

Research Article

Construction of Internet TV Industry Ecosystem Based on Data Mining Technology

Jie He 

Xijing University, Xi'an, Shaanxi 710123, China

Correspondence should be addressed to Jie He; 20190205@xijing.edu.cn

Received 5 January 2022; Revised 23 January 2022; Accepted 26 January 2022; Published 3 March 2022

Academic Editor: Xin Ning

Copyright © 2022 Jie He. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

While the Internet provides people with convenience, it also comes with security concerns. Users can more easily form groups, distort facts, and contribute to some sensitive topics on the internet. As a result, identifying and analyzing users' online behaviors are critical. This paper creates a new Internet TV industry ecosystem using DM (data mining) technology. The recommendation system model is established based on data from users' on-demand viewing behavior across the entire network, and the functions of various system modules and their coordination ability are described in detail. The evolution of users' online time is examined, providing data to support and explain the prediction analysis of users' click behavior and the analysis of users' search intent. The type of web page that the user clicks on can reveal the user's behavior tendency.

1. Introduction

With the development of digital technology, all services such as voice, image, and data transmission in three modern communication forms (telecommunication network, broadcast TV network, and computer Internet) can use common language, and therefore, telecom industry, broadcast TV industry, and computer industry with distinct industrial boundaries have common essential characteristics, that is, digital characteristics. As a result, the cores of the three networks converge and the boundaries are blurred, resulting in the phenomenon of "three networks merging" [1, 2]. In any case, cable TV operators have the best HD live channel communication capability. Simultaneously, most cable TV operators launch rich on-demand content services based on this model [3]. The rapid development of Internet information technology has resulted in the emergence of a slew of new media, including mobile media and mobile TV, which have had a significant impact on traditional media. It is critical to reform the traditional media.

Media integration, the general trend, and the integration of various media in technology, content, and market have brought great impact to traditional TV, which is facing a new reconstruction from the concept, mode, and channel

of communication. Convergence media not only changed the communication pattern but also gave birth to the new industrial chain and new profit model of media industry, which will bring new products and services and new market opportunities [4, 5]. Compared with traditional TV, Internet TV users can leave data such as identity characteristics, interest preferences, and behavior tracks when watching TV, which provides data basis for Internet TV operators to analyze users. Through DM (data mining) technology [6], users' related information can be found and analyzed. Similarly, through DM technology, need to know more about hobbies, education, occupation, gender, and so on of specific commodity buyers. Through this technology, targeted advertising for users can be realized, so that the reporting rate and effect of advertising can be improved, and the marketing investment of enterprises can get the best results. So as to have a deeper understanding of users' intentions and interests, tap potential users, predict users' behaviors, and serve users more pertinently [7]. Through users' personal characteristics and interests, it provides data support for personalized marketing strategies, so as to achieve better recommendation effect.

In general, smart TV users are younger and have a higher level of education, while their family income is

relatively high, implying that they have greater purchasing desire and power [8]. When “smart technology” and “Internet thinking” infiltrate the TV industry, they disrupt the inherent ecology of the industry’s stable structure, hastening the deconstruction and transformation of the traditional TV industry model. In the game of gene recombination and differentiation, the Internet TV industry ecology is constantly revealing its full outline [9]. Starting with the Internet TV industry ecosystem’s innovation, this paper investigates how TV media can break through traditional management modes in the context of convergence media, as well as the path and characteristics of current TV profit model innovation, in order to better guide TV practice.

2. Related Work

The value theory holds that the business model is the logic or way for enterprises to create and realize value. For example, literature [10] points out that business model describes the value creation logic behind the actual business process of a business system. Literature [11] holds that all enterprise activities are not only a series of value activities but also orderly “transaction” activities. Literature [12] holds that the basic meaning of business model is the basic logic of enterprise value creation. Literature [13] holds that business model is the only mixture of enterprise value flow, revenue flow, and logistics. Literature [14] holds that “new media is the communication medium for all people to interactively deliver personalized digital composite information to the public in real time.” Literature [15] indicates that “compared with traditional media, the essential feature of new media is the carrier of information dissemination by means of computers.” It is considered that “business model is the synthesis of enterprise value flow, income flow and logistics.” Literature [16] holds that business model is an institutional system for an organization to integrate many stakeholders to obtain excess profits on the premise of clarifying external assumptions, internal resources, and capabilities. All-media in a broad sense refers to the comprehensive application of media form, media generation, and media communication. All-media in a narrow sense refers to the communication concept based on the development of modern science and technology and media integration, which integrates traditional media and emerging media in many aspects [17].

The innovation of media profit model under the background of convergence media is the focus of current theoretical circles and press circles. Especially on the research of profit model, many foreign experts and scholars studied early and deeply and formed many original theories. Literature [18] holds that the core of enterprise profit is also the function of value chain design and application, and every enterprise is an assembly of various activities in the process of designing, producing, selling, distributing, and assisting its products. All these activities can be represented by a value chain. Literature [19] holds that a complete business model system includes profit model, positioning, business system, key resources and capabilities, free cash flow structure, and enterprise value.

People can now easily and quickly obtain the information they require on the Internet, thanks to the rapid development of the Internet industry and the increasing level of informatization in various fields in recent years. As the Internet has grown in popularity and information technology has advanced, a large amount of data has accumulated on the Internet. Literature [20] used crawling technology to collect more accurate Twitter data than previous studies and investigated the topology structure of the Twitter network, user ranking method, hot topic propagation mode, and other aspects, resulting in a quantitative analysis of the Twitter network and its information distribution. Literature [21] examines Twitter users’ emotions and attitudes toward television programming, as well as the feasibility of gathering TV viewers’ cognitive behaviors about programming from online social media. Literature [22] proposes a model of TV audience ratings and investigates the problem of viewing competition among multiple TV stations using this model. Literature [23] proposed a semi-Markov model-based method for describing user behavior. The main goal of these analysis and modeling studies is to better understand or predict the audience’s media contact behavior, as well as to improve services like program recommendation and content scheduling on television and the internet.

At present, under the condition of convergence media, the innovation research and development practice of TV media business model are still in the exploratory stage, so there are few theoretical documents. By combing the previous literature, the existing research results will not be able to provide good suggestions for the problems faced by China’s TV media industry in the profit model. Therefore, it is of great value and significance to explore the innovation of the Internet TV industry ecosystem under the condition of convergence media.

3. Research Method

3.1. Behavior Analysis of Internet TV Users Based on DM Technology. Driven by objective technical factors such as the popularization of Internet broadband, the improvement of network speed, and the popularization of intelligent networking equipment, the public’s media usage habits and consumer demand have undergone tremendous changes. Video content providers can directly provide video, games, and other information content services to users through the Internet without having to build a physical network. The media terminals are mainly smart TV integrated machines and smart set-top boxes.

At the same time, in such a big environment as China, traditional TV media also plays the role of political propaganda. Therefore, the problems of marketization and industry barriers have always existed, which has a certain blocking effect on the rapid development of traditional TV industry in China, and also restricts the change of business model. It is the existence of this kind of system that leads to the relatively fixed market model of TV media, its flexibility cannot keep up with the demand of the consumer market, the inaccurate positioning of some TV stations, and the lack of professional content of channels to attract consumers, which leads to the

low general sense of identity of consumers, and the participation of other new media in customer competition, which leads to a decline in attention and a serious loss of customers.

The study of DM technology is useful in a variety of fields, including the Internet, financial management, business decision-making, and retail business activities. DM and analysis technology are used in a variety of fields, but they are most commonly used in commercial fields such as banking, telecommunications, and the Internet. This chapter uses this as a starting point, building the DM algorithm model and providing scientific data support and basis for the company's business departments to make user personalized service decisions, based on the entire network data of an Internet TV operator, through the DM method, with the theme of analyzing users' browsing and viewing behavior, and with the purpose of focusing on user service and two problems to be solved. The capacity to acquire client data and other data via the Internet has become both a required and sufficient prerequisite for personalized intelligent recommendation. Currently, recommendation algorithms are evaluated based on the accuracy of recommendation results and other related evaluation criteria, but the entire intelligent recommendation system is not evaluated in the true sense, particularly in terms of personalized recommendation degree, user acceptance degree, persistence degree, and so on. The intelligent recommendation system's evaluation is one-sided, and there is no comprehensive evaluation.

As the foundation of intelligent recommendation system, data warehouse provides basic data support for data mart, analysis and statistics, DM, and other applications on the basis of integrating relevant business system data. IT is a very complicated IT system. In the process of modeling in reality, data models are established and subdivided in turn by using the thinking of multilayer modeling (Figure 1).

User's online time refers to the time when the user node joins the channel once and leaves the channel, which reflects the time when the node stays in the channel once. The measured data of this study is a series of snapshots that are continuous in time:

$$S = \{s_0, s_1, \dots, s_N\}. \quad (1)$$

s_i represents the i th snapshot. For each measured node p , find the corresponding snapshot subsequence:

$$S = \{s_n, s_{n+1}, \dots, s_{n+i}\}. \quad (2)$$

It means that the node p was found for the first time in the n th snapshot, and it was included in the subsequent i snapshots until the node disappeared in snapshot s_{n-1+i} .

Currently, Internet TV is primarily based on a peer-to-peer (P2P) architecture, with each node user acting as both a client and a server. It is necessary to create user nodes and accept data packets from neighboring nodes in order to obtain the network's overall topology. The number of users joining P2P PTV networks per unit time is referred to as the arrival rate of users. The user arrival rate is an important metric for analyzing P2P network dynamics and

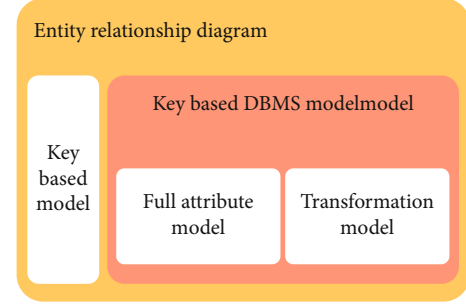


FIGURE 1: Hierarchical model of data warehouse.

user behavior. As shown in the formula, determine the user arrival rate at t time:

$$A_t = \frac{N_{t+1} - N_t}{T_{t+1} - T_t}, \quad (3)$$

where T_t represents the acquisition time of the t -th snapshot and N_t is the number of users in the t -th snapshot. In this paper, the time interval is in minutes, that is, A_t represents the arrival rate of users per minute.

In the P2P TV system, the user arrival behavior is highly related to the program arrangement of the channel. When the channel starts to play popular programs, a large number of user arrival events usually occur in a short time, which leads to the so-called flash crowd phenomenon. Remember that the coverage network corresponding to the snapshot is

$$G = (V, E), \quad (4)$$

where $V = \{v_i, w_i\}, i = 1, 2, \dots, N$ is the node set of the overlay network, w_i is the node weight, and the initial weight of each node is 1. Each node has a corresponding latitude and longitude coordinate attribute. e_{ij} represents the edge from node v_i to node v_j , and $E = \{e_{ij}\}$ is the edge set of the overlay network. D represents the distance operator between nodes, and K represents the merging threshold. The simplified algorithm is shown in Figure 2.

According to these training and working characteristics of DBN (deep belief network), this paper uses this algorithm to classify and model, and analyze users' web clicks and other behaviors. The essence of DBN classification prediction model is to improve the accuracy of classification and prediction by establishing a model with many hidden node layers and using a large number of unsupervised training data to learn the characteristics of data. To sum up, the technical means of multiple hidden layers in deep learning algorithm is adopted to improve the prediction accuracy.

SVM is a supervised classification algorithm based on statistical theory. It can improve the generalization ability of learning model by satisfying the minimum structural risk and minimize the empirical risk and confidence range, so as to obtain good statistical rules with less statistical samples.

Compared with DBN classification system, SVM (support vector machine) is a mature binary classification model [24]. SVM consists of input layer data and output layer.

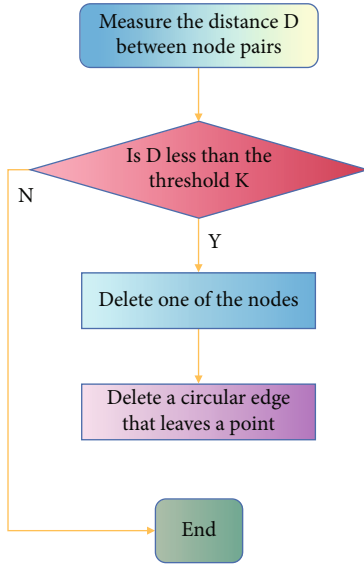


FIGURE 2: Topology pretreatment process.

Compared with the complex training process of DBN, SVM has a simpler and faster training process.

Suppose a binary classification problem with N training data samples, each data sample is represented as a binary $(x_i, y_i)(i = 1, 2, \dots, N)$, where x_i corresponds to the attribute set of the i th sample. At the same time, specify $y_i \in \{-1, 1\}$ as its class label, so the decision boundary of a linear SVM can be written as follows:

$$w \cdot x + b = 0, \quad (5)$$

where wb is the parameter of the model.

However, SVM puts forward a higher requirement based on this condition: the edge of its decision boundary must be the largest. To meet this requirement, then there is to minimize the following objective function:

$$f(w) = \frac{\|w\|^2}{2}. \quad (6)$$

For nonlinear SVM model, the kernel function is used to transform the given data set. Find the linear decision boundary after transforming the space, so that the corresponding class contains all data instances belonging to the class. In the new space, the linear decision boundary should have the following formula:

$$w \cdot \Phi(x) + b = 0. \quad (7)$$

SVM algorithm is the main comparison algorithm of user behavior analysis module, which is used to verify the accuracy of the model. At the same time, it is also the main modeling algorithm of the intention prediction module, which completes the user intention prediction.

Establish two binary classification models, DBN classification system and SVM, respectively. After the experimental verification, with the increase of categories to be classified,

the time efficiency and accuracy of “synchronous voting” are obviously higher, so this model is used here to realize multiclassification of “voting system.” The multiclassifier system is shown in Figure 3.

3.2. Construction of TV Industry Ecosystem. The innovation of TV industry ecosystem should be based on market demand and actively use technological changes, strive to overcome various adverse factors such as policy limitations and local monopoly, do a good job in its own market positioning, focus on combing the value chain, and then implement optimization and extension, build its own strong cooperation network, rationally design the cost and profit distribution structure, and finally realize the sustained win-win situation of enterprise value and customer value.

The implementation of regional management is an important way to solve this problem for the fragmented radio and television media industry. The first premise of the three networks’ integration is that radio and television will create a national network, establish a national network company, and manage regional networks. Acting according to the laws of the market economy is a basic principle in a market economy, and market-oriented means are unquestionably the most fundamental means of network integration. However, market regulation is lagging, and the means of complete marketization will have an impact on network integration efficiency. At the same time, a significant portion of network integration resistance stems from the obstruction or competition of government administrative forces at all levels and localities, so strong administrative intervention is a necessary means of breaking through resistance and promoting network integration quickly. Make all localities understand the value of lowering operating costs, improving network competitiveness, establishing proper operating strategies, improving industrial system and layout through the scale effect of network integration, and encourage them to join the regional network.

The diversification of consumer demand leads to the diversified development of TV program products. Professional channels should take this major as the theme program, the theme program occupies the main time period of the channel, and other programs can be relayed in other parts of the time period. If the audience market positioning of the professional channel is too large, it will lead to the channel being unspecific and easy to be understood as a “quasi-comprehensive” channel, which is contrary to the original intention of the professional channel. If the professional channel’s program production capacity and program content cannot satisfy the audience, it will still lead to the loss of the audience and the decline of attention. Therefore, it is necessary to improve the professional channel’s program production capacity and program quality, closely follow the theme content of the professional channel, form a certain brand influence, and provide more services for the audience.

The main elements of the Internet industry ecosystem gradually integrate each participant’s resources into high value-added services for consumers through interaction and close linkage and constantly expand the brand-new business model and industry commercial value. The

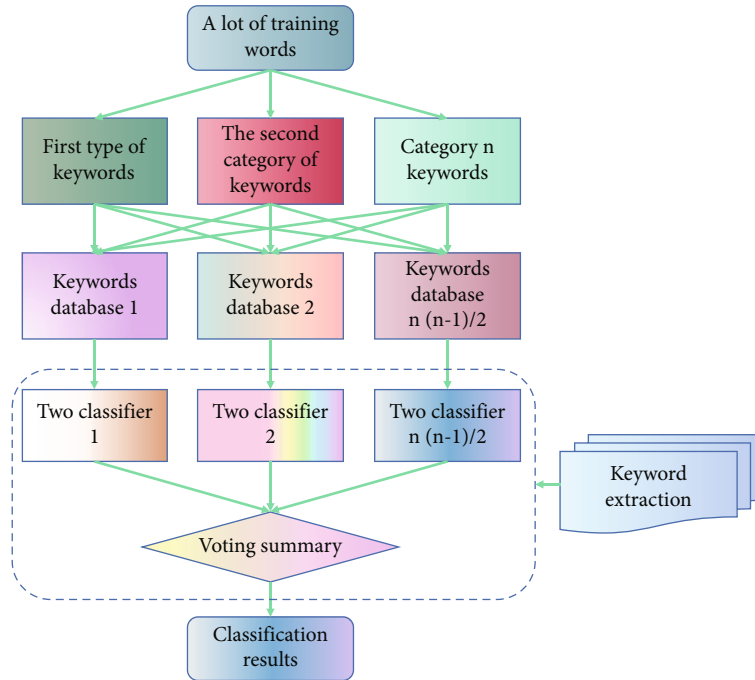


FIGURE 3: Multiclassification system.

“hardware manufacturing thinking” that originally pervaded the TV industry is completely subverted when Internet TV injects “Internet thinking” into the industry. In the future, Internet TV will only be able to gain more development opportunities in this massive wave of network society evolution if it exhibits more revolutionary courage, pays close attention to user needs, and relentlessly pursues the ultimate design and product innovation.

For TV brand columns, the higher the loyalty of viewers, the easier it is to turn into fans, and the business activities based on fans will also bring unexpected economic benefits to the columns. Build industrial chain and product chain, carry out offline activity marketing, or develop new formats and adjust industrial layout. The industrialized operation of TV media starts from program planning, production, and promotion and seeks profit opportunities in production process and output mode. By exploring local resources to find economic growth points, maximize the effective configuration of products, services, channels, audiences, and other internal media resources, and maximize the development of external media resources, so as to establish a business structure and business structure that are suitable for industrial management and profit-oriented. The higher the matching degree between internal and external profit elements of media, the greater the profit value.

4. Results Analysis and Discussion

The main work of the Internet intelligent recommendation system is to provide customers with recommendations on video programs, and the system serves VOD users. Users are mainly those who use the recommendation system, and they are also the main users in the viewing activities. Project

refers to providing video content and services to users in VOD (video on demand) activities and also being recommended users.

The average television viewing time in China has increased in recent years and now stands at around 270 minutes. The average audience reaching rate, on the other hand, has dropped dramatically, and the loss of loyal viewers will have an impact on the effectiveness and benefits of television commercials. As a result, there may be a decline in TV advertising revenue in China in the future. Figure 4 depicts the size of China’s television advertising revenue.

It can be seen that the advertising-oriented business model is no longer in line with the development of today’s times, and the advertising revenue crisis of traditional TV media industry is beginning to become prominent. It is urgent to change the business model, which can help enterprises to find profit space again and make the business income model tend to be diversified.

Users can be divided into three categories according to their online conversation duration: light viewers, moderate viewers, and heavy viewers. According to the average conversation distribution of users, users whose conversation duration is less than 10 minutes are defined as light viewers. Users between 10 minutes and 100 minutes are defined as moderate hand watchers. These users are normal viewers, most of whom have their own fixed preferences and are keen on their favorite programs.

Focusing on the potential needs of users with large screens in living rooms, the Internet not only provides high-definition TV programs but also provides users with a brand-new mode different from the experience of computer Internet and mobile Internet, making it the third-largest terminal entrance to the Internet after computers and

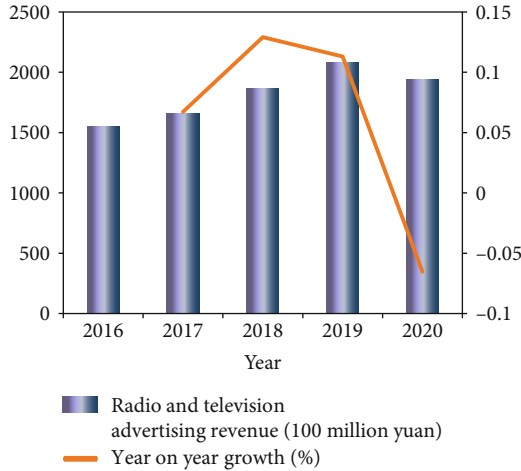


FIGURE 4: China's TV advertising revenue scale.

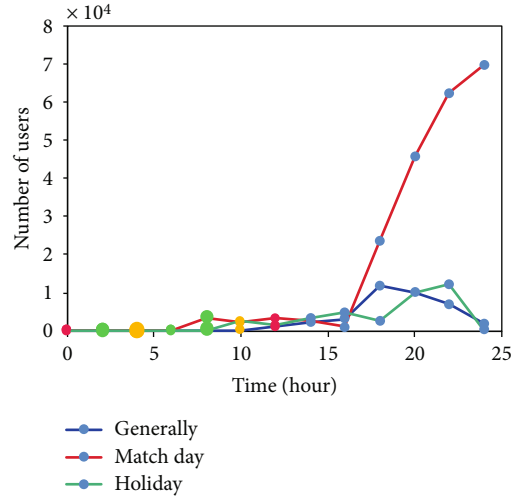


FIGURE 7: UUsee users' online time evolution.

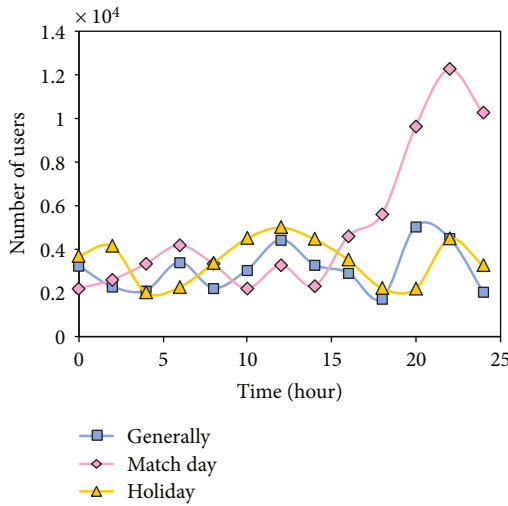


FIGURE 5: Evolution of online duration of PPLive users.

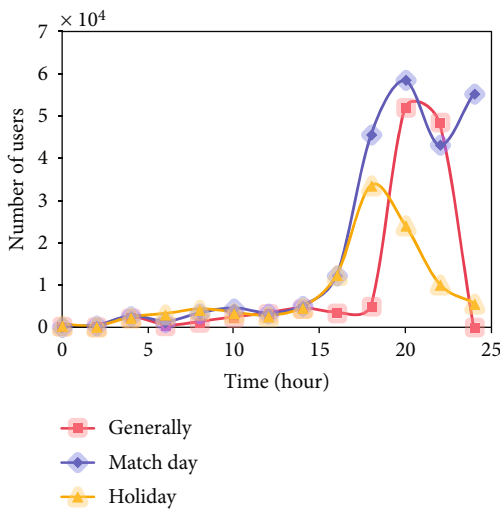


FIGURE 6: Evolution of online duration of PPStream users.

smartphones. This makes Internet TV have the ability to break free from the system shackles of traditional TV industry value chain, and with the strong social industry penetration force of the Internet, a brand-new value chain system with high openness and extensibility is constantly formed.

In order to study the users' attention content and attention time, this paper analyzes the evolution of users' online time, which is still based on the data of three platforms: PPLive, PPStream, and UUSee during the broadcast of a TV program. Describe the evolution of users' online duration as shown in Figure 5–7 with the characteristics of the number of users at each time point within 24 hours.

From Figure 7, available around 20:00 in the evening, the number of users increased slightly. In normal times or on match days, which reflects that around 20:00 is the prime time of a TV station. The evolution analysis of users' online time can help decision makers adjust their strategies. For example, if the TV station can know what TV content is popular and when advertising is best, in this example, therefore, it is best to advertise around 20:00.

Because of the complexity and diversity of Internet TV users' behaviors, it is very difficult for researchers to accurately analyze users' behavior characteristics, which involves the following processes: obtaining massive raw data, extract data that can be used for analysis from massive Internet traffic data; from these data, valuable characteristics and trends are obtained.

Characteristic analysis of user behavior refers to analyzing users according to their behavior characteristics when using the Internet. Usually, users are classified into different groups according to their characteristic data by using methods such as classification or clustering, and targeted services and marketing are carried out for users of different groups, which can not only maximize benefits but also improve users' experience.

DM is a very important part in the analysis of user's behavior and intention. It is our aim to use reasonable methods to DM the massive data so as to get accurate analysis and prediction results. In the following experiments and analysis, use various DM algorithms to try to get the most

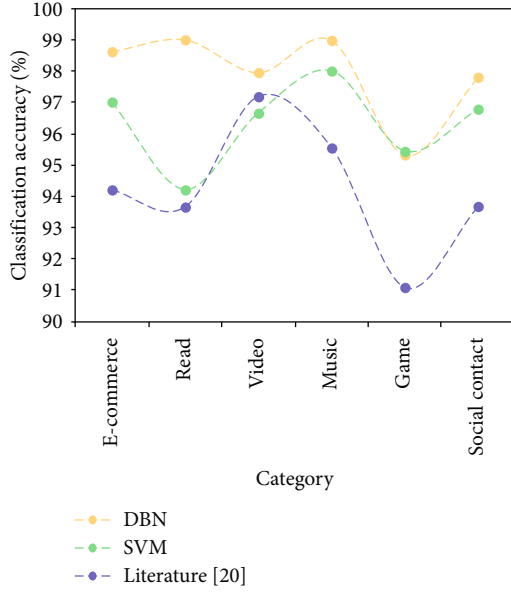


FIGURE 8: Two classification accuracy.

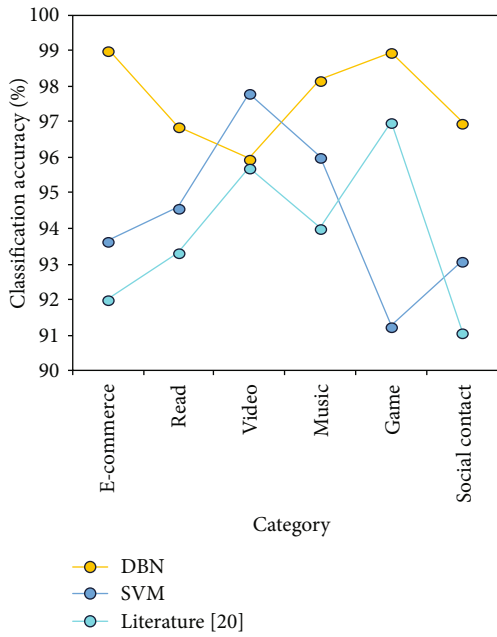


FIGURE 9: Triple classification accuracy.

accurate and meaningful results, such as deep learning algorithm, association rule algorithm, and SVM.

The webpage information generated by users clicking on webpages can be used to accurately classify webpages, and then the user's behavior tendency can be analyzed by the category of users clicking on webpages. Therefore, the classification and identification of web pages clicked by users is an important part of analyzing users' behaviors and intentions. Using the DBN classification prediction system, SVM model, and literature [20] method, the same training data and test data are modeled and predicted, respectively. For each binary model, 10 repeated experiments of randomly

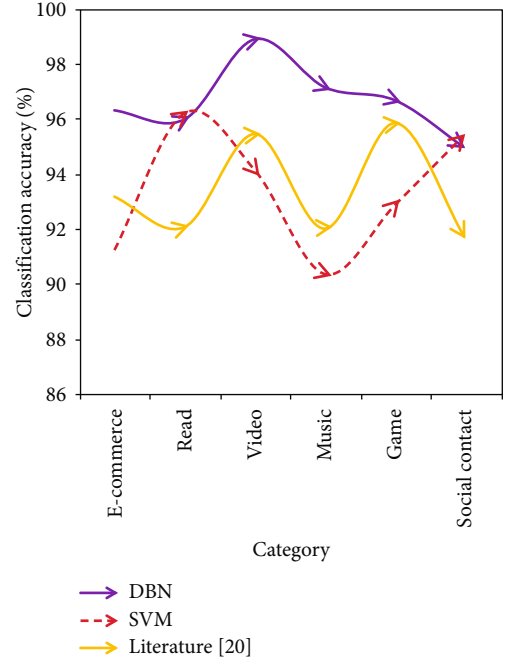


FIGURE 10: Four classification accuracy.

selected test data are conducted, and the prediction accuracy is counted. The data results obtained are shown in Figures 8–10.

From these figures, whether it is binary classification or multiple classification, DBN classification prediction system, SVM classification model, and literature [20] method can achieve very high accuracy, but the DBN classification prediction system is better.

At the same time, according to the previous introduction, the DBN system uses a large amount of unlabeled unsupervised data for model pretraining and only needs a few labeled data to fine-tune network parameters. For the massive Internet TV user data, the amount of data that can really get the tag is very small, and it is a process that requires a lot of labor costs. If we can reduce the cost of this part and improve the utilization degree and accuracy of the data by continuously optimizing the system model, we will be able to improve the efficiency of user data analysis.

The web page contains a large amount of text data, such as the title, keywords, descriptions, and links. Especially, the web pages of e-commerce websites contain a large amount of descriptive information of commodities, which plays a very important role for users who browse e-commerce websites and influence their online business behaviors. By analyzing these unstructured text information, model and predict the user's business intentions [4], and make user portraits, etc.

The accuracy of the DBN classification model is still very stable and efficient in the case of decreasing data volume. In the case of multiclassification, the final accuracy will fluctuate due to the change of accuracy of two classifiers, but the fluctuation is smaller when using the DBN model, and the classifier effect is more stable. According to the click records of each user, the types of the web pages clicked by that user can be summarized. According to experience, the number of

types of web pages clicked by each user is not unlimited, but usually concentrated in several specific types. The user's behavior tendency can be obtained by the type of web page clicked by the user.

5. Conclusion

It is not difficult to find that the explosive growth of Internet TV actually coincides with the inherent development law of the development of information society when we step outside of the TV industry ecosystem and examine the development trend of Internet TV in the macroinformation society. The intelligent recommendation system based on DM technology is primarily used in this paper to analyze the behavior of Internet TV users. A user behavior analysis based on the evolution of online duration is also proposed. We can clearly understand users' habits of watching programs at various time periods, what they care about, and so on, thanks to the intuitive chart. This mode gives users more accurate information, which is important for the radio station to adjust its broadcast content, broadcast time, and user behavior monitoring.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The author does not have any possible conflicts of interest.

Acknowledgments

This study was supported by Scientific Research Fund Project of Xijing University, 2020—Observing Tradition and Achieving Innovation in Chinese TV Variety Shows from the Perspective of Marxist Theory of Literature and Art (Grant no. XJ200207).

References

- [1] T. V. Alekseeva, "Features of modern internet media as a substantive component of training," *Science*, vol. 8, no. 42, pp. 10–13, 2019.
- [2] E. Y. Lo, "How social media, movies, and TV shows interacts with young adult literature from 2015 to 2019," *Publishing Research Quarterly*, vol. 36, no. 4, pp. 611–618, 2020.
- [3] Y. Yan, "Industry analysis of live streaming platform in internet celebrity time," *English on Campus*, vol. 14, p. 2, 2019.
- [4] J. P. Simon, "The global internet market(s): a reconstruction of the views of the industry," *Digital Policy, Regulation and Governance*, vol. 22, no. 2, pp. 109–133, 2020.
- [5] M. Kranz, "Why industry needs to accelerate IoT standards," *IEEE Internet of Things Magazine*, vol. 1, no. 1, pp. 14–18, 2018.
- [6] J. Kong, C. Yang, J. Wang et al., "Deep-stacking network approach by multisource data mining for hazardous risk identification in IoT-based intelligent food management systems," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 1194565, 16 pages, 2021.
- [7] F. Kollintza-Kyriakoulia, M. Maragoudakis, and A. Krithara, "Measuring the impact of financial news and social media on stock market modeling using time series mining techniques," *Algorithms*, vol. 11, no. 11, p. 181, 2018.
- [8] I. Malyavkina, "Linguistic peculiarities of media texts of financial sphere in the social media space," *Path of Science*, vol. 4, no. 12, pp. 4001–4005, 2018.
- [9] P. Vonderau, "The Spotify effect: digital distribution and financial growth," *Television & New Media*, vol. 20, no. 1, pp. 3–19, 2019.
- [10] L. I. Xuan, "Competition and political factors in the Russian media market during Putin's fourth president term," *Literature and Art Studies: English Version*, vol. 10, no. 11, p. 10, 2020.
- [11] S. C. Bilow, "Introduction: blockchain in media and entertainment," *SMPTE Motion Imaging Journal*, vol. 129, no. 1, pp. 20–21, 2020.
- [12] F. Onwuegbuche, J. M. Wafula, and J. Mung'Atu, "Support vector machine for sentiment analysis of Nigerian banks financial tweets," *Journal of Data Analysis and Information Processing*, vol. 7, no. 4, pp. 153–173, 2019.
- [13] A. Marais, "Considering the role and analysing the effectiveness of the media in covering issues of personal finance: the case of Huisgenoot (2007-2008)," *Global Media Journal African Edition*, vol. 3, no. 1, pp. 1–23, 2018.
- [14] S. Deng, Q. Cai, Z. Zhang, and X. Wu, "User behavior analysis based on stacked autoencoder and clustering in complex power grid environment," *IEEE Transactions on Intelligent Transportation Systems*, no. 99, pp. 1–15, 2021.
- [15] C. H. Lai and C. Y. Hsu, "Rating prediction based on combination of review mining and user preference analysis," *Information Systems*, vol. 99, article 101742, 2021.
- [16] D. Moon, H. Im, I. Kim, and J. H. Park, "DTB-IDS: an intrusion detection system based on decision tree using behavior analysis for preventing APT attacks," *The Journal of Supercomputing*, vol. 73, no. 7, pp. 2881–2895, 2017.
- [17] M. S. Chen, C. P. Hwang, T. Y. Ho et al., "Driving behaviors analysis based on feature selection and statistical approach: a preliminary study," *Journal of Supercomputing*, vol. 75, no. 4, pp. 2007–2026, 2019.
- [18] C. J. Lin, C. Wu, and W. A. Chaovaitwongse, "Integrating human behavior modeling and data mining techniques to predict human errors in numerical typing," *IEEE Transactions on Human-Machine Systems*, vol. 45, no. 1, pp. 39–50, 2015.
- [19] J. de J Costa, F. Bernardini, D. Artigas, and J. Viterbo, "Mining direct acyclic graphs to find frequent substructures – an experimental analysis on educational data," *Information Sciences*, vol. 482, pp. 266–278, 2019.
- [20] Y. Gu, X. Zhang, Z. Liu, and F. Ren, "BeSense: leveraging WiFi channel data and computational intelligence for behavior analysis," *IEEE Computational Intelligence Magazine*, vol. 14, no. 4, pp. 31–41, 2019.
- [21] I. H. Sarker, A. Colman, M. A. Kabir, and J. Han, "Individualized time-series segmentation for mining mobile phone user behavior," *The Computer Journal*, vol. 61, no. 3, pp. 349–368, 2018.
- [22] Y. Cui, "Intelligent recommendation system based on mathematical modeling in personalized data mining," *Mathematical*

Problems in Engineering, vol. 2021, Article ID 6672036, 11 pages, 2021.

- [23] S. Joo, K. Lu, and T. Lee, "Analysis of content topics, user engagement and library factors in public library social media based on text mining," *Online Information Review*, vol. 44, no. 1, pp. 258–277, 2020.
- [24] Z. Zhang, S. Chen, Y. Yu, and K. Tang, "A comparative study of aggressive driving behavior recognition algorithms based on vehicle motion data," *IEEE Access*, vol. 7, pp. 8028–8038, 2018.