Syntactic Relation	Semantics	Overall	Total
10	3.5	3.37	3.47	13.9
40	3.67	3.77	3.72	15.37

Table 3: Human Evaluation Table

Our evaluation at different training stages—10 and 40 epochs—reveals some improvements in its ability to generate Chinese couplets. The most noticeable improvement is observed in the semantics of the generated couplets. The semantic score increased from 3.37 at 10 epochs to 3.77 at 40 epochs, a jump of 0.40 points. The overall impression score saw a slight increase from 3.47 to 3.72, a more modest rise of 0.25 points compared to semantics. Although this is a less dramatic increase, it still reflects a general improvement in the quality of the couplets as perceived by human evaluators. This detailed analysis indicates that while all aspects of the model's output saw improvements, the rise in semantic understanding was particularly pronounced, underscoring the model's enhanced capability to handle the complexity of language in couplet generation. In contrast, the overall impression, although improved, did not see as large an increase, suggesting that while the couplets are becoming more accurate and contextually appropriate, there may still be room for improvement in terms of stylistic and creative aspects.

6 Conclusion

In conclusion, our Chinese Couplet Generator has demonstrated commendable performance in automating the creation of Chinese couplets. The system effectively utilizes advanced Natural Language Processing techniques to generate poetic lines that often meet the structural and thematic expectations associated with traditional couplets. However, despite its successes, there are notable areas where the model requires further refinement. A primary issue identified is the model's tendency to repeat characters in longer couplets. This repetition detracts from the quality of the couplets since it is common sense for Chinese if the first line avoids repetition, the second line should follow suit. Such repetitions have led to lower scores in our evaluations, highlighting a key area for improvement. Additionally, while the model occasionally exceeds expectations with creative outputs, it sometimes fails to generate meaningful or contextually appropriate responses. This inconsistency suggests a need for enhanced understanding and processing capabilities within the model.

Looking forward, the future improvement can be implementing an automatic evaluation system that assesses the overall thematic and structural coherence of the entire quatrain similar to humans. This will involve refining our model's ability to integrate and consider the content of all generated lines when producing each subsequent line, ensuring a cohesive and contextually relevant output. By improving the thematic integration and enhancing the model ' s handling of longer sequences, we anticipate significant advancements in the generator ' s ability to produce couplets that are not only structurally sound but also rich in cultural relevance and poetic value.

Our Chinese Couplet Generator shows great potential in preserving and celebrating the rich tradition of Chinese couplet culture, further improvements are essential to achieve the high standards set by traditional couplet celebration. Through continued research and development, we are optimistic about the future capabilities of this tool in the field of computational creativity and cultural preservation.

References

2019. Chinese couplets dataset without vulgar words.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. volume 1, Munich, Germany. Neural Network Research Press.

Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems*, pages 3104–3112. Curran Associates, Inc.

Huishuang Tian, Kexin Yang, Dayiheng Liu, and Jiancheng Lv. 2020. Anchibert: A pre-trained model for ancient chinese language understanding and generation. *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, pages 5998–6008. Curran Associates, Inc.

Qixin Wang, Tianyi Luo, Dong Wang, and Chao Xing. 2016. Chinese song iambics generation with neural attention-based model.

Table 4: Sample Results of the Chinese Couplet Generator

Upper Line	**Epoch 10 Answer**	**Epoch 40 Answer**	**Original Lower Line**
对月赏荷吟风韵	临风听雨听雨声	南海慈航有路长	东土师徒历难多
为官半世, 两袖清风, 留有美名传万古	处世千秋, 一身正气, 留存正气贯千秋	执政千秋, 一身正气, 长留浩气壮千秋	造福一方, 浑身正气, 何妨白发满千寻
临风披晓月	临水听清音	对酒对青山	抱石镇江涛
千古华堂奉君子	宗臣遗像肃清高	一生孝子传后人	一方净土护黎民
四野溢春声, 垄上林间莺语脆	一生存爱意, 心头梦里梦痕	九州添喜气，人间天上笑声甜	三山摇雨步，檐前屋后草花香
同涉蔺公, 论值或如和氏璧	共圆梦想, 成功不负汉唐风	共襄王母, 风流不让武侯王	堪称国砚，濡毫再振大唐风
月冷杯中, 一时误作出尘想	风来案上, 几度难为入梦人	风流笔底, 几度难消入梦思	蟹沉釜底, 何日敢于踏浪行
逍遥剑客邀名士	浪漫花仙觅故人	浪漫诗人醉酒仙	流放叶公采果实