

Digital Technology Empowers Modern Chinese Literature Research: An Analysis of Thematic Evolution in Classic Works Based on Text Mining

Yiyi Ru

Zhengzhou College of Finance and Economics, Zhengzhou 450000, Henan, China

Corresponding author: Yiyi Ru (e-mail: 13939086290@163.com)

ABSTRACT With the rapid development of digital humanities, traditional research on modern Chinese literature is constrained by limited samples and subjective interpretation, while text mining provides a quantitative and systematic analytical approach. This study takes 60 classic works of modern Chinese literature across three historical periods, namely 1919- 1949, 1949-1978, and 1978 to the present, as research objects and constructs an integrated text-mining framework covering data preprocessing, LDA topic modeling, and visual analytics. By combining computational analysis with traditional literary criticism, the study conducts an empirical investigation of cross-period thematic evolution. Eight core themes, including family-nation sentiment, human-nature exploration, and revolutionary narrative, are identified, and their evolutionary trajectories are systematically mapped. The results show that social-historical context, shifts in literary movements, and changes in authors' creative orientations are the main driving forces behind thematic transformation. The findings demonstrate that text mining can objectively and dynamically represent thematic evolution in modern Chinese literature. The study further argues that complementary integration of computational methods and humanistic inquiry is an important path for literary research in the digital age.

INDEX TERMS digital humanities, text mining, modern chinese literature, topic modeling, thematic evolution

I. INTRODUCTION

A. RESEARCH BACKGROUND AND SIGNIFICANCE

THE digital age has transformed the way in which information is accessed and disseminated. This transformation has reshaped traditional research methods and promoted innovation in the humanities and social sciences. This evolution allows for the systematic analysis of cultural data through the same precision-driven frameworks used to manage design databases and structural information [1]. Traditional modern Chinese literature research is limited by sample size and subjective perspectives, making it difficult to systematically sort out and dynamically track massive texts [2].

In the field of text analysis, text mining technology has become increasingly mature: word segmentation, word frequency analysis, topic modeling, social network analysis, word cloud, and visualization have realized the transformation of text analysis from human-driven to digital-driven to a certain extent [3].

Classic works of modern Chinese literature not only carry rich historical memories but also reflect the spiritual core

of the Chinese nation [4], and their thematic evolution is an important text carrier for interpreting China's modernization process. Data and text mining in the literary field are used for text analysis, theme identification, and style analysis. With the help of natural language processing, machine learning, and other technologies, text analysis has become more refined, and large-scale work research has become possible [5]. Applying text mining technology to the research on the thematic evolution of classic literary works mirrors the precision of computational design analysis, allowing researchers to overcome the limitations of traditional methods while promoting interdisciplinary innovation with significant academic and practical value. This systematic extraction of thematic threads functions similarly to the automated identification of fabric weave structures, providing a rigorous foundation for mapping the intricate continuity of cultural and industrial narratives.

B. REVIEW OF DOMESTIC AND FOREIGN RESEARCH STATUS

Traditional research on the themes of modern Chinese literature mainly focuses on case analysis and comparative research focusing on the thematic connotations of individual works, periods, or schools [6-9].

Although rich achievements have been accumulated, there are limitations such as limited samples and subjective perspectives. At present, there are many studies on digital humanities at home and abroad. Foreign digital humanities research started early, carrying out research on the evolution of literary schools [10] and the extraction of character traits in literary works [11] through text mining and other technologies; domestic scholars have applied text mining technology to the analysis of classical literature and modern and contemporary literature [12,13]. However, existing studies still have problems such as incomplete sample selection, insufficient technical adaptability, and disconnection between technical analysis and humanistic interpretation, and no systematic research on thematic evolution covering the entire cycle of modern Chinese literature has been formed.

C. RESEARCH CONTENT

Following the core idea of "technology empowerment - system construction - empirical analysis - reflection and prospect", this paper first clarifies the core concepts and theoretical basis, then constructs a text mining system adapted to modern Chinese literature texts, conducts empirical analysis on cross-period thematic evolution, and finally proposes the integration path of technology and traditional research.

D. RESEARCH INNOVATIONS

The innovations of this study include:

- (1) constructing a systematic text mining analysis framework with optimizing technical parameters;
- (2) selecting cross-period and multi-school samples to comprehensively depict the trajectory of thematic evolution;
- (3) achieving in-depth integration between computational analysis and traditional humanistic interpretation.

II. DEFINITION OF CORE CONCEPTS AND THEORETICAL BASIS

A. DEFINITION OF CORE CONCEPTS

Digital technology is a technical system that converts information into digital form for processing, storage, and transmission. Text mining, which can extract implicit, previously unknown, and potentially useful information from a large amount of text data [14], is the application of digital technology in the field of text analysis. The core technologies of text mining include text preprocessing (word segmentation, stopword removal, etc.), feature extraction (word frequency statistics, TF-IDF, etc.), topic modeling (LDA model, etc.), and visual presentation (word cloud, line chart, etc.), with quantitative and systematic characteristics [15]. The selection of classic modern Chinese literature works follows the criteria of literary value, temporal representativeness, and text accessibility, covering three historical periods: 1919-1949, 1949-1978, and 1978 to the present [16].

Literary theme refers to the core idea conveyed by a work. Thematic evolution refers to the continuous changes and phased characteristics affected by multiple factors, which are determined from three dimensions: changes in core keywords, changes in topic clustering, and deepening of thematic connotations.

B. THEORETICAL SUPPORT SYSTEM

Digital humanities theory advocates a "data-driven" research approach, providing a quantitative and visual perspective for literary research [17];

Textual criticism theory emphasizes close reading and interpretation of texts, laying the foundation for humanistic interpretation of thematic evolution analysis [18];

Evolutionary theory focuses on the dynamic changes of things in the time dimension, providing an analytical framework for revealing the internal logic of thematic evolution.

These three mutually support and form the theoretical basis of this paper.

C. ADAPTABILITY ANALYSIS OF TEXT MINING TECHNOLOGY

The semantic complexity, emotional richness, and rhetorical diversity of literary texts lead to adaptability difficulties in text mining, such as semantic ambiguity, difficulty in emotional recognition, and vague thematic boundaries.

To address these issues, this paper selects adaptive technologies and clarifies the application logic: In the text preprocessing stage, the jieba word segmentation tool, combined with a specialized stopword list and a custom dictionary for literary texts (including transitional vocabulary between semi-classical and modern vernacular), is used to improve word segmentation accuracy in response to the linguistic characteristics of literary works from the 1919-1949 period. In the feature extraction stage, a combination of word frequency statistics and TF-IDF weighting is adopted to highlight core thematic words.

In the topic modeling stage, an improved LDA model is used to enhance the interpretability of topic extraction by optimizing parameters such as the number of topics K and hyperparameters α and β .

In the result presentation stage, visualization methods such as word cloud and thematic evolution line chart are used to intuitively show the characteristics of thematic evolution.

At the same time, the technical mining results are combined with traditional textual criticism, and manual verification is carried out to ensure the accuracy of topic extraction.

III. CONSTRUCTION OF TEXT MINING SYSTEM FOR CLASSIC MODERN CHINESE LITERATURE TEXTS

A. SELECTION OF RESEARCH SAMPLES AND DATA PREPROCESSING

Based on the historical division and school characteristics, stratified sampling with comprehensive representativeness is adopted to select 60 classic works (20 works in each

historical period), covering major literary schools, genres (novels, poetry, prose, etc.), and representative writers. This sample size refers to the research norms of similar modern Chinese literature text mining studies [12,13], which focus on classic works rather than large-scale popular texts, ensuring the typicality and academic validity of the corpus. Texts are obtained from authoritative digital databases and unified into TXT format. After removing non-text content, standardized processing is performed.

The data preprocessing process is shown in Table 1. Python-related tools are used to implement denoising, word segmentation, stopwords removal, and word vector construction. These semantic associations between words lay a foundation for interpreting the evolution of humanistic themes such as "human nature exploration" from a quantitative perspective, bridging technical processing and literary analysis.

TABLE 1. Data Preprocessing Process

Preprocessing step	Processing method	Tools/standards	Purpose
Text denoising	Regular expression matching	Python re library	Remove interfering information such as special symbols and numbers
Word segmentation	Precise jieba word segmentation + custom dictionary	jieba word segmentation tool + literary field custom dictionary	Improve the accuracy of Chinese semantic recognition for works in the 1919-1949 period
Stopword removal	Dictionary matching and filtering	Harbin Institute of Technology stopwords list + literary specialized stopwords	Focus on core thematic words
Word vector construction	Word2Vec model training	Gensim library, window size 5, number of iterations 100	Capture semantic associations between words to bridge technical and literary analysis

Some representative samples selected are shown in Table 2, covering core schools of each period to ensure the representativeness and coverage of the samples.

B. TEXT MINING TECHNOLOGY PATH ORIENTED BY THEMATIC ANALYSIS

1) Extraction of Basic Text Features

This paper adopts a combination of word frequency statistics and TF-IDF weighting to extract core keywords:

(1) Obtain high-frequency words of works in each period through word frequency statistics to initially judge the thematic focus;

(2) Calculate the importance of words through TF-IDF weighting to screen core keywords of each period;

(3) Use the Word2Vec model to calculate the similarity of core keywords, construct a semantic association network, and explore the internal connections between thematic words.

2) Application of Topic Modeling Technology

This paper uses the LDA (Latent Dirichlet Allocation) topic model, which can automatically identify potential topics in texts and is suitable for thematic analysis of large-scale texts, for topic extraction.

Aiming at the characteristics of literary texts, this paper optimizes the parameters of the LDA model:

Determination of the number of topics K : Through grid search based on perplexity curve and topic coherence analysis, $K=8$ is finally determined. This is because $K=8$ achieves the lowest perplexity (128.6) and highest thematic coherence (0.78) among candidate values ($K=6-10$), and the extracted topics can fully cover the core thematic categories of modern Chinese literature (e.g., national survival, collective construction, individual development) without redundancy;

Optimization of hyperparameters α and β : α is set to 0.1 (enhancing the concentration of topics within documents), and β is set to 0.01 (enhancing the specificity of words within topics);

Selection of sampling algorithm: The sparse LDA algorithm is adopted to lay a foundation for future research expansion (e.g., increasing the sample size to 200+ works), although it is not necessary for the current dataset. Through the optimized LDA model, thematic modeling is performed on the preprocessed texts to extract core topics and thematic words of each period.

3) Visualization of Thematic Evolution

This paper adopts time series analysis and visualization technology to construct a thematic evolution map:

(1) Taking time as the horizontal axis and theme intensity (the frequency of themes appearing in works of each period) as the vertical axis, draw a thematic evolution line chart to show the intensity changes of core themes in the three historical periods;

(2) At the same time, use Gephi tool to construct a dynamic network of thematic associations, showing the association intensity and evolutionary relationship between themes in different periods. Through the visualization map, the overall situation and internal connections of the thematic evolution of modern Chinese literature are clearly presented.

C. VERIFICATION SYSTEM OF TEXT MINING RESULTS

1) Technical Level

Verification is realized through model fitting degree and result consistency test:

The perplexity index is used to test the fitting degree of the LDA model. A lower perplexity indicates a better model fitting effect. The perplexity of the optimized LDA model in this paper is 128.6, which is within the reasonable range of 120-140 reported in similar literary text LDA studies [12,13], confirming its reliability; The cross-validation method is adopted. The samples are divided into a training set (80%) and a test set (20%). The model is trained through the training set, and topic extraction is performed on the test set to verify the consistency of the results. The accuracy of

TABLE 2. Partial Representative Samples

Historical period	School type	Representative work	Author	Reason for selection
Enlightenment and National Salvation Period (1919-1949)	May Fourth Enlightenment Literature	Call to Arms	Lu Xun	A milestone of May Fourth Enlightenment Literature, criticizing national character
Enlightenment and National Salvation Period (1919-1949)	Crescent Moon School	Dead Water	Wen Yiduo	A representative work of the Crescent Moon School, combining formal beauty and ideological nature
Enlightenment and National Salvation Period (1919-1949)	July School of Poetry	Dayan River – My Nanny	Ai Qing	A classic work of anti-war literature, expressing patriotic feelings
Enlightenment and National Salvation Period (1919-1949)	Mandarin Duck and Butterfly School	Romance of Tears and Smiles	Zhang Henshui	A representative of popular literature, reflecting the life of citizens
Socialist Construction Period (1949-1978)	Revolutionary Historical Novel	Red Flag Spectrum	Liang Bin	Reflecting the revolutionary struggle process and shaping heroic images
Socialist Construction Period (1949-1978)	Agricultural Cooperation Novel	History of Entrepreneurship	Liu Qing	Showing the process of agricultural cooperation and embodying the spirit of collectivism
Socialist Construction Period (1949-1978)	Youth Literature	Long Live Youth	Wang Meng	Reflecting the spiritual outlook of new Chinese youth
Socialist Construction Period (1949-1978)	Industrial Theme Novel	Tempered into Steel	Ai Wu	Showing the upsurge of industrial construction and shaping the image of the working class
New Period to the Present (1978-Present)	Scar Literature	Scar	Lu Xinhua	The pioneering work of Scar Literature, reflecting on the Cultural Revolution
New Period to the Present (1978-Present)	Avant-garde Literature	To Live	Yu Hua	A classic of Avant-garde Literature, exploring the meaning of life
New Period to the Present (1978-Present)	Root-seeking Literature	White Deer Plain	Chen Zhongshi	A representative work of Root-seeking Literature, exploring the spiritual core of traditional culture
New Period to the Present (1978-Present)	New Era Literature	Baoshui	Qiao Ye	A typical text of New Era Literature, showing rural revitalization practice

cross-validation is 89.2%, indicating that the model results have high reliability.

2) Academic Level

Verification is realized through cross-comparison with traditional literary research results:

Compare the core themes extracted by text mining with the definition of literary themes in each period in authoritative literary history textbooks to verify the accuracy of topic extraction;

Invite 3 experts in the field of modern Chinese literature to conduct manual evaluation of the topic extraction results and thematic evolution analysis. The consistency of expert evaluation is 92%, ensuring that the research results comply with the academic norms of literary research.

IV. EMPIRICAL ANALYSIS OF THEMATIC EVOLUTION IN CLASSIC MODERN CHINESE LITERATURE BASED ON TEXT MINING

A. STAGES OF MODERN CHINESE LITERATURE DEVELOPMENT AND OVERALL SITUATION OF THEMATIC EVOLUTION

Through LDA topic modeling and word frequency analysis, this paper extracts 8 core themes (family and country sentiment, human nature exploration, social reality reflection, individual value, traditional culture inheritance, revolutionary narrative, era transformation, and emotional expression). By combining changes in theme intensity across the three historical periods, this study draws a line chart of illustrating the overall trajectory of thematic evolution (Figure 1) and summarizes the distribution of core thematic words in Table 3.

B. ENLIGHTENMENT AND NATIONAL SALVATION PERIOD (1919-1949)

The theme of this period is characterized by "national salvation as the core orientation with enlightenment as an important dimension". The "family and country sentiment" theme has the highest intensity (35.2%), which is higher than the "enlightenment narrative" (28.6%), reflecting that national survival was the primary appeal of the times. Literary works of this period formed a "family-country - individual" dual narrative structure, reflecting national crises through the portrayal of individual fates. For example, Lu Xun's *Call to Arms* criticizes national character and calls for national awakening through shaping individual images such as Ah Q and Kong Yiji.

C. SOCIALIST CONSTRUCTION PERIOD (1949-1978)

This period focuses on revolutionary narrative (theme intensity 42.5%) and collectivism (theme intensity 31.8%), strengthening the praise of national construction and recognition of revolutionary history. Works of this period, such as Wang Meng's *Long Live Youth*, show the collectivist enthusiasm of new Chinese youth and their love for the motherland through the growth process of a group of middle school students; Liang Bin's *Red Flag Spectrum* writes the legitimacy and inevitability of revolutionary history through the fate changes of a revolutionary family.

D. NEW PERIOD TO THE PRESENT (1978-PRESENT)

Since the New Period, themes have tended to be diversified, with individual value and traditional culture inheritance as the core, with theme intensities of 32.4% and 26.7% respectively. The definition of individual value here is multifaceted: it includes the individual spiritual exploration in Avant-garde Literature, the reflection on individual fate in

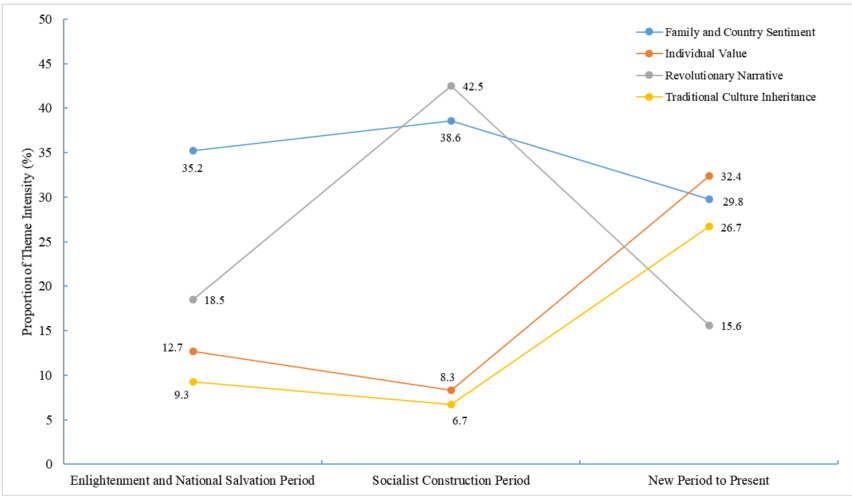


FIGURE 1. Overall Trend Chart of the Evolution of Core Themes in Modern Chinese Literature

TABLE 3. Core Thematic Words in Each Period

Historical period	Core themes	Theme intensity (%)	High-frequency thematic words	Thematic clustering characteristics
Enlightenment and National Salvation Period (1919-1949)	Family and country sentiment, enlightenment narrative, social criticism	35.2, 28.6, 22.3	National salvation, enlightenment, national character, anti-war, freedom	Focusing on the dual demands of national salvation and enlightenment, forming a “family-country-individual” dual narrative structure
Socialist Construction Period (1949-1978)	Revolutionary narrative, collectivism, construction	42.5, 31.8, 18.7	Revolution, collective, labor, construction, motherland	Centering on construction narrative, highlighting collectivism and revolutionary historical identity
New Period to the Present (1978-Present)	Individual value, traditional culture inheritance, era transformation, human nature	32.4, 26.7, 25.3	Individual, freedom, tradition, era transformation, human nature	Diversified themes, with individual value and traditional culture as core demands

Scar Literature, and the unity of individual development and social progress in New Era works such as Baoshui.

Works of this period, such as Chen Zhongshi’s *White Deer Plain*, explore the spiritual core of traditional culture through the rise and fall of the Bai and Lu families; Yu Hua’s *To Live* discusses the meaning of life and individual value through the personal fate of Fugui.

At the same time, the “era transformation” theme (25.3%) reflects literature’s attention to social changes since the reform and opening up, forming a pattern of diverse co-existence.

E. DYNAMIC EVOLUTION TRAJECTORY AND INTERNAL LOGIC OF CORE THEMES

1) Typical Theme 1: Evolution Path of Family and Country Sentiment and National Narrative

Family and country sentiment is a core theme of modern Chinese literature, and its evolution is deeply bound to China’s social and historical process.

In the Enlightenment and National Salvation Period, family and country sentiment centered on “national salvation first”, calling for awakening through depicting national crises. For example, Ai Qing’s *Dayan River – My Nanny* integrates personal emotions with national sufferings, expressing patriotic feelings.

In the Socialist Construction Period, family and country

sentiment combined with collectivism, focusing on the identification with the socialist motherland and enthusiasm for construction, and strengthening collective national identity by shaping model workers and revolutionary heroes.

Since the New Period, family and country sentiment has formed a multi-dimensional narrative of “individual-nation-world”, which not only connects individual and national development but also highlights the uniqueness of national culture from a global perspective. For example, Mo Yan’s works show the charm of Chinese national culture in the context of world literature through local narratives.

2) Typical Theme 2: Changes in the Expression of Human Nature Exploration and Individual Value

The evolution of the themes of human nature exploration and individual value reflects the continuous attention of modern Chinese literature to human subjectivity.

In the Enlightenment and National Salvation Period, human nature exploration served the enlightenment narrative, exploring the modernization of humans through criticizing and reflecting on national character. For example, Lu Xun’s *Diary of a Madman* exposes the oppression of human nature by feudal ethics through the awakening perspective of a “madman”.

In the Socialist Construction Period, individual value was overshadowed by collective value, and most works shaped

heroic images that obeyed the collective and sacrificed individuals, with weakened individual expression.

Since the New Period, human nature exploration has deepened, and works have focused on individual emotional needs and value pursuits. For example, Lu Yao's *Life* discusses the conflict between individual ideals and reality through Gao Jialin's choices.

3) Typical Theme 3: Evolution of Literary Reflection on Era Changes and Social Reality

Literary reflection on era changes and social reality is a core feature of modern Chinese literature, and its evolution presents a logic of "response - participation - reflection".

In the Enlightenment and National Salvation Period, literature responded to national crises and participated in enlightenment through social criticism. For example, Mao Dun's *Midnight* reflects the contradictions of semi-colonial and semi-feudal society through the description of Shanghai's industrial and commercial sector. In the Socialist Construction Period, literature actively participated in the construction process, serving the needs of the times through praising labor and revolution. For example, Liu Qing's *History of Entrepreneurship* guides farmers towards collectivization.

F. ANALYSIS OF INFLUENCING FACTORS OF THEMATIC EVOLUTION

Since the New Period, literature's reflection on reality has become more diversified, paying attention to both social progress brought about by reform and opening up and reflecting on problems in development. For example, Lu Yao's *Ordinary World* shows the struggle process of urban and rural youth while reflecting on class differentiation and value conflicts during the social transition period.

G. DRIVING ROLE OF SOCIAL-HISTORICAL CONTEXT

Social-historical context is the core driving force for the evolution of literary themes.

Social contradictions and era tasks in different periods directly determine the thematic orientation. For example, the national crisis in the Enlightenment and National Salvation Period, national construction in the Construction Period, and reform and opening up in the New Period constitute the core context of each stage respectively.

As a "mirror" of social reality, literature responds to the needs of the times through themes and participates in the construction of the spirit of the times.

H. LEADING EFFECT OF THE REPLACEMENT OF LITERARY TRENDS

The replacement of literary trends plays an important leading role in thematic evolution:

The May Fourth Enlightenment trend led the thematic orientation of the National Salvation Period; The socialist realism trend dominated the creation of the Construction Period;

Trends such as ideological emancipation and root-seeking since the New Period have promoted the diversification of themes.

The replacement of trends not only affects writers' creative concepts but also determines the expression paradigm and value orientation of themes.

I. INHERITANCE AND INNOVATION OF WRITERS' CREATIVE CONCEPTS

Writers' creative concepts are the internal driving force for thematic evolution. Based on their own experiences and pursuits, writers of different periods inherit and innovate themes:

Writers of the May Fourth Movement initiated the thematic tradition with the attitude of "enlighteners"; Writers of the Construction Period shaped the connotation of collectivism as "builders";

Writers since the New Period have promoted the return of themes to individual value and traditional culture from the perspective of "reflectors", continuing the spiritual context of modern Chinese literature while realizing the innovation of thematic expression.

V. REFLECTION AND PROSPECT OF MODERN CHINESE LITERATURE RESEARCH EMPOWERED BY DIGITAL TECHNOLOGY

A. ADVANTAGES AND LIMITATIONS OF TEXT MINING TECHNOLOGY IN LITERARY THEME RESEARCH

1) Advantages

The significant advantages of text mining technology are reflected in three aspects:

(1) Efficiently processing large-scale literary texts, breaking the limitation of traditional research samples, and realizing the overall grasp of thematic evolution;

(2) Reducing subjective biases through quantitative analysis, and accurately identifying core themes of each period with the help of word frequency statistics, LDA modeling, etc.;

(3) Intuitively presenting the trajectory and association of thematic evolution relying on visualization technologies (such as word cloud and emotional trend chart), helping to discover laws easily ignored in traditional research.

2) Limitations

There are still obvious limitations in the application of technology:

It is difficult to accurately capture the implicit emotional and aesthetic connotations of literary texts. The interpretation of rhetoric such as metaphors and symbols, as well as poetic "images", still needs to rely on traditional textual criticism;

Existing technologies are mostly adapted to ordinary texts, and are insufficiently adapted to the semantic complexity and rhetorical diversity of literary texts, which are prone to problems such as inaccurate word segmentation and biased topic extraction;

Text mining methods can reveal the observable patterns of thematic evolution, but they have limited capacity to explain the underlying humanistic and historical mechanisms behind these patterns. The interpretation of the humanistic connotations and historical logic behind themes still needs the support of traditional research.

B. INTEGRATION PATH OF MODERN CHINESE LITERATURE RESEARCH IN THE DIGITAL AGE

1) Complementary Integration of Technology Empowerment and Traditional Research Paradigms

The core path is the complementary integration of the two:

(1) Taking text mining technology as a tool to expand the sample scope and analytical perspective of traditional research, realizing quantitative analysis and dynamic tracking of themes;

(2) Taking traditional textual criticism as the foundation to deepen the humanistic interpretation of technical mining results.

For example, when analyzing *White Deer Plain*, text mining can be used to clarify the thematic positioning of "traditional culture inheritance", and then traditional criticism can be used to interpret its specific connotations and contemporary value.

2) Interdisciplinary Collaboration

Interdisciplinary collaboration is an important guarantee for the digital transformation of modern Chinese literature research:

Literary scholars clarify research directions and theoretical guidance;

Computer scholars optimize technical models to adapt to literary texts;

Historians provide contextual analysis to support thematic interpretation.

By constructing a closed loop research framework of "problem-technology-interpretation", interdisciplinary innovation in modern Chinese literature research is promoted.

3) Future Research Prospect

(1) In the future, it is necessary to further optimize the specialized text mining model in the literary field to improve technical adaptability;

(2) Expand technical application scenarios to cover multiple genres such as poetry and drama;

(3) Integrate multi-modal technologies to enrich analytical dimensions;

(4) Extend the research scope to non-classic works and regional literature, carry out international comparative research, and highlight the unique value of modern Chinese literature in the context of world literature.

VI. CONCLUSION

A. CORE LAWS OF THEMATIC EVOLUTION IN CLASSIC MODERN CHINESE LITERATURE WORKS

Through the integrated analysis of text mining technology and traditional literary research, this paper reveals the core laws of the thematic evolution of classic modern Chinese literature works by decoding the underlying systematic development of literary themes across historical periods:

Historically, the themes development of modern Chinese literature follows a clear evolutionary trajectory, progressing from Enlightenment and national salvation (1919-1949), through revolution and collectivism (1949-1978), to diversification and individualism-oriented expression since 1978, reflecting a shift from collective narrative to individual-oriented themes.

In terms of thematic connotation, modern Chinese literature consistently engages with family-nation sentiment, the exploration of human nature, and reflections on social reality, while continuously enriching their meanings of these themes in response to changing historical contexts, thereby forming an evolutionary logic jointly driven by "era context - literary trends - writers' concepts".

In terms of expressive paradigms, thematic expression has shifted from "instrumentality" to "humanism", from serving era tasks to focusing on human spiritual needs, highlighting the humanistic spiritual context of modern Chinese literature.

B. CORE VALUE AND IMPLEMENTATION PATH OF TEXT MINING TECHNOLOGY EMPOWERING LITERARY RESEARCH

Text mining technology provides a new technical support for modern Chinese literature research, and its core value lies in:

Breaking the sample limitation and subjective bias of traditional research, realizing quantitative analysis and dynamic tracking of massive literary texts, and discovering thematic associations and evolutionary laws easily ignored in traditional research.

The effective application of digital technologies in literary research depends on integrative strategies that combine computational analysis with traditional textual interpretation.

First, the integration of technology and traditional research paradigms, taking technical mining results as the foundation and traditional textual criticism as the deepening, realizing the organic unity of quantitative analysis and humanistic interpretation; Moreover, sustained interdisciplinary collaboration among literary studies, computer science, and history is essential for establishing a robust and sustainable digital research paradigm for modern Chinese literature.

AUTHOR CONTRIBUTIONS

Yiyi Ru designed, collected and analyzed the data, and drafted the manuscript. Yiyi Ru conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published.

Yiyi Ru participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

- [1] M. Meng, Z. Li, Z. Jiang, and Y. Cai, "The innovative development of humanities and social sciences in the digital age," *World Social Sciences*, no. 6, pp. 31-67, 2023, [Online]. Available: <https://ssaj.ajcass.com/Magazine/show/?id=90243>.
- [2] C. Schöch, "Big? smart? clean? messy? data in the humanities," *Journal of Digital Humanities*, vol. 2, no. 3, pp. 2-13, 2013, [Online]. Available: <https://hal.science/hal-00920254v1>.
- [3] Q. Yin and J. Zeng, "Form and meaning: a possible textual analysis theory from the perspective of digital humanities," *Shandong Social Sciences*, no. 11, pp. 69-77, 2021, doi: 10.14112/j.cnki.37-1053/c.2021.11.011.
- [4] J. Chen, "Exploration of the path for reinterpreting classic modern Chinese literature," *Jiaying Literature*, no. 24, pp. 16-19, 2024.
- [5] Y. Chen, "The application of artificial intelligence technology in literary research," *Journal of Dialectics of Nature*, vol. 47, no. 8, pp. 93-99, 2025, doi: 10.15994/j.1000-0763.2025.08.010.
- [6] M. Gu, "On the significance of Chinese literature to world literature — centered on Lu Xun's literary creation," *Social Sciences in China*, no. 8, pp. 183-193, 282, 2025.
- [7] D. Zhao, "Thematic evolution of "wife-selling" theme works in modern Chinese literature," *Journal of Shantou University (Humanities & Social Sciences Edition)*, vol. 31, no. 1, pp. 40-45, 96, 2015.
- [8] W. Zheng, "Pluralistic and turbulent modern Chinese literature — (III) Nationalist literature," *Journal of Jiamusi University (Social Sciences Edition)*, vol. 28, no. 4, pp. 68-70, 2010.
- [9] J. Wang and Z. Li, "The production mechanism and canonization of modern Chinese literature texts — an investigation based on the novel *Camel Xiangzi*," *Shandong Social Sciences*, no. 10, pp. 62-70, 2025, doi: 10.14112/j.cnki.37-1053/c.2025.10.006.
- [10] A. Peris, E. Meijers, and M. van Ham, "The evolution of the systems of cities literature since 1995: Schools of thought and their interaction," *Networks and Spatial Economics*, vol. 18, no. 3, pp. 533-554, 2018, doi: 10.1007/s11067-018-9410-5.
- [11] D. Skorinkin and A. Bonch-Osmolovskaya, "Text mining War and Peace: Automatic extraction of character traits from literary pieces," *Digital Scholarship in the Humanities*, vol. 32, no. suppl_1, pp. i17-i24, 2017, doi: 10.1093/llc/fqw052.
- [12] L. Wei and J. Lu, "Parsing Korean Classical Literature by Integrating Text Mining and Semantic Analysis," *Applied Mathematics and Nonlinear Sciences*, vol. 10, no. 1, 2025, doi: 10.2478/AMNS-2025-0517.
- [13] X. Zhang, "Comparing the Performance Characteristics and Commonalities of Chinese Modern and Contemporary Literature in Different Cultural Contexts by Combining Text Mining Techniques," *Applied Mathematics and Nonlinear Sciences*, vol. 9, no. 1, 2024, doi: 10.2478/AMNS.2023.2.00701.
- [14] D. Waegel, "The Development of Text-Mining Tools and Algorithms," Collegeville, PA, USA: Ursinus College; 2006.
- [15] F. Shao, "Research on the Mechanism of Knowledge Discovery in Emerging Technology Fields Based on Text Mining Methods [dissertation]," Tianjin, China: Tianjin Polytechnic University; 2025.
- [16] X. Wang, "Research on the Rewriting and Canonization of Famous Literary Works Since 1949 [dissertation]," Kunming, China: Yunnan University; 2023, doi: 10.27456/d.cnki.gydu.2023.000006.
- [17] C. Xia, "The theoretical evolution and paradigm shift of digital humanities: From "digital methods in humanistic research" to "humanistic research in the post-digital age"," *Journal of Library Science in China*, vol. 50, no. 5, pp. 83-101, 2024, doi: 10.13530/j.cnki.jlis.2024040.
- [18] X. Niu, "Modern criticism: The transformation of the "criticism" paradigm, the variation of "close text reading" and its possibilities — A review of *Literary Criticism: A Concise Political History and a discussion on current Chinese literary criticism*," *Beijing Literary Review*.