Review of User Behavior Analysis Based on Big Data: Method and Application

Mimi ZHANG [1, a], Yan WANG [2, b], Jianping CHAI [3, c]

[1, 2, 3] Communication University of China, Beijing ,100024, China

[a]email: 13121100807@163.com

Abstract. This paper reviews the existing user behavior analysis systematically and summarizes the latest progress of user behavior analysis. It introduces the user behavior analysis from multiple perspectives such as concept, data mining algorithm, application of different fields. It analyses deeply the advantages and disadvantages in different use environment. What's more, this paper particularly introduces the wide use of user behavior analysis in the field of telecommunication and Internet, and emphasizes the necessity of the user behavior analysis in the field of radio and television.

Introduction

User behavior analysis builds user information view and finds the rules of user behavior from massive user behavior data. It helps enterprises better understand the users' preferences, developing the value of users and ultimately bringing more benefits to enterprises[1].

The development of computer and cloud computing technology have promoted the production of big data. The annual national viewing data volume reaches 3.2 PB. Some enterprises through the full use of the data and digging in the battle to win. The reason which Amazon can win in the book industry is that it deeply mines and analyses the vast information of user behavior. User behavior analysis help Amazon learn more about the user's preference and provide more targeted services.

With the development of the technology of data, the multidimensional mass data acquisition and storage technology has become maturing, the structured and unstructured user behavior data has a large increase, making user behavior theory and application research especially important. Based on it, this paper reviews the existing user behavior analysis systematically and focuses on analyzing the theory and application in the fields of telecommunication, Internet, radio and television.

Model and Data Source

A. Model

Data mining is more and more widely used in the user behavior analysis.The current algorithm model of user behavior analysis has decision tree, clustering, association rules algorithm, sequential pattern mining, network model based on radio basis function and factor analysis.

The reference [2] describes the ID3 algorithm in the application of user behavior analysis in detail. First it generated the decision tree and training model through learning the existing data and pruning constantly. Then it made the prediction according to the training model.

The characteristics of models are shown in TABLEI.

TABLE I   Comparison of user behavior analysis model

Model	Core Algorithm	Model features
Classification algorithm model	Decision Tree	Easy to understand and implement; Classification results with strong interpretability; Does not need complex data preprocessing
Clustering algorithm model	K-means	Scalability; Efficiency; Object within the class is close but dissimilarity between class object
Association rule model	Association rules algorithm	It can find the potential connection between two objects[3]
Sequential pattern mining	Time series analysis	It can find the potential connection in the sequence with certain order
Neural network model	Network Model based on RBF	The network structure is simple; Learning speed is fast;
Factor analysis model	Principal component analysis/Factor analysis	Reduced the index dimension; But it can represent the 85% of the information of original index

B. Data Source

There are two kinds of data for user behavior analysis, one is structured data and the other is unstructured data. Structured data can be represented with a flat structure, such as number and symbol. Unstructured data can't be represented with unified structure, such as text, image and sound. Most of data mining is structured data. Text information in unstructured data will be involved in text semantic analysis. The types and characteristics of data are shown in TABLEII.

TABLE II   Types and Characteristics of data

Data Type	Data Characteristics
Structured Data	Basic attributes: user name, gender, age, education, income, etc; User behavior attributes: viewing behavior data, call log data, and web browsing behavior data, etc
Unstructured Data	The text reviews data, video image data, etc

Application of User Behavior Analysis

User behavior analysis is widely applied in many areas of industry, it's represented by telecommunication, Internet and radio and television.

A. Telecommunication User Behavior Analysis

User behavior analysis system in the field of telecommunication includes data mining analysis function layer, data mining process, model management layer and user behavior analysis function layer. Data mining analysis function layer provides the basic data mining analysis, including data preprocessing, classification and prediction, cluster analysis, social network analysis, association mining, sequence mining and figure mining. Data mining process and model management layer make the process of data mining analysis, and memorize the analysis model. User behavior analysis function layer invokes the results of the data mining process and model management layer for each application directly. These applications include user lifetime analysis, customer segmentation, customer churn analysis, cross/improve sales, customer complaint analysis, users owe cost forecast, fraud detection, customer credit rating model, competitor analysis and customer consumption behavior analysis. The applications of user behavior analysis in the field of telecommunication are

shown in TABLEIII.

TABLEIII   Applications in the Field of Telecommunication

Application	Model of data mining methods
Customer lifetime value analysis	Clustering analysis
Cross-selling	Frequent pattern mining/Association rules analysis
Customer loss prediction/Earnings forecast analysis	Decision tree/Time series analysis
Phone fraud	Clustering analysis
Customer business set up	Decision tree
Customer segmentation	Clustering analysis

B. Internet User Behavior Analysis

The primary task to provide personalized service is user behavior analysis. Internet service providers can understand the user's preferences better through tracking and learning user's behavior. And then they can formulate the corresponding solutions to distribute the flow rate more reasonably and provide content more appropriately.

What's more, user's requirements to the service quality of the site are higher and higher. Network service management is more scientific. It needs to analyze characteristics of user's group and user's different interests. Therefore, the network user behavior analysis and research is becoming more and more important.

The conventional methods of user behavior analysis in the field of Internet are user trajectory analysis, user basic motion analysis, correlation analysis, the target vector analysis method and TOP analysis. It mainly includes two kinds of analysis. One is normal and abnormal behavior detection based on the IP address[4]. The other is significant user access foundation based on the log. It can learn user's access pattern and find the page which user is interested in[5]. The applications of user behavior analysis in the field of Internet are shown in TABLEIV.

TABLEIV   Applications in the Field of Internet

Application	Model of data mining methods
Distance education system	Association rules analysis/Cluster analysis/Sequence pattern mining
Personalized recommendation system	Association rules analysis/Cluster analysis/Sequence pattern mining
Traffic monitoring	K-means Clustering
Normal/abnormal network behavior monitoring	Network model based on RBF

C. Radio and Television User Behavior Analysis

Radio and television field is currently in the era of big data. TV resources are richer. Digital cable TV system under the current coding and modulation formats can reach 500-600 series of transmission capacity[6]. In the case of such a huge amount of data, user behavior analysis is necessary in order to ensure the targeted advertising and the results of personality show

recommendation system more accurately.

The television audience viewing behavior analysis is the foundation of the television media operation. It analyzes user data, broadcast data and other supplementary data, it can find audience's behavioral characteristics and then establish user's portrait. It divides the viewers into different groups according to the specific business goals, and provides targeted services to the viewers who have different characteristics.

User behavior analysis includes television audience research, viewing content analysis, viewing motivation analysis, viewing mode analysis, viewing location analysis and viewing time analysis[7]. The dimensions of viewers' research include the number of viewers, sex structure, age structure, education level and income level. The dimensions of viewing content analysis include the type of program and the content of program. At present, the television classification of CSM system divides TV programs into 15 categories, 83 subclasses. The categories of programs include news/events, variety, sports, children's programs, music, TV shows, movies, economics, service life, the rule of law, etc. Personal viewing preferences, program promotion and viewing habits are the factors of viewing motivation. The following mainly introduces the research of viewing preferences and viewing habits.

User preference analysis includes the identification of personal viewing preference and the viewing preference classification research. User viewing habit analysis includes user daily viewing analysis, the identification of personal viewing habit and viewing habit classification research. User daily viewing analysis includes the trajectory of broadcast channel and pay channel. The important things of the viewing preference/habit classification research are to establish the optimal classification number and to describe the grading characteristics, finally it is concluded the preference/habit labels of program type.

The main algorithm in the process of user behavior analysis is factor analysis and K-means clustering. It selects user's viewing hours, the ratio of viewing time, broadcast time, days, viewing stability index and do factor analysis. Then it gives individual user preference model by calculating user's score on each factor and the comprehensive factor. It selects the original variables based on the results of the factor score. Then it is concluded the grade of user's preference/habit for the program. It uses RFM model to simulate the original index classification before clustering. The applications of user behavior analysis in the field of radio and television are shown in TABLEV.

TABLEV   Applications in the Field of Radio and Television

Application	Model of data mining methods
Radio and television public opinion analysis	Time series analysis/Clustering algorithm
Customer segmentation	RFM model/Kohonen neural network
Customer loss analysis	Clustering algorithm/Association rules analysis
Personality shows recommended	Based on the content recommendation algorithm/Collaborative filtering algorithm
EPG menu/VOD/Boot accurately targeted advertising	Time series analysis/Clustering algorithm
Channel/Program flow analysis	PageRank algorithm
Program evaluation	Text analysis/Hierarchical clustering method
Channel packaging sale	Association rules analysis

Conclusion

This paper introduces the concept of user behavior analysis, the common algorithms of data mining model and the application in different fields. User behavior analysis has the widespread application and achieves good results in fields such as telecommunication and Internet. Enterprises make the enormous growth of the business turnover and become stronger in the competition. But user behavior analysis in the field of radio and television has not yet built the complete theory system. The extension and application are still in the stage of exploration. The previous method of user behavior analysis is sampling survey. But today it is based on big data. With the improvement of the data collecting technology of the two-way user viewing data, the technology of collection and transmission based on the digital boxes have been achieved real-time viewing data acquisition and storage. User behavior analysis which uses real-time data of whole samples overcomes the one-sidedness of sampling survey and static resistance. It has important practical significance for the analysis of user's individual viewing preferences and habits, as well as the analysis of characteristics of user's group. User behavior analysis in the field of radio and television is at the start-up stage but grows in a high speed. Meanwhile, digital cable television fundamentally has changed the way television's profit model. Digital TV program must be guided by the user[8].The core to build the value-added services business model is user.

References

[1] Tengjiao Wang, Ziyu Lin. The application of data mining in the field of telecommunication customer behavior analysis. [J]. telecommuni cationmunication Technology. 2008,1.

[2] Jianjun Du, Haiyu Li, Rong Ma. Decision tree in data preprocessing and the application of the customer behavior analysis. [J]. Infortionmation Techonology.2008,12.

[3] Hongyan Wang, Daiwen Wu. Numerical attributes of association rule mining algorithm. [J] .Information Technology.2012,2.

[4] Xiaomei Zhang. Meteorological video information management system research and design of multimedia data. [D] .Inner Mongolia university. 2006.

[5] Fuqiang Dong, Li Ma,Bo Wu.A kind of user behavior classification method based on the Internet and research of the model. [J] .The modern electronic technology. 2002,22.

[6] Fengjie Li. Digital cable TV to business development in the future. [J].The cable TV of China. 2010,4.

[7] Bo Xu, Jiekui Zhang, The user interest model based on behavior analysis. [J]. Intelligence magazine. 2009,6.

[8] Wenqing Li. Television audience behavior analysis. [J]. Science and technology innovation review. 2012,28.