Innovative Applications of Big Data-Based Media Communication Analysis and Influence Evaluation: Postprint

Authors: Chen Jun, Chen Xinyi, Su Yu

Date: 2023-10-08T00:00:00+00:00

Abstract

How to trace the global real-time adoption and dissemination status of news reports in media convergence and big data environments, effectively evaluate their comprehensive influence, and assist editorial decision-making constitutes a new challenge confronting the development of media convergence. Through innovations in cross-media big data fusion technology, intelligent content comparison technology, cross-platform dissemination chain analysis technology, and media dissemination influence evaluation technology, realizing real-time global adoption tracking, dissemination analysis, and comprehensive influence assessment for converged media reports, and constructing an application system for media dissemination analysis and influence evaluation, will serve to guide editorial decision-making in the news media industry, enhance the dissemination capabilities of converged reporting and external reporting, and elevate overall media influence.

Full Text

Innovative Applications of Big Data-Based Media Communication Analysis and Influence Assessment

Abstract: In the converged media and big data environment, tracing the real-time global adoption and dissemination of news reports, effectively evaluating their comprehensive influence, and supporting editorial decision-making have become new challenges for media convergence development. Through innovations in cross-media big data fusion technology, intelligent content comparison technology, cross-platform communication link analysis technology, and media communication influence assessment technology, we can achieve real-time global adoption monitoring, communication analysis, and comprehensive influence evaluation for converged media reports. Constructing a media communication analy-

sis and influence assessment application system will help guide editorial decision-making in the news media industry, enhance converged reporting and foreign reporting capabilities, and elevate media influence.

Keywords: Big Data; Adoption Analysis; Media Communication Analysis; Influence Assessment
CLC Number: G206
Document Code: A
DOI: 10.19483/j.cnki.11-4653/n.2017.10.051
Authors: Chen Jun, Chen Xinyi, Su Yu

1. Challenges and Difficulties

Realizing real-time, automated, and comprehensive news reporting information dissemination analysis and influence assessment based on big data faces numerous technical challenges.

Challenge One: How to utilize distributed cloud computing technology to timely and accurately acquire multi-source, massive, heterogeneous, and dynamically updated media data information from the Internet. Converged media communication analysis requires full-media data, including both traditional media such as electronic newspapers and data from news websites, "two micros and one client" (Weibo, WeChat public accounts, and mobile news clients), and overseas social media platforms. Automatically and real-time monitoring and collecting massive Internet information from various platforms and media forms, parsing, cleaning, and processing millions to tens of millions of heterogeneous raw data daily, presents significant technical challenges in building a multi-source, massive, and dynamic converged media big data platform.

Challenge Two: How to leverage big data analysis and processing technologies to achieve automatic, intelligent, timely, and accurate analysis and calculation of news report adoption. Today's news reports come in rich varieties, including text, images, charts, videos, and multimedia formats in multiple languages. Identifying and accurately locating adoptions from vast and diverse heterogeneous media information requires designing and continuously optimizing adoption determination algorithms for different types of reports to intelligently handle various complex situations, which poses a major challenge to analytical technologies.

Challenge Three: How to employ intelligent content association analysis technology to bridge the platform communication gap between traditional and new media, enabling cross-media platform content communication analysis. In the all-media era, report information spreads across media platforms, requiring intelligent analysis technology to achieve information correlation and analysis to comprehensively grasp the full-media platform communication status of report information.

Challenge Four: How to design scientific communication influence evaluation indicators and assessment systems to quantitatively evaluate communication contributions. How to make media communication influence measurement more scientific, rational, comprehensive, and standardized? How to establish a scientifically reasonable quantitative indicator system to achieve quantitative analysis and evaluation of news report communication influence, thereby achieving comprehensive evaluation of both production "effort" and impact "achievement"?

2. Overall Architecture

A big data-based media communication analysis and influence assessment application system consists of "one platform, one knowledge base, seven technical layers, and six application functions," as shown in Figure 1 [Figure 1: see original paper].

2.1 "One Platform + One Media Knowledge Base"—Media Big Data Platform

The media big data platform is built using an advanced big data framework that integrates multiple types of media data, including traditional media (news websites and electronic newspapers), "two micros and one client" (Weibo, WeChat, and mobile news clients), and overseas social media platforms (Facebook, Twitter, and YouTube). It has accumulated substantial media foundation information to establish a media information database describing global media attributes.

2.2 Seven Technical Layers

The seven layers include: Data Bus Layer, Data Acquisition Layer, Data Access Layer, Data Integration Layer, Data Resource Layer, Business Analysis Layer, and Application Service Layer.

The **Data Bus Layer** implements the communication links for underlying hardware, data resources, technical components, and functional applications across the entire platform. Based on service bus technology, it addresses coordination, sharing, communication, and management among multiple data sources, modules, and applications, enabling service-oriented registration, management, and invocation of data, modules, and applications.

The **Data Acquisition Layer** achieves unified acquisition management of data from various channels, utilizing distributed cloud acquisition technology to ensure timeliness and constructing basic network facilities to guarantee high network availability and high reliability of proxy resources.

The **Data Access Layer** handles the access of news report data, Internet news data, "two micros and one client" data, and social media data.

The **Data Integration Layer** implements extraction, transformation, and fusion of multi-source heterogeneous data. It builds distributed data processing task queues to achieve ETL processes for massive tasks, performs preliminary structural processing of data, conducts precise field extraction and formatting for heterogeneous data, validates structured data legitimacy, filters garbage information, logs illegal information, and submits valid data for storage while parsing content according to source data structures and performing special field conversions.

The **Data Resource Layer** implements fusion storage management of heterogeneous data resources. Processed structured data is loaded onto relational databases, full-text retrieval databases, and distributed file systems within this layer based on different data types.

The **Business Analysis Layer** implements specific data analysis and processing such as content intelligent analysis, adoption analysis, communication analysis, influence assessment algorithms, and data statistics. To efficiently and real-time process massive Internet media data, a high-performance distributed real-time computing environment must be built. The Hadoop+Spark distributed computing framework can maximize the computing capabilities of software and hardware resources. On this foundation, "two analysis engines" are constructed: the Text Intelligent Analysis Algorithm Engine provides core algorithm support in natural language processing, including large-scale text clustering, entity information extraction, syntactic analysis, text classification, automatic summarization, spam information filtering, and text sentiment analysis; the Communication Analysis and Influence Analysis Engine implements business-level model algorithms, including original analysis, manuscript source identification, first-release identification, reprint/citation identification, communication analysis, and influence indicator system construction models.

The **Application Service Layer** implements various functional modules and integrated application services while providing external data services. Adopting a distributed microservice cluster architecture based on the data bus can address issues where business pressure increases, server capacity is difficult to estimate, and small service resources are wasted, thereby improving cluster utilization. Simultaneously, it enhances IT architecture flexibility, quickly responds to business environment changes and internal demands for business process optimization, maximizes reuse of existing IT resources, and avoids redundant construction.

2.3 Six Application Functions

Traditional Media Reporting Analysis Application: Implements global cross-platform real-time adoption data and interactive data monitoring for news reports in print media, online media, WeChat, and clients.

New Media Reporting Analysis Application: Implements real-time monitoring and analysis of Weibo and WeChat accounts, including follower counts,

post volumes, content, and interaction volumes.

Overseas Social Media Reporting Analysis Application: Implements account analysis on overseas mainstream social platforms, content and communication analysis of posts, comparative analysis with other mainstream media accounts, and monitoring of citations by overseas media.

Influence Assessment Indicator System: Implements a cross-channel media communication effect assessment indicator system for news reports, consisting of a network-wide influence index and different channel influence indices. For each communication channel, the indicator system covers three assessment granularities: reading, interaction, and adoption.

Topic Analysis Application: Implements event analysis, reporting analysis, and influence analysis for major thematic reports. It provides in-depth analysis of global media communication effects of major thematic event reports and comparative analysis with peer media reporting, analyzing event development trends, focus 脉络, main viewpoints, media and netizen attention, and public opinion sentiment development trends to achieve thematic reporting influence assessment.

Analysis Reports and Data Visualization: Implements automatic generation of multi-dimensional analysis reports for news reports with rich data visualization presentations.

3. Technological Innovations

3.1 Cross-Media Big Data Acquisition and Fusion Technology

Through technological and application innovation, we have constructed a media communication analysis and influence assessment application system based on big data. This helps guide editorial decision-making in the news media industry, enhances converged and foreign reporting capabilities, and improves media influence. The system achieves full-network cross-platform media monitoring, real-time global adoption, communication analysis, and comprehensive influence assessment for converged media reports, forming a scientific and reasonable news reporting evaluation indicator system. It solves the difficulty of real-time global monitoring and adoption statistics for foreign English reports, and through quantitative analysis, excavates communication characteristics of foreign English manuscripts to enable more targeted and effective editorial topic selection and more precise and effective communication of English manuscripts. The system strengthens monitoring of new media reports, achieves comprehensive and effective grasp of communication effects, and enables more targeted analysis by timely understanding of netizen focus and interaction characteristics, effectively guiding editorial decision-making. It also realizes overseas social media operation, analysis, and decision support functions through monitoring and analysis of overseas social media account information, post content, and interaction data, enabling real-time grasp of overseas social media operations,

tracking of international mainstream media communication hotspots, and rapid response.

The system employs distributed cloud acquisition technology, multi-source media data fusion technology, and media data cloud service technology to build a media big data platform that automatically collects global media websites, electronic newspapers, new media data, and overseas social media data, improving site coverage, data scale, and update speed. It deploys distributed cloud acquisition terminals that can collect global Chinese and English websites through local collection and overseas deployment with data relay. The system integrates self-collected data with multiple third-party data sources, fusing and utilizing multi-party data resources to form a "collection cloud" above multiple data clouds. It covers extensive media data types, enabling real-time dynamic collection of news, electronic newspapers, "two micros and one client," and overseas social media data, and provides basic data cloud services for various application systems through data fusion processing.

3.2 Stream Processing-Based Intelligent Comparison Technology for Article Adoption

Employing leading text semantic analysis technology combined with business rules, we have developed proprietary intelligent comparison technology for article adoption that can accurately locate the adoption of news reports in Chinese and English media. By summarizing business experience and rules and combining them with machine learning models, we continuously optimize and revise article adoption determination parameters to form business-approved article adoption determination rules and threshold settings. The algorithm achieves automatic adoption calculation for text and image reports, identifying not only explicit adoption (with explicit keywords marked) but also implicit adoption (without explicit keywords marked). By introducing the emerging Spark big data stream processing framework, the system achieves minute-level adoption result updates, providing near real-time adoption analysis data and enabling automatic adoption comparison and communication analysis for English manuscripts.

3.3 Cross-Platform Communication Link Analysis Technology

In the converged media communication era, manuscripts often spread across multiple media platforms, making it difficult to associate and discover transmission relationships between different platforms. By developing intelligent content association analysis technology, we have solved the problem of communication link identification between traditional and new media, achieving cross-media adoption landing for single manuscripts. The system accumulates a mainstream media knowledge base to automatically associate multiple publishing channels for the same manuscript and identify cross-platform communication generated by multi-platform publishing. Through text content similarity feature comparison technology, it identifies the same manuscript's adoption across different media platforms. Based on temporal features and content similarity features,

the system automatically identifies and associates cross-platform communication links for single manuscripts, tracing the manuscript's communication process.

3.4 Media Communication Influence Assessment Technology

Based on media big data information, the system forms a comprehensive manuscript influence evaluation model using content communication indicators and audience interaction indicators, providing multi-dimensional influence evaluation results for single manuscripts, circuits, and departments. Unlike existing single-indicator systems that rely solely on follower counts or attention numbers, this system forms a comprehensive influence evaluation indicator system incorporating quantitative evaluation indicators, content communication indicators, and audience interaction indicators. Based on fuzzy comprehensive evaluation methods, it provides quantitative evaluation and calculation models to quantify the influence assessment of manuscripts, circuits, and departments. Combined with cross-media communication link analysis results, the system fuses communication influence indices from multiple media platform dimensions to provide quantitative evaluation of converged and cross-media influence.

References

[?] Yang Weijie, Dai Ruwei, Cui Xia. A Method for Web News Impact Analysis Based on Information Retrieval Technology[J]. Journal of Software, 2009, 20(9): 2397-

[?] Wang Youzhong, Zeng Dajun, Zheng Xiaolong, et al. Internet News Media Analysis Based on Complex Network Theory[J]. Complex Systems and Complexity Science, 2009, 6(3):

[?] Wang Junze, Zeng Runxi, Du Hongtao. Analysis of Online Public Opinion Propagation Trends Based on Web Page Reprinting Relationship Identification[J]. Journal of Intelligence, 2015, 34(1): 144-

(Author Affiliation: Xinhua News Agency Communication Technology Bureau)

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