Business Intelligence-Based Data Analysis and Mining for News Editorial Workflows: A Post-print Analysis

Authors: Chen Xinyi, Kang Jie

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

Abstract

With the advent of the big data era, an increasing number of enterprises are adopting business intelligence technologies to extract valuable insights from data across production, sales, and other operational segments, thereby providing decision-making support for enterprise leadership. This paper applies key business intelligence technologies to news gathering and editing operations, focusing on the entire workflow of news production. Through data analysis and mining on topics including workflow status analysis, news productivity, and dissemination influence, it demonstrates the circulation process of news gathering, editing, and distribution, exhibits the production efficiency of editorial departments and personnel, and reveals the dissemination impact of published pieces. The paper elaborates on the operational principles of business intelligence and presents an in-depth investigation into data analysis and mining methodologies for news gathering and editing business processes.

Full Text

A Brief Analysis of Data Analysis and Mining for News Editorial Business Processes Based on Business Intelligence

Abstract: With the advent of the big data era, an increasing number of enterprises are adopting business intelligence technologies to “mine gold” from data across production, sales, and other operational segments, thereby providing decision-making support for corporate leadership. This paper applies key business intelligence technologies to news editorial operations, focusing on the entire news production workflow. Through data analysis and mining of themes such as workflow status, news productivity, and dissemination impact, we demonstrate the flow of news gathering, editing, and distribution processes; showcase the production efficiency of editorial departments and personnel; and reveal the

communication impact of published content. The paper introduces the working principles of business intelligence and conducts an in-depth study of data analysis and mining methods for news editorial business processes.

Keywords: Business Intelligence; Data Analysis; Data Mining
Classification Code: G210.7
Document Code: A
Article ID: 1671-0134(2017)12-118-03
DOI: 10.19483/j.cnki.11-4653/n.2017.03.030
Authors: Chen Xinyi, Kang Jie

In the big data era, the value of data is receiving increasing attention across all industries. The massive volume of business process data accumulated within enterprises urgently needs to be "mined for gold." Business intelligence serves as a powerful tool that meets this pressing need by transforming vast amounts of data into knowledge, helping to identify business trends from historical data and providing decision-making support for corporate leadership. According to Gartner surveys, business intelligence ranked as the top technology priority for global CIOs in both 2012 and 2013.

Traditional media today faces severe challenges in its transformation, and big data will serve as a powerful weapon for this transition. In the new media era, users require timely, accurate, and personalized news services. Data mining can help traditional media fully leverage their talent advantages, enhance brand competitiveness, and increase user stickiness. Within news editorial operations, substantial process data exists that has traditionally been overlooked in editorial systems. By applying business intelligence technologies to analyze and mine these business process data, we can improve the informatization level of editorial workflow management, and gain insights into reporting progress, personnel efficiency, content distribution, and communication effectiveness.

Business intelligence, also known as Business Intelligence (BI), involves collecting and managing data while providing a series of technologies and methods for analyzing various types of enterprise data. Business intelligence enables corporate leaders to grasp overall operations from a macro perspective, identify potential industry opportunities, and support decision-making.

2. The Significance of Business Intelligence Research

Business intelligence helps enterprises quickly identify problems and prompts managers to address them. Specifically for the news editorial industry, business intelligence addresses the urgent needs of media managers. Through analysis and mining of news dissemination impact, managers can clearly understand the communication effectiveness and influence of news content.

Business intelligence creates value for news and helps traditional media achieve a "user-centric" transformation and upgrading. By analyzing and mining user behavior, media organizations can classify or cluster user groups according to

factors such as gender, age, occupation, and geography, enabling audience segmentation and personalized content recommendations for different users. This helps media better understand their audiences.

3. Business Intelligence Key Technologies

3.1 OLAP

Online Analytical Processing (OLAP) provides a multidimensional data management environment that enables enterprise data analysts to model and analyze business problems from multiple dimensions.

3.3 Data Mining

Data mining involves extracting hidden, unknown, and valuable associations and patterns from large datasets, establishing models for decision-making, and providing functions for risk analysis and prediction.

4. Business Intelligence Architecture

The architecture first aggregates data dispersed across various enterprise systems, including both relational and non-relational data. Through Extract, Transform, Clean, and Load (ETL) processes, the data is ultimately loaded into a data warehouse according to a predefined data model. By analyzing enterprise data requirements, logical and physical models of the enterprise data warehouse are established to organize various types of enterprise data around analytical themes. On top of the data warehouse, multiple software tools are provided for end-user querying and report generation, including OLAP tools, data mining software, and reporting tools.

5.1 Data Sources

Data warehouse data collection requires obtaining data from various business application systems and management information systems, such as manuscript collection systems, editing systems, contribution systems, and OA systems, and storing it in the data warehouse according to unified data standards. This paper categorizes editorial business system data into two major types: static information data and dynamic information data.

Static information data refers to relatively stable information, primarily including static attribute data such as editorial departments, editorial personnel, and distribution channels. Dynamic information data collects continuously changing workflow data in editorial business systems, including collection, editing, approval, revision, distribution, contribution, and feedback processes. How to construct data models for news business data, particularly user behavior data, and analyze the flow of editorial content will be the focus of this paper.

5.2 Analysis Objectives

By collecting manuscript, workflow, personnel, and quality data, we can manage the entire editorial business process and understand reporting progress, personnel efficiency, content distribution, and communication impact. Data analysis and mining are conducted across three themes—workflow status, communication impact, and news productivity—to present the operational status of news production business.

News Productivity Analysis: Analyzes and presents the work efficiency of editorial departments and personnel over a period of time.
Communication Impact Analysis: Analyzes and presents the distribution and communication impact of published content.
Full Workflow Status Analysis: Analyzes the flow of manuscripts through various editorial stages.

5.3 Data Modeling

Data modeling primarily employs dimensional models. A metric is often related to multiple dimensions, and dimensional models express the relationships between data. For example, to understand the distribution of Chinese manuscripts on new media channels in January 2016, this distribution data relates to three dimensions: channel, time, and language. Dimensional modeling reflects data relationships from multiple angles and levels, reorganizing data across multiple dimensions to provide multidimensional views for decision-making.

Dimensional models consist of two types of tables with different properties: fact tables and dimension tables. Fact tables store business performance measurement values, where each fact represents a business metric (e.g., number of published manuscripts). Dimension tables provide perspectives for observing these measurements (facts), such as channel and time. The attribute columns of dimension tables (e.g., language including Chinese, English, French, Italian, Japanese, Arabic, Russian) serve as both constraints for data usage and cutting tools for data analysis. Therefore, the quality and depth of dimension tables directly affect the performance of the entire data warehouse.

Fact tables and dimension tables are typically integrated using star or snowflake schemas, with fact tables at the center surrounded by dimension tables.

[Figure 1: see original paper] Schematic Diagram of Snowflake Schema for News Publication Status

5.4 Key Indicator System

5.4.1 News Productivity In news productivity evaluation, metrics such as productivity rate, per capita manuscript output, and personnel input ratio can be used to assess departmental news productivity. Indicator values can be aggregated by year, quarter, month, bi-weekly, week, or day. Detailed manuscript

quantities can be viewed through dimensions such as media type, news category, contribution type, and manuscript processing workflow.

Table 1 News Productivity Indicators

Indicator	Definition
Total Processing Time	Total time consumed by a department (publishing department, approving department) to process manuscripts within a unit time
Manuscripts Processed per Unit Time	Number of manuscripts processed by a department (collecting/approving department) within a unit time
Time Consumed per Manuscript	Total time consumed divided by number of manuscripts
Number of Reporters	Number of personnel invested by the publishing department in manuscript processing
Number of Editors	Number of personnel invested by the editing department in manuscript processing
Number of Approvers	Number of personnel invested by the approving department in manuscript processing
Per Capita Manuscript Output	Number of manuscripts divided by total department personnel
Personnel Input Ratio	Total personnel invested in manuscript processing divided by total department personnel

5.4.2 Communication Impact In communication impact evaluation, this paper adopts traditional media impact index, online media impact index, domestic media impact index, overseas media impact index, overseas social media impact index, and whole-network impact index as primary indicators. The whole-network impact index is a weighted calculation of the other five indices.

Table 2 Online Media Communication Impact Indicators

Indicator	Definition
Adoption Volume	Adoption volume of a single manuscript on online media within a unit time

Indicator	Definition
Comment Volume	Comment volume of a single manuscript on online media within a unit time
Positive Comment Ratio	Ratio of positive comments to total comments for a single manuscript on online media within a unit time
Negative Comment Ratio	Ratio of negative comments to total comments for a single manuscript on online media within a unit time
Total Audience	Total audience size for a single manuscript on online media within a unit time
Reprint Volume	Reprint volume of a single manuscript on online media within a unit time
Reprint Depth	Reprint depth of a single manuscript on online media within a unit time

5.4.3 Full Workflow Status Analysis Full workflow status analysis can monitor changes in various indicators in real-time. Detailed indicator information can be viewed through dimensions such as manuscript processing stage (collection, storage, creation, new creation, editing, approval, etc.), manuscript media type, and manuscript contribution category.

Table 3 Manuscript Group Full Workflow Status Analysis Indicators

Indicator	Definition
Manuscripts Completed	Number of manuscripts completed from the previous processing node to the current node
Comprehensive Processing Time	Time consumed by a manuscript group from the previous processing node to completion of the current node
Comprehensive Cumulative Time	Cumulative time consumed from manuscript submission to present; i.e., accumulation of comprehensive processing time
Average Comprehensive Processing Time	Comprehensive processing time divided by number of manuscripts
Average Comprehensive Cumulative Time	Comprehensive cumulative time divided by number of manuscripts

Indicator	Definition
Stage Completion Percentage	Each business stage (collection, editing, distribution, contribution, feedback) has standard time quotas; this shows the completion time percentage within each stage
Full Workflow Completion Percentage	Completion time percentage of the manuscript group across the entire business process

5.5 Business Process Data Mining Algorithms

Guided by the "user-centric" philosophy, we fully utilize data mining technologies such as association rules, classification, and clustering to mine daily news business data. This paper employs the following methods to analyze news business user behavior data and full workflow status data.

5.5.1 Association Rules and Sequential Patterns Association rules are used to analyze user data and discover user behavior patterns. Association rules describe relationships between data items, deriving the presence of some items in a transaction based on the presence of others. The Apriori algorithm is a classic algorithm for association rules, originally proposed for market basket analysis to identify which products consumers frequently purchase together. In news operations, association rule mining can identify correlations between editorial staff personal characteristics and manuscripts, recommend relevant manuscripts based on staff interests, and assign reporting tasks to appropriate reporters or editors.

5.5.2 Time Series Analysis Time series analysis records event results at fixed time intervals. Examples in news business systems include the number of manuscripts processed during fixed daily periods, monthly manuscript processing volumes, and quarterly total publication volumes. Analyzing time series data can be aided by visualization methods such as bar charts and line graphs to observe phenomena and behavioral characteristics. Time series typically exhibit four main types of variations:

1. **Long-term or Trend Changes:** Reflect the overall direction of long-term changes, manifested as trend lines.
2. **Cyclic Movements:** Manifest as long-term oscillations along trend lines or curves, including periodic and non-periodic fluctuations.
3. **Seasonal Movements or Seasonal Variations:** Reflect events that recur annually, showing the same or similar patterns repeated in the same period across consecutive years.
4. **Irregular or Random Variations:** Changes caused by accidental or random events.

Applying data mining technology to news business process management significantly aids in the refinement of data summarization, analysis, and processing. By acquiring and analyzing user behavior patterns and historical editorial workflow data, we can comprehensively grasp the operational status of editorial business, understand the characteristics of editorial personnel, and achieve personalized and intelligent services.

6. Conclusion

Under the urgent situation of strategic transformation in traditional media, technological innovation is needed to enhance core competitiveness and communication impact. Big data serves as the core fulcrum for content, channels, and services, and is a powerful driver for traditional media transformation. This paper explores the application of data analysis and mining technologies to news editorial business process data within the general framework of business intelligence. Through analysis and mining of news productivity, communication impact, and full workflow status, we employ advanced methods and tools to streamline editorial business processes, identify points where behavioral data is generated and management controls are applied, and design indicator systems to help decision-makers grasp business development directions. With the development of the big data era, the application of business intelligence technologies will facilitate media convergence and provide strong support for the strategic transformation of traditional media.

References: [1] Zhang Liangjun, Chen Junde, et al. *Practical Case Analysis of Data Mining* [M]. Beijing: Mechanical Industry Press, 2013(7): 18-30.
[2] Chen Zhe. *Data Analysis: The Indispensable Assistant for Enterprises* [M]. Beijing: Mechanical Industry Press, 2015(5): 1-27.
[3] Ralph Kimball, Margy Ross. *The Data Warehouse Toolkit* (3rd Edition) [M]. Beijing: Tsinghua University Press, 2015(1): 5-11.

(Author Affiliation: Xinhua News Agency Technical Bureau)

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