Design Engineering

ISSN: 0011-9342
Issue: 1 | Pages: 33 - 41

Application of SPSS Data Analysis in Base Station Construction

Haojie Du, Kuohu Li*, Xiaoyan Song, Huanli Zhao, Ning Li, Xueqing Wang, Xichang Xue, Xianting Sun, Zexun Geng

Pingdingshan University

*Corresponding Author: Kuohu Li

Abstract:

This paper establishes the data analysis model through the actual communication data of the youth mobile phone, and carries out the SPSS simulation. The results show that analyzing the existing communication data can not only predict the development trend, but also avoid the harm and reduce the cost, and guarantee the high level with the minimum investment. Benefits can provide a useful reference for follow-up work and have important guiding significance.

Keywords: *SPSS, Data analysis.*

I. INTRODUCTION

With the rapid development of communication technology, how to update and replace the old communication facilities and expand new communication services is a difficult problem faced by many communication companies.

The main problems studied in this paper are as follows: 1) With the given call record information, a mathematical model is established to classify the subscribers. 2) If a new communication service needs to be launched, how to reasonably select some subscribers as the preferred promotion group and why. 3) Whether the construction of the company's communication facilities (such as base stations) is reasonable, and how to improve it.

II. MATHEMATICAL MODELLING

2.1 Hypotheses and Explanation of the Model

Hypothesis 1: Assuming that the load of each base station (30 base stations as an example) is measured by the number of on-line subscribers at the same time, and the maximum number of on-line subscribers allowed by 30 base stations at the same time is the same.

Hypothesis 2: 12:00 p.m. to 8:00 a.m. is a time period subscribers enjoys cheap rate for one day.

Hypothesis 3: The differences between roaming, long distance and short distance are ignored.

ISSN: 0011-9342
Issue: 1 | Pages: 33 - 41

Hypothesis 4: Base station with utilization rate below 1% or above 10% is considered unreasonable.

Hypothesis 5: Supposing that each base station has a signal impact on several nearby base stations.

2.2 Analysis Problem

For problem I: The subscribers are classified according to the call information [1], taking into account their minutes of calls per month, proportion of callings, proportion of inter-network calls, and proportion of calls with cheap rate, etc. Since the data given in this study do not involve the proportion of inter-network calls, three indicators are determined when classifying subscribers: telephone traffic, proportion of callings and proportion of calls with cheap rate.

EXCEL is used to process the data, calculate the three indicators of each subscriber within 10 days, based on which the subscribers are classified. In this paper, the period for cheap rate is defined from 12:00 p.m. to 8:00 a.m. Then, SPSS software and K-means fast clustering analysis were used to classify the subscribers and to test, and it was found that the classification was not significant when three indicators were considered at the same time. Elimination of variables was used to eliminate the proportion of calls with cheap rate. The other two indicators were used to classify, and the situations in which subscribers were divided into Class II, III and IV were compared and analyzed. Consequently, the optimal classification situation was obtained, and the classification criteria were obtained. According to the classification criteria, they were named as ordinary subscribers, value subscribers and gold subscribers.

For problem II: According to the three indicators put forward in problem I, through data analysis, it was found that 80% of the 150 people selected had calls with cheap rate. So a new business under consideration was Night Chatting Package, but the potential value of this business and its promotion needs to be determined. The time prediction method was used to forecast the night and day calls in the next five years, and to observe their changing trends, and then to judge whether the business had a development prospect. Based on the above analysis, some suggestions were given.

For problem III: Suggestions on whether the construction of the base station was reasonable were given [2]. Two indicators of rationality of base station construction: overall balance and base station utilization. Considering Newton's successive differential method, according to the 10-day traffic order of each base station, then the balance degree is defined by subtracting the ratio between the average and the volume of 15 base stations with larger call volume from the volume of 15 base stations with smaller call volume. Base station utilization rate is the ratio of 10-day utilization time to total time (best described in mathematical expressions).

Base stations with a utilization rate of less than 1% were considered likely to be dismantled (assuming that the dismantlement of the base station has no impact on subscriber communications). After deciding which base stations would be dismantled, some additional time should be taken to consider which base stations may be overburdened at present and build new base stations around. After the dismantlement of the base station and the construction of the new ones were determined accordingly, the utilization rate and overall equilibrium of each base station

ISSN: 0011-9342
Issue: 1 | Pages: 33 - 41

at this time would be compared to judge whether they were reasonable or not. If not, adjustment should be made according to the above standards till the construction of base station was reasonable.

2.3 Symbol Description

TABLE I. Symbol Description in this paper

T	Talk time
N	User number
M	Number of indicators determined by each use
xi	User i
xij	Ith user jth indicator
Yk1	Preferential call volume on day K
Yk2	Non-preferential call volume on day K
zp	Total call volume of base station P during June 1-10
A	Equilibrium of Call Quantity of Base Stations
Bp	Utilization Rate of the p-th Base Station
M	Maximum number of online users allowed by each base station at the same time
Npt	Maximum number of on-line users of the p-th base station in the t-th period

III. RESULTS OF SIMULATION AND EMULATION

3.1 Solution to Problem I

3.1.1 Concepts and Definitions

Telephone traffic A: A=Ct, Where, A=telephone traffic; C=number of calls; t=average duration of each call.

Proportion of callings= number of callings/ number of calls, number of callings + number of incoming calls= number of calls, number of calls= number of communications (if every call is connected). Lower proportion of callings indicates that the subscriber only uses the number passively, and he may be lost after the number conversion cost is reduced or the contract expires.

Proportion of calls with cheap rate= number of calls with cheap rate/ total number of calls [3]. There may be a period with cheap rate in some brands when subscribers enjoy a very cheap rate,

ISSN: 0011-9342
Issue: 1 | Pages: 33 - 41

but in other time, the proportion of calls is very small, which can be assumed that the rate in other time of such brand is not attractive to the subscribers who will not be subscribed if there is a cheaper service.

3.1.2 Statistical Description of Indicators

TABLE II. Statistical Description of Indicators

	MAX	MIN	AVERAGE	STANDARD DEVIATION	SUBJECT TO NORMAL DISTRIBUTION
TRAFFIC	6980	232	3775.72	1265.289297	Y
CALLING RATIO	1	0.065728	0.59156682	0.220086554	N
PREFERENTIAL TIME CALL RATIO	0.5	0	0.094043154	0.069733706	N

3.1.3 Application of K-Mean Clustering Method in this Problem

3 indicators were selected: telephone traffic, proportion of callings, proportion of calls with cheap rate, m=3; 150 subscribers were selected for cluster analysis, n = 150.

3.1.4 The Answer and Test of Problem 1

Three indicators were classified by SPSS software, and the results are as follows:

Table III. The results of three indicators

	CLUSTERING		ERROR		F	SIG.
	MEAN SQUARE	DF	MEAN SQUARE	DF		
CALLING RATIO	.670	2	.044	297	15.130	.000
TRAFFIC	1.965E8	2	288251.752	297	681.827	.000
PREFERENTIAL TIME CALL RATIO	.014	2	.005	297	2.935	.055

It can be seen from the table III that the significant level value of proportion of calls with cheap rate is 0.055 > 0.05, non-significant. Then, the proportion of calls with cheap rate was considered to be eliminated. The other two indicators were used to classify. The test results when they were classified into two classes, three classes and four classes are as follows:

It is found by comparing and analyzing the above three tables(IV, V,VI) that the error of proportion of callings when there are three classes is the minimum, in addition to the fact that the subscribers are usually divided into three classes by communication companies, the subscribers were divided into three classes herein.

TABLE IV. Divided into 2 categories

ISSN: 0011-9342
Issue: 1 | Pages: 33 - 41

ANOVA						
	CLUSTERING		ERROR		F	SIG.
	MEAN SQUARE	DF	MEAN SQUARE	DF		
CALLING RATIO	1.115	1	.045	298	24.864	.000
TRAFFIC	3.067E8	1	577079.217	298	531.498	.000

TABLE V. Divided into 3 categories

	CLUSTERING		ERROR		F	SIG.
	MEAN SQUARE	DF	MEAN SQUARE	DF		
CALLING RATIO	.670	2	.044	297	15.130	.000
TRAFFIC	1.965E8	2	288251.752	297	681.827	.000

TABLE VI. Divided into 4 categories

ANOVA						
	CLUSTERING		ERROR		F	SIG.
	MEAN SQUARE	DF	MEAN SQUARE	DF		
CALLING RATIO	.390	3	.045	296	8.679	.000
TRAFFIC	1.431E8	3	166689.648	296	858.574	.000

The clustering standard under such classification is:

The table VII shows that the subscribers in Class 1 have high proportion of callings and telephone traffic, those in Class 2 have the lowest proportion of callings and telephone traffic, hence Class 1 was identified as the golden clients and Class 2 as ordinary, and Class 3 as value.

TABLE VII. Clustering standard

ISSN: 0011-9342
Issue: 1 | Pages: 33 - 41

FINAL CLUSTER CENTER			
	CLUSTERING		
	1	2	3
CALLING RATIO	0.6910	0.5098	0.6083
TRAFFIC	5561.89	2420.70	3989.24

3.2 Solution to Problem II

Through the analysis of the data, it is found from the table VI that about 85.5% of the people have the experience of making phone calls at night, assuming that the new business is mainly aimed at night calls.

3.2.1 Establishment of Model II

The time series prediction method was used to predict the day and night traffic in the next 5 days. The time series prediction model is as follows.

Since the night and day traffic fluctuate at a certain level, the simple moving average method can be used to predict the night traffic in the next few days, so as to judge whether the service has a prospect.

Simple moving average method:

Given: the observation sequence is $y_1, y_2, \cdots y_T$, , number of moving average term N<T [4]. The formula for calculating the simple moving average is as follows:

$$M_t^{(1)} = \frac{1}{N}(y_t + y_{t-1} + \cdots + y_{t-N+1})$$

$$= \frac{1}{N}(y_{t-1} + \cdots + y_{t-N}) + \frac{1}{N}(y_t - y_{t-N}) = M_{t-1}^{(1)} + \frac{1}{N}(y_t - y_{t-N})$$

$$\hat{y}_{t+1} = M_t^{(1)} = \frac{1}{N}(\hat{y}_t + \cdots + \hat{y}_{t-N}), T = N, N+1, \cdots,$$

The standard error of prediction is as follows:

$$S = \sqrt{\frac{\sum_{t=N+1}^{T}(\hat{y}_t - y_t)^2}{T-N}}$$

When predicting night and daytime calls, N = 2, 3, 4 was chosen, and then the comparison was made to judge which prediction standard error is the smallest, which was the optimal number of moving items.

3.2.2 Solutions to Model II

It was predicted that from June 11 to 15 of a certain year, the number of calls with cheap rate was 9,394.5, 9,394.5, 9,742.5, 9,510.6 and 9,549.3, and that in other time period was 94,313, 95,420, 95,320, 95,018 and 95,253.

Within 10-15 days, the night call volume showed a slight upward trend while the day call volume showed a slight downward trend, so it was judged that this kind of service has a prospect. Therefore, it is possible to propose a similar "rate for night calls" for some subscribers with relatively high night calls.

ISSN: 0011-9342
Issue: 1 | Pages: 33 - 41

3.3 Solution to Problem III

3.3.1 Related Definitions

Equilibrium Q (the traffic of 15 base stations with larger traffic minus the traffic of 15 base stations with smaller traffic, divided by the average traffic of all base stations): If Q is smaller, it shows that the more average the traffic of each base station, the more reasonable the base station construction. When equilibrium Q=0, it shows that the same traffic of each base station is to the best scheme for construction of base stations.

Utilization rate P: The ratio of the base station's traffic from June 1 to 10 to the total ten-day time (in seconds).

Only when the base stations are very average can the Q value of equilibrium approach 0, and the quality of subscribers' calls be guaranteed.

3.3.2 Determination of Objective Function

Total call equilibrium of 30 base stations:

$$A = \frac{\sum_{q=16}^{30} z_{q_1} - \sum_{q=1}^{15} z_{q_1}}{\overline{z}}$$

Utilization rate of the p-th base station:

$$B_p = \frac{z_p}{t_{total}}$$

3.3.3 Determination of Constraint Conditions

A day is divided into 24 periods. The actual number of calls of a base station in each period must be less than the maximum number of calls allowed by the base station: $N_{pt} < M$, t=1……24; p=1……30.

In summary, the optimization model for Problem III is:

$$\begin{cases} \max : A = \dfrac{\sum_{q=16}^{30} z_{q_1} - \sum_{q=1}^{15} z_{q_1}}{\overline{z}} \\ \max : B = \dfrac{z_p}{t_{total}} \end{cases}$$

s.t $N_{pt} < M$, t=1……24; p=1……30

3.3.4 Solution to the Model

Base stations with utilization rate less than 1% belong to unreasonable utilization and can be considered for demolition. When the utilization rate of base station is more than 10%, the subscribers may exceed the load of base station in a certain period of time, so it is necessary to build a new base station around.

1) Calculate the utilization of each base station

TABLE VIII. Calculate the utilization of each base station

ISSN: 0011-9342
Issue: 1 | Pages: 33 - 41

1	2	3	4	5	6	7	8	9	10
0.05104	0.01112	0.0045	0.01933	0.00183	0.0205	0.01403	0.01514	0.03121	0.02796
11	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
0.13849	0.02186	0.05253	0.01222	0.07963	0.0529	0.02799	0.02109	0.01972	0.03231
21	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
0.02952	0.09501	0.02678	0.02495	0.02917	0.0064	0.02071	0.12785	0.16193	0.13333

2) Observe the base stations used by subscribers within 10 days, as follows:

TABLE IX. The base stations used by subscribers within 10 days

USERS USING BASE STATION 5	ALL USER AVAILABLE BASE STATIONS	RATIO OF USERS USING BASE STATION 5
61	5，28	0.125
77	5，11，15	0.041667
96	5,6,19	0.125
178	5，9，11	0.115385
201	5,9,11	0.045455

TABLE X. Users and results using base station 3

USERS USING BASE STATION 3	ALL USER AVAILABLE BASE STATIONS	RATIO OF USERS USING BASE STATION 3
52	3,18	0.5
127	3,20	0.535714

Table XI. Users and results using base station 26

USERS USING BASE STATION 26	ALL USER AVAILABLE BASE STATIONS	RATIO OF USERS USING BASE STATION 26
51	11,16,26	0.166667
104	7,26	0.666667
234	11,16,26	0.045455

ISSN: 0011-9342
Issue: 1 | Pages: 33 - 41

IV. CONCLUSIONS AND SUGGESTIONS

Based on the statistical analysis of the distribution of base stations from the table VII- XI , it is found that the utilization ratio of base stations 3, 5 and 26 is relatively low, and base stations 11, 28, 29 and 30 are overloaded. The suggestions are as follows:

1) Consider dismantling base stations with low utilization (No. 3, 5, 6);

2) No. 11, 28, 29 base stations are under high traffic pressure, so it is guessed that the area has a large volume of calls. We should consider adding some base stations in the area. Consider adding some base stations in the area. The calculation shows that adding 2-3 base stations in this area can alleviate the call pressure.

3) Some regional base stations have smaller coverage (e.g. base stations No. 5, 6 and 7). It is suggested that some other base stations should be added in sparse areas in the future.

It can be seen from the above analysis examples that SPSS analysis can find out the law of data operation from existing data, can reduce the cost, provide useful reference for the transformation of subsequent equipment, and has greater practicability and innovation.

ACKNOWLEDGEMENT

Training plan of young backbone teachers in Colleges and universities of Henan Province, Haojie Du; “Research on Key Technologies of Microgrid Control System” (Pingdingshan University Youth Fund), Numbering: PXYQNJJ2017004, directors: Ning Li; “Research on Control System of Glass Cutting Machine Based on Machine Vision” (Pingdingshan University Youth Fund), Numbering: PXYQNJJ2018001, directors: Huanli Zhao, “Research on Peanut Quality Inspection Based on Machine Vision” (Henan Education Department Project), Numbering: 17A413009, directors: Xueqing Wang.

REFERENCES

[1] Tang Shu. Research on prediction method of heterogeneous network access capability based on hidden markov model. Harbin Institute of Technology.

[2] Hongbin Ren (2014) Evaluation and suggestion of wcdma network based on mr data. Science and technology communication 1(5): 43-45.

[3] Xiao Xu (2006) Research on customer loss and customer value analysis method in mobile communication enterprises. Tianjin University.

[4] Xiaomei Deng (2006) Research on Telecom Customer Segmentation Model Based on Data Mining. Dalian University of Technology.