

Quantitative of Sociocultural Trends Based on the Analysis of Historical Journal Text Big Data

Jiakuo Yang¹, Qing Wang², Xinghua Zhang^{2,*}

¹Northeast Normal University, Jilin 130024, China

²Beijing Jiaotong University, Beijing 100044, China

772498383@qq.com, 21121592@bjtu.edu.cn, zhangxh@bjtu.edu.cn (Corresponding author)

Abstract. The vast amount of documentary materials preserved for modern and contemporary sociology and history makes comprehensive reading impractical. This study employs big data and artificial intelligence techniques to quantitatively analyze the journal "Chinese Culture" published in Yan'an in the 1940s. By treating all texts in the journal as data, correlations among vocabulary are systematically analyzed to identify distinct themes. Quantitative analysis of theme ownership across articles reveals manifestations of social and cultural conditions during the transition from old to new democratic revolution. Whereas traditional micro-perspective research lacks comprehensive analysis, this study leverages big data techniques for objective measurement of the entire journal. Findings provide important guidance on social regulation and research by uncovering correlations in the cultural discourse. This quantitative approach facilitates a holistic understanding of the ideological, political, and cultural climate reflected in the seminal journal.

Keywords: social condition; history studies; theme ownership; big data technique

1. Introduction

In contemporary society, we are confronted with vast and complex amounts of data that contain information about numerous societal states [Allardt,1969]. Societal states encompass various dimensions, including mental states, economic states, social structural states, educational states, and more. By collecting, analyzing, and interpreting historical documents or literature from specific periods, we can gain insights into aspects of social development, societal structures, and societal transitions, allowing for a more effective assessment and measurement of societal states. Social measurement data from different time periods reveal distinct characteristics of each era [Cheadle,2017; Chen,2020]. In the Soviet period, based on the study of books and documents on urban construction and urban studies, the relationship between urban planning in the Soviet Union is analyzed [Dubrow,2008]. Japanese historians have analyzed the state of Japanese society and the national mentality based on the defensive reaction of young scholars to their modernizing remarks [Ehrmann,2021]. Sara Wallace Goodman conducted research on the American citizenship handbook during the 1970s, illustrating the evolution of American citizenship norms over half a century and the impact of local wars on this identity [Garon,1994]. Joshua Kjerulf Dubrow studied political protest data from various European countries to investigate the effects of new and old democracies and income inequality on the development of social stability [Goodman,2021]. This measurement of societal states is of great significance for social governance and research, as understanding the societal state of a specific era enables us to better provide solutions to contemporary issues and offer guidance for the continuous development and progress of society [Guo,2018].

This paper focuses on the "Chinese Culture" journal, aiming to objectively measure and quantitatively study the intellectual and cultural state of society at that time using big data research methods. "Chinese Culture" (hereafter referred to as the journal) was the official publication of the Cultural Association of the Shaanxi-Gansu-Ningxia Border Region during the Yan'an period and was an important comprehensive theoretical and academic journal during that time. The journal was founded in 1940 and ceased publication in 1941, consisting of 3 volumes and 14 issues with a total of 148 articles. Its content covered various fields including politics, art, literature, history, Marxist theory, and even natural sciences. The journal included narrative essays, discursive essays, explanatory essays, and practical essays. The Yan'an period was one of the most challenging years of the Chinese anti-Japanese war. As a historically significant magazine, the articles it contained objectively reflected the intellectual and cultural state of society during the Yan'an period, encompassing literary creation, historical research, class struggle, social transformation, economic development, and war, among other aspects.

Research on the "Chinese Culture" journal is an important direction in the sociological research of China during the Yan'an period. Some scholars have started to comprehensively investigate the development history, thematic changes, and authorship of the journal from the perspective of literature [Hjørland,1995; Huang,2022; Jelodar,2019]. Researchers focus on the founding background, objectives, and editorial team of the journal to gain a deep understanding of its macroscopic positioning and academic orientation [Jin,2023; Lindstedt,2019]. In addition, they pay attention to the core content of the journal and important published articles, revealing the academic characteristics and development trajectory of the journal through the study of authors, citation analysis, and thematic trends. In general, existing research provides insights into the literary aspects of the "Chinese Culture" journal, and researchers have used comprehensive analysis from a microscopic perspective to reveal the journal's academic characteristics, development trajectory, and academic influence.

However, it should be noted that previous studies have mainly focused on the isolated interpretation of typical articles, relying on manual observation and analysis. This approach often falls short when it comes to objectively understanding and measuring the societal state of a specific era. Observations and analyses conducted solely by humans are often insufficient, and machine-based techniques are required to uncover patterns in historical data [Lounsbury,1979; Miyata,2020]. The formation of common linguistic characteristics and themes in historical materials requires in-depth research using methods from

information science to reveal the historical significance of the journal as a platform for promoting ideology, beliefs, and values [Nunan,2012].

This paper employs statistical methods to treat all the texts published in the journal as data. With the assistance of big data and artificial intelligence technology, we aim to detect the societal state of the time. By using Latent Dirichlet Allocation (LDA) modeling, we systematically analyze "Chinese Culture," automatically identifying several themes within the journal and specifying the keywords and features of each theme. These themes, as viewed from the perspective of feature pairs in research articles, provide insights into the societal characteristics imbued with thoughts and theories in literary works. The results of the data analysis corroborate the real societal state reflected in the results of data analysis and several representative articles, accurately reflecting the societal state of that era.

LDA (Latent Dirichlet Allocation) is a statistical model used in text mining and topic modeling. It was proposed by Blei, Ng, Jordan, and others and is a generative probabilistic model of a corpus, representing documents as weighted collections of latent topics, where topics are characterized by distributions over words [Pavlinek,2017; Phillips,2001]. LDA analyzes patterns of co-occurrence of words in text documents and infers latent topics from a collection of texts. It can extract topic information from a large amount of textual data, aiding in the understanding of the underlying meaning of text and facilitating text categorization and analysis. LDA has been widely applied in various fields, including text classification, information retrieval, and text summarization [Rani,2021;Scaccia,2021; Shams,2017]. As a powerful topic modeling method, LDA has a broad application prospect in journal literature analysis. For example, Xie et al. used the LDA model to summarize the outbreak of emerging research hotspots in the field of structural engineering in the past two decades [Sooryamoorthy,2020]. Shams and Baraani-Dastjerdi analyzed articles from multiple journals using the LDA model, discovering correlations and co-occurrence patterns between different topics [Suissa,2022]. Scaccia and Scott employed the LDA model to perform topic modeling analysis on abstracts from journals, revealing that certain topics were more likely to appear together with specific themes [Tijssen,2016]. Yosuke Miyat et al. identified topics in library and information science using the LDA model and visualized the knowledge structure in this field, detecting transformations in research areas [Xie,2022]. By applying the LDA model, hidden information in literature can be deeply explored, revealing associations between different topics and providing further insights and inspiration for academic research and knowledge discovery.

It is important to note that the societal state measured in this paper does not represent the current societal state in China but rather the societal state during the Yan'an period. Focusing on a specific region and historical context, this study emphasizes the intellectual and cultural state, particularly during the transition from the old democratic revolution to the new democratic revolution. Therefore, the unique perspective of this research provides crucial historical observations.

2. Materials and Methods

2.1. Data source

The data for this paper comes from all 148 articles published in "Chinese Culture." A dataset of the entire journal was established, including fields such as titles, authors, issue numbers, article content, genre types, and publication dates. After preprocessing the original text data by segmenting the text, removing stop words, and eliminating punctuation marks and special characters, a specialized noun dictionary database for "Chinese Culture" was constructed for data analysis.

ID	Title	Stylistic Type	Publication Date	Author	First Issue	Content
1	...	article	19400215	Mao Zedong	23	...
2	...	article	19400215	Esxy	25	...
3	...	article	19400215	Zhou Yang	34	...
4	...	article	19400215	Wu Yuzhang	40	...
5	...	article	19400215	He Sijing	47	...
6	...	article	19400215	Chen Boda	40	...
7	...	article	19400215	Xian Xinghai	52	...
8	...	article	19400215	Xiao SAN	56	...
9	...	article	19400215	Hu Man	63	...
10	...	random feeling	19400215	arbor	70	...

Fig.1: Sample dictionary database data

2.2. Topic identification

By establishing an LDA model to investigate word clusters and topics, the LDA model is based on the assumption of a probabilistic topic model. It assumes a two-stage process for generating words in a document:

- (I) Randomly select a topic distribution.
- (II) For each word in the document:
 - (i) Randomly choose a topic from the distribution obtained in (I).
 - (ii) Randomly select a word from the corresponding vocabulary distribution.

In practice, both the distribution of topics within documents and the distribution of words within topics are not known a priori; they are inferred from the observed literature or, in this context, the content of the acquired articles. Mathematically, the relationship between latent variables and observed variables is represented by the joint distribution as shown in Equation (1).

$$p(\beta_{1:K}, \theta_{1:D}, z_{1:D}, w_{1:D}) = \prod_{i=1}^K p(\beta_i) \cdot \prod_{d=1}^D p(\theta_d) \sum_{n=1}^N p(z_{d,n} | \theta_d) p(w_{d,n} | \beta_{1:K}, z_{d,n}) \quad (1)$$

$$p(\beta_{1:K}, \theta_{1:D}, z_{1:D} | w_{1:D}) = \frac{p(\beta_{1:K}, \theta_{1:D}, z_{1:D}, w_{1:D})}{p(w_{1:D})} \quad (2)$$

When applying the LDA model, the number of topics should be specified, and when the topic perplexity is lower, the corresponding number of topics of the model is better.

$$Perplexity(D) = \exp - \frac{\sum_{d=1}^M \log(p(w))}{\sum_{d=1}^M N_d} \quad (3)$$

refers to the probability of occurrence of each word in the test set and is calculated as follows:

$$p(w) = p(z|d) \times p(w|z) \quad (4)$$

In this context, $p(z|d)$ represents the probability of each topic appearing in a document, and $p(w|z)$ represents the probability of each word in the dictionary appearing under a certain topic.

Regarding the peak distribution of topics in the article's topic projection, the dataset is divided into two categories: single-topic distribution and multi-topic distribution. The single-topic distribution refers to articles where the probability of the first most likely topic is greater than 5 times the probability of the second most likely topic. Conversely, it is classified as a multi-topic distribution. The specific definition is as follows:

$x_m [i]$ represents the probability of the i -th topic appearing in the m -th article, where n is the number of predefined topics. For each article, the probabilities of topics calculated using the LDA model form the dataset. From each dataset, the first maximum value and the second maximum value, denoted as $max1$ and $max2$, are extracted to define single-topic and multi-topic distributions.

$$x_m [j] = max1, \quad x_m [k] = max2 \\ j, k = 0, 1, 2, \dots, n \quad m = 1, 2, 3, \dots, 148$$

"Single Topic": $max1/max2 > 5$, article corresponds to *Topic #j*;" Multi-topic": $max1/max2 \leq 5$, all $x_m [i] > \frac{1}{5} max1$, all peaks, article corresponding to *Topic #i* (all eligible i values).

3. Results

T01- Marxist Theory (Topic #3)	Philosophy thought Science struggle Dialectics Law Marxist world theory Materialism form class idealism Revolutionary Method Concept Basis Human things Material phenomena Spiritual bourgeoisie objective subjective logic movement process position theory	
T02- Thoughts (Topic #4)	Revolutionary bourgeois Movement Political Thought Three People's Principles New Culture National Imperialism Socialist Struggle New Culture Movement Democracy World People's State capitalism Communism New Democracy Victory Economy Chinese nation period Semi-colonial era Leading class democracy revolution colonialism	
T03- Art (Topic #5)	Art and literature form works Drama writer Popular literature reality content novel thought national creation Common people popular author Mass world people times folk characters influence story art artist Music readers use	
T04- The War (Topic #6)	National Movement in the border areas of the Anti-Japanese War Political People anti-Japanese national struggle representative workers Ideological Activities Cadres cultural circles Education of the enemy achievements Science and art People's public academic organizations organ Theory of the victory of democracy by the Japanese invaders behind the enemy lines	
T05- History (Topic #0)	Socialist productive forces Comrade Enemy of labor capitalism Economy Industry Communist peasant force team Eighth Route Army relations Technology State means eye workers land Congress soldier Power People's condition planning process movement form	
T06- Education (Topic #7)	Educational theory Political Economy Bourgeois Economy Marxist-Leninist Politics Student cadres School Principles Planning Process Capitalist Forces Eighth Route Army Method Works Capital of Interest Struggle Soldier Revolution Bright Content Phenomenon Error Analysis Democratic purpose	
T07- Class (Topic #2)	Slave nation Times Material Agricultural class Proof economic system State religion representation fact Human people relations Report Interest characteristics Process Scholar Chinese nation phenomenon Political peasant tool Aristocracy unified production ruler	
T08- Politics (Topic #1)	Words things condition Method Chinese characters world dialectics movement change Meaning Sound human language unity relationship View of the era of the phenomenon of the war of resistance Economic sentences represent the basic elements of the struggle against the laws of national imperialism	

Without using the LDA algorithm, we cannot accurately classify each article, and the boundaries between different topics remain ambiguous. This is because in journal articles, topics are not explicitly labeled but are represented by probability distributions of a series of words. The importance of the LDA algorithm lies in its ability to discover underlying topics in articles by statistically analyzing word frequencies, word co-occurrences, and contextual information. By calculating the word distribution of each topic and the document-topic distribution, LDA can assign articles to the most relevant topics, achieving precise text classification.

Through the application of the LDA method, we can delve deeper into and understand the various fields covered by the "Chinese Culture" journal from a more comprehensive perspective, appreciating the rich content and significance of each topic. To explore the depth and breadth of these topics, we will introduce each topic's keywords, characteristics, and typical articles one by one, providing a more comprehensive examination of the important concepts and research areas within each topic.

3.1.1. Marxist theory theme

In the "Chinese Culture" journal, the thought process of the Sinicization of Marxism is extensively dis

cussed. It integrates achievements from various fields such as philosophy, ideology, and science. The core concepts include materialism, dialectical materialism, historical materialism, and the theory of class struggle. These articles reveal the significant contributions made during the Yan'an period to the Sinicization of Marxism. From the Party education during the Yan'an Rectification Movement to the interpretation of Mao Zedong Thought and the determination of the Party's guiding ideology, the Sinicization of Marxism during the Yan'an period had a milestone significance. "Chinese Culture" published articles such as Ai Siqi's "On China's Uniqueness," a series of philosophical lectures on "What is Philosophy," and "Comments on Several Important Philosophical Thoughts Since the War of Resistance Against Japan" by Yang Song, all of which delved deeply into the Sinicization of Marxism. They emphasized the need to apply Marxism in Chinese society, or to Sinicize Marxism, by firmly adhering to Marxist viewpoints, using the scientific methods of dialectical materialism and political economics established by Marx and Engels to concretely and objectively study China's social and economic relations. The dissemination of Marxism and dialectical materialism in China was indeed rooted in China's own needs. If we can Sinicize Marxism or make Marxist scientific disciplines in China, we will not only enrich and develop Marxism but also promote the great development of Chinese culture and achieve a glorious position in world human culture.

3.1.2 Intellectual theme

In the "Chinese Culture" journal, the intellectual theme encompasses various areas such as revolution, the bourgeoisie, movements, politics, and more. In this field, people contemplate and seek ideological and theoretical solutions to social and political issues. The article "On the Study of Mr. Sun Yat-sen's Thoughts" discusses the influence of Mr. Sun Yat-sen's thoughts on the Chinese revolution. "Gorky's Socialist Aesthetic View (Continued)" delves into Gorky's views on socialist aesthetics. "Characteristics of the May Fourth Cultural Movement" explores the impact of the May Fourth Movement on Chinese society and thought. "Morality in Christianity and 'People's Morality'" compares the morality of Christianity with the morality of the people. "Lu Xun's Direction Is the Direction of the New Culture of the Chinese Nation" elucidates Lu Xun's contributions to the direction of the new culture of the Chinese nation. These articles reflect the changes and developments in Chinese thought and culture in different periods and areas, providing valuable insights for understanding China's intellectual and cultural landscape.

3.1.3 Artistic theme

"Chinese Culture" adheres to the creative view that politics and art are integrated, combining the richness of literary and artistic content with diverse forms. During the transformation of art, "Chinese Culture" played a leading role in the Sinicization and popularization of Marxist art in the Yan'an base area, guiding the artistic direction of the time. "Chinese Culture" covers various aspects of the artistic landscape during the Yan'an period, including drama, fine arts, music, comics, dance, and more. It discusses the transformation and utilization of traditional folk arts in the Shaanxi-Gansu-Ningxia Border Region, which not only provided strong spiritual motivation for the resistance against Japan but also enriched the ideological and cultural thought system of the New Democracy. The revolutionary musicians of Yan'an, in their new music creation based on the inheritance of traditional Chinese music, served as a basic means and effective form of educating the masses. This resulted in numerous outstanding musical works that deeply aroused the enthusiasm of the workers and peasants. The artists' in-depth participation in the actual production struggles of the masses allowed them to collect and organize a large amount of traditional artistic resources, which, combined with the actual situation of the anti-Japanese struggle, were used to transform and utilize traditional folk arts to create art for the proletarian masses. The transformation and utilization of traditional folk arts during the Yan'an period were the results of the exploration of "nationalization" and "popularization" of revolutionary art in the "spirit of Yan'an" guided

by the theoretical and practical thinking of the Yan'an Rectification Movement and "New Culture Movement." In terms of fine arts, art workers, as workers and peasants, deeply engaged in actual production and struggle, collecting and organizing a large number of traditional fine arts resources and combining them with the practical situation of the anti-Japanese struggle to transform and utilize traditional fine arts to create art for the proletarian masses. The transformation and utilization of traditional folk arts during the Yan'an period were results of the exploration of "nationalization" and "popularization" of revolutionary art in the "spirit of Yan'an," guided by the theoretical and practical thinking of the Yan'an Rectification Movement and "New Culture Movement." The literary and artistic workers in Yan'an carried out in-depth exploration and utilization of traditional literary and artistic resources and the transformation and utilization of the traditional literature and art of the Shaanxi-Gansu-Ningxia Border Region in the "spirit of Yan'an." This, combined with the actual situation of the anti-Japanese struggle, transformed and utilized traditional literature and art to create literary and artistic works for the proletarian masses.

Art often takes on various forms, and at the same time, art is a part of literature, which often embodies art. This way of expression is the highlight of "Chinese Culture," the embodiment of art through the form of literature. With the change in the nature of society, the form of art is constantly evolving, and its content becomes more satirical and reveals the darkness of society. This was the main style of the early period of the New Culture Movement. However, with the introduction of Marxism, especially in the Yan'an region, a vibrant form and content of art emerged. This was the embodiment of the Sini-cization of Marxism in art. Art originates from life and transcends life, which is the most authentic representation. It reflects the revolutionary struggles led by the Communist Party of China against imperialism and feudalism, as well as the passionate scenes of building a new China.

3.1.4 War theme

The resistance against Japanese aggression during World War II is a significant historical period that encompasses various domains, including border regions, national movements, politics, civilian engagement, anti-Japanese sentiment, ethnic identity, and struggles. It also involves key figures, activists, ideological movements, and officials across these domains. In the "Chinese Culture" journal, numerous articles have been published that delve into the theme of war. Some of these articles include "Cultural Movements in the Shaanxi-Gansu-Ningxia Border Region during the War of Resistance Against Japanese Aggression," "Documents from the Shaanxi-Gansu-Ningxia Border Region Cultural Association Representative Assembly," "Tasks of the Current Cultural Movement," "The Mission and Work of North China United University," "An Overview of Literary and Artistic Movements in Enemy-Occupied Areas," and "Historiography Since the War of Resistance Against Japanese Aggression."

The war of resistance against Japanese aggression represents the Chinese people's struggle against Japanese invasion, a challenging and arduous conflict that required unity and resistance from people across the nation. In the border regions and areas behind enemy lines, people from various fields actively participated in the struggle, engaging in political, cultural, educational, and theoretical activities. The war-themed articles in "Chinese Culture" provide profound insights and perspectives, helping us better understand the pivotal role played by the masses in achieving victory during the war and the aspiration of people to achieve democracy and freedom through the triumph of the war.

3.1.5 Historical theme

Academic debates and discussions in "Chinese Culture" have yielded significant achievements in historical research. The journal published the first articles by Fan Wenlan after his arrival in Yan'an, including "Debates on Ancient History Stages" and "Evolution of Chinese Classical Studies." Kefu's work on the "Origin of Art," Yin Da's studies on the "Origins of the Chinese Nation and Its Culture," debates on the nature of the Yin and Shang societies, and Peiyuan's "A Bird's-Eye View of Logic History" are just a few examples of the diverse historical themes covered in the journal. These articles span various

fields, including ancient Chinese history, art history, ethnic history, logic history, classical studies history, art history, and drama history, showcasing the broad scope of historical research during the Yan'an period. For instance, in the field of ancient history, scholars such as Fan Wenlan, Yin Da, Xie Hua, Ye Huosheng, and Li Ji participated in discussions on the periodization of ancient history, contributing significantly to the study of ancient history. The articles published in "Chinese Culture" provide valuable source material for historical research.

3.1.6 Educational theme

Education encompasses a wide range of topics, including educational theory, political economy, and Marxist-Leninist ideology. "Chinese Culture" features discussions related to genetics in the field of education. On June 20, 1941, Le Tianyu published an important article titled "A Discussion on the Correct Application of Genetics," which introduced and critiqued the genetic theory of Trofim Lysenko while challenging the Morganian genetic theory. This article marked the beginning of a debate on genetic theory in China, making it a significant event in the history of genetics in China. Although the depth of genetic research in the articles is relatively shallow compared to historical studies, their contributions in paving the way for future academic pursuits should not be underestimated.

3.1.7 Class theme

Articles in "Chinese Culture" cover class relations, a crucial aspect of human society's development, encompassing economics, politics, society, and more. "Drama and Religion" examines the relationship between religion and art, analyzing religion as a form of class representation. "Debates on Ancient History Stages" explores the evolution of socio-economic and class relations under the slave system. "The Origin of the Chinese Nation and Its Culture" studies the origins of the Chinese nation from a cultural perspective, revealing exchanges and relations between different classes. "Customs of the Miao People in Xiangxi" focuses on the production tools, customs, and economic conditions of the peasant class, shedding light on the characteristics and development of agrarian society. "Several Key Issues in the Debate on the Nature of the Yin and Shang Societies" delves into the relationships and interests between rulers and the ruled in Shang society. "Examinations and Elections" critically analyzes the class nature and political implications of the imperial examination system, discussing political relationships between rulers and the people. These articles allow for a comprehensive examination of the essential concepts and areas related to class, offering insights into the relationships, interests, and power distribution among different classes. Such a perspective is crucial for understanding Chinese history and culture and the development of human society as a whole.

3.1.8 Political theme

Chinese Culture "was an official newspaper of the Shaanxi Gansu Ningxia Border Region Cultural Association during the Yan'an period. It was an important political publication during the Yan'an period and an important record of disciplines and activities related to power, state, government, social organizations, and social relations. This also corresponds to the main themes of "country", "nation", and "relations" contained in the theme. The important position of the Yan'an period in modern Chinese history and the history of the Communist Party of China cannot be underestimated. Mao Zedong and other older generation proletarian revolutionaries personally created and nurtured the "Yan'an Spirit" of self-reliance, hard work, seeking truth from facts, and wholeheartedly serving the people. It has accumulated as a precious treasure in the treasure trove of the Chinese nation's spirit, and has become an important spiritual pillar for the national unity and unity in socialist modernization construction and the great rejuvenation of the Chinese nation, the "Chinese Dream". Chinese Culture "is the first publication of "On New Democracy", which fully demonstrates that" Chinese Culture "has become an important platform for discussing the theory of new democracy within the party and also provides an excellent platform for China's political development. Chinese Culture has published many articles on political research, with the top priority being the issue of cultural and ethnic forms. Progressive intellectuals such

as Zhang Wentian, Zhou Yang, Mao Dun, Ai Siqi, Guo Moruo, Chen Boda, and Wang Shiwei have engaged in intense debates and published articles on related issues, which is of great significance for the formation of national, scientific, and popular policies at the political leadership level.

3.2 Theme diversity analysis

With the vector of themes, we can project the content of each article onto the vectors of these 8 themes. If the projection of an article on a certain topic vector is much greater than that on other vectors, we call it a single topic article. If the projection values on multiple topic vectors cannot be ignored, we call it a multi topic article. The journal contains 58 articles on a single topic and 90 articles on multiple topics out of 148 articles. Multi theme articles mainly focus on two and three themes, with 54 articles on two themes and 26 articles on three themes. These multi theme articles are written in the form of literature and art to program the ideas of Marxism and the New Democratic Revolution. Some articles combine the influence and significance of culture on society in the form of art, presenting the contribution of the literary and artistic field to society during the Yan'an period; Some articles focus on educational enlightenment during the war period and the New Culture Movement, emphasizing the profound impact of education in various historical periods. Some articles combine economy with art, emphasizing that economy promotes the development of art, and art reflects economic prosperity. Some articles focus on discussing the economic, artistic, and cultural development of society during the Anti Japanese War. This form has an important impact on the formation of the context in which the first generation of leaders of the CPC understand in simple terms and lift the weight lightly. This article selects four typical single theme, dual theme, and multi theme classic articles as examples to demonstrate their unique historical significance.

3.2.1 Single Theme Literature Mao Zedong's "Politics and Culture of New Democracy"

Mao Zedong's article "Politics and Culture of New Democracy" was first published in "Chinese Culture". This important work on the theory and strategy of China's New Democratic Revolution was a speech delivered by Mao Zedong at the first representative conference of the Shaanxi Gansu Ningxia Border Region Cultural Association on January 9, 1940. It systematically elaborated on a series of major issues related to China's democratic revolution, This provides the main framework and foundation for the preliminary construction of the discourse system of the New Democratic Revolution. According to the LDA results, the article is classified as an ideological topic with a 99.964% probability, therefore it is a single topic article. The representative vocabulary of ideological themes includes revolution, bourgeoisie, politics, ideology, Three Principles of the People, new culture, ethnicity, imperialism, etc. From the content of the article, it can be seen that this book is a great achievement that combines the universal truth of Marxism Leninism with the specific practice of the Chinese revolution. Its theoretical foundation is the Marxist revolutionary view, which is also an important vocabulary appearing in the ideological theme keywords. At the same time, the core of the New Democratic Revolution is to solve the problems of national independence and people's liberation, so the keywords "nation" and "democracy" also appear in ideological themes.

In the article, Mao Zedong also mentioned keywords such as revolution, class, and socialism, which are representative vocabulary of ideological themes. Mao Zedong pointed out that the Chinese revolution is divided into two stages: the New Democratic Revolution and the Socialist Revolution, and the future of the New Democratic Revolution under the leadership of the working class is inevitably socialism. Therefore, socialism is also one of the important contents of ideological themes. Mao Zedong also mentioned concepts such as the Three Principles of the People, democracy, and new culture, which are also representative vocabulary for ideological themes. Mao Zedong critically inherited the Three Principles of the People and gave birth to New Democracy. During the period of the New Democratic Revolution, the Party must adopt a new democratic political, economic, and cultural program that is different from both capitalism and socialism. This is closely related to the keywords mentioned in ideological themes such as ideology, politics, and culture.

Overall, this article corresponds to the ideological theme and elaborates on the important concepts and topics involved in the ideological theme from multiple perspectives. Mao Zedong's "Politics of New Democracy and Culture of New Democracy" is a great achievement that combines the universal truth of Marxism Leninism with the specific practice of the Chinese revolution. It scientifically summarizes the historical characteristics and revolutionary experience of China, and depicts the development laws of the Chinese democratic revolution and socialist revolution. At the same time, Mao Zedong also implemented the concept of national and cultural confidence and proposed new directions for cultural development, which are closely related to the concepts involved in ideological themes. Therefore, through the correspondence between this article and ideological themes, we can better understand China's revolutionary process and cultural development.

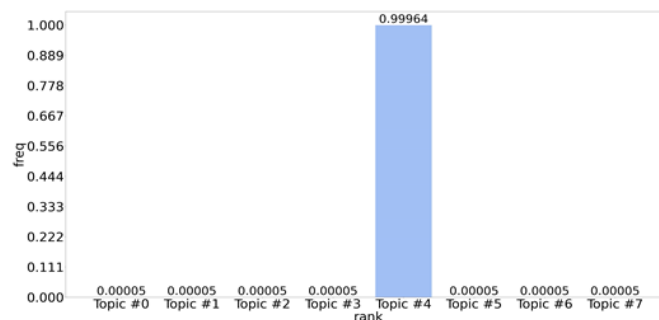


Fig.3: Thematic distribution of "Politics of New Democracy and Culture of New Democracy"

3.2.2 Dual-Themed Literature by Yang Zhou - Marxist Literary Theory

In the first volume, third issue of "Chinese Culture," the article "Reflections on the 'May Fourth' Literary Revolution" by Yang Zhou delves deeply into the issues of literary revolution and popularization, making significant contributions to the development of Marxist literary theory in China. The article exhibits a distinct dual-themed distribution, with the highest frequencies belonging to themes related to art and ideology, accounting for 66% and 22% respectively. The article primarily discusses the impact of the "May Fourth" literary revolution on Chinese literature and culture.

Within the realm of art-related themes, there is a higher frequency of occurrence, with characteristic keywords revolving around art, literature, creation, populism, reality, nationality, and thought. It focuses on the development, expression, and influence of literary and artistic works on a popular and national scale. The article discusses various aspects of literature, such as literary forms, drama, novels, fine arts, music, and analyzes how literary creation affects the popularization of content and form among the masses. Art-related themes also emphasize the influence of authors on the masses and the world, as well as how authors use timeless folk stories and characters to impact readers and society. All of these are representative of the content conveyed by art-related themes.

In comparison, the frequency of occurrence of ideology-related themes is slightly lower than that of art-related themes. Keywords in this category encompass revolution, ideology, politics, movements, and nationality, all closely related to politics, ideology, and social structural changes. Ideology-related themes represent the social and political backdrop of the "May Fourth" literary revolution movement discussed in the article. The article analyzes intellectual trends and movements, such as revolutionary thought, the Three Principles of the People, the New Culture Movement, democratic revolution, and communism, describing the victories, economic aspects, leadership, and class issues associated with them. It also sheds light on the oppression and humiliation faced by Chinese capitalism and socialism under the influence of semi-colonialism and imperialism. These elements are indicative of the content conveyed by ideology-related themes.

Through its analysis and content, this article demonstrates the characteristics of dual-themed distribution. Through discussions on art and literature, it reveals the impact and transformation of Chinese

culture in the context of the "May Fourth" literary revolution. Similarly, by delving into the social and political backdrop through ideology-related themes, the article examines how the "May Fourth" literary revolution, as a significant cultural and social movement, became an essential chapter in the development of Chinese culture.

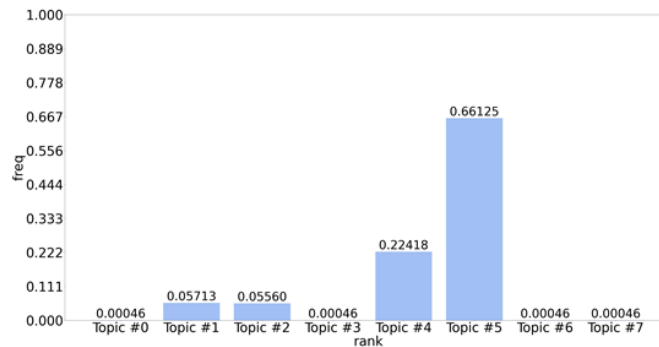


Fig.4: Distribution of themes in "The Sense of Two Thirty (Spirit) on the May Fourth Literary Revolution"

3.2.3 Dual-Themed Literature by Li Bozhao in "Overview of Enemy Rear Cultural and Artistic Movements" Development of Rural Drama

In the 3rd volume, 3rd issue of "Chinese Culture," renowned playwright Li Bozhao published an article titled "Overview of Enemy Rear Cultural and Artistic Movements," which caused a significant sensation. The article explores the development of theater in the Yan'an period, highlighting the role of local theater as a crucial chapter in the history of Chinese drama, an aspect that cannot be overlooked. This article exhibits a dual thematic distribution, encompassing both artistic and wartime themes. Based on the results of LDA (Latent Dirichlet Allocation), it is revealed that 58% of the article pertains to wartime themes, while 27% is dedicated to artistic themes. The artistic theme primarily focuses on characteristic words related to literature and the arts, such as novels, drama, fine arts, music, etc., while the wartime theme emphasizes terms related to resistance, civilians, and politics. The article places particular emphasis on the relationship between the war effort and cultural arts. It narrates how during the Enemy Rear Cultural and Artistic Movement, writers began to focus on writing realistic novels and dramas to promote national spirit, corresponding to the characteristics of the artistic theme. Additionally, the article underscores the role and impact of culture and the arts during the war, where literary works represented the people's sentiments, and artists created art accessible to the masses, aligning with the characteristics of the wartime theme. In the article, the author elaborates on the overview of the Enemy Rear Cultural and Artistic Movement, predominantly discussing aspects related to drama, music, fine arts, and the Enemy Rear New Cultural Movement, thus encompassing the vocabulary associated with both artistic and wartime themes. Therefore, this article effectively combines artistic and wartime themes, establishing a close connection between the two and conducting an in-depth analysis of the relationship between artistic creation and societal reality.

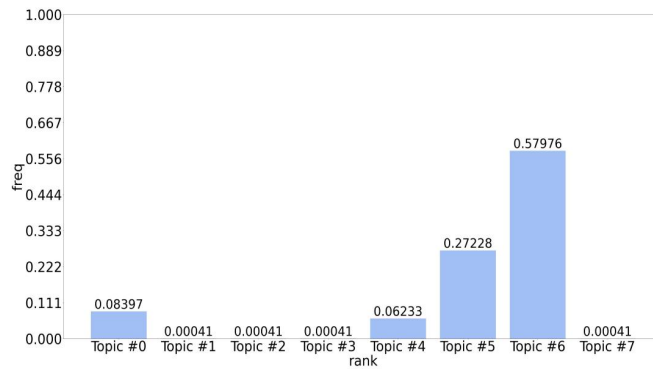


Fig.5: Thematic distribution of "Overview of the Literary Movement Behind Enemy Lines"

3.2.4 Multi-Topic Literature - Emile Yerushalmi's "On the Origin of Art" - The Origin of Art

In Volume 1, Issue 4 of "Chinese Culture," Emile Yerushalmi's article, "On the Origin of Art," primarily explores the question of the origin of art. Yerushalmi contends that Préhannov's definition is correct, a viewpoint endorsed by most Marxists, and it currently holds a dominant position in the Chinese literary and artistic community. The article encompasses discussions in various domains. According to the results of LDA (Latent Dirichlet Allocation), it can be observed that 37% of the article pertains to topics related to art, 22% to Marxist theory, 22% to education, 9% to class-related topics, and 8% to historical subjects. These topics span across multiple fields, including art, philosophy, education, economics, and socialism. Key terms associated with the art-related topics include "art," "literature," and "novel," reflecting the article's discussions on artistic creation and literary works, such as the analysis of ancient literature and ideas about novel writing. The Marxist theory-related topics feature terms such as "philosophy," "thought," and "materialism," which delve into the philosophical ideologies and the influence of materialism behind artistic creation. The education-related topics encompass terms like "education," "political economy," and "Marxism-Leninism," which highlight discussions on art education, political economics, and the application of Marxist-Leninist thought in artistic creation. Class-related topics incorporate terms such as "nation," "class," and "system," addressing the analysis of societal conditions during specific historical periods and pondering the relationship between artistic creation and social systems. Historical topics include terms like "socialism," "productivity," and "communism," exploring the positive impact of socialism and communism on artistic creation, as well as discussions on the relationship between productivity development and artistic creation. This article, due to its coverage of various fields such as art creation, philosophical thought, education, political economics, and social systems, exhibits a multi-topic distribution of thematic characteristics. Through the analysis of these five topics, a better understanding of the relationship between artistic creation and fields such as society and philosophy is gained.

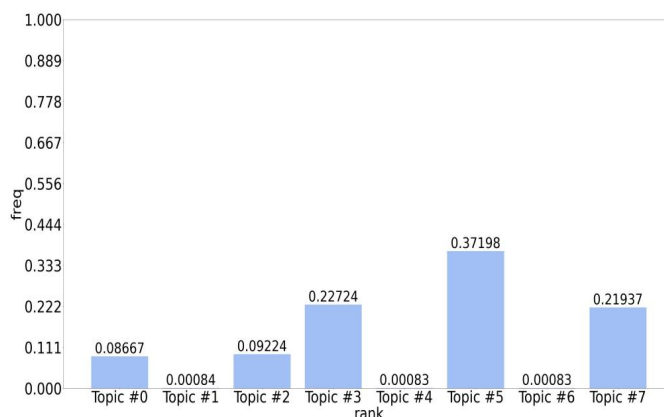


Fig.6: Thematic distribution of "Questions on the Origins of Art".

4. Conclusions

This study demonstrates the value of utilizing big data techniques to analyze journal texts for insights into historical societal conditions. The LDA topic model enabled systematic quantification of themes and vocabulary correlations in "Chinese Culture." Examining topic distribution and associations revealed the journal's academic scope and engagement across political, cultural, and ideological contexts during the Yan'an period.

While traditional literature offers limited micro-perspective analyses, this research provides comprehensive measurement of the entire journal corpus. By transitioning from keywords, to topics, to multi-topic articles, a thorough picture emerges of the diverse sociocultural landscape. The multiple topics reflect the journal's focus on complex societal issues and phenomena.

Quantitative attribution of article topics evaluates the journal's breadth and depth of impact. And analysis of weighted keywords uncovers specific historical backgrounds, dynamics, and concepts. Overall, this big data approach facilitates holistic understanding of the rich conditions underlying the seminal publication.

Connecting disparate data points enables simultaneously probing multiple dimensions and factors shaping past societies. The methodology's large-scale, multidimensional analyses advances measurement of sociocultural conditions. This demonstrates the technique's significant value for detecting nuances and comprehensively interpreting phenomena to inform contemporary social progress.

Acknowledgments

The authors thank Beijing Municipal Social Science Foundation (No. 21JCB093).

References

- Allardt, E. Evolutionary, structural, and institutional characteristics of societies. *Acta Sociologica* **1969**, *12*, 74-84. <http://www.jstor.org/stable/4193696>.
- Cheadle, C.; Cao, H.; Kalinin, A.; & Hodgkinson, J. Advanced literature analysis in a Big Data world. *Annals of the New York Academy of Sciences* **2017**, *1387*, 25-33. <https://doi.org/10.1111/nyas.13270>.
- Chen, X.; Zou, D.; Xie, H. Fifty years of British Journal of Educational Technology: A topic modeling based bibliometric perspective. *British Journal of Educational Technology* **2020**, *51*, 692-708. <https://doi.org/10.1111/bjet.12907>.

- Dubrow, J.K.; Slomczynski, K.M.; Tomescu-Dubrow, I. Effects of democracy and inequality on soft political protest in Europe: Exploring the European social survey data. *International Journal of Sociology* **2008**, *38*, 36-51. <https://doi.org/10.2753/ijso020-7659380302>.
- Ehrmann, M.; Hamdi, A.; Pontes, E.L.; Romanello, M.; Doucet, A. Named entity recognition and classification in historical documents: A survey. *ACM Computing Surveys*. *ACM Computing Surveys* **2021**, *56*, 1–47. <https://doi.org/10.48550/arXiv.2109.11406>.
- Garon, S. Rethinking modernization and modernity in Japanese history: a focus on state-society relations. *The Journal of Asian Studies* **1994**, *53*, 346-366. <https://doi.org/10.2307/2059838>.
- Goodman, S.W. Good American citizens : a text-as-data analysis of citizenship manuals for immigrants, 1921–1996. *Journal of Ethnic and Migration Studies* **2021**, *47*, 1474-1497. <https://doi.org/10.1080/1369183X.2020.1785852>.
- Guo, L.; Li, S.; Lu, R.; Yin, L.; Gorson-Deruel, A.; King, L. The research topic landscape in the literature of social class and inequality. *PloS one* **2018**, *13*, e0199510. <https://doi.org/10.1080/1369183X.2020.1785852>.
- Hjørland, B.; Albrechtsen, H. Toward a new horizon in information science: Domain-analysis. *Journal of the American society for information science* **1995**, *46*, 400-425. [https://doi.org/10.1002/\(SICI\)1097-4571\(199507\)46:6<400::AID-ASIS2>3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1097-4571(199507)46:6<400::AID-ASIS2>3.0.CO;2-Y).
- Huang, W.; Song, F.; Zhang, S.; Xia, T. Influence of deep learning-based journal reading guidance system on students' national cognition and cultural acceptance. *Frontiers in Psychology* **2022**, *13*, 950412. <https://doi.org/10.3389/fpsyg.2022.950412>.
- Jelodar, H.; Wang, Y.; Yuan, C.; Feng, X.; Jiang, X.; Li, Y.; Zhao, L. Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications* **2019**, *78*, 15169-15211. <https://doi.org/10.1007/s11042-018-6894-4>.
- Jin, S.; Stokes, G.; Hamilton, C. Empirical evidence of urban climate adaptation alignment with sustainable development: Application of LDA. *Cities* **2023**, *136*, 104254. <https://doi.org/10.1016/j.cities.2023.104254>.
- Lindstedt, N.C. Structural topic modeling for social scientists: A brief case study with social movement studies literature, 2005–2017. *Social Currents* **2019**, *6*, 307-318. <https://doi.org/10.1177/2329496519846505>.
- Lounsbury, J.W.; Roisum, K.G.; Pokorny, L.; Sills, A.; Meissen, G.J. An analysis of topic areas and topic trends in the Community Mental Health Journal from 1965 through 1977. *Community mental health journal* **1979**, *15*, 267-276. <https://doi.org/10.1007/BF00778706>.
- Miyata, Y.; Ishita, E.; Yang, F.; Yamamoto, M.; Iwase, A.; Kurata, K. Knowledge structure transition in library and information science: topic modeling and visualization, *Scientometrics* **2020**, *125*, 665-687. <https://doi.org/10.1007/s11192-020-03657-5>.
- Nunan, T.A. Ecologies of socialism: Soviet Gradostroitel'stvo and late soviet socialism. *Journal of Eurasian Studies* **2012**, *3*, 106-115. <https://doi.org/10.1016/j.euras.2012.03.002>.
- Pavlinek, M.; Podgorelec, V. Text classification method based on self-training and LDA topic models. *Expert Systems with Applications* **2017**, *80*, 83-93. <https://doi.org/10.1016/j.eswa.2017.03.020>.
- Phillips, C.S. Culture, social minds, and governance in evolution. *Politics and the Life Sciences* **2001**, *20*, 189-202. <https://doi.org/10.1017/S0730938400005475>.

Rani, R.; Lobiyal, D.K. An extractive text summarization approach using tagged-LDA based topic modeling. *Multimedia tools and applications* **2021**, *80*, 3275-3305. <https://doi.org/10.1007/s11042-020-09549-3>.

Scaccia, J. P.; Scott, V.C. 5335 days of Implementation Science: using natural language processing to examine publication trends and topics. *Implementation Science* **2021**, *16*, 1-12. <https://doi.org/10.1186/s13012-021-01120-4>

Shams, M.; Baraani-Dastjerdi, A. Enriched LDA (ELDA): Combination of latent Dirichlet allocation with word co-occurrence analysis for aspect extraction. *Expert Systems with Applications* **2017**, *80*, 136-146. <https://doi.org/10.1016/j.eswa.2017.02.038>.

Sooryamoorthy, R. Scientometrics for the study of sociology. *International Sociology* **2020**, *35*, 461-479. <https://doi.org/10.1177/0268580920957911>.

Suissa, O.; Elmalech, A.; Zhitomirsky-Geffet, M. Text analysis using deep neural networks in digital humanities and information science. *Journal of the Association for Information Science and Technology* **2022**, *73*, 268-287. <https://doi.org/10.1002/asi.24544>.

Tijssen, R. J.; Winnink. Twenty-first century macro-trends in the institutional fabric of science: bibliometric monitoring and analysis. *Scientometrics* **2016**, *109*, 2181-2194. <https://doi.org/10.1007/s11192-016-2041-z>.

Xie, Y.; Ning, C.; Sun, L. The twenty-first 'century of structural engineering research: A topic modeling approach. *Structures* **2022**, *35*, 577-590. <https://doi.org/10.1016/j.istruc.2021.11.018>.