

Research Article

Animation User Value Portrait Based on RFM Model under Big Data

Qifang Liu¹ and Wuying Deng² 

¹Haikou University Economics, Haikou, Hainan Province, China

²School of Art & Design, Changsha University, No. 98 Hongshan Road, Kaifu District, Changsha, Hunan Province, China

Correspondence should be addressed to Wuying Deng; dwy@ccsu.edu.cn

Received 27 April 2022; Revised 5 June 2022; Accepted 8 June 2022; Published 11 July 2022

Academic Editor: Naeem Jan

Copyright © 2022 Qifang Liu and Wuying Deng. 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.

The development and application of big data technology has expanded the sources and methods for enterprises in the animation industry to obtain data, which provided them with the opportunity to obtain more user samples and solved the computing and storage problems faced by enterprises with massive data. By using statistical analysis and data mining methods for modeling, the user portrait of each user is depicted in an all-round and three-dimensional manner. Based on this background, this study proposes the topic of user portrait modeling based on the Reference Forward Model (RFM) model under big data. In this study, the animation user portrait modeling based on the RFM model under big data is firstly summarized: collect and summarize the behavior data of animation client users, prepare the data, and use entropy and Pearson correlation coefficient methods for data processing. And, then the weight of RFM is calculated through AHP (Analytic Hierarchy Process), and the behavior data collection method for animation user portrait modeling is described. Finally, based on the model tag in the constructed user portrait tag system, the RFM model under big data is used to analyze and model from multiple dimensions. In particular, in the algorithm model of animation user recognition, the weight and value of RFM are calculated to obtain the user value, and the results of the animation user portrait are summarized. Experiments proved that based on RFM Model under Big Data can identify animation users more accurately.

1. Introduction

Users are becoming increasingly reliant on the network due to the rapid advancement of mobile Internet technology. The amount of organized, semistructured, and unstructured data in the network is growing at an exponential rate, ushering in the era of big data [1]. In the face of massive data in the era of big data, how to fully mine the value behind massive data, deeply understand user needs, and improve marketing efficiency has gradually become an important target of enterprises. Using the k-means algorithm of cluster analysis to build a customer of animation enterprises portrait model can improve the efficiency of customer management and reduce the management cost of enterprises [2, 3]. User portraits accurately reflect the attributes and preferences of target users, enabling enterprises to target the characteristics and preferences of target users. And, further adjust the

positioning of enterprise products in the animation industry, design more suitable products for users, and improve user experience [4]. By analyzing customers' consumption, behavior habits, and basic data, enterprises use user portraits to obtain users' consumption level, brand preference, and other information for accurate advertising to improve the marketing effect of animation enterprises [5]. At the same time, user portrait can also help enterprises improve product operation, optimize the process and experience of interaction with users, and enhance the user engagement of the platform [6, 7]. The user portrait accurately reflects the attributes and preferences of target users, enabling enterprises in the animation industry to further adjust the positioning of their products according to the characteristics and preferences of target users, design more suitable products for users, and improve users' experience [8]. At the same time, user portrait can assist animation companies in

TABLE 1: Indicators of RFM model and meanings.

Indicators	Category	Meanings of indicators
R	Recency	The last time a consumer made a purchase
F	Frequency	The number of purchases made by a consumer over a certain period of time
M	Monetary	The total amount of money purchased by a consumer over a given period of time

improving product operation, improving the process and experience of user interaction, and increasing platform user engagement. User portrait is based on the analysis of user behavior data to achieve accurate marketing of animation industry enterprises as the ultimate goal [9]. Enterprises need to be able to clearly understand user needs to guide the derivation. The customers of animation enterprises are mainly young people, and the needs of young people are characterized by instability and randomness, which requires animation enterprises to apply big data for user portrait, and improvement of products, and further implement personalized recommendation [10, 11]. Therefore, the construction of user portrait is of great significance to the steady development of animation enterprises. Its strong technical characteristics are also one of the important means for animation enterprises to master competitiveness and face peer competition.

The paper's organization paragraph is as follows: the RFM is presented in Section 1. Section 2 analyzes the big data collection of animation users of the proposed work. Section 3 discusses the RFM model of animation user value portrait in detail. Finally, in Section 4, the research work is concluded.

2. RFM Model

RFM model is a classical quantitative analysis model of customer relationship management. Hughes, an American database marketing expert, proposed it in 1994. He believed that customers might be classified into three groups based on their purchasing habits [12]. From a theoretical perspective, the smaller R means that such consumers have more purchasing interest and demand for enterprise products; the larger F means that consumers are more inclined to purchase enterprise products; the larger M value means that consumers have a higher contribution [13]. Indicators of the RFM model and meanings are shown in Table 1.

3. Big Data Collection of Animation Users

3.1. User Behavior Data. The behavioral data of animation users come from a specific animation platform. It collects and summarizes the behavioral data of animation client users [14] by using IMEI (International Mobile Equipment Identity) as the unique identifier in the way of embedded SDK (Software Development Kit) of animation client. According to statistics, there were around 9,352,700 active users on the client in the fourth quarter of 2021, and the entire amount of user behaviour data were approximately 2100TB. In order to support the collection of massive user behavior data, a data acquisition system with high concurrency and easy expansion is designed to realize the

collection and parallel processing of massive user behavior data [15]. The definition of event interaction interface parameters is shown in Table 2.

In order to improve the processing capacity of massive user behavior data, the load balancing server at the data receiving end routes the client requests to multiple HTTP Servers. Then, each HTTP Server implements the parsing, cleaning, and conversion of the parameters in the request. The user behavior data are saved in the local server according to the set entries, so as to realize the safe storage and convenient processing of massive requests [16]. Finally, the consumer consumes the user behavior data from the Kafka cluster and persists it to the distributed file system HDFS (Hadoop Distributed File System) [17].

The biggest feature of big data is massive data, in view of the massive collection of user behavior data, it is necessary to further summarize and count data of user behaviors [18]. In order to support the parallel processing of massive data, and make full use of the characteristics of high throughput and easy expansion of distributed computing. In order to ensure the robustness and accuracy of the final output strategy, this study summarizes and counts the user behavior data within 30 consecutive days to obtain a comprehensive description of the user behavior.

3.2. Data Preparation. Relevant user characteristics are extracted from the broad table based on animation business scenes and integrated with relevant research on the features of user behavior in animation scenes. Then, the original user information table I_0 is formed. Indicator description of the original user information table I_0 is shown in Table 3.

3.3. Data Preprocessing. As there are different correlations among the indicators in the original user information table I_0 , some noise indicators may have to be negated effects on the effectiveness of discrimination [19]. Therefore, the entropy method and Pearson correlation coefficient method are used to screen the core indicators. The specific steps are as follows.

3.3.1. Data Standardization. To avoid the influence of extreme values and different dimensions among indicators, logarithmic transformation of data is required first, as shown in the following formula:

$$x'_{ij} = \lg(x_{ij} + 1). \quad (1)$$

Here, x_{ij} is the j indicator value of the i user. Then, the forward range standardization method is used to standardize the data, as shown in formula:

TABLE 2: Definition of event interaction interface parameters.

Parameter	Data type	Mandatory or not	Description
OS	String	Yes	System model
UA	String	Yes	User agent
channelID	String	Yes	Channel number of the client
Version	String	Yes	Client version number
IMEI	String	Yes	Terminal device number
bearType	String	No	Network type
sessionID	String	No	Session ID
accountName	String	No	User name
recordTime	String	Yes	Event time
Params	String	Yes	Event parameters
exColumn	String	No	Extension field
eventType	Int	Yes	Event type
eventID	Long	Yes	Event ID

TABLE 3: Indicator description of the original user information table I_0 .

Indicator	Indicator description
SDT_LAUNCH_BUSY	Number of startup times per busy day
SDT_LAUNCH_FREE	Number of startup times per free day
SDT_ONLINE_BUSY	Total number of viewing times during busy hours
SDT_ONLINE_FREE	Total number of viewing times during free hours
SDT_DL_NUMS	Number of downloads per day
SDT_BUSY_DL_NUMS	Average number of downloads during busy hours
SDT_FREE_DL_NUMS	Average number of downloads during free hours
SDT_BUSY_GOOD_NUMS	Average number of likes during busy hours
SDT_FREE_GOOD_NUMS	Average number of likes during free hours
SDT_GOOD_NUMS	Total likes per day
SDT_BUSY_TOPIC_NUMS	Average number of topics published during busy hours
SDT_FREE_TOPIC_NUMS	Average number of topics published during free hours
SDT_ALL_TOPIC_NUMS	Daily number of topics created
SDT_BUSY_SHARE_NUMS	Average number of times of sharing during busy hours
SDT_FREE_SHARE_NUMS	Average number of times of sharing during free hours
SDT_SHARE_NUMS	Number of shares per day
SDT_BUSY_COMM_NUMS	Average number of comments during busy hours
SDT_FREE_COMM_NUMS	Average number of comments during free hours
SDT_COMM_NUMS	Total daily comments
SDT_ACFUN_NUMS	Total daily number of bullets
SDT_SHARE_PLATFORM_NUMS	Number of shared platform types per day
SDT_ACFUN_OPUS_NUMS	Daily number of shells
SDT_MONSIX_APPLY_PRODUCT_NUMS	Daily number of member package products ordered
SDT_JOIN_CIRCLE_NUMS	Number of circles joined daily
SDT_EACH_CARTOON_ACFUN_NUMS	Average number of bullets per work per day

$$x_{ij}^h = \frac{x_{ij}^h - \min(x_j^h)}{\max(x_j^h) - \min(x_j^h)} \quad (2)$$

A new data information table I_1 is formed. At this time, information table I_1 still contains n users and d_1 indicators.

3.3.2. Use Entropy Method. The entropy method is used to select indexes with large information content. Formula (3) is first used to calculate the probability $f(x_{ij}^h)$ of x_{ij}^h under index j .

$$f(x_{ij}^h) = \frac{x_{ij}^h}{\sum_{i=1}^n x_{ij}^h} \quad (3)$$

Then, formula (4) is used to calculate the entropy value e_j of index j .

$$e_j = \frac{-\sum_{i=1}^m f(x_{ij}^h) * \ln(f(x_{ij}^h))}{\ln(m)} \quad (4)$$

Since the amount of information is inversely proportional to the amount of entropy e_j , the part of an index with a lower entropy value is selected to obtain the new user information table I_2 .

3.3.3. Pearson Correlation Coefficient. Pearson correlation coefficient was used to eliminate some indexes with strong correlation [20]. Indicators are screened by calculating the

TABLE 4: Degree of linear correlation between indicators.

$ r $	$ r < 0.3$	$0.3 \leq r < 0.5$	$0.5 \leq r < 0.8$	$0.8 \leq r < 1$	$ r = 1$
Related degree	Minuent	Moderate to low	Moderate	High	Completely

TABLE 5: Definition of indicator in table I₃.

Indicator	Indicator description	Entropy
SDT_LAUNCH_BUSY	Number of startup times per busy day	0.972
SDT_LAUNCH_FREE	Number of startup times per free day	0.981
SDT_ONLINE_BUSY	Total number of viewing times during busy hours	0.736
SDT_ONLINE_FREE	Total number of viewing times during free hours	0.754

TABLE 6: Judgment matrix.

Judgment matrix	R	F	M
R	1	2	3
F	1/2	1	2
M	1/3	1/2	1

Pearson correlation coefficient, which is shown in the following formula:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (5)$$

Here, r is the correlation coefficient, x and y , respectively, represent two different indexes of the i sample in sample I₂, \bar{x} and \bar{y} are the observed mean values of indicators x and y , respectively. The degree of linear correlation between indicators is shown in Table 4.

Indexes with large entropy values in the highly correlated and perfectly correlated index groups ($|r| \geq 0.8$) are removed, and a new data information table I₃ is obtained. The definition of indicator in table I₃ is shown in Table 5.

The method of normal distribution random interpolation is used to fill the empty value of the field randomly [21]. Then, the new user information table I₄ is got.

4. RFM Model of Animation User Value Portrait

4.1. RFM Weight. The analytic hierarchy process (AHP) is used to determine the weight of the RFM index. 9 importance grades and their assigned values are given by Saaty were used to construct the discriminant matrix [22]. The judgment matrix is shown in Table 6:

The sum and product method is used to sort the discriminant matrix, and finally, the normalized vector W and the maximum characteristic root λ are obtained.

$$W = (0.5318, 0.2964, 0.1627)^T, \quad (6)$$

$$\lambda = 3.0075. \quad (7)$$

The consistency indicator CI is calculated.

$$CI = \frac{\lambda - n}{n - 1} = \frac{3.0075 - 3}{3 - 1} = 0.0049. \quad (8)$$



FIGURE 1: Data flow of animation user value analysis model.

TABLE 7: Scoring rules for all dimensions of RFM model.

Dimension	Score	Upper limit	Lower limit
R	1	>0	<1
	2	≥ 0	≤ 1
F	1	≥ 0	<0.11
	2	≥ 0.11	<0.18
	3	≥ 0.18	<0.25
	4	≥ 0.25	<0.50
	5	≥ 0.50	≥ 0.13
M	1	≥ 0	<0.05
	2	≥ 0.05	<0.08
	3	≥ 0.08	<0.13
	4	≥ 0.13	<0.25
	5	≥ 0.25	≥ 0.13

Because CI is not equal to 0, the consistency check fails. Therefore, it is necessary to further calculate the consistency ratio CR .

$$CR = \frac{CI}{RI} = \frac{0.0049}{0.596} = 0.0077. \quad (9)$$

The weights of R , F and M in the RFM model are 0.5318, 0.2964, and 0.1627, respectively.

4.2. RFM Value Degree. Based on the user's viewing data, the RFM node in SPSS Modeler software is used to calculate the user's RFM value. Set R , F , and M as three dimensions with weights of 0.5318, 0.2964, and 0.1627. The data flow of the animation user value analysis model is shown in Figure 1.

After the data flow runs, view the RFM analysis node. Scoring rules for all dimensions of the RFM model are shown in Table 7.

The RFM value v_i of the i user is calculated.

$$v_i = R_i w_R + F_i w_F + M_i w_M. \quad (10)$$

Here, R_i , F_i , and M_i are the scores of the user in R , F , and M dimensions, respectively. w_R , w_F and w_M are the weights of R , F , and M , respectively.

TABLE 8: Animation user value segmentation rules.

Serial number	R		F		M		User type
	R	Compared to the mean	F	Compared to the mean	M	Compared to the mean	
1	Long	>	Low	<	Low	<	Unworthiness
2	Long	>	High	>	Low	<	Generally retain
3	Long	>	Low	<	High	>	General user
4	Short	<	High	>	Low	<	General important
5	Long	>	High	>	High	>	Important retain
6	Short	<	Low	<	High	>	Important value
7	Short	<	Low	<	Low	<	Important development
8	Short	<	High	>	High	>	Important keep

TABLE 9: Summary of characteristics of clustering results.

Type	Label	Summary of the characteristics
1	Unworthiness	The time interval from the last viewing is large, the viewing frequency is small, and the total viewing time is long
2	Generally retain	The time interval from the last viewing and the total viewing time are small, and the viewing frequency is large
3	General user	The time interval from the last viewing, the total viewing time, and the viewing frequency are small
4	General important	The time interval from the last viewing, the viewing frequency, and the total viewing time are small
5	Important retain	The time interval from last viewing, viewing frequency, and total viewing duration are large
6	Important value	The time interval from the last viewing is small, the viewing frequency is small, and the total viewing time is large
7	Important development	The time interval from last viewing, viewing frequency, and total viewing duration are small
8	Important keep	The time interval from the last viewing is small, the viewing frequency, and the total viewing time are large

TABLE 10: RFM model intra-class distance and inter-class distance.

RFM model	
Intraclass distance	2684.53
Interclass distance	12904.11

4.3. *User Value.* After the value rating of animation users is obtained based on RFM model analysis, it is necessary to subdivide animation users according to their value, which is helpful for enterprises to carry out personalized precision marketing according to different user needs and user characteristics [23]. The standardized RFM index values of each user were compared with the overall mean value of this index. Animation user value segmentation rules are shown in Table 8.

4.4. *Result of Portrait.* A summary of characteristics of clustering results is shown in Table 9.

RFM model intraclass distance and interclass distance are shown in Table 10.

5. Conclusion

This study takes animation business as the background, based on the behavior data of animation users using the client to watch the animation, uses RFM under big data to conduct user portrait modeling and explores the method and detailed steps of user portrait construction.

- (1) The application of user time preference analysis in the new scene of the animation industry. The time

preference of animation client users is evaluated and modelled using past behavior data, and the preferred time period of each user is determined. The study of user time preferences is applied to the new field of animation big data. This enables enterprises to carry out marketing activities for individual users within the time range of user preference, which is beneficial to enterprises to improve user experience and marketing efficiency.

- (2) For enterprises in the animation industry, user portrait is the means, and precision marketing is the goal. The user portrait faithfully portrays each individual's entire image, while the label highly abstracts a specific attribute of the user. Enterprises can correctly identify target user groups by label selection and combination based on user picture labels created by themselves and conduct out marketing activities for target users, as part of the precision marketing process.

In this study, the modeling of animation user portrait is realized based on the business scene of animation, and the marketing strategy management platform is realized based on the user portrait. The construction of a tag system in user portrait modeling is mainly achieved through literature reading and market research, combined with the strategic goals and business scenes of the enterprise. However, the user portraits created in this study are not ideal. Because of limits in knowledge and ability, the interpretation of strategic objectives and comprehension of business logic are not always full, which inevitably leads to imperfect user pictures.

The process of creating user portraits is a long one. User portraits will be improved more in the future based on input from correct marketing based on user portraits.

Data Availability

The datasets used during the present study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

The study was supported by (1) Research on the Development of Hainan Animation and Comic Creative Industries—Based on Value Co-Creation Theory under Grant no. HNSK(ZC)21-160 and (2) Research on the Production and Communication of Network Animation Culture in the New Era From the Perspective of Value Cocreation, under Grant no. 2022YBA279.

References

- [1] B. Ma, F. Li, and G. Wang, "Stochastic RFM model and its application in retail customer value Identification," *Journal of Management in Engineering*, vol. 25, no. 1, pp. 102–108, 2011.
- [2] W. Wang, Y. Jin, and D. Chen, "Research on individual customer loyalty based on RFM model," *Financial BBS*, vol. 33, no. 3, pp. 75–80, 2012.
- [3] X. Lu, "Information intelligent matching based on supplier portrait," *China Tendering*, vol. 89, no. 05, pp. 165–167, 2022.
- [4] X. Xu, J. Wang, and H. Tu, "E-commerce customer segmentation based on improved RFM model," *Computer Applications*, vol. 32, no. 5, pp. 1439–1442, 2012.
- [5] Z. Liu and R. Yu, "Design of accurate advertising delivery system based on user portrait," *Jiangsu Science and technology information*, vol. 39, no. 10, pp. 69–72, 2022.
- [6] C. Liu, Q. Mei, and S. Cai, "Customer classification and value evaluation model based on RFM," *Technical Economics and Management Research*, vol. 12, no. 5, pp. 56–60, 2012.
- [7] M. Zhao and J. Qi, "Customer lifetime value based on RFM model of purchasing behavior and RFMP model of comment behavior," *Statistics and Information Forum*, vol. 29, no. 9, pp. 91–98, 2014.
- [8] B. Yang, "A two-stage customer association classification method based on RFM model data mining," *Statistics & Decisions*, vol. 25, no. 7, pp. 77–79, 2015.
- [9] L. Su, "Analysis of user behavior under big data," *Report of Science and Technology*, vol. 33, no. 5, pp. 129–133, 2017.
- [10] J. Yang, Z. Shi, and Z. Liu, "Simulation research on fast acquisition of terminal user information under big data analysis," *Computer Simulation*, vol. 35, no. 2, pp. 441–445, 2018.
- [11] Y. Zhao, F. Liu, and L. Luo, "Overview of user portrait research in the context of big data: knowledge system and research prospects," *Library Science Research*, vol. 63, no. 24, pp. 13–24, 2019.
- [12] A. Dudhia, "The reference forward model (RFM)," *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 186, pp. 243–253, 2017.
- [13] N. Zhang, C. Fan, and Y. Zhang, "A novel collaborative filtering personalized recommendation algorithm based on RFM model," *Telecom Science*, vol. 31, no. 9, pp. 103–111, 2015.
- [14] M. Hirai and K. Hiraki, "Inversion effect of human, robotic, and light-point walking animation: an ERP study(Summary of Awarded Presentations at the 24th Annual Meeting)," *Mechanics Research Communications*, vol. 28, no. 6, pp. 611–616, 2016.
- [15] J. Lv and F. Wei, "Analysis of network video user behavior under big data," *News Junkie*, vol. 07, no. 439, pp. 40–42, 2014.
- [16] W. Wang, X. Xia, and X. Li, "An interestingness model based on user behavior," *Computer Engineering and Applications*, vol. 48, no. 8, pp. 148–151, 2012.
- [17] Y. Ma, G. Li, and H. Xie, "Theoretical thinking on urban data portrait from the perspective of digital space," *Journal of Intelligence*, vol. 038, no. 001, pp. 58–67, 2019.
- [18] N. Guo, R. Wei, and Y. Shen, "Extraction of abnormal features in big data environment based on user portraits," *Computer Simulation*, vol. 37, no. 8, pp. 332–336, 2020.
- [19] X. Xue, "User portrait extraction and intelligent push system based on big data clustering," *Electronic Design Engineering*, vol. 30, no. 2, pp. 38–42, 2022.
- [20] D. Hao, W. Guan, and L. Zou, "Fast virtual grid matching localization algorithm based on Pearson correlation coefficient," *Computer Applications*, vol. 38, no. 3, pp. 763–768, 2018.
- [21] Y. Liu and T. Tao, "Study on the independence and irrelevance of random variables in normal distribution," *Journal of Shandong Agricultural University (Natural Science Edition): Natural Science Edition*, vol. 49, no. 2, pp. 339–341, 2018.
- [22] L. Chen and X. Yu, "A new sorting method of complementary judgment matrix," *Practice and Understanding of Mathematics*, vol. 46, no. 24, pp. 34–36, 2016.
- [23] N. Guo, Y. Shen, and Y. Han, "Simulation of dynamic cleaning method of inferior user data under big data analysis," *Computer Simulation*, vol. 45, no. 2, pp. 362–366, 2019.