

Exploring the trend of stylistic evolution of chinese modern and contemporary literary works based on big data analysis

Yiling Sun^{1,✉}

¹ Faculty of Art and Social Science, National University of Singapore, Singapore, 119077, Singapore

ABSTRACT

The study analyzes the stylistic evolution of contemporary Chinese literary works using the MONK project. Text mining tools in the project are used to analyze the thematic classification, emotional tendency and stylistic type changes of the works. Among them, LDA model and GBDT algorithm are used to identify the thematic classification of Chinese modern and contemporary literary works, SO-PMI algorithm is used to identify the emotional tendency in the works, and the vector space model can classify the style of the works. Based on the above methods, the theme and emotional changes of modern and contemporary Chinese literary works can be categorized into 3 stages: the awakening of Enlightenmentism at the beginning of the 20th century, the diversified presentation during the revolutionary period, and the diversified development after the reform and opening up. The styles of modern and contemporary Chinese literary works can be divided into epic style, lyrical style, rural theme style and intellectual theme style.

Keywords: MONK project, LDA model, SO-PMI algorithm, vector space model, style evolution trend

1. Introduction

Modern Chinese literature is a new literature formed under the conditions of historical changes within the Chinese society and the wide acceptance of the influence of world literature [14]. Contemporary Chinese literature, on the other hand, is the literature of China since 1949 or the literature that took place in a specific socialist historical context [21]. They are new creations in literary genres, narrative perspectives, lyricism and descriptive means in the corresponding social stages. There is a deep connection between Chinese modern and contemporary literature and Chinese excellent

✉ Corresponding author.

E-mail address: sunyiling2024@163.com (Yiling Sun).

Received 18 August 2024; Accepted 09 December 2024; Published Online 18 March 2025.

DOI: [10.61091/jcmcc124-31](https://doi.org/10.61091/jcmcc124-31)

© 2025 The Author(s). Published by Combinatorial Press. This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).

traditional culture, which is not only reflected in its borrowing and absorption of Chinese excellent traditional culture, but also in its inheritance and development of the essence of Chinese excellent traditional culture [9, 10]. Studying the evolution trend of the styles of modern and contemporary Chinese literary works will help us stand better in the forest of world culture and further strengthen our cultural confidence [6, 18].

Computational modeling of human language information is an important research field in the era of big data, and it is also the foundation and premise for human-computer interaction to be realized [8]. Human language often contains extremely complex and interrelated information, and rich rhetorical techniques and implied extra-verbal meanings are prevalent, which is not a small challenge for big data recognition and analysis [15, 7, 12]. Among them, the process of natural language processing technology starts from the text, which usually goes through word segmentation, lexical annotation, and syntactic analysis to realize the core task of semantic analysis [11, 1]. To understand more linguistic style information implied beyond the text, such as different literary genres, emotional tendencies, etc., the understanding of contextual structure and overall background puts forward higher requirements, which is also the key to realize the advanced semantic analysis and data analysis tasks [13, 2, 17, 4]. Utilizing big data analytics, natural language processing breakthroughs are focused on more specific linguistic analyses, in which the understanding and generation of different linguistic styles is key [5, 19].

The study first applies the MONK project to text mining of Chinese literary works resources from the early 20th century to the early 21st century, and preprocesses the data through the steps of data format conversion, text annotation, metadata addition, and data entry. Then the LDA model was applied to extract themes from the works, and the GBDT model was used to classify the themes. And the SO-PMI algorithm was used to automatically construct the domain emotion dictionary, and the emotion word vectors were obtained to delineate the tendency of emotion change in different stages of Chinese modern and contemporary literary works. Finally, based on the vector space model to transform the unstructured text data into highly structured text vectors, the text vectors are clustered and analyzed using the K-means clustering method to delineate the style types of Chinese modern and contemporary literary works.

2. MONK-based text mining of contemporary chinese literary works

As a digital humanities platform, the MONK project contains a large number of electronic resources in the field of humanities as well as tools for analyzing and mining these electronic texts. Using mathematical models such as the LDA theme model, the emotion lexicon and the vector space model in the MONK project, it is possible to conduct relevant explorations of the variations in the thematic categorization, emotional tendencies and stylistic types of modern and contemporary Chinese literary works.

2.1. *Pre-processing flow of MONK*

The MONK project is a combination of digital library and text mining project, where the preprocessing process includes the permanent storage and labeling of related digital resources. The platform provided by MONK allows users to select the corresponding processed resources of Chinese modern and contemporary literature to be mined without the need for preprocessing and other work. The text pre-processing of the project before this is a huge project, and the process of MONK text pre-

processing includes steps such as data format conversion, text annotation, metadata addition, and data warehousing. Data format conversion is mainly to convert various types of data into a uniform data format that can be exchanged. Text annotation is mainly to ensure the consistency of text in terms of spelling, word form, dialect, etc., and create document space. Metadata addition is mainly to add document-level attributes (author, date, genre, provenance, etc.), which are used to mine the relationship of various attributes. The annotated text data will be mapped to a relational database, and the text data, metadata, and preprocessed data of other text objects will be mined and analyzed by a coupled-structure “object model” written in JAVA.

2.2. Research on theme classification of literary works based on theme modeling

2.2.1. LDA model. The goal of the LDA model is to find the distribution of topics for each document and the probability distribution of words in each topic. First the number of topics of the synthesized document needs to be determined, denoted as K , and all distributions are expanded based on K topics [3].

LDA assumes that the prior distribution of document topics satisfies the Dirichlet distribution, i.e., for any document d , its topic distribution satisfies $\theta_d : \theta_d = \text{Dirichlet}(\vec{\alpha})$, where α is a hyperparameter of the distribution and is a K -dimensional vector.

LDA assumes that the prior distribution of words in topics is also Dirichlet distribution, i.e., for any topic k , its distribution of words β_k satisfies $\beta_k = \text{Dirichlet}(\vec{\eta})$, where η is the hyperparameter of the distribution and is an V -dimensional vector. V represents the size of the vocabulary list.

For the n th word in any document d , the posterior distribution of the topic distribution θ_d is:

$$\text{Dirichlet}(\theta_d | \vec{\alpha} + \vec{\eta}_d). \quad (1)$$

The posterior distribution of β_k is:

$$\text{Dirichlet}(\beta_k | \vec{\eta} + \vec{\eta}_k). \quad (2)$$

Since topic word generation does not depend on a specific document, document topic distribution and topic word distribution are independent.

For solving the topic distribution of each document and the distribution of words in each topic, each work or all the readers' comments of that work can be regarded as a document, and the number of topics is set to 1, then the topic words and their weights of that work or readers' comments can be extracted.

2.2.2. GBDT algorithm. GBDT is known as Gradient Boosting Decision Tree and can be used for regression or classification. With the continuous development of deep learning, it has been applied more in relation extraction tasks with its advantage of automatic feature extraction [20].

In the iteration of GBDT, assuming that the strong learner obtained in the previous round of iteration is $f_{t-1}(x)$ and the loss function is $L(y, f_{t-1}(x))$, our goal in this round of iteration is to find a weak learner $h_t(x)$ for the CART regression tree model to minimize the loss function $L(y, f_i(x)) = L(y, f_{i-1}(x) + h_i(x))$ in this round. That is, this round of iterations finds the decision tree to make the sample loss as small as possible.

By fitting the negative gradient of the loss function, we can solve classification and regression problems with GBDT by fitting the loss error, so that both classification and regression problems can be solved by fitting the negative gradient of their loss functions. The difference is simply that the negative gradient is different due to the different loss functions.

In our application, it is actually the multivariate (3 class labels) GBDT classification algorithm, assuming that the number of classes is $K = 3$, then the log-likelihood loss function at this point is:

$$L(y, f(x)) = - \sum_{k=1}^K y_k \log p_k(x), \quad (3)$$

where if the sample output category is k , then $y_k = 1$. The expression for the probability $p_k(x)$ of category k is:

$$p_k(x) = \exp(f_k(x)) / \sum_{k=1}^k \exp(fl(x)). \quad (4)$$

2.3. Research on emotional tendencies of literary works based on emotional lexicon

The most important core of the lexicon-based sentiment analysis method is the “lexicon plus scoring rules” method, which mainly adopts the manually organized and constructed sentiment lexicon to match with each sentiment word in the text, and sets the scoring rules of the sentiment score based on the location of the sentiment word, lexical nature, the combination of the degree of adverbs and the number of negatives, and then calculates the sentiment score of the work in accordance with the rules, so as to obtain the sentiment tendency of the work. The sentiment score of the literary work is calculated according to the rules, so as to get the sentiment tendency of the work. The sentiment dictionary in the MONK project uses a domain-specific dictionary constructed based on the SO-PMI algorithm, which can effectively improve the accuracy of the sentiment analysis [16].

2.3.1. SO-PMI algorithm. The Sentimental Orientation Point Mutual Information (SO-PMI) algorithm consists of two parts: the PMI and the SO, which is an extension of the Point Mutual Information (PMI) algorithm.

Therefore, firstly, the PMI algorithm is elaborated, the main use of PMI algorithm is to calculate the semantic similarity between words, its basic idea is to solve the possibility of A_word and B_word appearing in a text at the same time, i.e., $P(A_word \cap B_word)$, if $P(A_word \cap B_word)$ is larger, it indicates that the two words are more related to each other, and the formula for the PMI value between words is:

$$PMI(A_word, B_word) = \log_2 \left(\frac{P(A_word \cap B_word)}{P(A_word) P(B_word)} \right), \quad (5)$$

where $P(A_word \cap B_word)$ represents the probability of A_word and B_word occurring together in the whole corpus, and $P(A_word)$ and $P(B_word)$ represent the probability of A_word and B_word occurring individually in the corpus, when $P(A_word \cap B_word)$ is larger, which in turn indicates that the degree of association between the two words is higher, and vice versa, the degree of association is smaller. It is generally believed that:

- (1) At $PMI > 0$, the two terms are correlated and the greater the PMI the stronger the correlation.
- (2) At $PMI = 0$, the two terms are independent of each other.
- (3) At $PMI < 0$, the two terms are mutually exclusive.

Sentiment Orientation Point Mutual Information Algorithm (SOPMI) is introduced on the basis of PMI algorithm, and it can accurately capture the sentiment words by adding the sentiment orientation (SO) of the words. The steps of the algorithm are as follows:

Step 1: A set of positive and negative words with strong emotional tendency are selected as the benchmark words (seed words) and are represented by Pos_words and Neg_words respectively.

Step 2: For word A_word , subtracting the PMI of A_word with Pos_words from the PMI of A_word with Neg_words will result in an emotional tendency of A_word , i.e., SO-PMI.

Step 3: Calculate the SO-PMI value with 0 as the threshold, when the calculation result is greater than 0, then the word is a word with positive emotional tendency. If the calculation result is equal to 0, the word is a word with no significant emotional tendency. On the contrary, if the calculation result is less than 0, the word is a negative emotional word.

2.3.2. Construction of the domain sentiment lexicon. Different domains have different comment objects, and for a certain domain, there may exist certain words with more explicit emotional tendencies that are not found in the basic sentiment dictionary, so it is difficult to meet the individualized needs of different domains by relying only on the general sentiment dictionary. This paper constructs a domain sentiment dictionary based on the SO-PMI algorithm.

In order to use SO-PMI algorithm, we must first find words with obvious emotional tendency as seed words, which need to be customized, and generally find words with strong emotional tendency from the whole corpus, but it is too much workload to look for them one by one from the huge amount of text, so this paper extracts the keywords of the text with the help of TF-IDF algorithm, so as to construct the emotion seed words.

The TF-IDF algorithm is called Word Frequency-Inverse Document Rate. TF is the word frequency, which indicates how often a word occurs in the corpus. IDF is the Inverse Document Frequency Index, which is the number of texts in which the word occurs in n text. If the fewer the documents containing the word, then the larger the IDF, which indicates that the word is more representative. The TF-IDF algorithm is the product of the TF value and the IDF value, which is calculated as follows. The formula is as follows:

$$\begin{aligned} TF - IDF &= TF \times IDF \\ &= \frac{\text{The number of times a word occurs}}{\text{Total number of words in the article}} \times \log \left(\frac{\text{Total number of samples}}{\text{Number of documents containing the word}_{+1}} \right). \end{aligned} \quad (6)$$

TF-IDF is calculated using TfidfVectorizer on the text after jieba segmentation and sorted according to the size of its value. After the baseline word search is completed, the SO-PMI value between each word and the baseline word is calculated to determine whether a word is more inclined to appear with positive words or with negative words, by which the emotional tendency of the word is judged. Using the ChineseSoPmi function in the wordexpansion module, the SO-PMI values of the words are calculated and stored in two parts, the positive lexicon and the negative lexicon. The larger the SO-PMI value is, the more pronounced the emotional tendency is.

2.4. Research on style classification of literary works based on vector space modeling

2.4.1. Vector space model. The vector space model can transform a text into a vector representation consisting of feature terms and corresponding weights. In the experiment, the model is used to represent the style of a literary work.

For a text d_i , the feature term $t(j = 1, 2, \dots, n)$ is a different feature word, and by noting the weight of feature term t_j in text d_i as w_{ij} , the text d_i can be converted into the corresponding spatial feature vector as shown in Eq. (7):

$$V(d_i) = [(t_1, w_{i1}), (t_2, w_{i2}), \dots, (t_n, w_{in})]. \quad (7)$$

By transforming all the text into individual spatial vectors and then constructing them as a rectangle, a vector space model depicting all the text can be generated. Each row of this matrix represents

a vector representation of the text, and each column represents a feature or dimension. With such a vector space model, text data can be effectively represented and compared in a high-dimensional space, thus enabling structured organization of text, similarity calculation and classification analysis.

2.4.2. K-means clustering. Clustering is a process of pooling, classifying, and organizing data members that are similar in some ways, and is unsupervised learning. After converting documents into feature vectors, it is able to measure the differences between documents by calculating the approximation between the vectors, at which point the problem becomes how to calculate the degree of similarity between two spatial vectors. In the experiment, the cosine similarity is used as an evaluation index to measure the similarity between texts. The degree of similarity between two texts is quantified by calculating the cosine values of their corresponding feature vectors. Assuming that the two texts can be converted into multidimensional feature vectors A and B , the cosine value of the angle θ between A and B can be calculated using Eq. (8):

$$\cos \theta = \frac{\sum_{i=1}^n (A_i \times B_i)}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} = \frac{A \cdot B}{|A| \times |B|}. \quad (8)$$

From Eq. (8), it can be concluded that when the cosine value is close to 1, the closer the two vectors are, it indicates that the similarity between the two texts being compared is higher, and they can be clustered into one group.

3. Analysis of the trend of stylistic evolution of chinese modern and contemporary literary works

3.1. Thematic classification and evolution

3.1.1. Thematic classification and evolution of modern and contemporary chinese literary works. The MONK project applies the LDA model for theme extraction and the GBDT model for classification to categorize the theme identification of modern and contemporary Chinese literature into three stages. From the awakening of Enlightenmentism in the early 20th century (1949-1978), to the diversified presentation during the revolutionary period (1978-1989), to the diversified development after the reform and opening up (1990s to the present), the themes of modern and contemporary Chinese literature have been continuously expanded and deepened.

1) Results of theme identification of literary works in the first stage. In this stage, the LDA-GBDT model identifies three key themes of modern and contemporary Chinese literary works, and the theme identification results are shown in Table 1. The keywords under each identified theme, as well as the distribution of the proportion of each theme in the whole corpus and the final identification results are shown in Table 1. The three identified themes are Hard Times and People's Life (25.27%), Social Phenomena and Humanity Exploration (34.68%), and History and Personal Destiny (40.05%). For example, Lu Xun's *Diary of a Madman* and *The True Story of Ah Q* mostly focus on the people's difficult lives and struggles in wars, political movements and economic difficulties. Mo Yan's *The Red Sorghum Family* reflects the changes in the family, the countryside and traditional culture.

2) Theme identification results of literary works in the second stage. The results of theme identification of the second stage of Chinese modern and contemporary literary works are shown in Table

2. Literary works in this stage begin to pay attention to the new life and new atmosphere after the reform and opening up, such as the novel creation of Zhao Shuli and others. There are also some works reviewing and thinking about the revolutionary history, such as the novel creation of Liang Bin and others. In addition the theme of industrial-themed novels began to expand, reflecting the changes and developments in many aspects of society. A total of three themes were identified at this stage: description of new life (30.77%), remembrance of the revolution (44.08%), and expansion in other fields (25.15%).

Table 1. The first phase of the topic recognition results

| Theme | Key words | Thematic ratio | Document number | Identification result |
|--------|---|----------------|-----------------|--|
| Topic1 | People, War, Political movement, Economic hardship, Hard life, Strive | 25.27% | 1284 | The hard times and the people's life |
| Topic2 | Social change, Moral dilemma, Values conflict, Family change, Rural change, Traditional cultural change | 34.68% | 1762 | Social phenomena and human exploration |
| Topic3 | Family, Fates, Country, History, Time change | 40.05% | 2035 | History and personal destiny |

Table 2. The second phase of the topic recognition results

| Theme | Key words | Thematic ratio | Document number | Identification result |
|--------|--|----------------|-----------------|---------------------------------|
| Topic1 | Reform and opening up, Revitalization, New weather | 30.77% | 1863 | The description of the new life |
| Topic2 | Revolution, Left wing, Youth world, Civic life | 44.08% | 2669 | The pursuit of the revolution |
| Topic3 | Industry, Proletariat, Innovate, Life | 25.15% | 1523 | Other areas of expansion |

3) Results of theme identification of literary works in the third stage. The results of theme identification of literary works in the third stage are shown in Table 3. Three key themes were identified through the LDA-GBDT model, namely, repression and recovery (36.66%), transformation in change (27.27%), and pluralism (36.07%). For example, Jiang Zilong's novels reflect on the Cultural Revolution and reform. Literary works at this stage began to experiment with different forms and styles, such as growth and confusion, exploration of the body, and grotesque reality, which became new literary themes.

Table 3. The third phase of the topic recognition results

| Theme | Key words | Thematic ratio | Document number | Identification result |
|--------|--|----------------|-----------------|--------------------------|
| Topic1 | Reform and opening up, Revitalization, New weather | 36.66% | 2639 | Repression and recovery |
| Topic2 | Revolution, Left wing, Youth world, Civic life | 27.27% | 1963 | Transformation in change |
| Topic3 | Industry, Proletariat, Innovate, Life | 36.07% | 2596 | Diversity |

3.1.2. Comparative analysis of thematic coherence. Comparing the effectiveness of LDA, LDA-BERT-K-means, and LDA-GBDT models based on topic coherence evaluation metrics. In current text mining research, topic coherence is commonly used to measure the degree of consistency between words within a topic. In this study, topic coherence metrics are obtained by invoking the Coherencemodel model in the gensim natural language processing toolkit. Based on the thematic coherence evaluation index, the thematic coherence of the traditional LDA thematic model, the BERT combined with K-means clustering model, and the LDA-GBDT model in the corpus of modern and contemporary Chinese literature is calculated in three phases, and the comparison of thematic coherence of the three thematic modeling methods in different corpora is shown in Figure 1. The thematic coherence of the three thematic categorization methods in the MONK project is calculated in three phases. The coherence of each topic is 0.53, 0.529, and 0.55, respectively. The coherence of the topics of the other two models does not exceed 0.5.

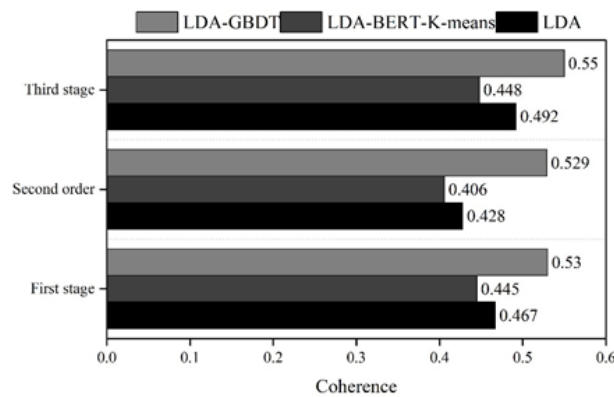


Fig. 1. Thematic coherence

3.2. Changes in emotional tendencies

3.2.1. Changes in the emotional tendencies of modern and contemporary Chinese literary works. The MONK project uses the SO-PMI algorithm to automatically construct an emotion lexicon for the textual domain of contemporary Chinese literature. The emotion information in the emotion dictionary is added to the word vector to get the emotion word vector.

The changes in the emotional tendency of modern and contemporary Chinese literary works can likewise be divided into three stages. Literary works in the period of Enlightenment awakening form a strong emotional impact through tragic emotions and tragic life, showing the people's hard encounters in the feudal society. In the revolutionary period, patriotic feelings are expressed more directly and fiercely in modern and contemporary literary works. The works after the reform and opening up emphasize values closer to human nature, express nature and show the personality of the works.

3.2.2. Sentiment word vector validity analysis. In order to validate the SO-PMI algorithm with the sentiment word vectors under the domain sentiment lexicon, we construct several validation models for experiments.

Model 1: Convolutional neural network model. The whole model has only one convolutional layer, using a combination of convolutional kernels (3,4,5) a total of 120 convolutional kernels work together. The pooling layer is connected after the convolutional layer and maximum pooling is used.

Model 2: Long Short Term Memory Network. The validation model uses a bi-directional long short term memory network with 50 LSTM units in each layer.

Model 3: GRU Neural Network. The validation model uses a bidirectional GRU neural network with 50 GRU units in each layer.

The accuracy evaluation of the three validation models is shown in Figure 2. Using the word vector training model obtained from skip-gram model training, the accuracy of the model obtained using ordinary word vector training is 88.59%, 86.32%, and 86.97%, respectively, as verified by convolutional neural network, long and short-term memory network, and GRU neural network. Taking this as a benchmark, the accuracy rates of the models obtained by training the emotion word vectors under the method of this paper are 90.26%, 90.41%, and 91.48%, respectively, which indicates that the method of automatically constructing the domain emotion dictionary using the SO-PMI algorithm, and adding the emotion information in the domain emotion dictionary to the word vectors is feasible and effective.

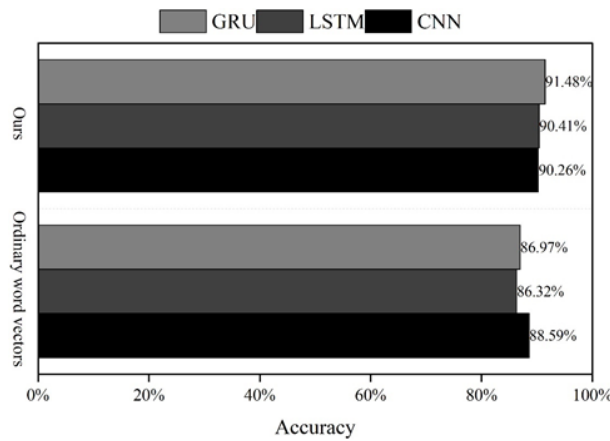


Fig. 2. The accuracy of the three verification models is evaluated

3.3. Analysis of style classification

The MONK project transforms a large amount of unstructured text data into highly structured text vectors by means of a text vector space model, and then analyzes the text vectors by clustering them using the K-means clustering method.

Words such as “revolution, war, hero, love, memory, nature, land, peasant, countryside, knowledge, education, thought” are frequently found in modern and contemporary Chinese literature.

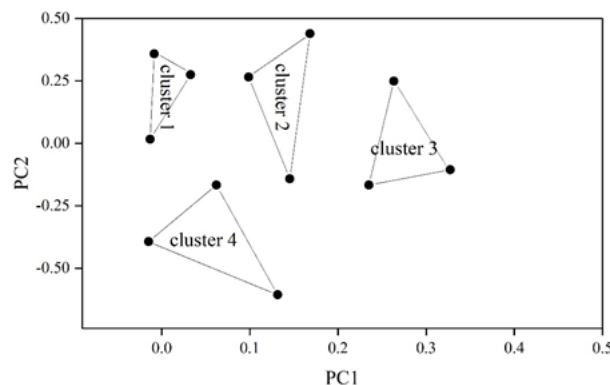


Fig. 3. K-means clustering results

The K-means clustering results are shown in Figure 3. The styles of contemporary Chinese literary works can be categorized into epic style (CLUSTER 1), lyrical style (CLUSTER 2), rural subject style (CLUSTER 3), and intellectual subject style (CLUSTER 4). The epic style is composed of "revolution, war, hero" and other related words, the lyrical style is composed of "love, memory, nature", "land, peasants, and countryside" constitute the rural theme style, and "knowledge, education, and thought" constitute the intellectual theme style.

4. Conclusion

The study uses text mining tools such as LDA theme model, sentiment lexicon and vector space model from MONK project to explore the theme, sentiment and stylistic changes of modern and contemporary Chinese literary works.

The coherence of each theme under LDA-GBDT is 0.53, 0.529, and 0.55, respectively, while the coherence of other theme classification models is below 0.5.

The SO-PMI algorithm is used to construct the domain sentiment dictionary, and the sentiment information in the domain sentiment dictionary is added to the word vectors, which are trained by convolutional neural network, long and short-term memory network, and GRU neural network, and the accuracies of the models obtained are 90.26%, 90.41%, and 91.48%, respectively, which are higher than the accuracies of the models obtained by ordinary word vector training.

Through K-means clustering, the styles of Chinese modern and contemporary literature can be divided into four styles: epic style, lyrical style, rural theme style and intellectual theme style.

References

- [1] D. Bamman, S. Popat, and S. Shen. An annotated dataset of literary entities. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2138–2144, 2019. <https://doi.org/10.18653/v1/N19-1220>.
- [2] E. M. Bender. Linguistic typology in natural language processing. *Linguistic Typology*, 20(3):645–660, 2016. <https://doi.org/10.1515/lingty-2016-0035>.
- [3] Q. Chen, R. Liu, Q. Jiang, and S. Xu. Exploring cross-cultural disparities in tourists' perceived images: a text mining and sentiment analysis study using lda and bert-bilstm models. *Data Technologies and Applications*, 58(4):669–690, 2024. <https://doi.org/10.1108/DTA-10-2023-0645>.
- [4] D. Gan, K. Numtong, H. Li, and S. Jiang. Exploring the application of the apriori algorithm in knowledge mining for linguistic data within chinese studies. *Eurasian Journal of Applied Linguistics*, 10(1):279–298, 2024. <http://dx.doi.org/10.32601/ejal.10124>.
- [5] X. Guo, S. Yu, S. Wang, and Y. Li. Study on the writing style of chinese vernacular novels based on svm. In *2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCI)*, pages 88–92. IEEE, 2017. <https://doi.org/10.1109/CIAPP.2017.8167186>.
- [6] E. Hyde and S. Wasserman. The contemporary. *Literature Compass*, 14(9):e12411, 2017. <https://doi.org/10.1111/lic3.12411>.
- [7] J. Kabbara and J. C. K. Cheung. Stylistic transfer in natural language generation systems using recurrent neural networks. In *Proceedings of the Workshop on Uphill Battles in Language Processing: Scaling Early Achievements to Robust Methods*, pages 43–47, 2016. <https://doi.org/10.18653/v1/W16-6010>.

- [8] D. Khurana, A. Koli, K. Khatter, and S. Singh. Natural language processing: state of the art, current trends and challenges. *Multimedia Tools and Applications*, 82(3):3713–3744, 2023. <https://doi.org/10.1007/s11042-022-13428-4>.
- [9] A. A. Lindsay. An overview of developments and trends in modern and contemporary chinese literature. *Emergence*, 62:77–81, 2019. <https://doi.org/10.7176/JLLL/62-08>.
- [10] B. S. McDougall. World literature, global culture and contemporary chinese literature in translation. *International Communication of Chinese Culture*, 1:47–64, 2014. <https://doi.org/10.1007/s40636-014-0005-7>.
- [11] I. Novakova and D. Siepman. *Phraseology and Style in Subgenres of the Novel: A Synthesis of Corpus and Literary Perspectives*. Springer Nature, 2019.
- [12] M. V. Redd and M. Hanumanthappa. Semantical and syntactical analysis of nlp. (*IJCSIT*) *International Journal of Computer Science and Information Technologies*, 5(3):3236–3238, 2014.
- [13] T. Shen, T. Lei, R. Barzilay, and T. Jaakkola. Style transfer from non-parallel text by cross-alignment. *Advances in Neural Information Processing Systems*, 30:6833–6844, 2017.
- [14] C. Sihe. A brief overview of chinese-language scholarship on modern chinese literature. *A Companion to Modern Chinese Literature*:465–482, 2015. <https://doi.org/10.1002/9781118451588.ch29>.
- [15] J. Worsham and J. Kalita. Genre identification and the compositional effect of genre in literature. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1963–1973, 2018.
- [16] A. M. Yang, J. H. Lin, Y. M. Zhou, and J. Chen. Research on building a chinese sentiment lexicon based on so-pmi. *Applied Mechanics and Materials*, 263:1688–1693, 2013. <https://doi.org/10.4028/www.scientific.net/AMM.263-266.1688>.
- [17] C. Yang, M. Sun, X. Yi, and W. Li. Stylistic chinese poetry generation via unsupervised style disentanglement. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3960–3969, 2018. <https://doi.org/10.18653/v1/D18-1430>.
- [18] Z. Yang and M. Dayong. Classicism 2.0: the vitality of classicist poetry online in contemporary china. *Frontiers of Literary Studies in China*, 12(3):526–557, 2018. <https://doi.org/10.3868/s010-007-018-0025-0>.
- [19] L. Yu. Research on automatic classification of writing styles based on knn algorithm. In *2023 IEEE 6th International Conference on Knowledge Innovation and Invention (ICKII)*, pages 555–558. IEEE, 2023. <https://doi.org/10.1109/ICKII58656.2023.10332718>.
- [20] T. Zhang, Y. Huang, H. Liao, and Y. Liang. A hybrid electric vehicle load classification and forecasting approach based on gbdt algorithm and temporal convolutional network. *Applied Energy*, 351:121768, 2023. <https://doi.org/10.1016/j.apenergy.2023.121768>.
- [21] Y. Zhu. Trend of contemporary chinese literature based on computerized multi-media integration. *Journal of Electrical Systems*, 20(9s):133–138, 2024.