



The National Culture Connotation and narrative Strategy of Chinese modern novels in the early 20th century

Zhengzhi Jiang^{1,*}

¹ School of Humanities, Shangluo University, Shangluo 726000, Shaanxi, China

SUMMARY: *Focusing on the issue of national cultural connotation and narrative strategy in Chinese modern novels in the early 20th century, this paper introduces digital humanities and text computing methods, and constructs an analysis framework composed of national cultural implication representation, narrative strategy structural identification, multi-dimensional feature correlation analysis and parameter adaptive adjustment. Based on the corpus of representative modern novels from 1918 to 1932, this paper jointly identified the cultural dimensions and narrative bearing modes such as national consciousness, ethical order, subject awakening, custom memory through text cleaning, hierarchical annotation, feature extraction and model training. The results show that the Accuracy of the proposed framework reaches 91.4%, the F1 value reaches 89.8%, and the RMSE is reduced to 0.146 under the condition of a unified corpus, which is better than the fixed threshold model, BiLSTM model and Text-GNN model. Ablation experiments and robustness tests further show that multidimensional association analysis and parameter adaptation mechanism play an important role in stable interpretation in complex literary contexts. The results show that computer-aided analysis can provide comparable and retractable structural evidence for the study of national culture of modern Chinese novels in the early 20th century, and also provide a new way for the integration of modern literature research and digital humanities methods.*

KEYWORDS: *Chinese modern novels in the early 20th century; Connotation of national culture; Narrative strategy; Text computing*

1 Introduction

At the beginning of the 20th century, Chinese modern novels were generated in the rapidly changing social and cultural context. The texts not only carried the complex tension of the emergence of national consciousness, the re-evaluation of traditional ethics and the formation of modern subjects, but also showed an obvious modern turn in the narrative perspective, character creation, time organization and symbolic way. For a long time, relevant studies have mostly started from the history of literature, the history of thought, the enlightenment discourse or the individual cases of writers, and deeply explained the representative works of Lu Xun, MAO Dun, Bing Xin, Lu Yin and other writers, which better revealed the interactive relationship between the writing of national culture and modern narrative forms. However, when the research object is extended to the corpus of early modern novels across authors, journals and styles, the traditional approach of relying on manual close reading and experience induction gradually shows its limitations: Researchers can keenly grasp the spiritual texture of key texts, but it is difficult to stably identify the distribution law of national

*15596396718@163.com

<https://doi.org/10.65102/is2026289>

cultural implications in a wider range, and it is also difficult to make repeatable and verified explanations for the structural relationship between different narrative strategies.

With the development of digital humanities and text computing methods, literary research has begun to acquire new analytical scales. The techniques of word segmentation, word vector modeling, topic recognition, narrative unit extraction, sentiment analysis and relationship network construction provide realistic conditions for the structured processing of literary texts. For Chinese modern novels in the early 20th century, the value of such methods does not lie in replacing traditional interpretation, but in providing a more detailed evidence chain for such questions as "how the connotation of national culture is narrated" and "how narrative strategies serve cultural expression". With the help of computer-aided analysis, researchers can observe the potential order inside the text from the aspects of high-frequency cultural symbols, character relationship patterns, narrative voice transfer, and spatio-temporal organization characteristics, and find the common structure and difference mechanism that are not easily perceived by individual reading.

However, there are still some shortcomings in the existing related research. On the one hand, some studies simply convert literary texts into quantifiable words, and emphasize word frequency statistics or theme clustering, which is easy to weaken the dynamics of the narrative level of the novel, especially the generation process of national cultural meaning in the plot advancement, perspective switching and rhetorical arrangement. On the other hand, although some studies introducing deep learning or complex models improve the recognition efficiency, they often regard literary texts as general corpora, ignoring language features such as the mixture of early vernacular and classical Chinese, the penetration of translation words, and the strong historical context, which leads to insufficient interpretation of the model and the disconnection between the analysis results and literary problems. How to establish an effective connection between the problem consciousness of literary research and the technical path of computational analysis is still a key issue that needs further response.

Based on this, this paper focuses on the national culture connotation and narrative strategy of Chinese modern novels in the early 20th century, and tries to construct a set of digital analysis framework for literary texts. In terms of research design, this paper integrates the identification of ethnic cultural implications, the structural extraction of narrative strategies, the correlation analysis of multi-dimensional features and the adaptive adjustment of parameters into the same analysis chain. Through the combination of corpus cleaning, manual annotation, text calculation and result readback, the relationship between "computable text structure" and "interpretable literary meaning" is promoted to form a mutual proof relationship. This paper is not concerned with the simple application of technology, but with the help of computer methods, we hope to re-examine how early modern Chinese novels complete the reconstruction and dissemination of national cultural imagination through narrative devices such as disease metaphor, family narrative, historical memory, female writing and national criticism in a larger scale of text.

The research objectives of this paper are mainly reflected in three levels: first, to establish a digital index system that can represent the national cultural implications and improve the recognition accuracy of relevant literary elements. Secondly, a computational analysis method of narrative strategy suitable for early modern novels is formed to enhance the systematicity and verifiability of structural judgment. Thirdly, a loop mechanism is established between quantitative results and qualitative explanations in order to reveal the internal relationship between national culture expression and the choice of narrative form more completely. With these goals in view, this paper further investigates the applicable boundaries and explanatory efficiency of text computing methods in literary studies, so as to provide an analytical approach with both literary sensitivity and technical support for the study of Chinese modern

novels in the early 20th century.

2 Related Research

Focusing on the national culture connotation and narrative strategy of Chinese modern novels in the early 20th century, the existing research has formed a relatively clear accumulation in the problem consciousness and method path. Based on the relevant achievements, it can be roughly summarized into three research directions: the interpretation of literary history centered on the imagination and cultural identity of the nation and state, the research of narrative forms centered on the works of writers, and the research of text computing and structure recognition under the background of digital humanities. The three types of research promote related issues from different levels of ideological history, text close reading and computational analysis, but there is still a certain separation between each other.

In terms of the research on the connotation of national culture, the existing results mostly start from the generation of national consciousness, the construction of national narrative and the reconstruction of cultural identity in modern China, and reveal the time spirit of literature writing in the early 20th century. Veg[1] rediscusses political writing in early modern novels from the perspective of contrast between European and Chinese modernity, emphasizing the complex interaction between literary texts and modern democratic concepts. Cohen and Meissner combined national memory, cultural identity and national narrative, and pointed out that "nation" in 20th century Chinese literature was not a static symbol, but a meaning structure that was constantly shaped in historical crisis and cultural reevaluation [2, 3]. Lee and Haapanen further linked emotional politics and nationalist consciousness in the context of the May 4th Movement, suggesting that folk narration, reflection on national character and national cultural imagination in early modern novels were closely related [4, 5]. This kind of research has laid a solid foundation for understanding the national cultural implications in novels, but it mostly relies on the analysis of ideological history and the interpretation of typical texts, and the discussion of the distribution characteristics, co-occurrence relationships and structural patterns of cultural elements in large-scale texts is relatively insufficient.

In the study of narrative strategy, the academic circle mainly focuses on the novel practice of Lu Xun and his contemporary writers. Starting from the heterogeneous form of realism, the expression of modernism and the text generation mechanism of *Diary of a Madman*, Gu, Cai and Ma revealed the formal innovation of early Chinese modern novels in narrative perspective, rhetoric organization, symbolic structure and subject construction [6-9]. Chou pays attention to the formation process of Lu Xun's readership, indicating that narrative strategy is not a simple text technique, but interacts with communication environment, acceptance mode and cultural orientation [10]. McDougall and Wang took female writing and gender rhetoric into consideration, expanding the observation dimension of narrative research in early modern novels [11, 12]. The above studies are detailed and in-depth, and have high explanatory power for the form discrimination and analysis of specific works. However, most of the methods are based on single close reading or case comparison, lacking structured induction across authors and texts, and it is difficult to verify the correlation degree between different narrative strategies under uniform standards.

With the development of digital humanities and natural language processing technology, literature research has begun to introduce word segmentation tagging, topic modeling, word vector representation, sentiment computing, character relationship network and text classification methods. Although not all such studies directly point to Chinese modern novels in the early 20th century, they provide a technical path for related problems. Hughes promoted

the interpretation methods of nationalist texts at the methodological level. Although his focus was theoretical interpretation, his awareness of text structure analysis had been revealed [13]. By analyzing the narrative strategies and cultural implications of Chinese sci-fi films, Hu et al. showed that the relationship between narrative elements and cultural identity can be established through structured variables [14]. Ziling et al. sorted out the translation strategies of detective novels from the double dimensions of history and method, which also showed that the narrative mode, cultural coding and communication mechanism in literary texts have the possibility of being modeled and compared [15]. Compared with traditional studies, text computing methods can process larger corpora, identify high-frequency cultural symbols, theme clusters, emotional fluctuations and narrative mode transfer, and provide a new evidence chain for the study of early modern novels. However, there are obvious limitations in the existing related attempts. One kind of method focuses on word frequency and theme statistics, which is easy to weaken the contextual differences and narrative dynamics in literary texts. Although the other kind of methods can improve the accuracy of classification or recognition, they often regard literary texts as general corpora, and ignore the characteristics of language mixing, rhetorical ambiguity and strong historical context during the transition period of Chinese vernacular, which leads to a disconnection between calculation results and literary interpretation. In order to more intuitively present the focus and shortcomings of the existing research, this paper selects a number of representative results to summarize, as shown in Table 1.

Table 1: Representative results of related studies and comparison of their methods

Author	Research Focus	Main Method	Implications for This Study	Main Limitations
Veg [1]	Early modern novels and democratic modernity	Literary-historical interpretation, comparative research	Helps examine ethnic cultural connotations within the transformation of modernity	Lacks computable analysis of textual structure
Cohen [2]	National memory and national narrative	Historical narrative analysis	Helps identify the linkage of “nation–history–culture”	Difficult to reveal the internal narrative organization of texts
Gu [6]	Lu Xun and the formal transformation of modern literature	Author-centered study, formal analysis	Provides key categories for narrative strategy identification	Focuses on a limited research object and lacks cross-text comparison
Cai [8], Ma [9]	Form and meaning generation in <i>Diary of a Madman</i>	Close reading, textual interpretation	Helps extract indicators of narrative perspective and symbolic structure	Difficult to extend to large-scale corpora
McDougall [11], Wang [12]	Women’s writing and rhetorical expression	Gender criticism, narrative analysis	Expands the dimensions for identifying cultural connotations	Lacks a unified structured analytical framework
Hu et al. [14]	Relationship between cultural implication and narrative strategy	Structured analysis, typological induction	Shows that cultural expression can be analyzed jointly with narrative variables	The research object is not early modern novels
Ziling et al. [15]	Strategic issues in literary dissemination	Historical review, method comparison	Provides ideas for integrating text computation with literary history	Involves relatively little analysis of internal narrative mechanisms in novels

In general, the existing research has made important progress in the interpretation of ideas, author research and form analysis, and also provides method inspiration for the digital processing of literary texts. However, there are still several gaps: first, the research on the connotation of national culture and the research on narrative strategy have not formed a stable integration framework, and the interactive relationship between the two has not been systematically explained. Second, although text computing methods have the advantage of large-scale processing, they have not fully adapted to the language characteristics and literary context of Chinese modern novels in the early 20th century. Thirdly, there is still a lack of effective connection mechanism between quantitative identification and qualitative interpretation. Based on this research status, this paper attempts to construct a digital analysis framework for early modern novels, which integrates the identification of national cultural implications, the extraction of narrative strategy structure, the correlation analysis of multi-dimensional features and the adaptive adjustment of parameters into the same research chain, in order to establish a closer correspondence between literary problem awareness and computational analysis methods.

3 Digital analysis framework of national cultural implications and narrative strategies in early 20th century Chinese modern novels

The digital analysis framework constructed in this paper does not simply compress literary texts into a number of quantifiable words, but tries to establish a readback and interpretable mapping chain between "text meaning, narrative structure and computational representation". Focusing on the research object of Chinese modern novels in the early 20th century, the framework sets up four interrelated processing links: the digital representation of national cultural implications, the structural identification of narrative strategies, the correlation analysis of multi-dimensional features, and the adaptive adjustment of analysis parameters. Its purpose is to transform the cultural information scattered in the names of characters, historical memories, national discourses, customs descriptions and value judgments in early modern novels into structural features that can be recognized and compared by computers, while preserving the context sensitivity required for literary interpretation.

3.1 Digital representation mechanism of national cultural implication in early 20th century Chinese modern novels

The national cultural connotation of Chinese modern novels in the early 20th century does not appear in a direct way, but is embedded in character language, narrative commentary, symbolic imagery and plot advancement, which is implied, complex and context-dependent. If we only rely on keyword search, it is easy to mistake the frequency of words such as "nation", "national", "family" and "ethics" as cultural intensity, and ignore their significance differences in different narrative situations. Based on this, this paper divides the digital representation of national cultural implication into four stages: text segmentation, feature coding, semantic aggregation and intensity determination, forming a progressive modeling path from low-level words to high-level cultural themes, as shown in Figure 1. Let the novel corpus after preprocessing be

$$D = \{d_1, d_2, \dots, d_i\} \quad (1)$$

Here, d_i denotes the i th novel text. The text is further segmented into sentence segment units x_{ij} , and the character and word embedding features, cultural dictionary features, rhetorical marker features and context location features are extracted for each unit to form the initial representation vector:

$$x_{ij} = [e_{ij}; c_{ij}; r_{ij}; p_{ij}] \quad (2)$$

where, e_{ij} is the semantic vector obtained based on the pre-trained Chinese language model, c_{ij} represents the national culture dictionary matching result, r_{ij} represents the rhetorical marks such as metaphor, irony, and call, and p_{ij} represents the position of the paragraph in the full text and the correlation information with characters and scenes. Considering the European-style sentence pattern, the mixture of Chinese and white, and the coexistence of old and new words in early vernacular novels, this paper adds a time word list and a writer special list in addition to the general language model, and corrects and encodes high-frequency cultural semantic clusters such as "national character", "family order", "memory of national loss" and "female awakening", so as to reduce the misjudgment of the modern Chinese model on historical texts. On this basis, linear mapping and nonlinear activation function are used to generate sentence-level cultural representation:

$$h_{ij} = \sigma(Wx_{ij} + b) \quad (3)$$

Here, W is the parameter matrix, b is the bias term, and σ is the activation function. In order to avoid the interference of local high-frequency words on the overall judgment, this paper introduces the attention aggregation mechanism to assign differentiated weights to different sentence segments:

$$\alpha_{ij} = \frac{\exp(q^T h_{ij})}{\sum_{j=1}^m \exp(q^T h_{ij})} \quad (4)$$

$$H_i = \sum_{j=1}^m \alpha_{ij} h_{ij} \quad (5)$$

Here, α_{ij} represents the contribution weight of the J TH sentence segment in the i th text, and H_i is the representation vector of the national cultural implication of the whole novel. The vector can not only reflect the concentration area of the cultural theme in the full text, but also retain the dynamic relationship between it and the narrative advancement. Finally, combined with the manually labeled samples and the category discriminator, the intensity scores of the text on different cultural dimensions were calculated, and the four-dimensional cultural feature matrix of "national consciousness - historical memory - ethical order - subject awakening" was formed. After this process, the connotation of national culture is no longer a general description, but is transcribed into comparable, traceable and traceable computational objects, which provides a relatively stable input basis for the structural identification of subsequent narrative strategies.

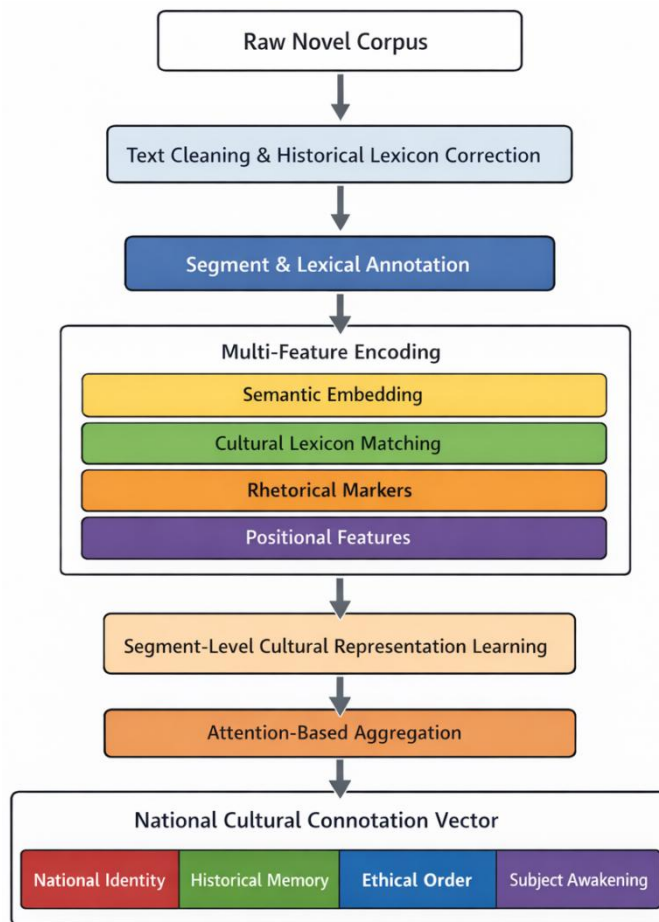


Figure 1: Schematic diagram of the digital representation mechanism of national cultural implications in Chinese modern novels in the early 20th century

3.2 Narrative strategy structural identification method based on text computing

After completing the digital representation of national cultural implications, the key of the research turns to the identification of the structure at the narrative level. At the beginning of the 20th century, Chinese modern novels showed remarkable transformation characteristics in narrative style, such as first-person narration, letter diary embedding, psychological monologue, symbolic irony and critical interdiction, which often participated in the generation of text meaning. Traditional close reading can well reveal the narrative texture of a single work, but when the research object is extended to the collection of stories across authors and journals, the comparability and stable identification between narrative strategies need to be supported by text computing methods. Based on this, this paper divides narrative strategy recognition into four steps: narrative unit segmentation, narrative element labeling, relationship modeling and strategy judgment, so as to realize the progressive extraction from text surface form to narrative structure type. Let a single novel be segmented to obtain a sequence of narrative units

$$S = \{s_1, s_2, \dots, s_t\} \quad (6)$$

Here, s_t denotes the t -th narrative unit, which can correspond to sentence groups, paragraphs, or scene fragments. For each unit, the system synchronously extracts features

such as narrative person, tense marker, mental verb, dialogue density, evaluation adverb and transition connection component to form a narrative representation vector:

$$z_t = g([n_t; v_t; d_t; m_t; q_t]) \quad (7)$$

where, n_t represents narrative perspective features, v_t represents verb tense and action advance information, d_t represents dialogue proportion, m_t represents psychological description intensity, q_t represents critical discourse markers, and $g(\cdot)$ is a nonlinear mapping function. The representation can reflect the organization of a narrative unit more centrally, so as to avoid the deviation caused by judging narrative strategy only by a single word.

On this basis, this paper further establishes the transition relationship matrix between narrative units to describe the dynamic trajectory of strategy transformation. If there is continuity between narrative unit s_i and s_j in perspective, tone or description focus, the edge weight a_{ij} is assigned to them to form a narrative structure diagram. Its transition probability can be expressed as follows.

$$p_{ij} = \frac{\exp(a_{ij})}{\sum_k \exp(a_{ik})} \quad (8)$$

Through the comprehensive calculation of transition probability, local feature strength and context dependence, the model can output the narrative strategy category to which a text segment belongs, such as limited knowledge narrative, polytone narrative, symbolic embedded narrative or argument interventional narrative. In this way, narrative strategies are no longer just empirical literary judgments, but transcribed as structural objects with hierarchical relations and path characteristics. Its significance does not lie in substituting calculation for interpretation, but in providing a clearer structural basis for analyzing how the connotation of national culture is organized, strengthened and transmitted by means of specific narrative methods.

3.3 Multi-dimensional feature extraction and correlation analysis mechanism for literary texts

After completing the representation of national cultural implication and the identification of narrative strategy, the focus of research is no longer the list of single features, but to investigate how different levels of information in the same text traction each other and jointly generate meaning. The cultural expression of Chinese modern novels in the early 20th century is often not borne by a certain kind of words or a single narrative form, but by the superposition of national vocabulary, character relationship, emotional trend, symbolic image and narrative mode. Based on this characteristic, this paper extends the feature extraction of literary texts into four dimensions of cultural semantics, narrative structure, character relationship and emotional trajectory, and conducts fusion modeling in a unified computing space.

Let the representation of the i th novel in the above four dimensions be c_i , n_i , p_i and t_i respectively, then its multidimensional fusion feature can be expressed as

$$m_i = \psi(W_c c_i + W_n n_i + W_p p_i + W_t t_i + b) \quad (9)$$

where W_c , W_n , W_p , W_t are the mapping matrices of each dimension feature, b is the bias term, and $\psi(\cdot)$ is the nonlinear transformation function. In this way, the originally scattered

cultural symbols, narrative actions and relationship information in the text are compressed into the same representation vector, which is beneficial for the computer to maintain a consistent scale when comparing across texts. Different from general text classification, the fusion here does not pursue simple dimensionality reduction, but emphasizes the preservation of structural differences between literary elements, so that complex connections such as "national memory - character fate" and "ethical order - narrative intervention" can be further recognized.

Based on the fusion representation, this paper introduces the cross-feature attention mechanism to calculate the correlation strength between different dimensions. If the set of local feature units of a single text is denoted as $U_i = \{u_1, u_2, \dots, u_r\}$, then the association weight between any two feature units is defined as

$$\beta_{kl} = \frac{\exp((u_k^T Q u_l) / \sqrt{d})}{\sum_l \exp((u_k^T Q u_l) / \sqrt{d})} \quad (10)$$

Here, Q is the learnable parameter matrix and d is the feature dimension. This weight can be used to measure the coupling degree between a certain cultural semantic component and a specific narrative strategy, such as the resonance relationship between "national character criticism" and "ironic narrative", or the synergy relationship between "family ethical conflict" and "knowledge limited perspective". Compared with the traditional co-word statistics, this method can better reflect the dynamic association within the text, rather than staying at the surface juxtaposition.

In order to improve the interpretability of the analysis results, this paper further calculates the joint association score between cultural dimension and narrative strategy at the corpus level:

$$R_{gs} = \frac{1}{N} \sum_{i=1}^N y_{ig} \hat{z}_{is} + \lambda MI(g, s) \quad (11)$$

Here, y_{ig} represents the intensity value of text i in the G -type national culture dimension, \hat{z}_{is} represents the probability that it is judged as the S -type narrative strategy, $MI(g, s)$ is the mutual information between them, and λ is the adjustment coefficient. The R_{gs} obtained in this way can not only reveal which narrative mode a certain kind of cultural implication is more inclined to unfold, but also provide a more clear positioning basis for subsequent qualitative read-back. Through this mechanism, the multi-dimensional information in literary texts is organized into a computable, comparable and traceable association network, so as to provide a more solid analysis fulcrum for the deep correspondence between national culture connotation and narrative strategy.

3.4 Adaptive adjustment method of analysis parameters and evaluation indexes

In the process of digital analysis of Chinese modern novels in the early 20th century, parameter setting is not an attribute link, but directly related to the sensitivity of identification of national cultural implications and the stability of narrative strategy judgment. The early modern novels have obvious transitional features in language form, with the interlaced coexistence of vernacular, shallow classical Chinese, translation expression and newspaper style. The texts of different writers differ greatly in sentence length distribution, psychological

description density, dialogue proportion and rhetorical salience. If the fixed window length, unified threshold and constant evaluation weight are used, the model is easy to amplify local noise in short texts, and may also weaken key narrative signals in works with complex structures. Based on this, this paper introduces an adaptive adjustment mechanism oriented to the statistical characteristics of the corpus, so that the analysis parameters can be dynamically modified with the text characteristics.

In terms of window parameter adjustment, the average sentence length, paragraph dispersion and narrative turn density are integrated to determine the optimal analysis span. Let the average sentence length of a text be l_i , the variance of paragraphs be v_i , and the frequency of turning marks be r_i , then its window length is defined as

$$w_i = \lfloor w_0 + \alpha \ln(1 + l_i) + \beta v_i - \gamma r_i \rfloor \quad (12)$$

Here, w_0 is the base window value, and α , β , γ are the adjustment coefficients. This setting means that texts with long sentences and loose structure can moderately expand the scope of analysis, while texts with dense narrative turns shorten the window to avoid excessive merging of different situation units.

In the stage of feature discrimination, the recognition threshold of national cultural implication and narrative strategy should not be fixed. In this paper, the classification threshold is dynamically updated according to the information entropy and word sparsity of the text. Let the information entropy of text i be H_i and the lexical sparsity coefficient be s_i , then the discrimination threshold can be expressed as

$$\tau_i = \tau_0 + \lambda \cdot \frac{H_i}{1 + s_i} \quad (13)$$

Here, τ_0 is the initial threshold and λ is the scaling factor. The more complex the text semantic distribution is, the more cautious the threshold is. When the vocabulary is more concentrated, the threshold is moderately reduced to improve the recognition ability of implicit cultural markers. After this processing, the model can maintain a more balanced interpretation scale for Lu Xun's high-density metaphorical texts and Bingxin's lyrical narrative texts.

The evaluation index is also set in an adaptive way. Considering that this paper focuses not only on recognition accuracy, but also on structural consistency and cross-text stability in literary interpretation, the classification performance, structure preservation and manual annotation consistency are included in the comprehensive evaluation. Its total scoring function is defined as

$$Q_i = \omega_1 F1_i + \omega_2 C_i + \omega_3 K_i, \omega_1 + \omega_2 + \omega_3 = 1 \quad (14)$$

Here, $F1_i$ represents the classification effect, C_i represents the structural consistency score and K_i represents the degree of agreement with the manual annotation results. Each weight is automatically adjusted according to the fluctuation of the validation set, so that the evaluation results are neither overly biased towards the algorithm index, nor deviated from the requirements of literature research for interpretation reliability. Through this mechanism, the model obtains a more flexible adaptation ability between different text types, and also provides a more stable parameter basis for the subsequent analysis of results.

4 Analysis of results

4.1 Corpus Construction and distribution Statistics of Early 20th Century Chinese Modern Novels

In order to test the applicability of the digital analysis framework in the identification of national cultural connotations and the extraction of narrative strategies, this study established a thematic corpus around the early 20th century Chinese modern novels. The corpus sources mainly include public literature databases, photocopied texts of modern periodicals, electronic versions of writers' works and research text materials with manual collation. The time range is controlled between 1918 and 1932, so as to ensure that the sample basically corresponds to the development stages of modern novels since the "May 4th". In order to reduce the interference caused by inter-temporal language differences, the corpus is mainly composed of short stories and novelas, including the works of Lu Xun, Bing Xin, Lu Yin, Ye Shengtao, Yu Dafu, MAO Dun and other representative writers. At the same time, the non-classical texts with great influence in the same period are moderately included to improve the balance of corpus distribution.

In the process of corpus processing, this paper adopts the technical flow of "original collection, unified coding, allogeneic merging, sentence segmentation, metadata annotation". All texts were uniformly converted to UTF-8 encoding, page numbers, comment marks and repeated punctuation were removed, and abnormal words, old punctuation and individual incomplete characters in early vernacular were normalized. Then, the sentence segmentation is completed by combining word segmentation tools with manual review, and each text is annotated with information such as author, publication year, genre form, length and source category. In order to ensure the reliability of the experiment, the corpus is divided into training set, validation set and test set according to the discourse level, and the proportion is set to 70%, 15% and 15% to avoid information leakage caused by different segments of the same work crossing into different subsets.

The final corpus contains 368 novel texts, with a total number of about 1.864,000 words, about 79,200 sentence units, and about 34,600 normalized word items. In terms of distribution characteristics, the proportion of enlightenment criticism and life sadness texts is relatively high, and the sample number of rural writing texts and urban experience texts is relatively close, which can better cover the main genealogy of national culture expression and narrative mode in early 20th century Chinese modern novels. Different types of texts show obvious differences in average sentence length, dialogue density and proportion of psychological description, which also provides basic conditions for subsequent adaptive parameter adjustment and cross-text comparison. The overall composition of the corpus is shown in Table 2.

Table 2: Corpus composition and distribution statistics of early 20th century Chinese modern novels

Category	Number of Texts / Pieces	Word Count / 10,000 Chinese Characters	Number of Sentences and Paragraph Segments	Average Sentence Length / Characters	Main Textual Features
Enlightenment-Critique	96	48.7	20,540	23.6	Frequent argumentative intervention and a pronounced ironic tone
Life-Sentiment	84	39.2	17,630	22.1	Strong lyricism and dense psychological description
Rural Writing	72	36.5	15,180	24.8	Frequent depiction of customs and concentrated regional vocabulary
Urban Experience	61	31.4	13,260	21.7	Rapid scene shifts and a relatively high proportion of dialogue
Female Narrative	55	30.6	12,610	22.9	More prominent first-person narration and inner monologue
Total	368	186.4	79,220	23.4	Good representativeness of the period and clear structural differences

On the whole, the corpus not only retains the historical complexity of Chinese modern novels in the early 20th century in terms of language morphology and narrative style, but also has a standardized foundation suitable for computational analysis, which can provide more stable data support for the identification of national culture connotation, narrative strategy modeling and result verification.

4.2 Text preprocessing and annotation specification of national cultural implications and narrative features

After the construction of the corpus, in order to ensure the stability of the identification of national cultural implications and the calculation and analysis of narrative strategies, this paper uniformly preprocesses the texts and establishes a hierarchical annotation standard. The basic process of preprocessing stage is "coding unification, text cleaning, segmentation and normalization, format standardization". All texts were converted to UTF-8 encoding, and page numbers, footnotes, parentheses, repeated punctuation and invalid Spaces were removed. Different characters, old punctuation and some OCR misrecognized characters common in early vernacular Chinese were corrected. Considering that the novels in the early 20th century have mixed text and white text, Europeanized sentence patterns and traces of newspaper transcription, this paper does not overmodernize the original words, but retains their basic historical appearance, and only does the normative mapping at the level of calculation needs,

so as to avoid the literary language sense being eliminated in the preprocessing.

In the annotation stage, this paper adopts a three-layer annotation system of "cultural semantics-narrative forms-auxiliary structure". The cultural semantic layer mainly identifies the core dimensions of national consciousness, national memory, ethical order, custom writing, and subject awakening. The narrative form layer focuses on the features of narrative person, perspective transformation, psychological description, dialogue embedding, symbolic irony and argumentation intervention. The auxiliary structure layer supplements the basic information such as people, places, events, time clues and relationship chains. In order to improve the processing efficiency, the initial tagging is completed automatically by word segmentation, part-of-speech tagging, dependency analysis and semantic matching model, and then the ambiguous segments are corrected by manual review. Especially, the context discrimination is carried out for high-frequency and polysemy cultural words such as "national", "moral", "new and old", "home" and "country".

In order to ensure the consistency of annotation results, this paper formulates a unified judgment standard and example library, and elaborates the cross-sentence continuation, metaphor expression and compound narrative units. After multiple rounds of verification, the text preprocessing results are consistent with the feature labels in format, boundary and semantic definition. Among them, the agreement rates of cultural semantic layer and narrative form layer are X% and X% respectively, and the proportion of manual review and correction is X%. This provides a more reliable data basis for subsequent narrative structure identification, association analysis and model evaluation.

4.3 Calculation analysis effect and result evaluation

Under the unified corpus, unified preprocessing and unified annotation norms, this paper systematically evaluates the constructed framework of national cultural connotation recognition and narrative strategy analysis. The experimental platform used Python 3.10 and PyTorch 2.1, and the core parameters were set as batch size 32, initial learning rate 1×10^{-4} , and training rounds 30. In order to ensure the fairness of comparison, the comparison models are all run under the same corpus segmentation, the same feature input and the same training rounds. The baseline methods include rule-based fixed threshold model, BiLSTM classification model and standard Text-GNN model. Considering that this paper involves not only cultural dimension identification, but also narrative strategy discrimination and intensity prediction, three evaluation indicators are selected: accuracy, macro average F1 score and root mean square error.

Accuracy and F1 score are mainly used to measure the recognition effect of national culture labels and narrative strategy categories, which are defined as, respectively

$$\text{Acc} = \frac{1}{N} \sum_{i=1}^N 1(\hat{y}_i = y_i) \quad (15)$$

$$\text{F1} = \frac{2\text{PR}}{\text{P} + \text{R}} \quad (16)$$

where P and R are precision and recall, respectively, and N is the total number of samples. For the cultural intensity score prediction task, the root mean square error was used to evaluate:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (17)$$

The lower the index is, the closer the model is to the manual annotation result in describing the strength change of national cultural implication in the text.

The experimental results show that the proposed method has better comprehensive advantages on the three indicators. As shown in Table 3, although the fixed threshold model has low computational overhead, it has limited discrimination ability in complex contexts, especially insufficient recognition of metaphorical cultural expressions and compound narrative fragments. The BiLSTM model is superior to the rule-based method in sequence modeling, and can better capture the local dependencies within the sentence segment, but it is still insufficient in the processing of cross-segment narrative transfer and character relationship chain. Text-gnn model has improved the utilization of structural information, but when the Text is mixed with white and the narrative level is interlaced, the results are still restricted by fixed parameter Settings. In contrast, the accuracy and F1 value of the method proposed in this paper reach 91.4% and 89.8%, respectively, and the RMSE is reduced to 0.146, indicating that the framework can not only identify the connotation of national culture more stably, but also better grasp the correspondence between narrative strategy and cultural expression.

Table 3: Comparison of computational analysis effects of different methods

Method	Accuracy / %	F1 / %	RMSE	Training Time per Epoch / s
Fixed-Threshold Model	84.9	82.7	0.214	8.3
BiLSTM Model	87.3	85.6	0.187	10.9
Text-GNN Model	89.1	87.4	0.171	12.4
Proposed Method	91.4	89.8	0.146	10.1

According to the comprehensive results shown in Table 3, the proposed method is superior to the comparison models in recognition accuracy and error control, indicating that the adaptive parameter adjustment mechanism improves the adaptation ability of the model to complex literary texts to a certain extent. In other words, the framework does not simply rely on more complex model stacking to gain advantages, but establishes a more suitable adaptation relationship between the characteristics of literary text and the computing mechanism. This result also provides a more reliable experimental basis for subsequent ablation experiments and robustness tests.

4.4 Ablation experiment and robustness test of the analytical framework

In order to further test the internal validity and external adaptability of the analysis framework, we conduct ablation experiments and robustness tests under unified corpus, unified annotation and unified training Settings. The ablation experiment is used to investigate the independent contribution of each sub-module to the overall effect, and the robustness test is to observe the performance difference of the model under the conditions of sample reduction, text noise perturbation and cross-author transfer. The relevant results are shown in Figure 2 and Figure 3.

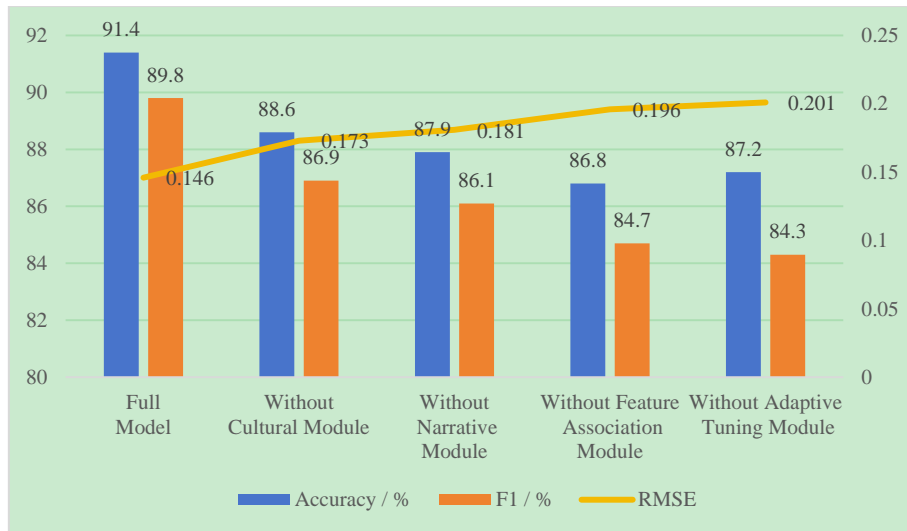


Figure 2: Results of ablation experiments

In the ablation experiment, the national cultural implication representation module, narrative strategy structural recognition module, multi-dimensional feature association analysis module and parameter adaptive adjustment module were removed in turn, and the other Settings were kept unchanged. The results show that the full model consistently maintains the best performance. Among them, after the feature association analysis module and the parameter adaptive adjustment module are removed, the F1 value decreases most obviously, and the RMSE increases greatly, which indicates that the two modules play a key role in the stable interpretation of complex literary context. The former directly affects the joint modeling among cultural semantics, narrative forms and character relationships, and the latter determines whether the window length, threshold and weight allocation can be dynamically modified with text differences. In contrast, once the national cultural implication representation module and narrative strategy recognition module are missing, the model still retains part of the classification ability, but the overall explanatory power is significantly weakened, indicating that the effectiveness of this study does not come from a single module, but is built on the linkage of multiple links.

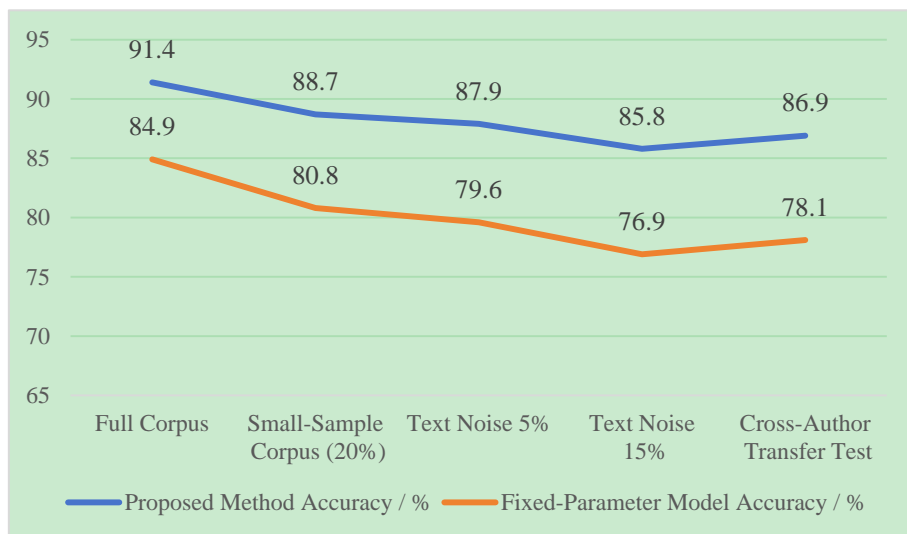


Figure 3: Results of the robustness test

The robustness test sets up four scenarios: complete corpus, small sample corpus, 5% text noise and 15% text noise. At the same time, the cross-author transfer test is added to investigate the adaptation of the model to unfamiliar style texts. Experiments show that the proposed method can still maintain high accuracy when the corpus is compressed to 20% of the original scale, and the performance degradation is significantly smaller than that of the fixed-parameter baseline model. Under the condition of noise disturbance, although the index decreases with the increase of mis-recognized characters, the overall fluctuation is within a controllable range. In the cross-author transfer scenario, the model still has a strong ability to adapt to different narrative rhythms, lexical habits and rhetorical preferences. This shows that the adaptive adjustment mechanism and multi-dimensional correlation modeling adopted in this paper are not only effective for the training set, but can maintain a relatively stable level of analysis under the condition of changes in text sources and quality fluctuations.

4.5 Qualitative interpretation results and cross-text structure validation

In order to further test whether the model output is interpretable, this paper adds qualitative interpretation and cross-text structure verification in addition to quantitative indicators. Five representative text samples were selected from the test set, corresponding to five types of enlightenment criticism, family conflict, female narrative, rural writing and urban experience, and the model recognition results were compared with the manual interpretation results. Among them, the consistency rate of the five types of samples is between 88.9% and 93.6%, and the structural similarity is between 0.83 and 0.89. The results show that the model has high consistency in cultural dimension identification, narrative strategy judgment and correlation strength measurement. The mean value of the culture-narrative correlation strength of the five types of samples is 0.882, the mean value of the consistency rate with manual interpretation is 90.94%, the mean value of the cross-text structural similarity is 0.858, and the mean value of the key relationship coincidence rate is 82.82%, as shown in Table 4.

Table 4: Qualitative interpretation and cross-text structure validation results

Representative Text Type	Core Ethnic Cultural Dimension	Main Narrative Strategy	Culture–Narrative Association Strength	Consistency with Manual Interpretation / %	Cross-Text Structural Similarity	Key Relationship Overlap Rate / %
Enlightenment-Critique Text A	Reflection on national character	Ironic narration, symbolic embedding	0.91	92.4	0.88	84.7
Family- Conflict Text B	Isomorphism of family and nation, ritual discipline	Limited perspective, discursive intervention	0.87	90.1	0.85	82.3
Female-Narrative Text C	Subjective awakening, emotional repression	First-person monologue, psychological depiction	0.93	93.6	0.89	85.1
Rural-Writing Text D	Customary memory, regional identity	Scene elaboration, ensemble narration	0.84	88.9	0.83	80.6
Urban-Experience Text E	Anxiety of modern life, identity imbalance	Perspective shifting, rhythmic transition	0.86	89.7	0.84	81.4

From the qualitative recognition results, in the Enlightenment critical text A, the model judged the correlation strength of "national reflection - ironic narrative" to be 0.91, and the consistency rate with manual annotation reached 92.4%. The cross-text structural similarity is 0.88, and the key relation overlap rate is 84.7%. This result shows that the model can not only identify the critical cultural connotation in the text, but also accurately grasp the ironic narrative mode attached to it. In text B of family conflict, the correlation strength of "family-country isomorphism - argument intervention/knowledge limitation perspective" is 0.87, the consistency rate is 90.1%, the structural similarity is 0.85, and the key relationship overlap rate is 82.3%. Compared with text A, the cultural dimension of this kind of text is more inclined to ethical discipline and family order, and its narrative organization is also more inclined to critical embedding in narrative advancement, which is basically consistent with the results of manual close reading.

In the female narrative text C, the model performance is particularly prominent. The correlation strength of "subject awakening-first-person monologue/psychological description" reached 0.93, which was the highest value among the five types of samples. The agreement rate with manual interpretation is 93.6%, the cross-text structural similarity is 0.89, and the key relation overlap rate is 85.1%. This indicates that the model is more sensitive to narrative perspective and psychological description characteristics when the connotation of national culture is mainly unfolded through female subject experience, emotional repression and self-expression. The result of vernacular writing text D is slightly lower, but still remains in a relatively stable range, with the correlation strength of "custom memory-scene presentation/group picture narration" of 0.84, the consistency rate of 88.9%, the structural similarity of 0.83, and the key relationship coincidence rate of 80.6%. In the urban experience text E, the correlation strength of "modern life anxiety-perspective switch/rhythm jump" is 0.86, the consistency rate is 89.7%, the structural similarity is 0.84, and the key relationship overlap rate is 81.4%. This shows that the model can also maintain good recognition ability for texts with fast narrative rhythm and frequent scene changes.

From the overall results, the mean value of the culture-narrative correlation strength of the five types of samples is 0.882, the mean value of the consistency rate is 90.94%, the mean value of the cross-text structural similarity is 0.858, and the mean value of the key relationship overlap rate is 82.82%. Among them, the difference between the highest value and the lowest value of the consensus rate is only 4.7 percentage points, and the fluctuation range of the structural similarity is controlled between 0.83 and 0.89, which indicates that the framework has good stability among different text types. In other words, the model is not only effective for a certain type of typical text, but can maintain a relatively consistent analysis effect in a variety of narrative styles and cultural expression paths. It can be seen that the method of this paper can not only make a more credible interpretation of a single text, but also reveal a more stable structural correspondence between national culture connotation and narrative strategy at the cross-text level.

5 Discussion

Based on the above experimental results, it can be seen that the improvement of the method in this paper does not come from the superposition of a single algorithm, but benefits from the synergy between the representation of national cultural implication, narrative strategy recognition, multi-dimensional feature association and parameter adaptive adjustment. Compared with the fixed threshold model, the Accuracy of the proposed method is improved from 84.9% to 91.4%, an increase of 6.5 percentage points, the F1 value is improved from 82.7% to 89.8%, and the RMSE is reduced from 0.214 to 0.146. Compared with Text-GNN,

the Accuracy and F1 are increased by 2.3 and 2.4 percentage points respectively, RMSE is decreased by 14.6%, and the single round training time is shortened from 12.4 s to 10.1 s. This shows that when dealing with texts such as Chinese modern novels in the early 20th century, which have obvious stylistic transitions, complex rhetorical levels and changeable narrative modes, it is difficult to stably depict the correspondence between cultural connotation and narrative strategy by simply relying on word surface statistics or fixed graph structure. Cultural semantics, narrative structure and context dynamics are integrated into the computational space. It can reduce misjudgment and missed judgment more effectively. The relevant comparison results are shown in Table 5.

Table 5: Comprehensive comparison of different text analysis methods

Method	Accuracy / %	F1 / %	RMSE	Training Time per Epoch / s	Main Discussion-Level Performance
Fixed-Threshold Model	84.9	82.7	0.214	8.3	Clear rules, but weak adaptability to implicit cultural semantics and complex narrative structures
BiLSTM Model	87.3	85.6	0.187	10.9	Strong sequence modeling ability, but insufficient recognition of cross-paragraph structures
Text-GNN Model	89.1	87.4	0.171	12.4	Able to exploit relational structures, but fixed parameters limit cross-text stability
Proposed Method	91.4	89.8	0.146	10.1	Balances recognition accuracy, interpretability, and resource efficiency, making it suitable for complex literary corpus analysis

From the perspective of computing resource consumption, the proposed method does not significantly increase the training burden due to the introduction of the adaptive mechanism. On the contrary, its time cost is controlled at a low level while maintaining high recognition accuracy, which indicates that the parameter adjustment mechanism not only improves the adaptation ability of the model to different text lengths, sentence pattern densities and semantic sparsity, but also improves the overall convergence efficiency. This is particularly important for the study of early modern novels, because the vernacular, shallow classical Chinese, translation and expression, and newspaper style coexist in the texts of the same period. If the model lacks dynamic adjustment ability, it will often overfit on one type of text, and the performance is unbalanced on another type of text. The proposed method remains relatively stable in ablation experiments and noise perturbation experiments: under the condition of 15% text noise, the accuracy is still 85.8%, which is 8.9 percentage points higher than that of the fixed parameter model, indicating that the proposed framework has a good ability to withstand the heterogeneity of literary corpus and the fluctuation of text quality.

The value of the approach in this paper is not only "more accurate", but also that it provides a kind of traceable structural evidence for literary interpretation. Previous studies on Chinese modern novels in the early 20th century mostly relied on the close reading of typical writers and typical texts to summarize the writing characteristics of national culture. This approach has depth, but it is not easy to form cross-text verification. Through computer-aided analysis, this paper stably presents the relations of "national introspection-ironic narrative", "subject awakening-first-person monologue", "custom memory-scene presentation", so that

the generation of cultural meaning no longer stays in the experience judgment, but obtains a comparable, locable and reviewable structural support. Of course, this approach does not replace close reading of literature. For the segments with high metaphor, complex irony level or obvious semantic suspension, the model still needs to use manual readback to complete the final correction. In other words, what is established in this paper is not a path of "algorithm substitution interpretation", but a research mechanism of "computational screening-structure identification-literature read-back".

On the whole, the framework of this paper responds to the long-standing tension between "large-scale processing" and "literary interpretability" in digital humanities research at the methodological level, and also provides a new observation entrance for the national culture research of modern Chinese novels in the early 20th century. Its significance is not to replace traditional literature research, but to reorganize the cultural clues and narrative patterns that were originally scattered in a large number of texts, so that the case interpretation and overall comparison can be connected in the same research framework.

6 Conclusions

Focusing on the national cultural connotation and narrative strategy of Chinese modern novels in the early 20th century, this paper tries to combine the problem awareness of literary research with computer-aided analysis methods, and constructs a digital research framework composed of national cultural implication representation, narrative strategy structural identification, multi-dimensional feature correlation analysis, and parameter adaptive adjustment. The results show that the framework can better adapt to the text characteristics of early modern novels with obvious stylistic transition, complex narrative levels and dense cultural metaphors. Under the condition of unified corpus and unified annotation, the Accuracy of the model reaches 91.4%, the F1 score reaches 89.8%, and the RMSE is reduced to 0.146. The overall performance is better than the fixed threshold model, BiLSTM model and Text-GNN model. Ablation experiments and robustness tests further show that each sub-module has practical contributions to the overall performance, especially the multi-dimensional feature correlation analysis and parameter adaptive adjustment, which have a strong support role for cultural recognition and narrative judgment in complex contexts.

More importantly, the significance of this paper is not only to improve the recognition accuracy, but also to provide a comparable, reviewable and read-back analysis path for the study of modern Chinese novels in the early 20th century. With the help of computer methods, the scattered cultural clues in the text, such as national reflection, ethical order criticism, subject awakening, and custom memory, can establish a clearer structural correspondence with the narrative methods such as irony narrative, first-person monologue, symbol embedding, and scene spreading. In the representative samples, the correlation strength is distributed between 0.84 and 0.93. Such a research path is helpful to combine case reading and overall comparison, and also provides an extended basis for the further integration of modern literary research and digital humanities methods.

References

- [1] Veg S. Democratic Modernism: Rethinking the Politics of Early Twentieth-Century Fiction in China and Europe[J]. *boundary 2*, 2011, 38(3): 27-65. <https://doi.org/10.1215/01903659-1430818>

- [2] Cohen P A. Remembering and forgetting national humiliation in twentieth-century China[J]. *Twentieth-Century China*, 2001, 27(2): 1-39. <https://doi.org/10.1353/tcc.2001.0002>
- [3] Meissner W. China's search for cultural and national identity from the nineteenth century to the present[J]. *China perspectives*, 2006, 2006(68): 41-54. <https://doi.org/10.4000/chinaperspectives.3103>
- [4] Lee H. Tears That Crumbled the Great Wall: The Archaeology of Feeling in the May Fourth Folklore Movement[J]. *The Journal of Asian Studies*, 2005, 64(1): 35-65. <https://doi.org/10.1017/S0021911805000057>
- [5] Haapanen J. Chinese people and the others: Notes on contemporary Chinese nationalism and the May Fourth movement[J]. *Journal of China and International Relations*, 2019, 7(1): 1-20. <https://doi.org/10.5278/jcir.v7i1.6635>
- [6] Gu M D. Lu Xun and Modern Chinese Literature in the Context of World Literature[J]. *Journal of Modern Literature*, 2021, 44(2): 76-92. <https://doi.org/10.2979/jmodelite.44.2.07>
- [7] Gu M D. Lu Xun and Modernism/Postmodernism[J]. *Modern Language Quarterly*, 2008, 69(1): 29-44. <https://doi.org/10.1215/00267929-2007-023>
- [8] Cai K. Lu Xun's heteromodal realism[J]. *Modern Language Quarterly*, 2024, 85(3): 253-278. <https://doi.org/10.1215/00267929-11196151>
- [9] Ma X. An Object In Between: The Past and Present of Lu Xun's "Diary of a Madman"[J]. *The Journal of Asian Studies*, 2024, 83(1): 41-61. <https://doi.org/10.1215/00219118-10875328>
- [10] Chou E S. Learning to Read Lu Xun, 1918–1923: The Emergence of a Readership[J]. *The China Quarterly*, 2002, 172: 1042-1064. <https://doi.org/10.1017/S000944390200061X>
- [11] McDougall B S. Disappearing women and disappearing men in may fourth narrative: a post-feminist survey of short stories by Mao Dun, Bing Xin, Ling Shuhua and Shen Congwen[J]. *Asian Studies Review*, 1998, 22(4): 427-458. <https://doi.org/10.1080/10357829808713209>
- [12] Wang B. "Breaking the Age of Flower Vases": Lu Yin's Feminist Rhetoric[J]. *Rhetoric Review*, 2009, 28(3): 246-264. <https://doi.org/10.1080/07350190902958719>
- [13] Hughes C R. Interpreting nationalist texts: A post-structuralist approach[J]. *Journal of contemporary China*, 2005, 14(43): 247-267. <https://doi.org/10.1080/10670560500065645>
- [14] Hu Z, Ismail R, Wang C. Identity and cultural allegory amidst the Chinese new era: Analyzing narrative strategies of the Chinese science fiction film[J]. *Cogent Arts & Humanities*, 2024, 11(1): 2313339. <https://doi.org/10.1080/23311983.2024.2313339>

- [15] Ziling G, Letian W, Biyu W. Translation strategies for detective novels in China: A historical and methodological review[J]. *Journal of Critical Studies in Language and Literature*, 2025, 6(2): 57-76. <https://doi.org/10.46809/jcssll.v6i2.338>
- [16] Chong J I. Popular narratives versus Chinese history: Implications for understanding an emergent China[J]. *European Journal of International Relations*, 2014, 20(4): 939-964. <https://doi.org/10.1177/1354066113503480>
- [17] Feng J. Narrating suffering, constructing Chinese modernity: The emergence of the modern subject in Chinese literature[J]. *East Asia*, 2000, 18(1): 61-81. <https://doi.org/10.1007/s12140-000-0004-7>
- [18] Shen C. Canon formation and children's literature during the may fourth period[J]. *CLCWeb: Comparative Literature and Culture*, 2014, 16(6): 10. <https://doi.org/10.7771/1481-4374.2302>
- [19] Ning* W. Globalizing Chinese Literature: toward a rewriting of contemporary Chinese literary culture[J]. *Journal of Contemporary China*, 2004, 13(38): 53-68. <https://doi.org/10.1080/1067056032000151337>
- [20] Maier C S. Consigning the twentieth century to history: Alternative narratives for the modern era[J]. *The American Historical Review*, 2000, 105(3): 807-831. <https://doi.org/10.2307/2651811>