Postprint: Design of a Mainstream Media Short Video Data Analysis and Visualization Platform

Authors: Zhang Fan, Liu Wei, Meng Jing

Date: 2025-07-09T00:00:00+00:00

Abstract

【Objective】 This paper introduces the development of a short video data analysis and visualization platform for mainstream media at Hebei Daily Newspaper Group, aiming to explore and analyze the dissemination mechanisms and evaluation systems of short videos within mainstream media contexts, thereby promoting the digital transformation of traditional media. By investigating the dissemination patterns of short videos, it provides media organizations with a foundation for dissemination strategies and content optimization. 【Methods】 This study employs technical methodologies including data analysis, data mining, and data visualization, aligned with the platform' s practical requirements. Through the collection and processing of massive short video datasets, it analyzes the dissemination effectiveness and influence of short video content, while exploring the underlying patterns and trends in short video dissemination. 【Results / Conclusion】 The study successfully developed a data analysis system that integrates short video monitoring, dissemination power prediction, and daily ranking display, enabling comprehensive surveillance of trending short videos and high-impact accounts. Through data modeling and algorithmic analysis, the system establishes scientific evaluation criteria for short videos and demonstrates predictive capabilities for dissemination power. The daily ranking provides creators with precise feedback, significantly enhancing management and operational efficiency. This system furnishes mainstream media with more intelligent and accurate support for short video content, thereby fostering innovation and development in the new media era.

Full Text

Design of a Data Analysis and Visualization Platform for Mainstream Media Short Videos

Authors: Zhang Fan, Liu Wei, Meng Jing
Affiliation: Hebei Daily Newspaper Group, Shijiazhuang, Hebei 050022

Abstract

This paper introduces the construction of a data analysis and visualization platform for mainstream media short videos developed by Hebei Daily Newspaper Group. The platform aims to explore and analyze the dissemination mechanisms and evaluation systems of short videos within mainstream media, thereby facilitating the digital transformation of traditional media. By investigating the propagation patterns of short videos, it provides media organizations with evidence for communication strategies and content optimization. The system employs technical approaches including data analysis, data mining, and data visualization, tailored to the platform' s practical requirements. Through the collection and processing of massive short video datasets, it analyzes the dissemination effectiveness and influence of short video content, uncovering underlying patterns and trends. The successful development of this integrated data analysis system, which combines short video monitoring, propagation prediction, and daily ranking display, enables comprehensive tracking of trending short videos and high-impact accounts. Through data modeling and algorithmic analysis, the platform establishes scientific evaluation criteria for short videos and possesses certain predictive capabilities for propagation potential. The daily ranking feature provides creators with precise feedback, significantly enhancing management and operational efficiency. This system offers more intelligent and accurate support for short video content management in mainstream media, driving innovation and development in the new media era.

Keywords: short videos; dissemination mechanism; evaluation system; data analysis; data visualization
Classification Code: G223
Document Code: A
Article ID: 1671-0134(2025)04-69-05
DOI: 10.19483/j.cnki.11-4653/n.2025.04.013
Citation Format: Zhang Fan, Liu Wei, Meng Jing. Design of a Data Analysis and Visualization Platform for Mainstream Media Short Videos [J]. China Media Technology, 2025, 32(4): 69-73.

Introduction

Mainstream media outlets such as China Media Group have actively promoted short video innovation across various domains. As President Xi Jinping emphasized: "Wherever the people are, that is where the focus of propaganda and ideological work should be. Cyberspace has become a new realm for people' s production and daily life, and thus it should also become a new space for our Party to build consensus" [1]. The Decision of the Central Committee of the Communist Party of China on Further Deepening Reform Comprehensively to Advance Chinese Modernization proposes to "establish a work mechanism and evaluation system adapted to all-media production and dissemination, and promote systematic transformation of mainstream media." It also stresses the need to "explore effective mechanisms for integrating culture and technology,"

identifying technological integration as an inevitable path for media industry development and providing new directions for innovation while highlighting the strategic value of culture-technology convergence.

This paper introduces the design and application of a data analysis and visualization platform for mainstream media short videos developed by Hebei Daily Newspaper Group. The core objective is to enhance the dissemination efficiency and management capabilities of short video content through technologies such as data analysis, data mining, and visualization. Specifically, the system comprises four primary subsystems: a data capture and storage subsystem, a short video monitoring and evaluation system, a short video propagation prediction system, and an in-house media daily ranking system. The data capture and storage subsystem is responsible for real-time crawling and storage of short video platform data, providing the foundation for subsequent analysis. The monitoring and evaluation system focuses on displaying the dissemination patterns of trending short videos across platforms. The propagation prediction system employs techniques such as negative quadratic regression to forecast short video dissemination trends. The daily ranking system generates daily rankings based on content performance, helping creators and decision-makers understand real-time popularity and user feedback. This platform provides more intelligent and precise support for short video content management and dissemination in mainstream media, fostering innovation and development in the new media era [2-4].

Funded Project: 2024 Shijiazhuang Science and Technology Plan Project "Multi-Modal Based Mainstream Media Short Video Propagation Data Analysis and Visualization System" (Project No.: 2457906501).

1. System Design

1.1 System Architecture Design

This system decomposes the short video data analysis and decision-making process into multiple interdependent yet interconnected subsystems. The architecture adopts a modular design, with each subsystem sharing data and collaborating through interfaces to achieve efficient information flow and processing [5-8]. The system architecture can be divided into four primary layers: data collection, data processing, analysis and prediction, and decision support.

The **data collection layer** is responsible for acquiring raw data from short video platforms. This layer encompasses web scraping technologies, API interface calls, data cleansing, and storage functions to ensure data timeliness and accuracy.

The **data processing layer** handles, stores, and manages the collected data through big data technologies. It utilizes database management systems to categorize and store data, providing support for subsequent analysis and processing.

The **analysis and prediction layer** employs data mining and machine learning

algorithms for in-depth data analysis, generating visual charts and propagation trend forecasts.

The **decision support layer** transforms analytical results into actionable decision-making information and visualization dashboards, providing data support for short video content producers and managers to help them formulate more effective strategies.

1.3 System Platform and Integration

The system employs a microservices architecture, integrated with Dubbo and Spring Boot to enable modular independent deployment. Each microservice focuses on specific business domains and interacts through RESTful APIs. Frontend data is transmitted to the backend via APIs, where it is processed and stored in MySQL. Meanwhile, Spark handles large-scale data computation, storing results in Hive, while Apache Kafka enhances data real-time processing capabilities. The frontend utilizes Vue.js combined with ECharts for visualization and Axios for asynchronous communication. The backend is built on Spring Boot and Dubbo, with Spring Security managing authentication, MyBatis handling data persistence, and Shiro overseeing permission management. Service scheduling and management are handled by Dubbo to ensure high availability. For deployment, the system uses Docker containerization and relies on Kubernetes for cluster management and automated scaling. Frontend-backend communication employs SSL/TLS encryption, with sensitive information such as passwords stored in encrypted form. The overall architecture balances efficiency, security, and scalability, supporting high concurrency and big data processing. Additionally, a comprehensive logging and monitoring system ensures stable platform operation and facilitates timely identification and resolution of potential issues.

1.4 Key Technologies

1.4.1 Frontend Technology Frontend technology is crucial for system visualization and user experience. In short video data analysis, presenting clear and intuitive trends, rankings, and user feedback is key to system success. Both Vue.js and React are widely adopted frameworks in modern frontend development, suitable for building complex, dynamic user interfaces. For short video data visualization systems, selecting these frameworks enables more efficient construction of highly interactive and responsive frontends. ECharts is a powerful open-source visualization library supporting various chart types including line charts, bar charts, and heatmaps. In short video data analysis, it facilitates the implementation of video propagation trend graphs, daily ranking displays, and user interaction analysis [9-10].

1.4.2 Data Processing Technology Hive is a data warehouse infrastructure built on Hadoop that primarily provides a SQL-like query language, HiveQL, for operating large-scale datasets. Its design goal is to simplify big data processing,

enabling non-programmers to query data using SQL-like syntax. Apache Spark is a fast and general-purpose big data processing engine supporting both batch and stream processing. Unlike Hadoop' s MapReduce, Spark loads data into memory for computation, thereby reducing disk I/O and typically achieving faster processing speeds than Hadoop [11-12].

1.4.3 Negative Quadratic Regression In regression analysis, a model containing a quadratic term with a negative coefficient is referred to as negative quadratic regression. This approach is commonly used to describe phenomena with parabolic trends and is suitable for data exhibiting inverted U-shaped or normal distributions. It finds extensive application in data analysis for short video platforms and social media, particularly for predicting short video propagation potential and trends. Applied to short video propagation prediction, negative quadratic regression can effectively forecast video dissemination patterns [13-16].

1.4.4 OpenCV OpenCV is an open-source computer vision library widely used in image and video processing, machine learning, and related fields. It provides a robust set of tools for image processing, object detection, feature matching, and deep learning integration. In this system, it is primarily utilized to automatically extract visual features from short videos [17-18].

2. Subsystem Functions

2.1 Data Capture and Monitoring Subsystem

The data capture and monitoring subsystem serves as the foundation of the entire system, responsible for real-time crawling and storage of data from major short video platforms and social media. Its objective is to ensure timely and accurate data acquisition. The captured data includes video titles, URLs, cover images, publishing accounts, release timestamps, view counts, likes, comments, shares, and other metrics. Following data capture, the system standardizes data from different platforms. Since some platforms adjust their data to varying degrees, a data monitoring and alerting mechanism is essential to ensure data timeliness and accuracy. By monitoring crawling frequency, success rates, and platform response statuses, the system triggers alerts to notify administrators when anomalies occur, such as crawling failures or platform interface updates.

2.2 Short Video Monitoring and Evaluation System

The short video monitoring and evaluation system is a comprehensive analytical tool designed to track and assess the dissemination effectiveness and influence of trending short video content and impactful accounts published by mainstream media. It comprises four primary functional modules: a propagation heat ranking, comprehensive account propagation analysis, media influence ranking, and cross-platform content performance analysis.

The **propagation heat ranking** calculates a composite index for each short video by configuring weights for interaction metrics such as likes, shares, and comments through the backend. This index is used to display the top 50 works across platforms, with the ranking refreshed every 10 minutes to reflect the latest trends and popular content. Works with significant changes in their composite index receive special annotations to facilitate timely tracking of their dynamic variations. Additionally, the system invokes sentiment analysis and text semantic matching interfaces to deeply analyze the emotional tone of videos. Works with strong emotional expression are flagged to help users identify content likely to evoke intense emotional resonance. For identical works appearing across multiple platforms, annotations differentiate them, alerting users to monitor each platform' s performance and audience response.

The **comprehensive account propagation analysis** module displays trends in the number of works, follower counts, and average composite indices for short video platform accounts within custom time periods. The average composite index is calculated as the sum of all composite indices for works published by an account in a day divided by the number of works published that day. These trends are presented in chart form, enabling users to easily identify outstanding accounts within specific periods and promptly detect fluctuations in their propagation power.

The **media influence ranking** provides users with a more holistic perspective, showcasing the comprehensive performance of a single media organization across multiple platforms. Specifically, the ranking incorporates key indicators including the number of works published within a defined period, follower counts, and the comprehensive influence index of those works. The comprehensive influence index is calculated as the sum of all works' composite indices divided by the total number of works, representing the media' s aggregated influence across platforms. This module helps users understand a media organization' s overall performance across platforms, offering comparative cross-platform data analysis for accurate assessment of dissemination effectiveness across different channels.

The **cross-platform content performance analysis** module helps users deeply analyze dissemination effectiveness differences between platforms by comparing the performance of identical content types across different platforms. Through charts and visualized data, it clearly presents audience interaction patterns and preferences on each platform. This module not only helps media organizations identify which platforms are best suited for different content types but also provides profound insights into cross-platform content propagation differences. For creators, this functionality offers decision-making support for developing cross-platform content strategies, enabling them to precisely reach target audiences.

2.3 Short Video Propagation Prediction System

Propagation prediction systems typically employ machine learning techniques including regression analysis, time series forecasting, and deep learning to construct various predictive models. Regression analysis models predict future video propagation effectiveness by analyzing relationships between historical video data and dissemination volumes. Considering the cost, quantity of historical samples, and characteristics of short videos, this system selects negative quadratic regression to build its prediction model. Through modeling historical samples, short video instances are extracted across 35 dimensions including visual features, auditory features, and publishing characteristics. Regression analysis ultimately identifies 35 indicator items for model construction: contrast, clarity, sharpness, dominant colors, dominant color proportion, music genre, presence of human voice, presence of dialect, content characteristics, regional characteristics, production characteristics, emotional characteristics, whether published on a workday, publication time slot grouping, video duration, video duration grouping, account follower count, presence of celebrities, whether an anchor appears on screen, among others. Based on the regression coefficients of these extracted indicators, the system configures relevant parameters for its current prediction model.

When users upload a video for analysis and complete the basic classification information, interfaces for visual, auditory, and publishing feature analysis are triggered to extract short video data features. These features are then input into the regression model to generate propagation predictions and provide optimization recommendations across the 35 extended indicator dimensions.

2.4 In-House Media Daily Ranking System

The in-house media daily ranking system is a data analysis tool specifically designed for creators to help them understand the performance of their short video content. Due to authorization privileges, the system can access far more analytical data than frontend metrics such as likes, shares, and comments. With support from short video platform data interfaces, the system can also retrieve multi-dimensional backend data including completion rate, 5-second completion rate, cover click-through rate, 2-second bounce rate, average playback duration, homepage visits, follower growth, and shares via chat and social circles. This functionality enables creators to clearly visualize each video' s interaction patterns on the platform. Additionally, the system displays information about outstanding creators within the media organization, incentivizing continuous improvement. Decision-makers can also leverage this system' s analysis to quickly grasp current content publishing situations, enabling more scientific content decisions that drive creator growth and platform content optimization.

3. Key Considerations in Platform Construction

3.1 Data Collection Timeliness and Standardization

In the current short video ecosystem, content updates and dissemination occur at extremely rapid speeds, with platform interfaces and presentation formats continuously evolving. Therefore, only through real-time monitoring, adjusting data ingestion methods according to interface changes, and promptly collecting dynamic data from all platforms can we accurately grasp content popularity fluctuations and trends. To maintain both data integrity and platform-specific characteristics, we must establish robust and consistent data ingestion rules during collection, particularly concerning the standardization of comparative metrics. The core objective of data standardization is to transform the data structures and presentation formats of various platforms into a unified standard format to facilitate subsequent analysis and modeling. This not only reduces errors arising from disparate data structures but also enhances the efficiency and accuracy of data analysis.

3.2 Prediction Model Timeliness

During prediction model construction, most features can be efficiently extracted through automated technologies such as computer vision and sentiment analysis. However, for more complex dimensions involving video content classification, narration style analysis, and shot transition patterns, manual annotation remains necessary to ensure data accuracy. Consequently, prediction model parameters require regular updates to adapt to the continuously evolving trends in short video content. As technology advances and algorithms optimize, achieving comprehensive automation of feature extraction becomes key to model upgrading and optimization. At that point, ingested data can directly drive real-time model updates, thereby improving prediction precision and adaptability, enabling more accurate reflection of the dynamic changes in short video propagation power.

3.3 System High Availability and Scalability

The volume of ingested short video data is enormous, particularly during traffic peaks or when hot events occur, requiring the system to rapidly process multiple video streams concurrently. Distributed architecture and load balancing can enhance the system' s concurrent processing capabilities, ensuring efficient response even during data surges. As short video platforms evolve, new functional requirements and data sources may continuously emerge. System flexibility and scalability can support future demand changes, preventing architectural limitations during business expansion. The design incorporates a modular architecture to support subsequent functional extensions. Through loose coupling design, the system can quickly integrate new data sources and functionalities without affecting existing operations.

3.4 Multi-Dimensional and Intuitive Analysis Presentation

For data visualization systems, providing user-selectable, clear, and intuitive analytical results is a core requirement. Short video dissemination effectiveness is multi-dimensional, and presenting data through various formats helps users analyze video performance from multiple perspectives, enabling more precise decision-making. Multi-dimensional displays enrich analytical results and enhance decision support accuracy. Visualization tools such as charts, heatmaps, and scatter plots can be used to present different data dimensions. Clear interface design not only improves user operational efficiency but also reduces information overload, allowing users to focus on the most critical data. Interface design should remain concise, avoiding excessive redundant information. Through data charts, pie charts, bar charts, and other visualizations, complex data is presented in ways that enable users to quickly identify underlying trends and patterns.

Conclusion

The platform has successfully designed and developed an integrated data analysis and visualization system that combines short video monitoring, propagation prediction, and in-house media daily ranking display. It enables comprehensive tracking of trending short videos and high-impact accounts in mainstream media. By integrating data modeling and algorithmic analysis, the system establishes a scientific and rational evaluation standard for short videos and possesses certain predictive capabilities for propagation potential. Furthermore, the platform' s daily ranking feature provides creators with rapid and precise feedback, significantly enhancing short video management and operational efficiency. The system offers more intelligent and accurate support for short video content management and dissemination in mainstream media, driving innovation and development in the new media era.

References

[1] Xi Jinping. Accelerating Media Convergence Development and Building an All-Media Communication Landscape [J]. Qiushi, 2019(6): 4-8.
[2] Zhao Bing. Enhancing the "Presence Capability" of Mainstream Media on Mobile Internet with Internet Thinking [J]. Chinese Journalist, 2024(2): 12-15.
[3] Zhao Bing. Grasping Communication Laws in the All-Media Era and Striving to Improve Positive Publicity Quality and Efficiency [J]. News Front, 2024(6): 4-8.
[4] Zhang Yuhao. Practice and Reflection on Driving Communication Power Enhancement Through All-Media Production and Dissemination Mechanisms—A Case Study of Hebei Daily Newspaper Group' s Communication Power Dashboard [J]. Interview, Writing, and Editing, 2025(1): 9-11.
[5] Li Jinling, Yuan Xin, Yang Biao. Design and Implementation of an Infectious Disease Data Visualization Platform Based on Web Technology [J]. Computer

Applications and Software, 2023(10): 101-106, 173.
[6] Zhang Fan, Du Yaru, Zhang Xindong, et al. Design of an Integrated Business and Finance Platform in a Media Convergence Environment [J]. China Media Technology, 2024(4): 84-88.
[7] Shen Enya. Big Data Visualization Technology and Applications [J]. Science and Technology Review, 2020(3): 68-83.
[8] Li Xiangming. Research on Data Visualization for Finance Companies Under Treasury System Construction [J]. Finance and Accounting, 2023(15): 64-66.
[9] Zhao Jun. Implementation of Epidemic Statistics Charts Based on Vue.js [J]. Computer Programming Skills and Maintenance, 2020(3): 144-147.
[10] Li Ping, Li Yong, Fan Quanrun. Design and Implementation of a University Teaching Status Data Visualization and Analysis Platform [J]. Experimental Technology and Management, 2020(5): 46-51, 10.
[11] Li Junli. Parallelization of Mutual Information Calculation for Categorical Data on the Spark Platform [J]. Computer Engineering and Applications, 2021(7): 95-100.
[12] Wang Yanyan, Wang Yanning, Liu Jiaxin, Ren Jiadong. Research on Big Data Applications for Port Logistics Based on Hadoop [J]. Journal of Yanshan University, 2023(3): 216-220, 228.
[13] Zhu Maoran, Ma Xiaoyi, Gao Song, et al. The Impact of Emotional Divergence on Information Re-dissemination in Social Media—A Case Study of Weibo [J]. Journal of Intelligence, 2024(5): 143-151.
[14] Peng Yanni. Research on Key Technologies for Temporal Event Sequence Data Visualization [D]. Changsha: Central South University, 2022.
[15] Huang Yangkun, Chen Changfeng. Automation of Visual Communication and Aesthetic Construction of National Image—A Computational Aesthetic Investigation Based on Twitter Social Bots [J]. Modern Communication—Journal of Communication University of China, 2023(8): 96-104.
[16] Sun Zhenghui, Zheng Jianping, Wang Youwei. Research on the Impact of Short Video Visual, Auditory, and Content Features on E-commerce Marketing Effectiveness [J]. Journal of Marketing Science, 2023(4): 1-21.
[17] Li Xianfeng, Xu Sen, Hua Yiming. Maritime Video Ship Fire and Smoke Detection Technology Based on OpenCV Computer Vision [J]. Ship Science and Technology, 2021(22): 202-204.
[18] Guan Lu, Zhou Baohua. Application of Computer Vision Technology in Journalism and Communication Research [J]. Contemporary Communication, 2022(3): 20-26.

Author Biographies:

Zhang Fan (1991—), male, from Meixian County, Shaanxi Province, holds a master' s degree and serves as a Senior Engineer in the Technology Support and R&D Department of Hebei Daily Newspaper Group. His research focuses on media convergence, data analysis, and informatization.

Liu Wei (1980—), male, from Ningjin County, Shandong Province, holds a bachelor' s degree and serves as a Chief Editor at Hebei Daily Newspaper Group.

His research focuses on media convergence.

Meng Jing (1993–), female, from Xingtai City, Hebei Province, holds a master' s degree and serves as an Assistant Editor in the Chief Editor' s Office of Hebei Daily Newspaper Group. Her research focuses on media convergence, data analysis, and informatization.

(Copy Editor: Li Yansong)

Note: Figure translations are in progress. See original paper for figures.