Research on Factors Influencing User Experience of Beijing Mobile Based on GBTD Classification Model

Xiaojun Xu *

College of mechanical and electrical engineering, Nanjing Tech University Pujiang institute, Nanjing, China, 211222

* Corresponding Author Email: 1600879748@qq.com

Abstract. To improve the quality of user experience for Beijing Mobile, this article focuses on user evaluation, evaluates the influencing factors of user satisfaction, and determines its degree of impact. Firstly, this article preprocesses existing data and classifies it into scenarios. Secondly, descriptive statistical analysis is conducted on the data, and Outlier are studied. Finally, the GBTD classification model optimized by genetic algorithm is introduced and trained to determine the main factors of satisfaction and the importance percentage of each factor. The study found that the main factors affecting the satisfaction of Beijing Mobile's voice service are: "Total GPRS traffic", "MOU in the first three months", "ARPU in the current month", and "Network issues"; The main factors that affect the satisfaction of mobile internet business are: "Poor network signal/no signal", "redirect times", and "GPRS resource usage for the month".

Keywords: GBTD Classification Model, Genetic Algorithm Optimization, descriptive statistics, Mobile user experience satisfaction.

1. Introduction

With the rapid development of mobile communication technology, people are increasingly inseparable from the various conveniences brought by mobile communication technology. As network coverage continues to improve. Operators are also increasingly valuing their customers' online user experience, especially in today's era of information transparency and product homogenization. Customer satisfaction, as an indicator reflecting the difference between customer expectations and actual perceived product services, has become a crucial reflection of the market operation status of major operators. Based on customer complaints, solving problems that affect the user experience point by point is a traditional method of improving customer satisfaction. However, with the significant increase in the number of users, the variety of mobile products is becoming increasingly diverse, and customer demand is increasing. Traditional methods are no longer effective in improving customer satisfaction.

Under this condition, this article analyzes the influencing factors of user satisfaction based on subjective scenario data [1-2] and objective indicators such as GPRS total traffic [3], predicts user satisfaction scores, and provides quantitative analysis of the degree of influence of each factor.

2. Materials and Methods

2.1. Data sources and analysis

This paper uses the data *(https://www.saikr.com/vse/bigdata2022)* provided by Question B of MathorCup College Mathematical Modeling Challenge 2022 Big data Competition for analysis and research.

2.2. Data preprocessing

2.2.1. Classification of scene data described by users

In the given data, users have selected the option of "other" in their description of their own scenario classification and described it in text, but some of them are included in the previous 7 scenarios. For example, the scenarios described by the user are subway and high-speed rail, and this should be made earlier, and such situations should still be numbered 5 (subway) and 7 (high-speed rail); Some are complaints with emotional venting nature and cannot extract information about the scene, such as "not remembering", "many regions", "expensive fees", etc. These are uniformly numbered as 100 and these data are removed.

Classify the remaining 'other' scenarios in detail according to the user's description, with labels ranging from 8 to 28, as shown in Table 1. Based on this classification standard, re classify the 'other' scenario data described by the user.

Table 1. Corresponding Table of Label Meaning during Detailed Classification.

Number	Lable	Meaning
1	8	underground
2	9	elevator
3	10	On the way
4	11	of one's own unit
5	12	mountain area
……	……	……
17	24	Sports venues
18	25	Army Camp
19	26	restaurant
20	27	stairs
21	28	Under the bridge

2.2.2. Data deletion and supplementation

There are many missing values for 'Redirect Times' and' Redirect Residency Market ', which represent unknown situations and missing data [4]. Therefore, this field is deleted; the blank values of "whether it is a caring user" and "whether it has been to the business hall" represent "no" and should be supplemented completely; the blank values for indicators such as "poor internet quality times" and "disconnection times" are 0. Therefore, in Excel, use Ctrl+G to locate, fill in the blank values as 0, and then use Ctrl+Enter to replace all.

Except for the classification description of scenarios, which has significant classification value, the subjectivity of other user phenomenon descriptions is strong, and the majority of missing values are considered as not having classification statistical value. Therefore, they are deleted [5], similar to "APP category comments" and "APP subcategory video comments" The "APP subcategory game notes", "APP subcategory online notes", "Number resources - card issuance time", and "Number resources - activation time" have also been deleted due to too many missing values.

2.2.3. Conversion of categorical data

In order to facilitate later modeling and computation, it is necessary to quantitatively convert categorical data such as "4G" and "5G" [6]. Use Ctrl+F to search and replace in Excel, and the conversion rules are listed in Table 2 below.

Table 2. Corresponding Table for Conversion Rules of Classified Data and Quantitative Data.

Fields	Classified Data	Quantitative Data
4\5G User	2G	1
	4G	2
	5G	3
VIP	NO	1
	YES	2
4G network customer	NO	1
	YES	2
5G network customer	NO	1
	YES	2
Terminal Type	Realme	1
	Bubukao	2
	Hisense	3
	Black Shark	4
	Huawei	5
	LeTV Mobile	6
	Unicom	7
	Lenovo	8
	Meizu	9

2.2.4. Missing value processing

After completing the quantitative transformation of the categorical data, attachments containing all numbers were obtained, but it was found that there were still some empty values, which accounted for a small proportion and were unknown cases of missing data. Filling them in as an average can reduce the impact of the empty values.

2.3. Descriptive statistical analysis of data

Import the processed data into SPSSPRO for descriptive statistics, and the analysis results are shown in Table 3.

Table 3. Descriptive Statistical Analysis Results Table.

Variable name	Sample size	Max	Min	Average	Standard deviation	Median	Variance	Kurtosis	Skewness	Coefficient of variation(CV)
Network coverage and signal strength	5431	10	1	8.34	2.42	9	5.85	2.10	-1.69	0.29
Voice call stability	5431	10	1	8.43	2.37	10	5.63	2.49	-1.79	0.28
Voice call clarity	5431	10	1	8.63	2.20	10	4.85	3.65	-2.02	0.26
No Signal	5431	1	-1	-0.47	0.88	-1	0.78	-0.87	1.06	-1.88
Unable to dial with signal	5431	2	-1	-0.47	1.14	-1	1.30	0.91	1.71	-2.41
Sudden interrupt during the call	5431	3	-1	0.04	1.75	-1	3.07	-0.79	1.10	50.34
crossed line	5431	5	-1	-0.76	1.17	-1	1.36	20.44	4.74	-1.53
One party cannot hear during the call	5431	6	-1	0.87	3.10	-1	9.58	-0.89	1.06	3.57
Times of mos quality difference	5431	110	0	4.76	7.52	2	56.60	30.78	4.17	1.58
Number of disconnected and dropped calls	5431	86	0	0.86	2.59	0	6.70	271.58	11.74	3.03
ARPU of the current month	5431	892.7	0	85.08	68.50	68	4691.5	11.90	2.37	0.81
MOU of the current month	5431	29493	0	463.30	657.36	298	432127	722.89	18.45	1.42
MOU in the first three months	5431	31066	0	463.17	632.85	307	400494	1012.6	22.33	1.37
Total GPRS traffic	5431	1.89×10^9	0	1.80×10^7	3.24×10^7	1.20×10^7	1.05×10^{15}	2062.3	36.52	1.80

PS: MOU means average monthly call time per household. GBTD means gradient elevation decision tree. ARPU means average income per user which is used to measure telecom operators' business revenue. MOS represents mean opinion score.

Observing the coefficient of variation in the last column, we found that the coefficient of variation of the three items is less than 0.15, except for "crossed line", "unable to dial with signal" and "No signal", Outlier may exist in other influencing factors. However, these Outlier cannot be eliminated because Outlier still exist in the training model and subsequent prediction. If outliers are eliminated in the training model, the trained model will be effective only when there are no Outlier, and the prediction with outliers will be more inaccurate. Therefore, these Outlier cannot be eliminated but retained.

3. Model establishment and solution

3.1. GBTD Classification Model

The GBTD classification model is a lifting algorithm that adds up different decision trees and continuously reduces the deviation value to achieve the effect of gradient lifting [7]. Each new tree in the following is a negative gradient direction fitting the Loss function.

For the Boosting model, it is strengthened through weak learning and trained through weighting. The calculation formula is as follows:

$$F_m(x) = F_{m-1}(x) + \arg\min_h \sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) + h(x_i)) \tag{1}$$

The previous round of model $F_{m-1}(x)$ plus the previous model $h(x_i)$ need to reduce the loss energy before adding this new decision tree. Therefore, every additional decision tree will definitely be stronger than the previously trained one [8]. However, when the number of decision trees increases to a certain amount, the result basically fluctuates around the stable value [9].

3.2. User Satisfaction Analysis

Import the processed data into SPSSRPO for analysis, and select the GBTD classification method optimized by genetic algorithm for model training [10-11]. The parameter configuration of the model is shown in Table 4, and the feature importance is shown in Figure 1. From it, it can be found that the main factors affecting overall satisfaction with voice calls are: total GPRS traffic [15.1%], MOU in the first three months [11.7%], ARPU in the current month [11.7%], and Network issues [11.1%].

Table 4. Model parameter configuration table.

Parameter Name	Value
Training time	1min28.955s
Data segmentation	0.7
data shuffle	No
Cross validation	5
Loss function	deviance
Node splitting evaluation criteria	friedman_mse
Number of base learners	100
Learning rate	0.1
No return sampling ratio	1
Maximum feature ratio considered during partitioning	None
Minimum number of samples for internal node splitting	2
Minimum number of samples for leaf nodes	1
Minimum weight of samples in leaf nodes	0
The maximum depth of the tree	10
Maximum number of leaf nodes	50
Threshold for impure node partitioning	0

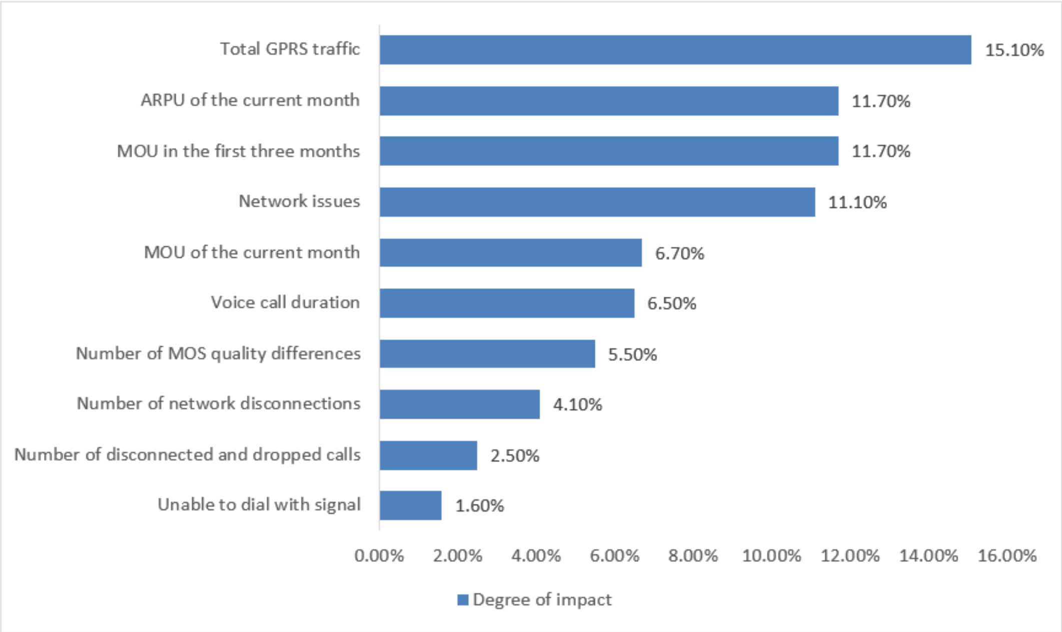

Figure 1. Bar chart of the importance of features affecting overall satisfaction with voice calls.

The model configuration shown in Table 4 is still used, and it can be found that the main factors affecting overall satisfaction with mobile internet access are: poor network signal/no signal [5.9%], Redirects [4.60%], and GPRS resource usage for the month [4.1%], as shown in Figure 2

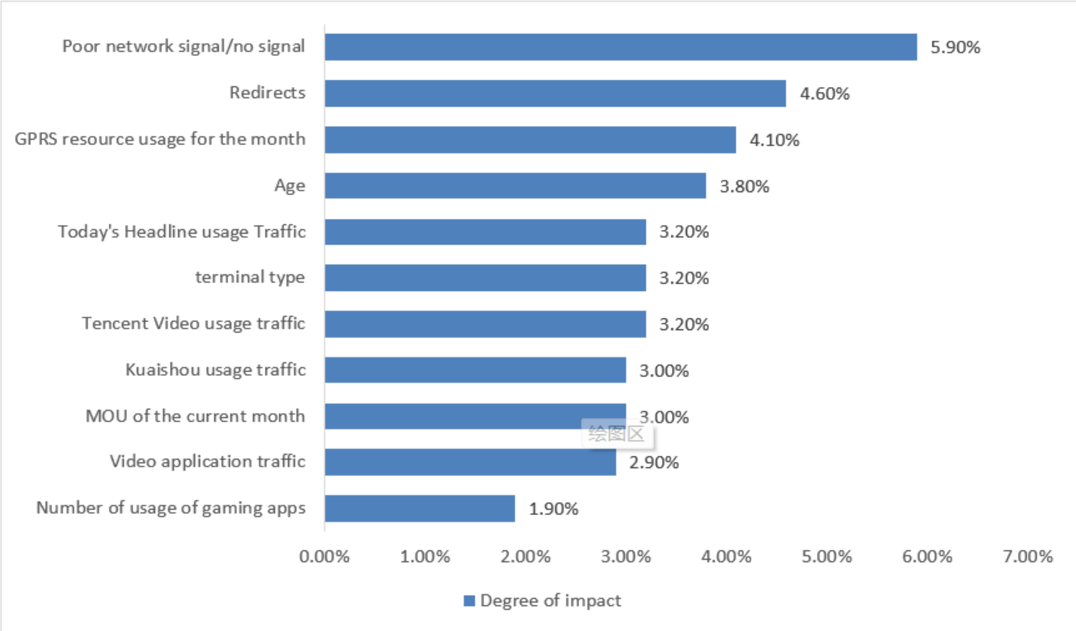

Figure 2. Bar chart of the Importance of Features Affecting the Satisfaction of Mobile Internet Service.

4. Conclusions

In the era of big data, the demand of mobile phone users for mobile communication has further expanded, and the use scenarios have become more complex. The traditional satisfaction survey improvement methods can no longer meet the current user needs. Under these conditions, this article takes into account various factors such as the subjective description of scene data by users and the objective indicator of GPRS total traffic, and establishes a GBTD classification model. On this basis, the initial population number is set to 50, the maximum number of iterations is 150, the mutation probability is 0.1, and the crossover probability is 0.5. Genetic algorithm optimization is carried out and model training is conducted. Finally, the main factors that affect satisfaction and the importance percentage of each factor were obtained. The main factors that affect Beijing Mobile's voice service satisfaction were determined to be: "Total GPRS traffic ", "MOU in the first three months", "ARPU in the current month", and "Network issues "; The main factors that affect the satisfaction of mobile internet business are: "poor/no network signal", "Redirects ", and "GPRS resource usage for the month". By analyzing the user experience, this model can guide operators to make targeted improvements and improve user satisfaction. At the same time, this model has a certain generalization ability, because it does not excessively eliminate Outlier, it has advantages in handling Outlier, and is competent for real scenarios. This is of great significance for predicting user satisfaction and improving the experience in a timely manner in the future.

References

[1] Yuan Xiaohua, Li Mingxia, Wang Lihong, etc Descriptive Statistical Analysis of Blood Composition Quality Sampling Data [C]//Chinese Blood Transfusion Association. Paper Album of the 9th Blood Transfusion Conference of the Chinese Blood Transfusion Association. Paper Album of the 9th Blood Transfusion Conference of the Chinese Blood Transfusion Association, 2018:64

[2] Zhang Dingding. Descriptive Statistical Analysis and SPSS Implementation [J]. Journal of Union Medical College, 2018,9 (05): 447

[3] Qin Yu GPRS data traffic analysis and its equivalent model research [D]. Beijing Jiaotong University, 2013

[4] Zhai Ruyu Descriptive statistics and visualization of flight delay data [D]. Heilongjiang University, 2019.

[5] Chen Mirong Emphasize the value of descriptive statistics in social science research [N] China Social Science Journal, 2022-05-25 (002).

[6] Zhang Wei, Yang Weisen, Bai Qian, et al. Study on the correlation between motor vehicle ownership and urban air quality in Xi'an based on descriptive statistics and grey correlation model [J]. Transportation Research, 2022,8 (03): 111-119.

[7] He Shijian Research on recommendation algorithm based on gradient lifting decision tree and deep belief network fusion [D]. Guangxi Normal University, 2017

[8] Qu Hongquan, Wang Zhengyi, Sheng Zhiyong, et al. Fiber intrusion signal classification based on gradient lifting decision tree algorithm [J]. Progress in Laser and Optoelectronics, 2022,59 (23): 106-113

[9] Ou Dongxiu, Zhang Xinyin, Zhao Yuan, et al. Prediction of Train Emergency Delay Time in Urban Rail Transit Based on Gradient Enhancement Decision Tree Cascading Classification Method [J]. Urban Rail Transit Research, 2022-25 (10): 65-70.

[10] Ma Longfei, Xiao Hanmin, Tao Jingwei, et al. An Intelligent Classification Method for Lithology Based on Gradient Enhancement Decision Tree Algorithm [J]. Oil and Gas Geology and Recovery, 2022, 29 (01): 21-29.

[11] Liu Chunlei, Liang Ruisi, Di Yuanhao. Research on Short Text Classification Based on TFIDF and Gradient Enhancement Decision Tree [J]. Science and Technology Wind, 2019 (24): 231.