# A Multi-task Framework with Enhanced Hierarchical Attention for Sentiment Analysis on Classical Chinese Poetry: Utilizing Information from Short Lines


**Quanqi Du**
LT3, Ghent University
Groot-Brittanniëlaan 45, 9000 Gent
Flanders, Belgium
quanqi.du@ugent.be

**Veronique Hoste**
LT3, Ghent University
Groot-Brittanniëlaan 45, 9000 Gent
Flanders, Belgium
veronique.hoste@ugent.be



## Abstract

Classical Chinese poetry has a long history, dating back to the 11th century BC. By investigating the sentiment expressed in the poetry, we can gain more insights in the emotional life and history development in ancient Chinese culture. To help improve the sentiment analysis performance in the field of classical Chinese poetry, we propose to utilize the unique information from the individual short lines that compose the poem, and introduce a multi-task framework with hierarchical attention enhanced with short line sentiment labels. Specifically, the multi-task framework comprises sentiment analysis for both the overall poem and the short lines, while the hierarchical attention consists of word- and sentence-level attention, with the latter enhanced with additional information from short line sentiments. Our experimental results showcase that our approach leveraging more fine-grained information from short lines outperforms the state-of-the-art, achieving an accuracy score of 72.88% and an F1-macro score of 71.05%.


## 1 Introduction

Classical Chinese poetry generally refers to Chinese poetry composed in certain traditional forms and written in classical Chinese, a special terse, rhythmic and musical version of Chinese that is different from the modern Chinese today and mainly used in classical literature. As one of the earliest forms and most important genres of traditional literature, classical Chinese poetry is a crucial carrier of the thoughts and emotions of ancient Chinese literati. [1] In fact, it is widely believed that "emotion" is the essence of classical Chinese literary culture (Xia, 2021). By investigating the emotion expression in classical Chinese poetry, we are able to have a glimpse of the personal and social issues involved in the poetry, through which we can gain more insights into the ancient Chinese culture (Hou and Frank, 2015; Wei et al., 2020; Zhang et al., 2023).

[1] https://www.zgbk.com/ecph/words?SiteID=1&ID=389600&Type=bkztb&SubID=683

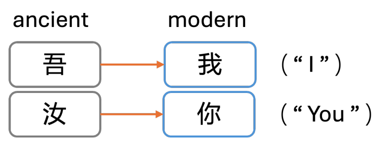


Figure 1: Pronoun changes from ancient Chinese to modern Chinese.

However, compared with the study of emotion in other text genres we encounter in modern life, such as news articles (Nemes and Kiss, 2021) and reviews (Yi and Liu, 2020), there is less research on poetry (Ahmad et al., 2020a), not to mention on classical Chinese poetry (Tang et al., 2020). Actually, it is more challenging to process ancient Chinese texts than modern Chinese text. On the one hand, the classical Chinese has some old vocabulary and phrases that have been replaced or fallen out of usage in modern Chinese. For instance, the pronouns in ancient Chinese are quite different from those in modern Chinese, as shown in Figure 1. On the other hand, the grammar of ancient Chinese is more concise and flexible compared to modern Chinese. As for classical Chinese poetry, it is considered as a more complicated synthesis of imagery and symbolism, allusions, quotations and derivation, and antithesis (Liu, 2022). All these factors make it a difficult task to conduct sentiment analysis in classical Chinese poetry.

To analyse sentiment in classical Chinese poetry, previous studies have explored different methods, for example, constructing sentiment lexicons (Hou and Frank, 2015; Zhang et al., 2023), transferring knowledge from modern Chinese (Zhao et al., 2014), or extracting imagery words (Shen et al., 2019; Su et al., 2023). Although these studies improved the general performance for the task



*Proceedings of the 4th International Conference on Natural Language Processing for Digital Humanities*, pages 113–122
November 16, 2024. 

of sentiment analysis in classical Chinese poetry, by utilizing special words in the poems or drawing upon knowledge beyond the poems, they did not consider the compositional structure of the poems. Usually, a classical Chinese poem comprises several short lines, which may show different emotions, and in return, contribute to the overall emotion expression of the poem. Thus in this paper, for the task of sentiment analysis of classical Chinese poetry, we propose to take the sentiment of short lines into consideration by using a multi-task framework with a hierarchical attention network, which includes the sentiment analysis task of both the overall poem and the short lines of which the poem is comprised. We will show that, by leveraging the sentiment information from the short lines, we can outperform the current state-of-the-art in sentiment analysis of ancient Chinese poetry.

## 2 Related Work

### 2.1 Sentiment analysis for computational poetry studies

Sentiment analysis using computational methodologies is receiving increasing attention in literary studies (Kim and Klinger, 2021). In 2012, Kao and Jurafsky applied sentiment lexicons to investigate whether the affect expressed in a poem makes it more beautiful or not. Later, Delmonte et al. (2013) introduced the SPARSAR system for the syntactic, semantic and prosodic analysis of poetry, which also includes a sentiment analysis module. In the comparison of formal and informal texts, Kaur and Saini (2014) found that taking into account genre-specific features helped to improve sentiment analysis performance on formal texts (including poetry). More recently, the advantages of machine learning further promoted the task of sentiment analysis in poetry. Promrit and Waijanya (2017) used Convolutional Neural Networks (CNNs) for category classification and sentiment analysis for Thai poems. Ahmad et al. (2020b) proposed an attention-based C-BiLSTM model to classify poetry texts into different emotional states, while Rajan and Salgaonkar (2020) employed a Naïve Bayes classifier on Konkani Poetry. Similar researches were also carried out on classical Chinese poetry (Tang et al., 2020), Punjabi poetry (Kaur and Saini, 2020) and Latin poetry (Sprugnoli et al., 2022).

### 2.2 Sentiment analysis for classical Chinese poetry

Although classifying the sentiment in classical Chinese poetry helps to investigate the ancient cultural and emotional life, this task is not as popular as in texts of other fields. Most studies in this field focus on the poems from the Tang Dynasty, which is believed to be the Gold Age of poetry, with over 50,000 poems created and surviving to this day.[2] On the basis of the Complete Anthology of Tang Poetry, Hou and Frank (2015) proposed a novel graph-based method to create a sentiment lexicon for classical Chinese poetry, with which they analyzed the association of sentiments with different poets and a variety of topics. To take full advantage of deep learning and linguistic knowledge, Zhang et al. (2023) combined supervised sentiment term extraction and classification to incorporate linguistic knowledge into deep learning models for the task of sentiment lexicon construction.

Besides the construction of sentiment lexicons, scholars also investigated extracting more information from the poems. Li and Li (2018) introduced the Frequent Pattern Growth Algorithm with the Term Frequency-Inverse Document Frequency to capture the hidden relationships between each word. (Shen et al., 2019) explored ways to extract sentimental imageries at the levels of both character and word and integrated this information in the task of sentiment analysis. To utilize the imagery words, Su et al. (2023) introduced related visual modality information and obtained better performance in sentiment analysis of classical Chinese poetry.

In addition to the information present in classical Chinese poetry, some scholars also tried to incorporate external knowledge into the task of sentiment analysis. Zhao et al. (2014) introduced the modern translations of ancient texts and transferred this knowledge in the translation to the classical Chinese poems. Liu et al. (2020) incorporated the knowledge of classical Chinese poetry from appreciation/translation annotations into the knowledge graph construction.

Furthermore, more recently, the development of large language models also aided in the promotion of research on classical Chinese literature. In 2021, Tian et al. released AnchiBERT, a pre-trained language model based on the architecture of BERT

[2] https://www.zgbk.com/ecph/words?SiteID=1&ID=272482&Type=bkdzb&SubID=808

and trained on large-scale ancient Chinese corpora. Then, based on BERT-base-Chinese from Google [3] and Chinese-BERT-wwm (Cui et al., 2021), Wang et al. (2022) designed the pre-trained model SikuBERT and SikuRoBERTa specifically for ancient Chinese. There are also other similar models for ancient Chinese, e.g., BERT-ancient-Chinese (Wang and Ren, 2022), GuwenBERT, [4] GujiBERT (Wang et al., 2023), and BERT_CCPoem, [5] which is particularly designed for classical Chinese poetry.

### 2.3 Sentiment analysis with multi-task learning and a hierarchical attention network

Traditional sentiment analysis is approached as an independent single task, but recent research has shown that it can also be considered as one of the sub-tasks in multi-task learning. Balikas et al. (2017) incorporated two sentiment analysis tasks (one with a 3-category label set and the other with a more fine-grained 5-category label set) into a multi-task model and demonstrated benefits by jointly learning the two sub-tasks for tweets. Aspect-based sentiment analysis is also a field where it is a common practice to simultaneously model the aspect term extraction and the corresponding sentiment classification (He et al., 2019; Zhao et al., 2023; Wu et al., 2023; Rani and Jain, 2024). Scholars also investigated combining with sentiment analysis other emotion-related tasks, e.g., complaint identification (Singh et al., 2022) and sarcasm detection (Tan et al., 2023). Chauhan et al. (2020) even proposed an all-in-one multi-task framework that incorporates five sub-tasks, including humour, sarcasm, offensive content, motivational content detection and sentiment analysis.

Following the “Attention is all you need" credo from Vaswani et al. (2017), the attention mechanism has been investigated in many variants and the hierarchical attention network (HAN) (Yang et al., 2016) is one of them. HAN contains the attention mechanisms at different levels, such as the word-level and sentence-level, capturing important words/sentences in the task context. It has been shown that HAN performs well with long-sequence information, especially when the information is ordered in a certain structure, for example in document-level classification (Pappas and Popescu-Belis, 2017) and translation (Werlen et al., 2018). HAN is also applied in sentiment-related tasks. Cheng et al. (2017), for example, designed a network with aspect attention and sentiment attention for aspect-level sentiment analysis. Mirroring the structure of the social media content, Cheng et al. (2019) proposed a network with HAN for cyberbullying detection on Instagram. HAN is also used to detect depression from transcripted clinical interviews (Mallol-Ragolta et al., 2019). More recently, Chanaa et al. (2021) exploited HAN for the task of E-learning text sentiment classification.

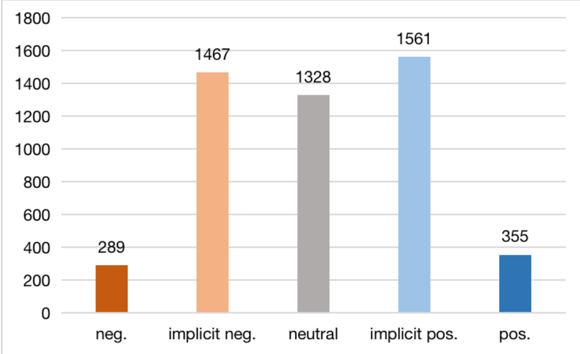


Figure 2: Sentiment label distribution in FSPC, with *negative* and *positive* abbreviated as *neg.* and *pos.* respectively.

As mentioned above, previous studies suggest that multi-tasking helps to improve the performance of sentiment analysis, but there is little such research in the field of classical Chinese poetry, which is mainly investigated by constructing auxiliary lexicons or by fusing knowledge from sources beyond the poetry. On the other hand, as classical Chinese poetry is highly structured and information-condensed, we hypothesize it is promising to apply HAN for sentiment analysis of this special genre, which has never been done before. In this paper, we aim to utilize the sentiment knowledge of the short lines to help predict the overall sentiment of the poetry with a HAN-based multi-task framework.

## 3 Dataset

Building a classical Chinese poetry corpus for sentiment analysis is not trivial, since full comprehension of the poems is one of the most important preconditions for sentiment labelling, which sets high expectations for annotators. Thanks to the work of Chen et al. (2019), we now have a fine-grained sentimental poetry corpus (FSPC), which we are going to utilize in our experiments.

FSPC is composed of 5,000 classical Chinese

[3]https://huggingface.co/google-bert/bert-base-chinese
[4]https://github.com/ethan-yt/guwenbert
[5]https://github.com/THUNLP-AIPoet/BERT-CCPoem

| lines | sentiment | overall sentiment |
|---|---|---|
| 诸侯分楚郡<br>The lords divide the land of Chu | neutral | negative |
| 饮饯五溪春<br>Feasting and bidding farewell in the spring of Five Rivers | positive | |
| 山水清晖远<br>The distant beauty of mountains and waters shines clear | neutral | |
| 俱怜一逐臣<br>All pity this one exiled courtier | negative | |

Figure 3: An example of the different sentiment labels at the line level and the overall sentiment of a given poem. Note that the English translation comes from ChatGPT and works only as reference.

poems. Each poem and each line in the poem are manually annotated by experts in Chinese literature with five classes, ranging from *negative* and *implicit negative*, over *neutral*, and to *implicit positive* and *positive*, as shown in Figure 2. Compared with *negative* and *positive*, implicit sentiments refer to the emotions that are suggested or hinted at but not directly stated. Due to the unbalanced label distribution, the implicit sentiments are merged into *negative* or *positive*, respectively, in later experiments. It should also be noted that the sentiment of each line can be different from each other, as shown in Figure 3, with the sentiment of the first line being neutral or implicit while the last line is often aligned with the overall sentiment of the poem (Chen et al., 2019). This suggests each line of the poem may have a varied contribution to the holistic sentiment of the whole poem, and it would be beneficial if we take into consideration the unique sentiment label of each line in the sentiment analysis of the whole poem.

## 4 Method

To investigate whether the task of sentiment analysis for the whole poem can benefit from the introduction of short line information, we use a framework that fine-tunes a pre-trained model with labels of both the lines and the overall poem, as shown in Figure 4.

**Data pre-processing** Each poem consists of 4 lines, and along with the poem, sentiment labels of both the lines and the overall poem are used for the pre-trained model fine-tuning. The use of "|" as separators between short lines in FSPC makes it easy to identify the line border.

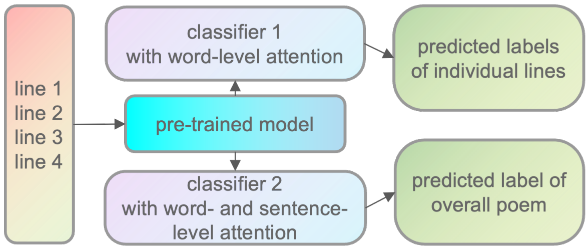


Figure 4: Framework to fine-tune the pre-trained model to predict labels of individual lines and the overall poem.

**Feature extraction and fine-tuning** The pre-trained model is used both in the feature extraction and fine-tuning stage. In the former case, the architecture of the pre-trained model is preserved, taking in the texts and outputting the feature vectors for the next stage. In the fine-tuning stage, two linear transformation layers are added at the end of the pre-trained model as classifiers for sentiment in the short lines and the overall poem, respectively. Both classifier 1 and classifier 2 use word-level attention, and sentence-level attention is also applied in classifier 2. For both the lines and the overall poem, the cross-entropy loss between the predicted logits and the true labels is calculated. The two loss values are then added to get a combined loss, based on which backpropagation is performed to compute gradients for model parameter updates.

It should be noted that for the short lines, each of them is independently encoded, and the encoding of the whole poem and the short lines are conducted separately.

**Enhanced sentence-level attention** To exploit the short line label information in the sentiment prediction task of the overall poem, additional information derived from the line labels is integrated into the sentence-level attention scores.

**Output evaluation** For the evaluation of the model predictions, both accuracy and F1-macro score are applied. Accuracy is one of the most intuitive performance evaluation metrics, but it may be misleading due to its sensitivity to class imbalance, which is why we also report the F1-macro score.

## 5 Experiment and Results

### 5.1 Experimental set-up

To evaluate the influence of sentiment analysis for short lines as a sub-task in the multi-task framework, we take the single task of sentiment analysis for the overall poems as the baseline.

We choose the dataset FSPC in our experiments. Considering the imbalanced distribution of the original five category labels, as shown in Figure 2, we merge the label *implicit negative* into *negative*, and *implicit positive* into *positive*. As the prediction difficulty of each poem may vary, ten-fold cross-validation (Kohavi, 1995) is applied in the experiment, providing a more reliable estimate of the model's performance compared to a single train-test split. The results reported later are the average results of the ten folds.

We designed two experiments: experiment 1 compares the influence of the framework complexity on the model prediction, while experiment 2 compares the performance of frameworks based on different pre-trained models.

**Experiment 1** In this experiment, two modes are designed with the pre-trained model SikuBERT, namely the multi-task mode and the single-task mode, with the latter including only the sentiment analysis task of the overall poem, used for the ablation study. For the multi-task mode, HAN is added and then additional information from the short line labels is combined with HAN to the network. It is expected that HAN is able to better capture the hierarchical structure of classical Chinese poetry.

**Experiment 2** Apart from SikuBERT, we also investigated other pre-traind models, i.e. BERT-base-Chinese, SikuRoberta, BERT-ancient-Chinese and BERT_CCPoem within the best framework from Experiment 1.

**Hyperparameters** For a fair comparison, all experiments are performed with the same dataset and hyperparameter settings. The optimizer is AdamW, with a learning rate of 1e-6. The epoch is set to 10 in experiment 1 and 20 in experiment 2, while the batch size is always set to 8. The mechanism of early stopping is applied in the experiment and the monitor is set as evaluation loss while the patience is set to 3. It should be noted that in both experiment 1 and 2, the early stopping is always triggered before the training reaches the epoch limit.

### 5.2 Experimental results

**Benchmark** To the best of our knowledge, there are recently 2 published studies on 3-class sentiment prediction with FSPC, evaluated with accuracy or F1 scores, and the best model achieved an F1-macro score of 64.8%, as shown in Table 1. Zhang et al. (2022) proposed to extract word vectors with two pre-trained models separately and then fuse them to enrich text vector information, while Hong et al. (2023) built a hybrid corpus of classical Chinese poetry with its modern Chinese translation, on the basis of which they fine-tuned a ChineseBERT model (Sun et al., 2021). Both these studies experimented with overall poem sentiment only.

| Studies | Acc | F1 |
|---|---|---|
| Zhang et al. (2022) | NA | 64.38 |
| Hong et al. (2023) | 67.10 | 64.80 |

Table 1: Benchmark results on the sentiment analysis task trained and evaluated with the FSPC dataset only.

| Method | Acc | F1 |
|---|---|---|
| single-task | 69.00 | 66.27 |
| multi-task | 69.32 | 66.50 |
| multi-task + HAN | 70.06 | 67.49 |
| **multi-task + HAN + line label** | **70.96** | **68.51** |

Table 2: Accuracy and F1-macro scores on the overall poem sentiment analysis task on the FSPC dataset based on the SikuBERT pre-trained model.

**Experiment 1** As shown in Table 2, the sentiment analysis performance improves with the increase in model complexity. To start with, the accuracy and F1-macro of the single-task model are 69.00% and 66.27% respectively, which already outperforms previous models in Table 1, which suggests the advantage of pre-trained models specifically designed for ancient Chinese, as the studies of both Zhang et al. (2022) and Hong et al. (2023) are based on pre-trained models for modern Chinese only.

Compared with the single-task model, the multi-task model achieves slightly higher scores in both accuracy (+0.32%) and F1 (+0.23%), suggesting the positive influence of the additional short line information on the performance of the overall poem sentiment prediction. Moreover, when the HAN is applied to the framework, a more noticeable improvement is witnessed, with an increase of 0.74% in accuracy and 0.99% in macro F1, indicating the advantage of HAN in dealing with structured texts. Furthermore, the introduction of short line labels has a positive influence on the model performance, reaching an accuracy score of 70.96% and an F1-macro score of 68.51%. We could thus conclude that leveraging both the poem structure and the fine-grained sentiment information at the line level

| Method | Acc | F1 |
|---|---|---|
| BERT_CCPoem | 67.54 | 65.24 |
| BERT-base-Chinese | 69.60 | 67.33 |
| BERT-ancient-Chinese | 70.28 | 68.31 |
| SikuBERT | 70.96 | 68.51 |
| SikuRoBERTa | **72.88** | **71.05** |

Table 3: Accuracy and F1-macro scores of the overall poem sentiment analysis task using different pre-trained models in combination with the best framework in Experiment 1 (multi-task+HAN+line label).

is advantageous for sentiment classification of the classical Chinese poems.

**Experiment 2** Results of other pre-trained models with the best framework in Experiment 1 are shown in Table 3. The F1-macro scores range from 65.24% to 71.05%, the latter being obtained by SikuRoBERTa.

BERT-base-Chinese is the only model that is not pre-trained specifically for ancient Chinese. Considering the difference between ancient Chinese and modern Chinese, it is not surprising that BERT-base-Chinese does not perform better with classical Chinese poems than most of the other models, except the model BERT_CCPoem. Although BERT_CCPoem is the only model pre-trained specifically for Chinese ancient poems in the experiments, it shows worse accuracy and macro F1-scores compared to the other models. One possible reason could be that the training data of BERT_CCPoem is limited to poems, resulting in a smaller vocabulary of 11,809, while other models, such as BERT-base-Chinese and BERT-ancient-Chinese have a larger vocabulary of 21,128 and 38,208 respectively.

The SikuRoBERTa-based model within the proposed framework (multi-task+HAN+line label) in our experiments achieves the best performance, with an accuracy of 72.88% and an F1-macro score of 71.05%, outperforming previous studies, as shown in Table 1.

## 6 Prediction Analysis

With the ten-fold cross-validation in Experiment 2, we obtained predictions for all 5,000 poems using the SikuRoBERTa model in the multi-task framework with line-label-enhanced hierarchical attention, as shown in Figure 5.

First of all, it is clear that the model performs better with poems with clear sentiment polarities,

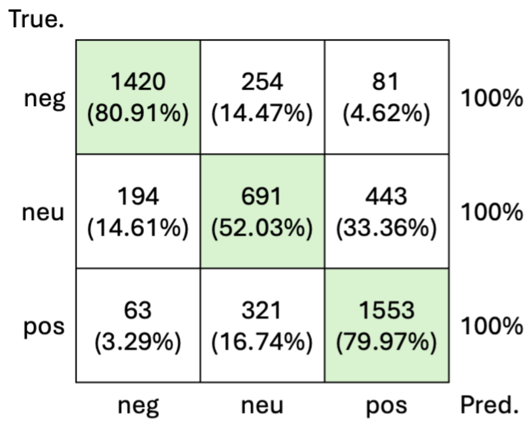


Figure 5: Sentiment prediction distribution of SikuRoBERTa-based multi-task framework with enhanced hierarchical attention. The X-axis represents the predicted labels, while the Y-axis stands for the true labels

i.e., negative and positive, reaching an accuracy of around 80%, while the model has only fair predictions with neutral poems, with an accuracy of about 52%. For the negative and positive poems, more wrong predictions are distributed on their nearest sentiment neighbour, i.e., neutral, rather than the far/opposite sentiment neighbours, i.e., positive and negative respectively, suggesting the model's ability to differentiate between positive and negative poems.

As for the neutral poems, the model achieves an accuracy of 52.03%, which is far less than the performance with negative and positive poems, indicating that differentiating between neutral and implicit sentiment is still challenging. Although only about half are correctly predicted, around one-third of the wrong predictions are predicted as negative and the rest as positive, with the latter outnumbering and even doubling the former, indicating that the model tends to have more non-negative predictions on neutral poems. To gain more insights into this tendency, we further investigated the short line labels of 194 false negative poems and 443 false positive poems which are labelled as neutral by the annotators as shown in Figure 5. For both the false negative and false positive groups of poems, we plotted the sentiment label distribution of each line group (i.e. all first lines are joined into one group, the same holds for the second lines, etc.), and the results are shown in Figure 6 and 7.

As shown in Figure 6, for the 194 neutral poems that are predicted as negative by the model, line 1 and line 2 seem to hold sentiment labels that

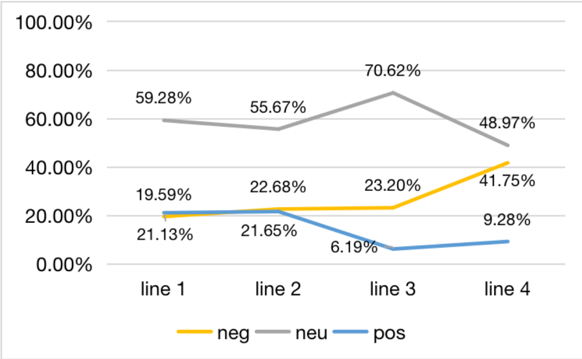


Figure 6: True sentiment label distribution across short lines in 194 neutral poems that are predicted as negative.

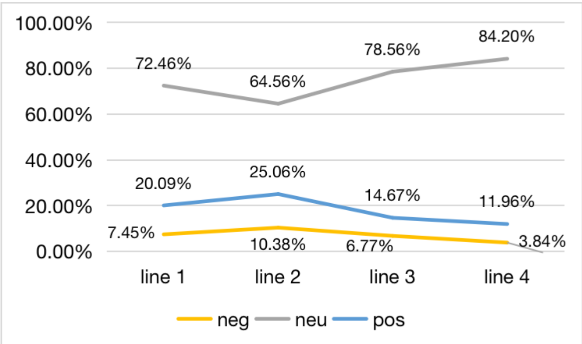


Figure 7: True sentiment label distribution across short lines in 443 neutral poems that are predicted as positive.

contain almost equal percentages of negative and positive labels. In line 3, however, there is more negative sentiment labelled than positive and this tendency continues in line 4, where the negative sentiment reaches 41.75%, which is more than four times of the number of positive annotations. On the other hand, as shown in Figure 7, when we consider the sentiment annotations at the line level of the 443 neutral poems that were predicted as positive by the model, we see an opposite tendency with always more positive sentiment labels than negative ones. Based on the sentiment distribution across short lines, we hypothesize that although the model fails to make correct predictions for about half of the neutral poems, it still senses the more implicit sentiment polarities in the poems, leading to negative predictions for the more negative poems and positive predictions for the more positive poems.

Moreover, we also investigate specific poems where there is a *significant difference* between the true labels and the predictions. By *significant difference*, we refer to the cases where the poem is classified with the complete opposite polarity of its labeled polarity, e.g. the overall sentiment of

<table>
<tr><th rowspan="2">Poem</th><th colspan="2">True labels</th><th colspan="2">Prediction</th></tr>
<tr><th>line</th><th>overall</th><th>line</th><th>overall</th></tr>
<tr><td>珠树森森秀阮林<br>The pearl trees flourish, lush as in the Ruan Forest</td><td>neutral</td><td rowspan="4">negative</td><td>positive</td><td rowspan="4">positive</td></tr>
<tr><td>高堂有母各欢心<br>In the grand hall, the mother's heart is filled with joy</td><td>positive</td><td>positive</td></tr>
<tr><td>兰陔娱养时多暇<br>Amidst the fragrant garden, time for care is ample</td><td>neutral</td><td>positive</td></tr>
<tr><td>勿遣平安阙嗣音<br>Let not the news of peace be left unheard by the heir</td><td>negative</td><td>neutral</td></tr>
</table>

Figure 8: Comparison of the true labels and predictions of one poem. Note that the English translation comes from ChatGPT and works only as reference.

the poem is predicted as positive while the true label is negative, or vice versa. Of the 5,000 poems, there were in total 144 poems classified with this opposite polarity, and we found that this opposite polarity labelling happens both at the line level and the poem level. Figure 8 shows an example poem with its manual annotations and model predictions for the short lines and the overall poem. For the short lines, the model predicted positive instead of neutral for the first and the third lines, and predicted neutral instead of negative for the last line, generally upgrading the short line sentiment toward the polarity of positive, which might lead to the positive overall prediction. However, if contextual or topic information of the poem, which is "farewell", would be provided, and if the model learns from other poems that "farewell" is usually related to negative emotions, the possibility of the model predicting this poem as overall negative would increase.

## 7 Conclusion

In this paper, we introduced a multi-task framework with enhanced hierarchical attention for sentiment analysis on classical Chinese poetry. This multi-task framework consists of two sub-tasks, the sub-task of sentiment analysis on the short lines in the poem, and the sub-task of sentiment analysis on the overall poem. For the latter task, a hierarchical attention network composed of word- and sentence-level attention was applied. Furthermore, to further utilize the information from short lines, additional information from short line sentiment was introduced to the sentence-level attention.

Experiments on the FSPC dataset show that our framework, compared with the single-task setup and based on the pre-trained model SikuBERT, yields a increase of macro F1 from 66.27% to 68.51%. We also experimented with other pre-trained models, and the best performance is demonstrated by SikuRoBERTa, with an accuracy of 72.88% and an F1-macro of 71.05%, thus largely outperforming the state-of-the-art with an increase 5.78% in accuracy and 6.25% in F1-macro score (Hong et al., 2023).

We also investigated the SikuRoBERTa-based model predictions in more detail. We found that the model achieves an accuracy of about 80% with negative and positive poems, but only an accuracy of around 50% with neutral poems, which might be related to the "implicit emotion" writing style favoured in classical Chinese poems. Moreover, a further look at the predictions on the neutral poems suggested that the incorrect predictions are subtly aligned with the more prominent positive or negative sentiment polarities in the short lines, which indicates that regardless of the incorrect predictions on the neutral poems, the model still senses whether the poems are more positive or negative.

## 8 Future Work

The multi-task framework proposed in this paper has demonstrated potential in the task of sentiment analysis on classical Chinese poetry. It would also be interesting to introduce the task of emotion recognition in the framework, although this might require more annotations. Moreover, as the introduction of short line labels helps to improve the model performance, it is also promising to include additional information, such as the background or the topic of the poem to make more accurate predictions. We will also investigate how these different types of information can be optimally fused.

## 9 Limitation

In the experiment of this paper, the results are limited to the dataset FSPC which contains 5000 instances. A larger dataset would help to produce a stronger statement.

## 10 Acknowledgements

This research received funding from the Flemish Government under the Research Program Artificial Intelligence (174E07824). We would also like to thank the anonymous reviewers for their valuable and constructive feedback.

## References

Shakeel Ahmad, Muhammad Zubair Asghar, Fahad Mazaed Alotaibi, and Sherafzal Khan. 2020a. Classification of poetry text into the emotional states using deep learning technique. *IEEE Access*, 8:73865–73878.

Shakeel Ahmad, Muhammad Zubair Asghar, Fahad Mazaed Alotaibi, and Sherafzal Khan. 2020b. Classification of poetry text into the emotional states using deep learning technique. *IEEE Access*, 8:73865–73878.

Georgios Balikas, Simon Moura, and Massih-Reza Amini. 2017. Multitask learning for fine-grained twitter sentiment analysis. In *Proceedings of the 40th international ACM SIGIR conference on research and development in information retrieval*, pages 1005–1008.

Abdessamad Chanaa et al. 2021. E-learning text sentiment classification using hierarchical attention network (han). *International Journal of Emerging Technologies in Learning (iJET)*, 16(13):157–167.

Dushyant Singh Chauhan, SR Dhanush, Asif Ekbal, and Pushpak Bhattacharyya. 2020. All-in-one: A deep attentive multi-task learning framework for humour, sarcasm, offensive, motivation, and sentiment on memes. In *Proceedings of the 1st conference of the Asia-Pacific chapter of the association for computational linguistics and the 10th international joint conference on natural language processing*, pages 281–290.

Huimin Chen, Xiaoyuan Yi, Maosong Sun, Cheng Yang, Wenhao Li, and Zhipeng Guo. 2019. Sentiment-controllable chinese poetry generation. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, Macao, China.

Jiajun Cheng, Shenglin Zhao, Jiani Zhang, Irwin King, Xin Zhang, and Hui Wang. 2017. Aspect-level sentiment classification with heat (hierarchical attention) network. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 97–106.

Lu Cheng, Ruocheng Guo, Yasin Silva, Deborah Hall, and Huan Liu. 2019. Hierarchical attention networks for cyberbullying detection on the instagram social network. In *Proceedings of the 2019 SIAM international conference on data mining*, pages 235–243. SIAM.

Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, and Ziqing Yang. 2021. Pre-training with whole word masking for Chinese BERT. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:3504–3514.

Rodolfo Delmonte et al. 2013. Computing poetry style. In *Proceedings of the First International Workshop on Emotion and Sentiment in Social and Expressive*

*Media: approaches and perspectives from AI (ESSEM 2013)*, volume 1096, pages 148–155. CEUR Workshop Proceedings (CEUR-WS. org).

Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2019. An interactive multi-task learning network for end-to-end aspect-based sentiment analysis. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 504–515.

Jie Hong, Tingting He, Jie Mei, Ming Dong, Zheming Zhang, and Xinhui Tu. 2023. A hybrid corpus based fine-grained semantic alignment method for pre-trained language model of ancient Chinese poetry. In *2023 IEEE International Conference on Big Data (BigData)*, pages 4794–4801. IEEE.

Yufang Hou and Anette Frank. 2015. 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.

Justine Kao and Dan Jurafsky. 2012. A computational analysis of style, affect, and imagery in contemporary poetry. In *Proceedings of the NAACL-HLT 2012 workshop on computational linguistics for literature*, pages 8–17.

Jasleen Kaur and Jatinderkumar R Saini. 2014. Emotion detection and sentiment analysis in text corpus: a differential study with informal and formal writing styles. *International Journal of Computer Application, ISSN*, pages 0975–8887.

Jasleen Kaur and Jatinderkumar R Saini. 2020. Designing Punjabi poetry classifiers using machine learning and different textual features. *Int. Arab J. Inf. Technol.*, 17(1):38–44.

Evgeny Kim and Roman Klinger. 2021. A survey on sentiment and emotion analysis for computational literary studies. *Zeitschrift für digitale Geisteswissenschafte*.

Ron Kohavi. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI'95, page 1137–1143, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Gang Li and Jie Li. 2018. Research on sentiment classification for Tang poetry based on TF-IDF and FP-growth. In *2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, pages 630–634. IEEE.

James JY Liu. 2022. *The art of Chinese poetry*. Routledge.

Yutong Liu, Bin Wu, and Ting Bai. 2020. The construction and analysis of classical Chinese poetry knowledge graph. *Journal of Computer Research and Development*, 57(6):1252–1268.

Adria Mallol-Ragolta, Ziping Zhao, Lukas Stappen, Nicholas Cummins, and Björn W. Schuller. 2019. A hierarchical attention network-based approach for depression detection from transcribed clinical interviews. In *20th Annual Conference of the International Speech Communication Association (INTERSPEECH 2019)*, pages 221–225.

László Nemes and Attila Kiss. 2021. Prediction of stock values changes using sentiment analysis of stock news headlines. *Journal of Information and Telecommunication*, 5(3):375–394.

Nikolaos Pappas and Andrei Popescu-Belis. 2017. Multilingual hierarchical attention networks for document classification. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1015–1025.

Nuttachot Promrit and Sajjaporn Waijanya. 2017. Convolutional neural networks for Thai poem classification. In *Advances in Neural Networks-ISNN 2017: 14th International Symposium, ISNN 2017, Sapporo, Hakodate, and Muroran, Hokkaido, Japan, June 21–26, 2017, Proceedings, Part I 14*, pages 449–456. Springer.

Annie Rajan and Ambuja Salgaonkar. 2020. Sentiment analysis for Konkani Language: Konkani Poetry, a case study. In *ICT Systems and Sustainability: Proceedings of ICT4SD 2019, Volume 1*, pages 321–329. Springer.

Somiya Rani and Amita Jain. 2024. Aspect-based sentiment analysis of drug reviews using multi-task learning based dual BiLSTM model. *Multimedia Tools and Applications*, 83(8):22473–22501.

Yabo Shen, Yong Ma, Chunguo Li, Shidang Li, Mingliang Gu, Chaojin Zhang, Yun Jin, and Yingli Shen. 2019. 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*, pages 283–290. Springer.

Apoorva Singh, Sriparna Saha, Md Hasanuzzaman, and Kuntal Dey. 2022. Multitask learning for complaint identification and sentiment analysis. *Cognitive Computation*, 14(1):212–227.

Rachele Sprugnoli, Francesco Mambrini, Marco Passarotti, Giovanni Moretti, et al. 2022. Sentiment analysis of Latin poetry: First experiments on the Odes of Horace. *Computational Linguistics CliC-it 2021*, page 314.

Chang Su, Shupin Liu, and Chalian Luo. 2023. MISC: A multimodal approach for sentiment classification of classical Chinese poetry. In *International Conference on Intelligent Computing*, pages 432–442. Springer.

Zijun Sun, Xiaoya Li, Xiaofei Sun, Yuxian Meng, Xiang Ao, Qing He, Fei Wu, and Jiwei Li. 2021. ChineseBERT: Chinese pretraining enhanced by glyph and Pinyin information. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 2065–2075, Online. Association for Computational Linguistics.

Yik Yang Tan, Chee-Onn Chow, Jeevan Kanesan, Joon Huang Chuah, and YongLiang Lim. 2023. Sentiment analysis and sarcasm detection using deep multi-task learning. *Wireless personal communications*, 129(3):2213–2237.

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

Huishuang Tian, Kexin Yang, Dayiheng Liu, and Jiancheng Lv. 2021. 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.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Dongbo Wang, Chang Liu, Zhixiao Zhao, Si Shen, Liu Liu, Bin Li, Haotian Hu, Mengcheng Wu, Litao Lin, Xue Zhao, et al. 2023. GujiBERT and GujiGPT: Construction of intelligent information processing foundation language models for ancient texts. *arXiv preprint arXiv:2307.05354*.

Dongbo Wang, Chang Liu, Zihe Zhu, Jiangfeng Liu, Haotian Hu, Si Shen, and Bin Li. 2022. Sikubert and sikuroberta: Construction and application of pretrained models of Siku Quanshu in orientation to digital humanities. *Library Tribune*, 42(6):31–43.

Pengyu Wang and Zhichen Ren. 2022. The uncertainty-based retrieval framework for ancient Chinese CWS and POS. In *Proceedings of the Second Workshop on Language Technologies for Historical and Ancient Languages*, pages 164–168.

Yuting Wei, Huazheng Wang, Jiaqi Zhao, Yutong Liu, Yun Zhang, and Bin Wu. 2020. GeLaiGeLai: a visual platform for analysis of classical Chinese poetry based on knowledge graph. In *2020 IEEE International Conference on Knowledge Graph (ICKG)*, pages 513–520. IEEE.

Lesly Miculicich Werlen, Dhananjay Ram, Nikolaos Pappas, and James Henderson. 2018. Document-level neural machine translation with hierarchical attention networks. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2947–2954.

Zhaozhen Wu, Guoyi Cao, and Wanghao Mo. 2023. Multi-tasking for aspect-based sentiment analysis via constructing auxiliary self-supervision acop task. *IEEE Access*.

Chen Xia. 2021. Poetry and emotion in classical Chinese literature. In *The Routledge Handbook of Chinese Studies*, pages 289–303. Routledge.

Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical attention networks for document classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*, pages 1480–1489.

Shanshan Yi and Xiaofang Liu. 2020. Machine learning based customer sentiment analysis for recommending shoppers, shops based on customers' review. *Complex & Intelligent Systems*, 6(3):621–634.

Lingli Zhang, Yadong Wu, Qikai Chu, Pan Li, Guijuan Wang, Weihan Zhang, Yu Qiu, and Yi Li. 2022. SA-model: multi-feature fusion poetic sentiment analysis based on a hybrid word vector model. In *2022 5th International Conference on Pattern Recognition and Artificial Intelligence (PRAI)*, pages 984–988.

Wei Zhang, Hao Wang, Min Song, and Sanhong Deng. 2023. A method of constructing a fine-grained sentiment lexicon for the humanities computing of classical Chinese poetry. *Neural Computing and Applications*, 35(3):2325–2346.

Guoshuai Zhao, Yiling Luo, Qiang Chen, and Xueming Qian. 2023. Aspect-based sentiment analysis via multitask learning for online reviews. *Knowledge-Based Systems*, 264:110326.

Huidong Zhao, Bin Wu, Haoyu Wang, and Chuan Shi. 2014. Sentiment analysis based on transfer learning for Chinese ancient literature. In *2014 International Conference on Behavioral, Economic, and Socio-Cultural Computing (BESC2014)*, pages 1–7. IEEE.