Social Sciences and Humanities
ISSN Print: 3104-4328
ISSN Online: 3104-4336
CD ISSN: 3104-4344
https://doi.org/10.63313/SSH.9080

AI-Enhanced Big Data Content Analysis of Weibo for Digital Cultural Heritage: Insights into Sichuan Opera

Min Liu*. Xinran Chen. Wei He and Fubo Liu
Chengdu University of Technology, Chengdu 610059, China
Email: 15882402089@163.com

How to cite this paper: Liu, M., Chen, X. R., He, W., & Liu, F. B. (2026). AI-enhanced big data content analysis of Weibo for digital cultural heritage: Insights into Sichuan Opera. *Social Sciences and Humanities*, 3(3), 15-27. ISSN Print: 3104-4328; ISSN Online: 3104-4336.
https://doi.org/10.63313/SSH.9080
Published: 2026-04-07

Copyright © 2026 by author(s) and Erytis Publishing Limited.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
CC BY 4.0

Abstract

Sichuan Opera, a prominent Chinese intangible cultural heritage, is celebrated for its diverse genres, distinctive vocal and acrobatic performance styles, and rich musical traditions. However, in the digital era it faces challenges such as audience decline and difficulties in cultural transmission due to competition from new media and globalization. Preserving Sichuan Opera is crucial for maintaining cultural diversity and fostering social innovation. In this study, we leverage digital technologies to analyze and support its preservation. Weibo, a leading Chinese social media platform, was used as a data source to capture public discourse related to Sichuan Opera. After collecting and preprocessing thousands of relevant Weibo posts, we applied natural language processing (NLP) and machine learning techniques to encode, cluster, and analyze the text data. Through combined qualitative and quantitative analysis, we identify key thematic topics, geographic distribution patterns, and publication trends in Sichuan Opera's online presence. Our AI-driven big data analysis reveals current strengths and gaps in the Opera's social media dissemination. The results provide actionable insights for enhancing audience engagement and informing digital strategies to sustain Sichuan Opera, demonstrating the potential of social media analytics and AI in the knowledge management and conservation of intangible cultural heritage. The methodology could also guide digital preservation efforts for other traditional operatic art forms and cultural media.

Keywords

Weibo analytics; Digital cultural heritage; Big data analysis; Sichuan Opera preservation

1. Introduction

Chinese opera, a folk-art tradition combining singing, recitation, acting, and dancing, is a cornerstone of China's cultural heritage. As a distinctive regional form, Sichuan Opera, evolving from the fusion of traditional opera and Bashu cultural elements over nearly 300 years, was inscribed on China's inaugural national Intangible Cultural Heritage list in 2006. However, contemporary diversified entertainment and changing media have reduced its audience and hindered its development, demanding new strategies for its revitalization and inheritance.

The digital era and the proliferation of social media have opened new avenues for cultural dissemination. Unlike traditional media outlets that produce centrally curated content, social media allows users to create, edit, and share information interactively (McDermott, 2010; Mergel, 2012). As China's leading microblogging platform, Weibo's data helps reveal cultural topic propagation trends and public engagement (Nian et al., 2022), making it a promising path for Sichuan Opera promotion.

This paper presents a novel approach that combines qualitative and quantitative analyses of Sichuan Opera discourse on Weibo. A dataset of Sichuan Opera-related Weibo posts was compiled and preprocessed using Python. The text data were then analyzed with natural language processing techniques: sentiment analysis (using the ROSTCM6 tool) to assess public attitudes toward Sichuan Opera, social network analysis to examine how discussions propagate among users, and Latent Dirichlet Allocation (LDA) topic modeling (via scikit-learn and pyLDAvis visualization) to uncover the underlying thematic structure of online discussions. These insights can inform strategies for the preservation and revitalization of Sichuan Opera, helping to safeguard this intangible cultural heritage in the digital age.

2. Theoretical model

2.1. Social network analysis

Social network analysis (SNA), which refers to studying and mapping social structures through graph theory, has been widely used in many social science fields (Zaefarian et al.,2022). Emerging in the 1930s and developed in the 1970s from sociology, anthropology and mathematics, SNA is a proven useful tool for researchers (Verdú et al., 2021). Its commonly used software includes ROSTCM6, UCINET, Pajek, etc. Due to the explosive rise of online social networks, SNA has emerged as a significant academic field in recent years (Singh et al.,2024). In a social network, nodes are individual participants and links are relationships between participants (Alwash& Levine,2019). Degree centrality is based on the degree of a node, that is, the number of arcs directly connected to it (Zhang et al.,2011). The degree centrality formula is:

$$C_{degree}(v) = \frac{degree(v)}{N-1} \tag{1}$$

where v is node. degree(v) is the degree of node v. N-1 in the formula represents the total number of nodes except itself.

Betweenness Centrality is a measure of how well nodes are mediated between them as shortest paths in the network. The betweenness centrality formula is:

$$C_{betweenness}(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{2}$$

where σ_{st} is the shortest number of paths from node s to node t.$\sigma_{st}(v)$ is the shortest number of paths through node v.

Closeness Centrality measures the average shortest path length from a node to other nodes in the network. The closeness centrality formula is:

$$C_{closeness}(v) = \frac{N-1}{\sum_{\mu \neq v} d(v,\mu)} \tag{3}$$

The higher the closeness centrality, the shorter the path from the node to other nodes and the higher the efficiency of information propagation. $d(v,u)$ represents the distance from node v to node u (Zhang&Luo,2017).

2.2. Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) is a popular topic modeling technique for hidden semantic discovery of text data (Zhao et al.,2020). Topic modeling methods based on LDA have been applied to natural language processing, text mining, and social media analysis, information retrieval (Jelodar et al., 2019; Tong&Zhang,2016). In the LDA model, a document is generated as follows(Blei, Ng& Jordan,2003)：

(a) Document i is generated by sampling from the Dirichlet distribution α, which has a subject distribution of θ_i.

(b) Generate the j-th word of document i by sampling from the polynomial distribution of the topic θ_i. Its theme is $z_{i,j}$.

(c) Dirichlet distribution β sampling generate subjects $z_{i,j}$. Its word distribution is $\emptyset_{z_{i,j}}$.

(e) From the polynomial distribution $\emptyset_{z_{i,j}}$of words, the final word generated by sampling is $w_{i,j}$.

(f) Repeat the above steps to eventually generate a document with N words.

(g) The result is M documents on K topics.

Therefore, the joint distribution of all visible and invisible variables in the whole model is as follows:

$$p(w_i, z_i, \theta_i, \emptyset|\alpha,\beta) = \prod_{j=1}^{N} p(\theta_i|\alpha) p\big(z_{i,j}|\theta_i\big) p(\emptyset|\beta) p\left(w_{i,j}\middle|\emptyset_{z_{i,j}}\right) \tag{4}$$

The maximum likelihood estimate of the word distribution for a document can be integrated by combining $\theta_i$and $\emptyset$ in the above equation. Finally, z_i is summed:

$$p(w_i|\alpha,\beta) = \int \theta_i \int \emptyset \sum_{z_i} p(w_i, z_i, \theta_i, \emptyset|\alpha,\beta) \tag{5}$$

3. Data Preprocessing

3.1. Text filtering and data mining

Both web crawlers and APIs are nowadays widely used technologies to collect data from social media (Xie, Chu, Chiu&Wang,2022), including Weibo (Liu&Hu,2018). Since the number of microblog posts on the official account is too small, Sichuan opera actors and fans are also taken as data collection objects. Use Python to crawl Weibo users related to Sichuan Opera and their original Weibo posts from 2013~2023. The standard for crawling is Sichuan Opera-related elements. In the end, 137 Weibo accounts and 10,057 comment texts were obtained.

3.2. Data preprocessing

Data preprocessing is a major and essential stage whose main goal is to obtain final data sets which can be considered correct and useful for further data mining algorithms (García et al.,2016). This process mainly includes text tokenization, stop word processing, and data cleansing.

Data cleansing is a crucial step in text preprocessing. It involves identifying and removing extraneous characters, duplicate data, erroneous data, and formats that can't be processed from the original text sample, so as to ensure data accuracy and consistency, and improve the effectiveness of the model (Gharatkar et al. 2017).

3.3. Text encoding

This paper intends to conduct statistical analysis by digitally encoding the content of self-media posts and converting them into quantitative indicators. Therefore, the invisible indicators for manual coding of Weibo accounts are nature of the subject, post contents, content relevance, degree of multimedia utilization, entertaining degree and originality. The partial coding table is as follows:

Table 1. The partial coding table

Number	Category	Indicator description	Encoding
1	Total number of posts	Total number of articles published by self-media	The total number of posts:1~50=1; The total number of posts:51~100=2; The total number of posts:101~150=3; The total number of posts:151~200=4; The total number of posts:201~250=5; The total number of posts:251~300=6
2	Annual number of articles	Average annual number of articles published by self-media	Annual number of articles:0~1=1; Annual number of articles:2~3=2; Annual number of articles:4~5=3; Annual number of articles:6~7=4; Annual number of articles:8 and above=5
3	Average number of likes	Average number of likes of self-media articles	Average number of likes:0=1; Average number of likes:1=2; Average number of likes:2=3; Average number of likes:3=4; Average number of likes:4=5;

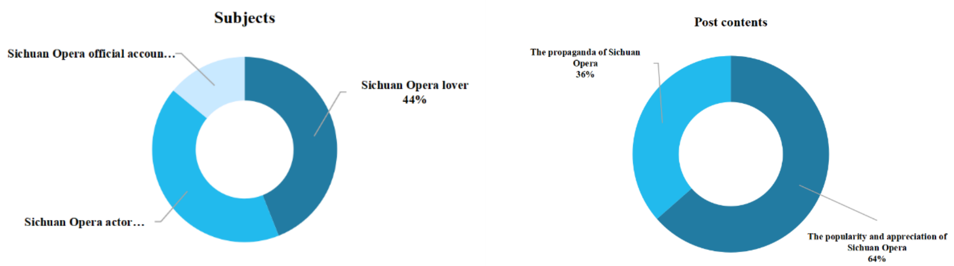
Number	Category	Indicator description	Encoding
			Average number of likes:5=6; Average number of likes:6 and above=7
4	Average number of forwards	Average number of reposts of self-media articles	Average number of forwards:0=1； Average number of forwards:1=2； Average number of forwards:2=3； Average number of forwards:3=4； Average number of forwards:4=5； Average number of forwards:5=6； Average number of forwards:6 and above=7
5	Average number of comments	Average number of comments of self-media articles	Average number of reviews:0=1; Average number of reviews:1=2; Average number of reviews:2=3; Average number of reviews:3=4; Average number of reviews:4=5; Average number of reviews:5=6; Average number of reviews:6 and above=7

3.4. Reliability and validity test of coding results

In statistics and psychology, reliability generally refers to the degree of consistency and stability of the results obtained by a measurement tool when measuring the same object (Ahmed& Ishtiaq,2021). Currently, there are about 39 different indicators of consent. Commonly used in the field of communication is Holstis Coefficient Reliability, Cohens kappa(k), Scotts pi(π)， Krippendorffs alpha(α) (Śmietańska& Podziewski,2019). The Holstis Coefficient Reliability is used for the coding in this article.The Holstis Coefficient Reliability is calculated as: H=0.9236. Because the confidence level is more than 0.9, the confidence level of the encoding is high.

3.5. Analysis of text encoding results

By digitally encoding the content of self-media posts and converting them into quantitative indicators, statistical analysis is carried out. The results are shown in the Figure 1.



(a) (b)

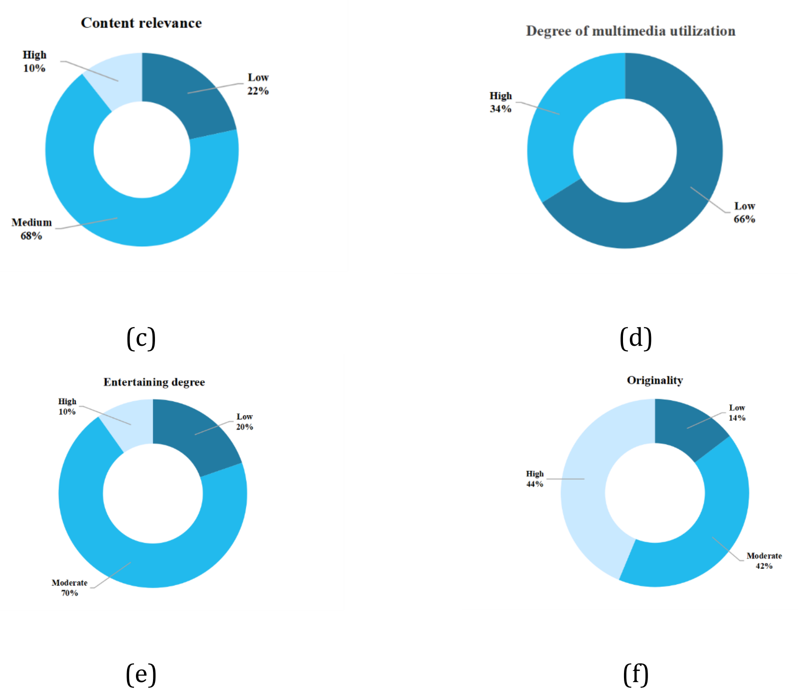


Figure 1. The result of digitized encoding:(a) Subjects, (b) Post contents, (c) Content relevance , (d) Degree of multimedia utilization, (e) Entertaining degree, (f) Originality

4. Results and discussion

4.1. Interpretation of Invisible Metrics

After encoding the text, the project was able to explore its internal logical structure, and then we can clarify the content and quality of the text published by the self-media accounts. The coding table of the self-media platform is as follows:

Table 2. The coding table of the self-media platform

The total number of posts	The number of self-media accounts	Percentage	Annual number of articles	The number of self-media accounts	Percentage
1-50	43	31.4%	0-1	0	0.0%
51-100	20	14.6%	2-3	58	42.3%
101-150	11	8.0%	4-5	34	24.8%
151-200	30	21.9%	6-7	27	19.7%
201-250	21	15.3%	≥8	7	5.1%
251-300	12	8.8%			
Average number of likes	The number of self-media accounts	Percentage	Average number of forwards	The number of self-media accounts	Percentage
0	9	6.6%	0	32	23.4%
1	17	12.4%	1	11	8.0%
2	10	7.3%	2	28	20.4%
3	11	8.0%	3	17	12.4%
4	22	16.1%	4	8	5.8%

5	19	13.9%	5	6	4.4%
Above 6	49	35.8%	6	35	25.5%
Average number of comments	The number of self-media accounts	Percentage	Number of followers	The number of self-media accounts	Percentage
0	34	24.8%	0-2000	17	12.4%
1	21	15.3%	2001-4000	48	35.0%
2	30	21.9%	4001-6000	25	18.2%
3	13	9.5%	6001-8000	29	21.2%
4	7	5.1%	≥8000	18	13.1%
5	24	17.5%			
6	8	5.8%			
Subjects involved	The number of self-media accounts	Percentage	Publish content	The number of self-media accounts	Percentage
Sichuan Opera official account	1467	14.59%	The popularity and appreciation of Sichuan Opera	6390	63.54%
Sichuan Opera actor	4195	41.71%	Publicity and promotion of Sichuan Opera	3667	36.46%
Sichuan Opera lover	4395	43.70%			
Content relevance	The number of self-media accounts	Percentage	Multimedia degree	The number of self-media accounts	Percentage
Low correlation	2176	21.64%	Low multimedia	6648	66.10%
Medium correlation	6812	67.73%	High multimedia	3409	33.90%
High correlation	1069	10.63%			
Degree of interest	The number of self-media accounts	Percentage	Originality	The number of self-media accounts	Percentage
Low level of interest	1983	19.72%	Low level of originality	1471	14.63%
Moderate level of interest	7087	70.47%	Moderate level of originality	4187	41.63%
High level of interest	987	9.81%	High level of originality	4399	43.74%

Based on the analysis of Table 2, three preliminary findings are observed:

1) Most self-media accounts show insufficient activity, primarily because their high-quality and highly resonant content encourages users to prefer liking and forwarding over commenting, and potential platform restrictions on commenting may also play a role;

2) Official accounts and Sichuan Opera actor accounts attract more followers, benefiting from their authoritative status or the cultural specificity that enhances their influence;

3) The content posted by most self-media accounts is generally weak and urgently needs improvement, particularly in terms of relevance, multimedia utilization, entertainment value, and originality.

4.2. Sentiment Analysis

Sentiment analysis on social networks mines users' opinions, emotions, and attitudes to derive useful insights into community opinions (Chen et al.,2022). In general, sentiments are classified into three categories: positive, neutral and negative (Ray&Chakrabarti,2017). The sentiment analysis table of Sichuan opera texts is as follows:

Table 3. Sentiment analysis table of Sichuan opera texts

Type	Number of records	Percentage
Positive emotions	1881	18.71%
Neutral emotions	6993	69.53%
Negative emotions	1183	11.76%
The positive sentiment segment statistics are as follows：		
General positive emotions（0—10）	1312	13.05%
Moderately positive emotions（10—20）	535	5.32%
High positive emotions（20 and above）	34	0.34%
The results of the negative sentiment segment are as follows：		
General negativity（-10—0）	949	9.44%
Moderate negativity（-20—-10）	214	2.15%
High negativity（-20 and below）	0	0.00%

Comprehensive analysis shows that the emotional tendency of Sichuan opera online texts is mainly neutral, while the positive and negative emotions are relatively less. This may reflect the fact that in online texts, the discussion or description of Sichuan Opera presents more objective and neutral attitudes, while the emotional color is bland and presents a "not too concerned" attitude.

4.3. Social network analysis (SNA)

Combined with the results of sentiment analysis of Sichuan Opera online texts, this paper further explores the social network analysis of Sichuan Opera online texts. In the social network analysis of Sichuan Opera online texts, this paper will focus on the following two points. One is the nodes. The nodes of the network represent individuals. By analyzing the number, importance and connection of nodes, we can understand the influence and status of words in Sichuan opera online texts. The other one is

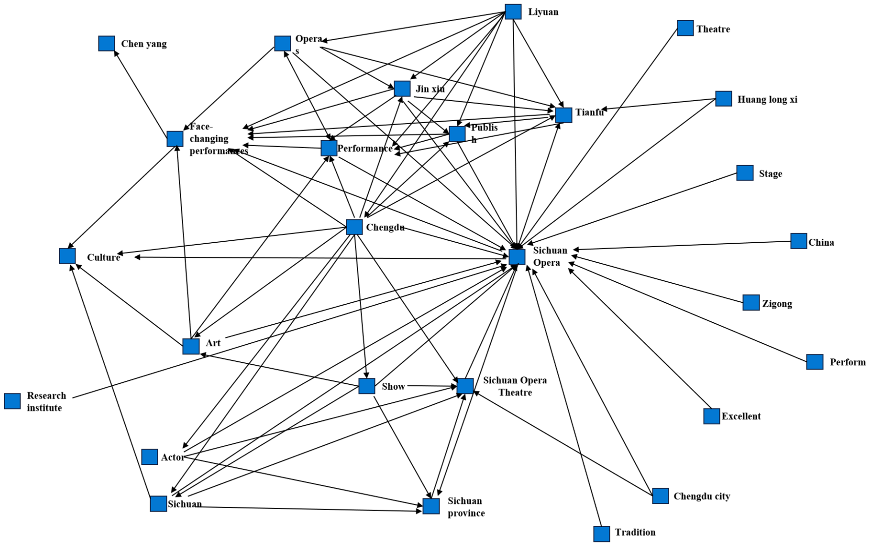


Figure 2. Analysis of the social network of Sichuan Opera texts

The edges. Edges represent relationships between individuals. By analyzing the density, intensity, and direction of edges, we can reveal the communication patterns and information transmission paths between different words in Sichuan Opera online texts. Figure 2 shows the results of text analysis on Sichuan Opera networks.

4.4. Latent Dirichlet Allocation (LDA)

In the analysis of Sichuan Opera online texts, the LDA can discover the topics. Using the LDA topic model, the potential themes or topics in these texts can be automatically discovered, helping people understand the internal structure and association of Sichuan Opera online texts.

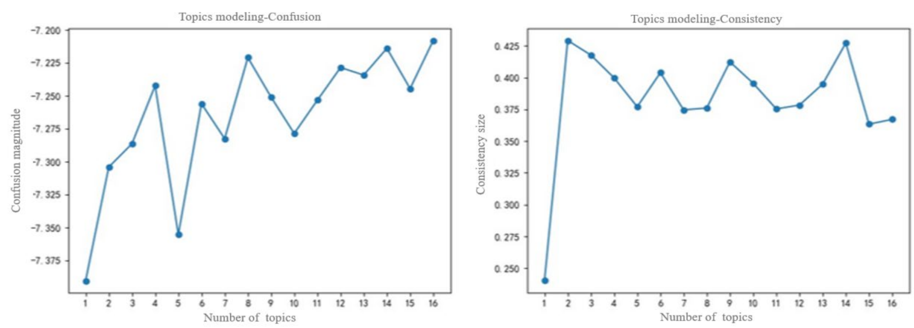


Figure 3. LDA theme clustering of Sichuan opera texts

Due to the increasing confusion output of the LDA model of the texts, the number of topics selected in this part is selected consistently. By calculating the puzzle of the topics from 1 to 16, the final number of topics is 5 for clustering (Figure 3).

In this study, LDA topic modeling is carried out on the texts data according to the sklearn in Python, and the visualization results of topic clustering of comments are obtained based on the pyLDAvis, as shown in the Figure 4. And based on the above results, the topics are summarized as shown in the Table 4.

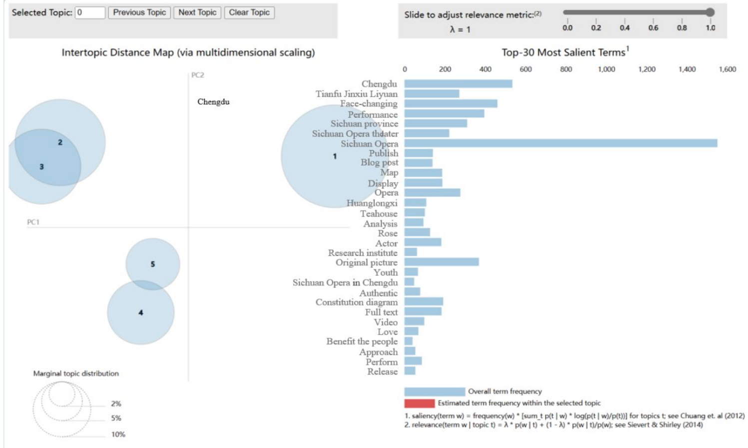


Figure 4. LDA visualization results

Table 4. LDA clustering text table

Topics	The name of the cluster	Words
Topic 1	Sichuan opera inheritance	0.118*Sichuan Opera +0.055*Chengdu +0.044*Face-changing performances +0.04*Performance +0.029*Tianfu Jinxiu Liyuan +0.025*Opera +0.021* Original picture +0.015*Publish +0.015*Blog post +0.015*Art
Topic 2	Sichuan opera publicity	0.022*Map +0.021*Display +0.018* Original picture +0.017*Constitution diagram +0.017*Sichuan Opera +0.011*Art +0.010*Video +0.008*performance +0.008*Face-changing performances +0.007*Love
Topic 3	Sichuan Opera venue	0.038*Sichuan Province +0.033*Sichuan Opera +0.029*Sichuan Opera Theater +0.021*Performance +0.014*Art +0.013*Map + 0.013*Display +0.010*Full text +0.009*Culture +0.009*Perform
Topic 4	Sichuan Opera Studies	0.023*Sichuan Opera +0.020*Actor +0.014*Art +0.013*Full text +0.011* Research institute +0.011*Youth + 0.010* Sichuan Opera in Chengdu +0.008*Sichuan province +0.007*Country +0.007*China
Topic 5	The development of Sichuan opera	0.027*Sichuan Opera +0.015* Original picture +0.011*Face-changing performances +0.009*Full text +0.008*Sichuan Province +0.008*Constitution diagram + 0.007*Approach +0.006*Art +0.006*Opera+ 0.005*Figure

4.5. Recommendations

In light of these findings, the following strategies are proposed to improve platform engagement, account management, and content quality:

1) Enhancing Platform Engagement:

a) Improve content quality: Ensure that posted content is of high quality and aligned with audience interests. This will boost user engagement and increase the overall volume of content.

b) Promote user interaction: Create discussion groups and host online events to stimulate communication among users and further enhance engagement.

c) Utilize data analytics: Analyze user data to understand preferences and behavior patterns. This information can be used to optimize platform features and content recommendation strategies.

2) Strategies for Account Types

a) Official accounts: Strengthen interactions with users by providing valuable content and services, thereby building a strong brand image and gaining greater user trust and attention.

b) Sichuan Opera performers: Leverage personal influence to share more engaging and in-depth content. Actively engage with fans to expand influence and popularity on social media platforms.

3) Content Improvement Strategies

a) Increase topic relevance: Ensure that content aligns with the interests and needs of the target audience. By researching audience preferences and characteristics, content creators can select topics that resonate more strongly with their followers.

b) Enhance multimedia usage: Present content using multiple media formats (images, videos, audio, etc.). Multimedia content is more likely to attract user attention and improve the overall user experience.

c)Foster audience interaction: Actively engage with followers by gathering feedback and suggestions. Understanding user needs allows content creators to adjust their strategies and continuously improve content quality.

5. Conclusion and Future work

In this study, Weibo was employed as a data collection platform, and the collected text data were preprocessed. Based on principles of qualitative analysis and natural language processing, the text was encoded, clustered, and mined. By combining qualitative and quantitative approaches, big data mining and analysis were conducted on key aspects of Sichuan Opera dissemination. From this analysis, the following conclusions can be drawn:

a) The activity levels of most self-media accounts remain relatively low, as evidenced by a small total number of posts and comments. However, the high number of likes and reposts suggests that the user base is engaged and exhibits a high degree of loyalty.

b) In terms of account characteristics, most self-media accounts have comparable follower counts, and the majority are either official organization accounts or accounts belonging to Sichuan Opera performers.

c) The content published by most self-media accounts requires significant improvement. Specifically, much of this content exhibits low relevance to audience interests, limited use of multimedia elements, minimal engagement appeal, and a lack of originality.

Acknowledgements

This work was supported by project grants from Key Laboratory of Digital Protection and Intelligent Sharing of Traditional Local Opera Resources (Chengdu University of Technology), Sichuan Provincial Department of Culture and Tourism (Grant No.24XQYJS06).

References

[1] McDermott, P. (2010). Building open government. *Government Information Quarterly*, *27*(4), 401-413.

[2] Mergel, I. (2012). The social media innovation challenge in the public sector. *Information polity*, *17*(3-4), 281-292.

[3] Nian, F., Li, J., Diao, H., & Yu, X. (2022). Weibo core user mining and propagation scale predicting. *Chaos, Solitons & Fractals*, *156*, 111869.

[4] Zaefarian, G., Misra, S., Koval, M., & Iurkov, V. (2022). Social network analysis in marketing: A step-by-step guide for researchers. *Industrial Marketing Management*, *107*, A11-A24.

[5] Verdú, M. J., De Castro, J. P., Regueras, L. M., & Corell, A. (2021). MSocial: practical integration of social learning analytics into Moodle. *IEEE Access*, *9*, 23705-23716.

[6] Singh, S. S., Muhuri, S., Mishra, S., Srivastava, D., Shakya, H. K., & Kumar, N. (2024). Social Network Analysis: A Survey on Process, Tools, and Application. *ACM Computing Surveys*, *56*(8), 1-39.

[7] Alwash, N., & Levine, J. D. (2019). Network analyses reveal structure in insect social groups. *Current opinion in insect science*, *35*, 54-59.

[8] Zhang, H., Fiszman, M., Shin, D., Miller, C. M., Rosemblat, G., & Rindflesch, T. C. (2011). Degree centrality for semantic abstraction summarization of therapeutic studies. *Journal of biomedical informatics*, *44*(5), 830-838.

[9] Zhang, J., & Luo, Y. (2017, March). Degree centrality, betweenness centrality, and closeness centrality in social network. In *2017 2nd international conference on modelling, simulation and applied mathematics (MSAM2017)* (pp. 300-303). Atlantis press.

[10] Zhao, F., Ren, X., Yang, S., Han, Q., Zhao, P., & Yang, X. (2020). Latent dirichlet allocation model training with differential privacy. *IEEE Transactions on Information Forensics and Security*, *16*, 1290-1305.

[11] Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia tools and applications*, *78*, 15169-15211.

[12] Tong, Z., & Zhang, H. (2016, May). A text mining research based on LDA topic modelling. In *International conference on computer science, engineering and information technology* (pp. 201-210).

[13] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993-1022.

[14] Xie, R., Chu, S. K. W., Chiu, D. K. W., & Wang, Y. (2021). Exploring public response to COVID-19 on Weibo with LDA topic modeling and sentiment analysis. *Data and Information Management*, *5*(1), 86-99.

[15] Liu, X., & Hu, W. (2019). Attention and sentiment of Chinese public toward green buildings based on Sina Weibo. *Sustainable cities and society*, *44*, 550-558.

[16] García, S., Ramírez-Gallego, S., Luengo, J., Benítez, J. M., & Herrera, F. (2016). Big data

preprocessing: methods and prospects. *Big data analytics*, *1*, 1-22.

[17] Gharatkar, S., Ingle, A., Naik, T., & Save, A. (2017, March). Review preprocessing using data cleaning and stemming technique. In *2017 international conference on innovations in information, embedded and communication systems (iciiecs)* (pp. 1-4). IEEE.

[18] Ahmed, I., & Ishtiaq, S. (2021). Reliability and Validity: Importance in medical research. *Methods*, *12*(1), 2401-2406.

[19] Śmietańska, K., & Podziewski, P. (2019). The use of the Krippendorff's coefficient in determining intra-rater reliability in human visual quality control of furniture manufacturing processes. *Annals of Warsaw University of Life Sciences SGGW Forestry and Wood Technology*, *108*, 97-103.

[20] Chen, J., Song, N., Su, Y., Zhao, S., & Zhang, Y. (2022). Learning user sentiment orientation in social networks for sentiment analysis. *Information Sciences*, *616*, 526-538.

[21] Ray, P., & Chakrabarti, A. (2017, February). Twitter sentiment analysis for product review using lexicon method. In *2017 International Conference on Data Management, Analytics and Innovation (ICDMAI)* (pp. 211-216). IEEE.