Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/39/20230619

Thematic classification of ancient Chinese poetry using TwinEmbedAttentionNet

Haoxiang Chang[1,5], Beitong Li[2,6], Fengwei Wang[3,7], and Yue Bin[4,8]

[1] School of Mathematics Science, Hebei Normal University, Shijiazhuang 050010, China
[2] Department of Information System, City University of Hong Kong, Hong Kong, China
[3] Tandon School of Engineering, New York University, New York 11101, USA
[4] Guanghua Qidi College, Shanghai 200439, China

[5] 18134158182@163.com
[6] libeitong0304@gmail.com
[7] rexwang519@163.com
[8] vincentbyue@outlook.com

Abstract. Ancient Chinese poetry, a reflection of China's rich cultural and philosophical fabric, encapsulates the evolving socio-cultural nuances of its historical epochs. Despite its cultural significance, there remains an evident lacuna in comprehensively classifying its recurring themes, due in part to the conciseness and polysemy intrinsic to the language and the essentiality of embedded cultural and historical contexts. Addressing this challenge, this study introduces TwinEmbedAttentionNet, a pioneering method tailored for the thematic classification of ancient Chinese poetry. This approach synergistically integrates pretrained word and sentence embeddings with an attention mechanism, ensuring the nuanced representation of the poetry's intricate details. Our results showcase its superior performance over existing models. Furthermore, an in-depth examination of model components offers insights into their respective thematic categorization efficacies. This research not only advances the academic understanding of ancient Chinese poetry but also underscores the potential of innovative neural networks in processing historically rich textual data.

Keywords: Ancient Chinese Poetry, Thematic Classification, TwinEmbedAttentionNet

1. Introduction

Ancient Chinese poetry, spanning across several dynastic periods, embodies the rich tapestry of China's cultural, philosophical, and aesthetic legacy [27, 15, 29]. These poems offer invaluable insights into the human experience of the Chinese milieu. Rooted in diverse traditions like Confucianism, Daoism, and Buddhism, the themes interwoven in these poems echo the beliefs, sentiments, and sensibilities of their times [12, 3, 20]. This rich poetic tradition was not merely a product of its temporal context; rather, it bore witness to the transformation of Chinese society through various epochs. Whether it was the celebration of nature's beauty or the profound introspection of life and existence, the thematic diversity of Chinese poetry encapsulates the evolving socio-cultural landscape of ancient China [30].

Despite its undeniable significance, there exists a considerable gap in comprehensively classifying and analyzing the themes prevalent in ancient Chinese poetry [24]. While several scholars have delved deep into the stylistic nuances [9, 8] and coarse-grained sentiments [28, 19] of these poems, there is a growing need for a robust and fine-grained thematic framework that delineates the recurring motifs and underlying narratives that have shaped this poetry.

The thematic classification of ancient Chinese classical poetry does not only cater to the academic quest for understanding but also serves a broader purpose. However, the intricacies and richness of classical Chinese poetry, characterized by its conciseness and profound symbolism, present significant challenges in this task. First and foremost, the conciseness and polysemy of the language used in these poems often mean that each character or word can carry multiple layers of meaning [5, 22]. This characteristic lends unique beauty and emotional depth to the poetry but simultaneously introduces variability and ambiguity in interpretation. The challenge here is to accurately capture and interpret such multifaceted meanings, especially for automated classification systems. Secondly, the embedded cultural and historical contexts are pivotal to understanding these poems [16]. Without a profound grasp of these backgrounds, accurate thematic and emotional categorization becomes a daunting task. This depth of cultural and historical comprehension poses challenges not only for modern readers but also for automated tools.

To tackle this challenge, we propose a novel method named **TwinEmbedAttentionNet (A Twin Embedding Attention Network for ancient Classical Chinese classification)**, which utilises advanced pretrained language models and attention mechanisms. Firstly, word embeddings that transform ancient Chinese texts into vectors capturing their foundational semantics; and secondly, a pre-trained BERT model that proffers contextually rich embeddings for the textual data. Building upon this foundation, the model seamlessly incorporates a multi-head self-attention mechanism, reinforcing its capability to apprehend and represent intricate interdependencies within sentences.

Overall, our contribution is threefold. Firstly, we introduce TwinEmbedAttentionNet, a novel methodology specifically tailored for ancient Chinese classification. This approach adeptly blends both pretrained word sentence embeddings with multi-head attention, enabling it to capture intricate thematic details effectively. Secondly, we conducted a comprehensive evaluation against strong baseline models, the results of which underline the superior performance of our proposed method. Lastly, through ablation experiments and meticulous analysis of thematic category misclassifications, we provide a nuanced understanding of each component's significance in the model and its efficacy across specific thematic categories.

The structure of our study is organized as follows: Section 2 reviews the recent advancements in ancient Chinese classical poetry classification tasks. Section 3 delves into the intricacies of our proposed method. Section 4 outlines our experimental results, key findings, and discussions. Lastly, Section 5 provides a conclusion to our study.

2. Literature review

Over the years, researchers have adopted various techniques to understand, analyze, and classify Chinese poetry.

Shen et al. [18] launched a pioneering effort in the domain of sentiment analysis of Tang poetry, employing a sentiment classification model based on imagery-aided and classification ensemble. Their work was distinct as it mined sentiment at both character and word levels, yielding significant improvements over traditional methods. Similarly, Hou and Frank [8] introduced a weakly supervised sentiment lexicon based on Weighted Personalized PageRank. Their innovative graph-based approach displayed better performance than a previously popular PMI-based method. Further enriching the sentiment analysis domain, Wu et al. [25] proposed a transfer learning model, CATLPCO, which expanded feature vectors of ancient texts and introduced three classifiers for sentiment analysis.

Hu and Zhu [9] proposed a text classification model focused on categorizing Tang poetry into seven thematic categories. By employing the Vector Space Model, Naive Bayes and SVM, they achieved compelling classification results. Their research also highlighted the significance of poetry titles, authors, and types in determining poetry themes. On a similar note, Zhang and He [28] introduced a BiLSTM-based approach for the style classification of ancient Chinese poetry, leveraging both poetry words and word vectors. Their approach showed a marked improvement in classification over traditional models. Tang et al. [19], pushing the boundaries further, proposed a deep learning model combining CNN and GRU. Their multi-channel processing model reshaped the feature vector of sentences and outperformed three other methods in terms of accuracy. Furthermore, Wang et al. [24] proposed the confidence-based syntax encoding network (cSEN) for ancient Chinese poetry theme classification, addressing the noise and incompatibility between ancient and modern Chinese. Bridging the gap between imagery and emotion in Tang poetry, Fu et al. [4] emphasized the limitations of traditional research methods that depend heavily on expert knowledge. By applying aspect-level sentiment analysis, they employed multiple mainstream emotion classification models to discern the sentiment in poetic imagery.

Most research on Tang poetry has centered around coarse-grained sentiment analysis, with notable exceptions like [24]. There's a significant gap in delving deep into finer thematic nuances and emotional layers within these poems. This study pioneers in this direction, aiming for a precise fine-grained thematic classification. It employs a novel model that synergizes word embedding with pre-trained sentence embedding. This fusion not only captures intricate themes but also links them to corresponding sentiments.

3. Methodology

This section introduces a novel approach for thematic poetry classification. Our method synergistically leverages pre-trained word and sentence embeddings, providing a robust and rich representation of poetic text. As shown in Fig 1, the model integrates a multi-head self-attention module, enabling it to glean diverse information from the dual embeddings through parallelized attention computations. These dual-weighted embeddings, combined with the multi-head self-attention outputs, culminate in a unique fused embedding used for classification. For the classification phase, we utilize a variant of the RCNN (Recurrent Convolutional Neural Network) [13] architecture for efficient encoding and classification.

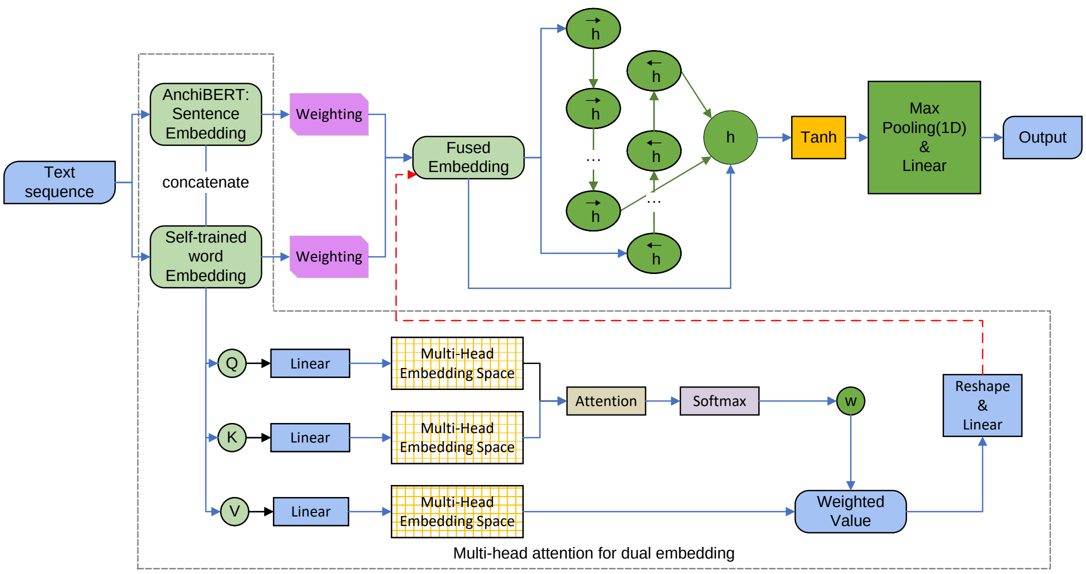

Figure 1: Schematic of our proposed approach for thematic poetry classification.

Subsequent subsections offer a detailed exploration of each component of the model.

3.1. Dual embedding

The sentence embeddings leveraged in our study originate from AnchiBERT [21]. AnchiBERT is a pre-trained language model built upon BERT [10], specifically tailored for and trained on expansive ancient

DOI: 10.54254/2755-2721/39/20230619

Chinese literature. Given the absence of large-scale pre-trained word embeddings for ancient Chinese corpora, we employed the Word2Vec model [17] to train word embedding vectors on our curated corpus of ancient Chinese poetry.

Let us consider a poetry sequence X comprising tokens $X = \{x_1, x_2, \ldots, x_n\}$. Token x_i is initially mapped to a word embedding through:

$$e_{w_i} = \text{W2V}(x_i) \in \mathbb{R}^{d1},$$

where $\mathbb{R}^{d1}$ is the embedding space determined by Word2Vec. Consequently, the poetry sequence X can be represented as an embedding matrix $\mathbf{E}_{wv} \in \mathbb{R}^{d1\times n}$. Concurrently, the sequence undergoes transformation via the BERT model, yielding BERT embeddings:

$$\mathbf{E}_B = \text{BERT}(X) \in \mathbb{R}^{d2\times n},$$

where $\mathbf{E}_B$ symbolizes the sequence of BERT embeddings, defined as $\mathbf{E}_B = \{e_{b_1}, e_{b_2}, \ldots, e_{b_n}\}$. Here, $d1$ and $d2$ represent the dimensions of the Word2Vec and BERT embeddings, respectively.

To amalgamate information from both the Word2Vec and BERT embeddings, we introduce a learnable parameter vector $\boldsymbol{\omega} = [\omega_1, \omega_2]$. This vector undergoes normalization through the softmax function:

$$\hat{\boldsymbol{\omega}} = \text{softmax}(\boldsymbol{\omega}).$$

Using the normalized weights, the fused embedding for each token x_i can be computed as:

$$e_{c_i} = \hat{\omega}_1 \cdot e_{w_i} + \hat{\omega}_2 \cdot e_{b_i}.$$

3.2. Multi-head attention mechanism

To encapsulate intricate dependencies within the dual embedding sequences and to concurrently attend to disparate regions of the input space, a multi-head attention mechanism is employed.

Initially, we concatenate the embeddings $\mathbf{E}_{wv}$ and $\mathbf{E}_B$ to formulate a new composite embedding vector $\mathbf{E}_f$. This vector belongs to $\mathbb{R}^{d\times n}$, where $d = d1 + d2$ signifies the cumulative dimension of both embeddings.

For the composite embedding $\mathbf{E}_f$, we define three distinct linear transformations to derive the Query vector (Q), Key vector (K), and Value vector (V) matrices. These transformations can be articulated as:

$$\begin{aligned}
\mathbf{Q} &= \text{Linear}(\mathbf{E}_f), \\
\mathbf{K} &= \text{Linear}(\mathbf{E}_f), \\
\mathbf{V} &= \text{Linear}(\mathbf{E}_f),
\end{aligned}$$

where Linear$(\cdot)$ represents the linear transformation operation.

Subsequently, to better discern diverse patterns across the input space, the transformed queries, keys, and values are divided into multiple attention heads. Each attention head (h) projects the input space into a lower-dimensional subspace. The operations leading to this partitioning can be articulated as:

$$\begin{aligned}
\mathbf{Q}_h &= \text{Reshape}(\mathbf{Q}), \\
\mathbf{K}_h &= \text{Reshape}(\mathbf{K}), \\
\mathbf{V}_h &= \text{Reshape}(\mathbf{V}),
\end{aligned}$$

where Reshape$(\cdot)$ modulates tensor dimensions, resulting in the shape $\mathbb{R}^{H\times n\times \frac{D}{H}}$. Here, H denotes the number of attention heads and D signifies the aggregate embedding dimension spanned by all heads.

For each attention head, the scaled dot-product attention mechanism determines the attention scores. Specifically, these scores are derived from the dot product of the query and key, subsequently normalized by the square root of the dimensionality of each head, $\sqrt{\frac{D}{H}}$:

$$\alpha_h = \frac{\mathbf{Q}_h \cdot \mathbf{K}_h^\intercal}{\sqrt{\frac{D}{H}}}.$$

Subsequent to score computation, we determine the weighted values by scaling the value matrices with the aforementioned attention scores:

$$\mathbf{W}_h = \alpha_h \cdot \mathbf{V}_h.$$

Lastly, to merge insights from all attention heads, we concatenate these weighted values. They are then mapped back to the original input space via a linear transformation:

$$\mathbf{E_H} = \text{Linear}(\text{Concatenate}(\mathbf{W}_1, \ldots, \mathbf{W}_H)).$$

Here, Concatenate$(\cdot)$ serves to amalgamate the individual weighted representations from each head.

3.3. Encoder and classifier

To assimilate sequential interdependencies, both the attention output, $\mathbf{E_H}$, and the fused embedding, $\mathbf{E_f}$, are processed through a bidirectional LSTM [6]. This operation can be mathematically represented as:

$$\mathbf{V}_{\text{LSTM}} = \text{concat}\left[\overrightarrow{LSTM}(\mathbf{E_H} + \mathbf{E_f}), \overleftarrow{LSTM}(\mathbf{E_H} + \mathbf{E_f})\right].$$

Here, $\overrightarrow{LSTM}$ and $\overleftarrow{LSTM}$ signify the forward and backward LSTM operations, respectively. Furthermore, $\mathbf{V}_{\text{lstm}}$ denotes the sequence of LSTM outputs.

Subsequently, the LSTM outputs are amalgamated with the attention output. This composite representation undergoes a transformation mediated by a linear layer, followed by the hyperbolic tangent activation function:

$$\mathbf{V}_{\text{linear}} = \tanh\left(\mathbf{W}_{\text{linear}} \cdot [\mathbf{V}_{\text{LSTM}}, (\mathbf{E_H} + \mathbf{E_f})]\right),$$

where $\mathbf{W}_{\text{linear}}$ represents the matrix facilitating the linear transformation.

Lastly, to condense the sequential information present in $\mathbf{V}_{\text{linear}}$ into a fixed-dimensional representation, max pooling is employed over its sequence dimension:

$$v = \text{MaxPool}(\mathbf{V}_{\text{linear}}).$$

Subsequent to this pooling, the derived vector v is routed through the feed-forward layer to facilitate prediction of the target class labels:

$$\hat{Y} = \text{FC}(v).$$

Herein, $\hat{Y}$ denotes the predicted labels corresponding to the input poetry text X.

4. Experiment

4.1. Dataset

The dataset utilized for this study was introduced by [24]. It comprises 20,360 ancient Chinese poems designated for training and an additional 800 poems set aside for validation. Each poem has been meticulously annotated by humans and categorized into one of ten thematic classes. Detailed metadata for this dataset is provided in Table 1.

Table 1: Metadata of the Ancient Chinese Poetry Classification Dataset

Item	Value
Total training samples	20,360
Total validation samples	800
Number of classes	10
Total characters in training samples	5,446
Average character count per sample	46.57
Average word count per sample	73.69
Language	Chinese

Ancient Chinese poetry encompasses a diverse array of themes, each reflecting specific emotional undercurrents and historical contexts. For a deeper understanding of the dataset used in our study, we provide a detailed description of these themes, highlighting the emotions they typically convey and their broader meanings, as shown in Table 2.

Table 2: Themes and Emotions in Ancient Chinese Poetry

Theme	Emotions Expressed	Meaning and Context
Homesickness	Nostalgia, longing	Feelings of missing one's homeland and yearning for familiar places and memories.
Longing for a Distant Love	Love, sorrow, frustration	Emotions tied to love and heartache for a distant or unattainable love interest.
Landscape	Appreciation of nature, tranquility	Celebrating the beauty and serenity of natural landscapes.
Yearning for a Loved One	Affection, longing	Deep feelings and longing for an absent person, often separated by circumstances.
Ode to Objects	Admiration, reflection	Poetic praise of objects, often unveiling deeper symbolic meanings.
Farewell	Farewell, good wishes	Expressions of both sadness and well-wishing during a departure.
Warfare	Valor, sorrow, patriotism	Narratives on the heroism, tragedy, and impact of wars on individuals and society.
Pastoral Life	Simplicity, contentment	Depictions of harmonious life in rural settings with a connection to nature.
Lament for the Deceased	Grief, remembrance	Mourning and paying tributes to the deceased, contemplating life and mortality.
Nostalgia for Antiquity	Reminiscence, reflection	Reflecting upon historical times, events, and iconic figures of the past.

4.2. Baselines

To evaluate the effectiveness of our proposed method, we compared it against multiple well-established baselines, encompassing both traditional and state-of-the-art approaches.

- **cSEN (Confidence-based syntax encoding network)** [24]. cSEN aims at alleviating the negative impact of syntax derivation noise and facilitating better ancient Chinese text encoding and decoding.

- **BERT (Bidirectional Encoder Representations from Transformers)** [10]. A pre-trained model based on the Transformer architecture, BERT has been influential in various NLP tasks, including text classification.
- **XLNet** [26]. An advanced pre-trained Transformer model, XLNet incorporates a permutation language modelling objective, enhancing its capability to grasp the bidirectional context.
- **ALBERT** [14]. As a streamlined version of BERT, ALBERT optimizes both model size and training duration through techniques like parameter sharing, ensuring efficient performance.
- **AnchiBERT** [21]. AnchiBERT adopts the same architecture as BERT but is specifically pretrained on ancient Chinese literature. The training corpus spans Chinese anthologies written between 1000 BC and 200 BC.
- **LEGACT (A Lightweight Ensemble Learning for Glyph-Aware Chinese Text Classification)** [7]. LEGACT integrates classical shallow neural networks as its weak classifiers with XGBoost serving as a strong classifier. Empirical evidence suggests that LEGACT outperforms pretrained language models specifically in the realm of Chinese-orientated NLP tasks.
- **TextCNN** [11]. Employing Convolutional Neural Networks (CNN) for text tasks, TextCNN uses diverse convolutional kernel sizes to capture and understand local nuances within texts.
- **BI-GRU-ATTENTION**: By implementing the attention mechanism on the output of a bidirectional Gated Recurrent Unit (GRU), this model emphasizes different sequence words based on their significance, yielding a weighted sequence representation for classification.
- **S2SAN (Sentence-to-sentence attention)** [23]: Focused on extracting sentence-level representations and relationships, S2SAN is particularly effective for processing online user-generated content.

For the transformer-based models, namely BERT, XLNet, ALBERT and AnchiBERT, we incorporated an additional linear layer subsequent to their core architectures to facilitate classification. The pretrained checkpoints utilized for BERT and XLNet are `BERT-wwm-ext` and `XLNet-base` respectively, as provided by [2].

In contrast, the checkpoint for ALBERT is `albert-base-chinese`, sourced from [1]. For the remaining models that rely on word embeddings, we initialized their embedding layers using our pretrained word embedding vectors.

4.3. Evaluation metric

In our evaluation, both micro F1 scores and macro F1 scores are utilized. Micro F1 takes into account the total counts of true positives, false positives, and false negatives across all classes, proving beneficial in the presence of class imbalances. Conversely, Macro F1 calculates the F1 for each individual class and then averages these scores, offering insights into model performance on specific classes, especially minority ones.

The metrics are defined as:

$$\text{Micro F1} = \frac{2 \times \text{Micro P} \times \text{Micro R}}{\text{Micro P} + \text{Micro R}},$$

where:

$$\text{Micro P} = \frac{\Sigma\text{True Positives}}{\Sigma\text{True Positives} + \Sigma\text{False Positives}},$$
$$\text{Micro R} = \frac{\Sigma\text{True Positives}}{\Sigma\text{True Positives} + \Sigma\text{False Negatives}}.$$

For Macro F1:

$$\text{Macro F1} = \frac{1}{N}\Sigma(\text{F1}_i),$$

where F1_i represents the F1 value of the i-th category and N is the total number of categories.

4.4. Experimental results

Table 3 presents the performance of various models evaluated on a specific task. This includes contemporary state-of-the-art models as well as traditional approaches. The selected performance metrics, Micro F1 and Macro F1, are especially valuable for tasks characterized by imbalanced datasets or those involving multi-label classification.

Table 3: Performance Evaluation

Model	Micro F1	Macro F1
cSNE	0.938	0.919
XLNet	0.925	0.889
ALBERT	0.924	0.895
AnchiBERT	0.934	0.909
BERT	0.913	0.877
LEGACT	0.929	0.897
TextCNN	0.885	0.829
BI-GRU-ATTENTION	0.828	0.732
S2SAN	0.914	0.876
TwinEmbedAttentionNet (Our Method)	**0.949**	**0.930**

Our proposed method emerges as the superior performer, registering the highest Micro F1 of 0.949 and a Macro F1 of 0.930. Such scores reflect a robust performance across both dominant and less prevalent classes in the dataset. The cSNE model closely tails our method, notably in the Macro F1 score, positioning itself as the second-best in a comprehensive performance. Among the models founded on the Transformer architecture, such as XLNet, ALBERT, AnchiBERT, and BERT, commendable scores are generally observed. AnchiBERT, notably, outperforms the more generalized BERT and ALBERT but falls slightly behind XLNet in the Macro F1 metric.

Regarding traditional models based on word embeddings, LEGACT exhibits a surprisingly robust performance, outstripping both TextCNN and BERT in the Micro and Macro F1 scores. This underscores that despite the advancements in recent models, traditional approaches like LEGACT can remain potent contenders, particularly when hyperparameters are optimally adjusted.

We further visualized the classification results generated by TwinEmbedAttentionNet, as depicted in Fig 2.

- **Ode to Objects:** Achieved a precision of 91% and a recall of 87%. Out of 47 instances, 41 were correctly classified. Misclassifications primarily encompassed themes such as Landscape, Nostalgia for Antiquity, and Farewell.
- **Landscape:** Both precision and recall stood at 98%, correctly identifying 43 out of 44 instances. A solitary misclassification pertained to the Farewell theme.
- **Yearning for a Loved One:** Both precision and recall were at 98%, accurately classifying 88 out of 91 poems. Occasional misclassifications were noted in the Homesickness and Longing for a Distant Love themes.
- **Nostalgia for Antiquity:** Precision was at 88% with a recall of 93%, correctly classifying 37 out of 40 poems. Misclassifications were seen in the Homesickness and Farewell themes.
- **Homesickness:** Representing the largest category, it recorded a precision and recall of 99%, correctly classifying 350 out of 357 instances.
- **Lament for the Deceased:** The model exhibited flawless performance with a 100% precision and recall, accurately categorizing all 18 instances.

DOI: 10.54254/2755-2721/39/20230619

- **Warfare:** Secured a precision of 83% and a recall of 97%, correctly classifying 60 out of 62 poems. Certain poems were misdirected to the Yearning for a Loved One theme.
- **Pastoral Life:** Achieved a 100% precision with a recall of 91%, correctly identifying 43 out of 47 poems. Misclassifications were directed towards the Ode to Objects theme.
- **Farewell:** Precision stood at 82% and recall at 77%. Out of 60 instances, 46 were accurately identified. This theme exhibited the most diverse misclassifications relative to others.
- **Longing for a Distant Love:** Displayed a precision of 89% and a recall of 97%, correctly classifying 33 out of 34 poems.

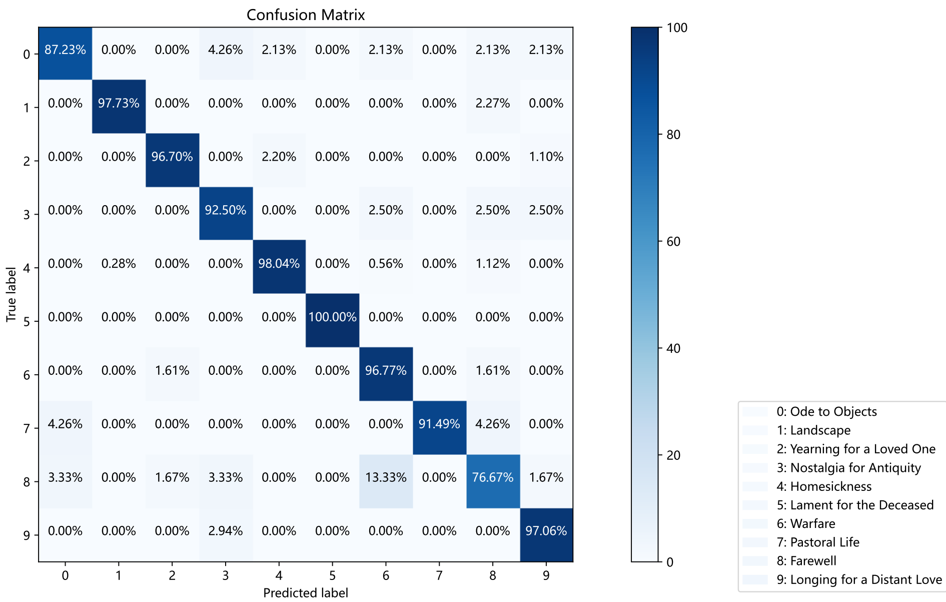

Figure 2: Confusion matrix illustrating thematic poetry classification.

In summary, while most themes demonstrated impressive classification results—particularly ”Lament for the Deceased”—the ”Farewell” theme indicated a potential avenue for further model enhancement and optimization. Insights derived from the confusion matrix further support these observations, offering guidance for augmenting the model’s effectiveness across specific thematic categories.

4.5. Ablation study

4.5.1. Module ablation To thoroughly evaluate the advantages of pre-trained models specialized for classical Chinese, we compared the foundational anchiBERT model, which was pre-trained on ancient Chinese, against the BERT-wwm model, pre-trained on modern Chinese. Moreover, taking into account the distinctive linguistic characteristics of the Chinese language, we explored the distinctions between word embeddings (obtained post-tokenization) and character embeddings (that operate at the level of individual characters, circumventing tokenization). The results are listed in Table 4.

Table 4: Results of the Ablation Study

Configuration	Micro F1	Macro F1
Full model	0.949	0.930
Without Word Embedding	0.933	0.909
Without AnchiBERT Embedding	0.898	0.856
AnchiBERT replaced with BERT-wwm	0.925	0.897
Word Embedding replaced with Character Embedding	0.948	0.929

Exclusion of Word Embedding: A slight reduction in performance was noted, implying that word embeddings do augment the model's efficacy. Nevertheless, their omission doesn't critically impair the model.

Exclusion of AnchiBERT Embedding: The marked decline in performance, particularly in the Macro F1 score (a drop of 0.074), accentuated the pivotal role of anchiBERT, reaffirming its adeptness in discerning classical Chinese subtleties.

Replacement of AnchiBERT with BERT-wwm: There was a noticeable dip in performance, although it was less significant than entirely excluding anchiBERT. This bolsters the claim that anchiBERT's pre-training on ancient Chinese grants it a competitive edge in classifying classical Chinese poetry.

Replacement of Word Embedding with Character Embedding: The observed performance was on par with that of the full model, indicating that, for the task of classifying ancient Chinese poetry, the granularity of embeddings—whether token-based or character-based—has a minimal impact on the results.

In conclusion, anchiBERT stands out as a quintessential component, highlighting the value of domain-specific pre-trained models. The granularity of embeddings, whether token or character-based, showed minimal variation in terms of influencing model performance.

4.5.2. Module freezing in ablation study The second series of ablation experiments adopted a different approach, concentrating on the model's robustness and efficacy when certain core components, specifically anchiBERT and word embeddings, were rendered static or "frozen", as shown in Table 5. The primary motivation was to discern the contribution levels of these elements towards the model's adaptability during the training process.

Freezing anchiBERT: When anchiBERT was frozen, a tangible decrease in both Micro and Macro F1 scores was evident. This outcome underscores the indispensability of fine-tuning anchiBERT, emphasizing the rich semantic and contextual information it gleans during training.

Freezing Word Embedding: Preventing updates to word embeddings resulted in a modest degradation in performance. While this decline wasn't as significant as in the case of freezing anchiBERT, it indicates that the model, beyond leveraging initial word embeddings, benefits from continuous adjustments during training, especially for classifying ancient Chinese poetry.

Dual Freezing: Concurrently freezing both anchiBERT and word embeddings culminated in a more pronounced drop in scores, showcasing the synergistic advantage of dynamically updating both components.

In essence, the adaptability of both anchiBERT and word embeddings augments the model's capabilities. Conversely, when these components are rendered inert, the model's prowess diminishes.

Table 5: Outcomes of Component Freezing in the Ablation Study

Configuration	Micro F1	Macro F1
Full model	0.949	0.930
anchiBERT Frozen	0.914	0.882
Word Embedding Frozen	0.939	0.915
Both Components Frozen	0.919	0.886

4.6. Discussion

The experimental results presented in this study paint a nuanced picture of the performance capabilities and potential pitfalls associated with varying model configurations and ablation strategies in the domain of ancient Chinese poetry classification. From the initial model performance results, it was evident that the TwinEmbedAttentionNet consistently outperformed baseline classifiers. This superior performance underscores the effectiveness of the approach adopted, especially when compared to widely acknowledged architectures such as BERT, XLNet, and AnchiBERT. While the proposed model exhibited promising results, the nuanced insights derived from the ablation studies indicate potential areas of exploration and optimization. This could involve further exploration of embedding granularities, the inclusion of additional pre-trained models, or more advanced fine-tuning strategies to continually push the boundaries of what's achievable in ancient Chinese poetry classification.

5. Conclusion

Ancient Chinese poetry stands as a treasured cornerstone of Chinese culture. While many NLP researchers have delved into this rich literary tapestry, the nuanced task of fine-grained thematic classification has often been overlooked. Given the distinctive linguistic style inherent to classical Chinese literature, we introduced TwinEmbedAttentionNet—a pioneering method crafted explicitly for this thematic classification. By seamlessly integrating both pretrained word and sentence embeddings, coupled with a multi-head attention module, our approach adeptly captures the thematic essence woven into the poetry. The experimental results attest to the distinct advantage and effectiveness of TwinEmbedAttentionNet. TwinEmbedAttentionNet, while effective, is heavily influenced by the sample number and quality of the learned data. Thus, it might falter when faced with poetic styles or themes outside the training spectrum. Also, the model's computational demands, given its architecture, could be challenging when scaling.

There's a plethora of avenues for exploration. Delving into cross-cultural applications, the model could be tested on literature from various eras and regions. By incorporating advanced embeddings or attention mechanisms, the model's thematic understanding might be further refined. Transfer learning techniques offer an avenue to boost model performance, especially in data-scarce situations. Collaborations with literary scholars can marry computational prowess with deep literary insights. Lastly, crafting an interactive platform for real-time thematic classification can be both a scholarly and public engagement tool.

References

[1] CKIPLab. Albert base model for Chinese on hugging face model hub. `https://huggingface.co/ckiplab/albert-base-chinese`, 2023. Accessed: 2023-08-26.

[2] Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, Shijin Wang, and Guoping Hu. Revisiting pre-trained models for Chinese natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*, pages 657–668, Online, November 2020. Association for Computational Linguistics.

DOI: 10.54254/2755-2721/39/20230619

[3] Nico H Frijda and Louise Sundararajan. Emotion refinement: A theory inspired by Chinese poetics. *Perspectives on Psychological Science*, 2(3):227–241, 2007.

[4] Chaoshuai Fu, Pan Li, Yadong Wu, Guiyu Zhang, and Jing Gao. Research on poetry sentiment analysis based on imagery. *Modern Computer*, 28(12):9–16, 2022. In Chinese.

[5] Sean Golden. No-man's land on the common borders of linguistics, philosophy & sinology: Polysemy in the translation of ancient Chinese texts. *The Translator*, 2(2):277–304, 1996.

[6] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[7] Jingrui Hou and Ping Wang. Assemble the shallow or integrate a deep? toward a lightweight solution for glyph-aware Chinese text classification. *Plos one*, 18(7):e0289204, 2023.

[8] Yufang Hou and Anette Frank. Analyzing sentiment in classical chinese poetry. In *Proceedings of the 9th SIGHUM workshop on language Technology for Cultural Heritage, social sciences, and humanities (LaTeCH)*, pages 15–24, 2015.

[9] Renfen Hu and Yuchen Zhu. Automatic classification of Tang poetry themes. *Acta Scientiarum Naturalium Universitatis Pekinensis*, 2:262–268, 2015.

[10] Jacob Devlin Ming-Wei Chang Kenton and Lee Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*, pages 4171–4186, 2019.

[11] Yoon Kim. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751, Doha, Qatar, October 2014. Association for Computational Linguistics.

[12] Charles Kwong. Nature and religion in ancient Chinese poetry. *Journal of Asian History*, 37(1):33–69, 2003.

[13] Siwei Lai, Liheng Xu, Kang Liu, and Jun Zhao. Recurrent convolutional neural networks for text classification. In *Proceedings of the AAAI conference on artificial intelligence*, volume 29, 2015.

[14] Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. ALBERT: A lite BERT for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942*, 2019.

[15] Yanfang Liu. A preliminary study on the dramatic factors of ancient Chinese poetry under the changing geographical environment. *Journal of Environmental and Public Health*, 2022, 2022.

[16] Xing Lu. *Rhetoric in ancient China, fifth to third century BCE: A comparison with classical Greek rhetoric*. Univ of South Carolina Press, 2022.

[17] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[18] Yabo Shen, Yong Ma, Chunguo Li, Shidang Li, Mingliang Gu, Chaojin Zhang, Yun Jin, and Yingli Shen. Sentiment analysis for Tang poetry based on imagery aided and classifier fusion. In *Artificial Intelligence for Communications and Networks: First EAI International Conference, AICON 2019, Harbin, China, May 25–26, 2019, Proceedings, Part II 1*, pages 283–290. Springer, 2019.

[19] Yongrui Tang, Xumei Wang, Peng Qi, and Yan Sun. A neural network-based sentiment analysis scheme for Tang poetry. In *2020 International Wireless Communications and Mobile Computing (IWCMC)*, pages 1783–1788, 2020.

[20] Naikan Tao. Subjectivity and innovation in contemporary Chinese poetry (part 2). *Journal of the Oriental Society of Australia*, 44:41–66, 2012.

[21] Huishuang Tian, Kexin Yang, Dayiheng Liu, and Jiancheng Lv. AnchiBERT: A pre-trained model for ancient Chinese language understanding and generation. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2021.

[22] Kui Shin Voo and Larry Hovee. The lamb of god hidden in the ancient Chinese characters. *CEN Technical Journal*, 13(1):81–91, 1999.

[23] Ping Wang, Jiangnan Li, and Jingrui Hou. S2SAN: A sentence-to-sentence attention network for sentiment analysis of online reviews. *Decision Support Systems*, 149:113603, 2021.

[24] Ping Wang, Shitou Zhang, Zuchao Li, and Jingrui Hou. Enhancing Ancient Chinese understanding with derived noisy syntax trees. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop)*, pages 83–92, Toronto, Canada, July 2023. Association for Computational Linguistics.

[25] Bin Wu, Jia Ji, Lin Meng, Chuan Shi, Huidong Zhao, and Yiqing Li. Sentiment analysis of Tang poems and Song lyrics based on transfer learning. *Journal of Electronics*, 11:2780–2787, 2016. In Chinese.

[26] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. XLNet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, 32, 2019.

[27] Michelle Yeh. “There are no camels in the Koran”: What is modern about modern Chinese poetry? In *New Perspectives on Contemporary Chinese Poetry*, pages 9–26. Springer, 2008.

[28] Hang Zhang and Zhongshi He. Research on poetry style classification technique based on BLSTM. *Modern Computer*, 02:12–17+23, 2020. In Chinese.

[29] Jiaqi Zhao, Yuting Wei, and Bin Wu. Analysis of the social network and the evolution of the influence of ancient Chinese poets. *Social Science Computer Review*, 40(4):1014–1034, 2022.

[30] Zhenjie Zhao and Xiaojuan Ma. Shadowplay2. 5d: A 360-degree video authoring tool for immersive appreciation of classical Chinese poetry. *Journal on Computing and Cultural Heritage (JOCCH)*, 13(1):1–20, 2020.