New Media User Behaviour Research Based on Big Data Analysis

Zhixuan Zhu*
King's College London Strand, London, WC2R 2LS, United Kingdom

*Corresponding author: zhixuanzhu@kcl.ac.uk

Abstract. With the rapid development of diversified business models and market segments in the era of big data, many industries are facing the dual challenges of high customer acquisition costs and churn rates. How to accurately obtain user behavior data has become a top priority. This article uses big data technology to scientifically analyze the ratings and behavior data of new media users. Through new media user behavior analysis, we can accurately discover user behavior characteristics, optimize users' personalized experience, and bring better new media business operations—basis for decision-making.

1. Introduction
In recent years, big data analysis has become an important technology favored by more and more companies, governments, and various organizations. The analysis of enormous data traffic can effectively depict users' habits, hobbies, consumption levels, and Internet access. Busy time and other information can further construct the user's behavior model to predict consumer behavior better and design a more optimized and convenient system.

How to timely and accurately count and analyze data useful to new media companies from this information, and dig out the hidden laws and values behind this information, has become a strategy that restricts new media companies' decision-making development, expands the market, and expands their customer base. Important factor. New media companies’ current mainstream approach is to introduce third-party modules like Zhuge IO and Baidu Statistics or to develop their back-end statistics modules [1]. To this end, this article designs a user behavior analysis platform, uses big data analysis technology, Lucene full-text search cross-analysis algorithm, custom embedding technology, eliminates the risk of accessing third-party services for user analysis, and performs new media users on data Channel conversion, new addition, retention, and multi-dimensional cross-analysis can well discover hidden laws, predict the future trend of data, and display it in the form of graph visualization, which primarily meets the needs of corporate decision-makers.

2. New media user behavior analysis platform construction

2.1. System principle and architecture
The new media system platform includes hardware parts and software part. The hardware part includes: encoder, database server, storage array, live broadcast server, transcoding server, streaming media server, on-demand server, content management server, upload/audit workstation, management workstation,
Routers, load balancing equipment, network switches, data security gateway servers, strip servers, transcoding workstations, firewalls, and CDN distribution platforms, providing live broadcast, on-demand, news, and smart city convenience services, with multiple channels to release information, numerous functions such as code stream publishing information, using CDN to publish information nearby and publishing information purposefully. It can realize the unified management and release of content, realize the integration and unified login of users of various applications; in a real sense, realize the independent multi-service management, multi-terminal operation, unified, centralized data and business application management of the entire radio and television new media system platform. The physical topology of the system is shown in Figure 1.

![Physical topology diagram of the new media system](image)

As shown in Figure 2 below, when a user operates the website, user behavior data will be sent to this system's receiver. The receiver will filter and filter the data through algorithms and then store the data warehouse's valid data. When the data warehouse receives the data, the system uses a series of big data users behavior analysis algorithms such as user activity, terminal analysis, and geographic analysis to perform weighted scores based on the user's web browsing behavior, and then compares globally to analyze the active and silent users. At the same time, it will analyze the trend of the calculated results to analyze whether the user is retained, and finally display it with big data analysis chart technology, and return the result to the user [2]. Simultaneously, due to a large amount of data that affects the system's retrieval speed, the system uses a full-text retrieval algorithm to solve search retrieval efficiency fundamentally. Finally, the data feedback will be displayed on the system page to provide users with real-time use. Simultaneously, to solve the problem that the acquired user data may be wrong due to the iteration of the website version or the transmission process, the verification and detection module is designed and developed. This module uses crawlers and recursive tag algorithms to check the data's correctness, ensuring the statistical data. Completeness.
2.2. Hardware platform construction

The new media system's hardware platform includes encoders, various servers, workstations, and network security access equipment. If you build it yourself, you will need a lot of construction funds. To save the cost of system construction, the new media system's hardware platform is leased from Tencent Cloud. All service servers and network access bandwidth are directly leased, and only the corresponding lease fees need to be paid each year. The TV station itself only needs to configure the live encoder, upload/audit workstation, and background management workstation.

The information security system guarantees the efficient and safe operation of mobile applications of the new media system. The security measures of the platform must be considered from all aspects; firewalls and isolation devices are deployed at the boundary with the external system, and intrusion detection is deployed in the core switch of the host system system; The system deploys a network antivirus system and anti-trojan horse system to prevent viruses and malicious code from threatening the system; adopts a security audit system to conduct comprehensive security protection and audits on intranet servers, hosts, and equipment; secure database management systems Reinforce, deploy the corresponding security strategy, improve the security of the database management system through the excellent configuration of security attributes; strengthen the security of the computer and network equipment within the system, and deploy the corresponding security strategy [3]. Since the new media system is constructed by renting a cloud platform, the primary hardware platform is the operator's cloud platform, and there is a sound security protection system. TV stations only need to set up a firewall system to protect their upload/audit workstations and back-end management workstations. Antivirus and safety are enough.

2.3. Software design

The functional software modules of the established new media system primary platform include: streaming media live broadcast platform software, streaming media-on-demand platform software; streaming media backstage management platform software; broadcasting control and stripping system software; streaming media transcoding system software; content management System software; mobile APP client software [4]. The main functions of the primary platform of the new media system include: realizing the simultaneous network live broadcast and on-demand broadcasting of TV stations and radio stations; realizing primary picture and text news push; introduction of beautiful scenery, food, superior

---

**Figure 2.** User behavior analysis platform architecture
products, etc.; living payment, ticket booking, government and enterprise Convenience services such as direct connection, service inquiry, and life services; with all-media interactive functions; support for multi-channel and multi-screen publishing of program content on TVs, computers, smartphones, and tablets.

3. The analysis method of user behavior data

The behavior of users in the platform can be characterized by analyzing, disassembling, and reorganizing network data packets' data fields. Useful standard fields are given in Table 1. According to the information provided in Table 1, you can describe when and where a data packet is sent from, where, and what kind of data is requested or responded to. According to actual needs, you can also further analyze the body content of the HTTP packet. Collection and analysis (such as extracting the title of the visited page through the <title> <title> tag) to make the analysis of user behavior more refined. The following is a row based on the fields defined in Table 1

For the analysis method.

| Host (HTTP package) | The network host and port number of the requested resource |
|---------------------|----------------------------------------------------------|
| User-Agent (HTTP package) | Contains the user's device information, browser version, operating system version, and other information |
| Content—Type (HTTP package) | Return the type of resource |
| Content (HTTP package) | HTTP packet body |
| Src / Dst (TCP packet) | Source and destination IP address |
| Src / DstPort (TCP packet) | Source destination port |
| Src/Dst (Ethernet II package) | Source and destination MAC address |
| Title (HTML body) | Page title |
| Keywords (HTML body) | Search content |

According to the above data definition, the relationship model shown in Figure 3 can be summarized: the overall analysis of users is divided into two categories; one is user attribute description; this type of analysis describes some specific attributes of users, such as consumption Level, most concerned content, etc.; the other is user behavior description, this type of analysis describes a specific type of user's trend or habits, such as the time and space distribution of data packets, the group distribution of access objects, etc. The first type of description can help analysts portray the image of the user (individual or group) and then classify or summarize the user (group); the second type of description dynamically displays the stage characteristics of the user (group), which can help analyze the teacher understands the user's behavior patterns [5]. In specific behavior analysis, the two types of descriptions need to be combined to obtain a complete user behavior description result.
Figure 3. User (group) behavior analysis relationship model

3.1. Analysis method for the entire network platform
The whole platform analysis can describe the behavior trajectory of one or more user groups or obtain the behavior pattern of the user group by summarizing the data characteristics of the majority group. Analysis purpose: Through the statistics of the content of the webpages visited by the users in the platform, the hot content that the current user group is most concerned about is obtained. Analysis strategy: By aggregating the Title field, the results are sorted in descending order to get the titles of all visited pages and filtered through regular expressions to remove the results such as "404 Not Found" and "302 Found" in the results, which are equal to the meaningless results of the analysis.

3.2. Network analysis based on clustering algorithm
In addition to the description of group user behavior for the entire platform, the research object can also be converted to a single object for behavior description by specifying the source, destination IP address, source, and destination MAC address. In this way, a specific user’s complete behavior pattern can be accurately and effectively depicted, thereby achieving more targeted and accurate user behavior analysis. Use the clustering algorithm to divide the obtained feature data into n ordered data sets. Based on the data set's spatial density, in the data set with interference data, mine the key data that affects user behavior and allows These data have different sizes and different attributes [6]. In the sample set, if the object is in the domain of the core object q, then a is directly accessible from q; when there is q = q,a = q in the sample set, then q,a it is directly accessible from q, and then called a from q. The density is up to.

\[ Q = \frac{1}{k} \sum_{i=1}^{r} \sum_{u=0}^{n} \left| \mu - \bar{u} \right|^2 \]  

Where: Q represents the convergence result of multiple feature data; r represents the number of feature data clusters; u represents feature data; \( \bar{u} \) represents the average value of i feature data; k represents the total number of feature data; \( W_i \) represents the attribute of feature data set. Based on the above clustering results, cross-analyze the inherent properties of library mobile user characteristic data, including cross-analysis of user age, cross-analysis of library usage, and cross-analysis of age and library usage patterns. According to the analysis results, the behavior keywords of mobile users of different age groups are set, and the user behavior analysis method of big data mining technology is completed.
4. Conclusion
The analysis of big data can be divided into two types, in essence. One is to summarize the data that has occurred now and find the rules; the other is to summarize the past data and predict the future development trend. Big data analysis has become the industry’s consensus due to its large volume, complete data, and brief analysis, which is an inevitable future development trend. This article introduces an effective data analysis method based on the analysis of network data messages. Through intuitive chart display, network user groups' behavior mode and behavior trend prediction are further obtained. To improve the government's livelihood governance, corporate business operations, and network Provide valuable information reference in security management areas.

References
[1] Qiao, Y., Xing, Z., Fadlullah, Z. M., Yang, J., & Kato, N. Characterizing flow, application, and user behavior in mobile networks: A framework for mobile big data. IEEE Wireless Communications, 25(1) (2018) 40-49.
[2] Li, S., Jiao, F., Zhang, Y., & Xu, X. Problems and changes in digital libraries in the age of big data from the perspective of user services. The Journal of Academic Librarianship, 45(1) (2019) 22-30.
[3] Xie, J. Research on Weibo user behavior system for subjective perception and big data mining technology. Journal of Intelligent & Fuzzy Systems, 38(2) (2020) 1225-1234.
[4] Hong, S. New Authentication Methods based on User's Behavior Big Data Analysis on Cloud. Journal of Convergence Society for SMB, 6(4) (2016) 31-36.
[5] Parwez, M. S., Rawat, D. B., & Garuba, M. Big data analytics for user-activity analysis and user-anomaly detection in mobile wireless network. IEEE Transactions on Industrial Informatics, 13(4) (2017) 2058-2065.
[6] Cassavia, N., Mascieri, E., Pulice, C., & Sacca, D. Discovering user behavioral features to enhance information search on big data. ACM Transactions on Interactive Intelligent Systems (TiiS), 7(2) (2017) 1-33.