参赛队员姓名：曹凌微 钱成 田肇阳

中学：清华大学附属中学

省份：北京

国家/地区：中国

指导教师姓名：吴昊 王殿军

论文题目：Evaluation and Prediction of Cell Phone Sales Based on Various Techniques

Abstract

With the progress of our society as well as the technology, online shopping gradually becomes a trend increasingly preferred by young people. This work mainly speculates on the sales of cell phones as a representative, aiming to construct a model capable of analyzing which are the most crucial factors and traits promoting the success of certain types of cell phones.

To begin with, we use Information Entropy to extract the most crucial factors: Comment Count, Good Comment Count and Search Count. We also employ Principal Component Analysis to complete the same goal. The top significant factors are Display Resolution, Recording Definition, RAM and ROM. Next, we apply the results above to Linear Regression and Analytic Hierarchy Process for the modeling, in pursuit of further detailed conclusion. The method of AHP yields straightforward graphs by using qualitative analysis, providing further insight to which specific traits contribute more to the success of the sales volume of that certain type of cell phone.

Furthermore, we optimize all these models with three different methods and employing BP Neural network, Principal Component Regression and Bayes Distinction respectively for quantitative analysis, also concerning which specific traits are more crucial to the sales volume. For the last step of optimization, BOOST algorithm is applied to produce more reliable and stable results. Model's feasibility and sensitivity are finally tested using the data in the testing set, establishing the model’s application value.

In a word, the model constructed not only yields the ranking of individual variables' significance related to the phones' sales volume but also gives insight about which particular traits contribute more to sales volume. It also enables the manufactures to predict sales volume, given its related features, and they can be more informed of the customers' needs and thus maximizing their profits. The testing of the model proves its stability as well as reliability, making it accessible and valuable for the further application in real life. Besides the practical application, the mathematics methods applied to the model are also better than the previous researches, which yield inconclusive and vague results. Therefore, we believe that the optimized model proposed is a huge improvement both in application and methodology, which fills in the vacancy in a nowadays major economic domain and will yield significant social value.

Key Words: Information Entropy, Principle Component Regression, Bayes Distinction, BP Neural Network Fitting, BOOST algorithm

Contents

1 Background

1.1 Research Background

With technological advancement and social development, the use of the Internet has gradually become widespread around the world. The Internet now has developed to provide a platform for uses ranging from completing daily demands to conducting research. With respect to completing daily demands, the Internet has provided a possibility for online shopping. Given the fact that people nowadays have overwhelmed schedules and heavy workloads due to the fast pace of our society, more and more people prefer to shop online instead of going to department stores and supermarkets in person. However, online shopping possesses deficiencies and inconvenience despite its advantages. Shortcomings like being unable to see the products in person have become the greatest worry among customers as they may risk purchasing low-quality products due to lack of key information presented online. On the other hand, producers also suffer from the worry of selling their products. As a result, determining what characteristics of products are crucial to sales volume is the main challenge for online companies. To solve this problem, we choose a specific kind of product——cell phone——to analyze what kinds of cell phones have the highest sale volume.

1.2 Current Research Status

Current Research mainly focuses on several key factors which are considered to influence sales volume. Abroad, Judith Chevalier et al [1] discovers that positive comments are crucial to customers' purchase choices by examining online comments on Amazon. Christy M.K. Cheung, based on the dual process theory, constructs the model of receiving information to study the factors that influence the online consumer information receiving and finds that comprehensiveness and correlation are the most important factors. Kelly o. Cowart conducts a questionnaire survey of 357 sample of university students in the United States through consumer decision-making form. He finds that in online purchase of clothing, quality consciousness, brand consciousness, fashion consciousness, hedonism, impulsivity, and brand loyalty are positively correlated to consumer buying behavior, while price sensitivity is a negative correlation. Michael d. Smith et al [2] by comparing the shopping network of 20268 valid samples for empirical research, finds that goods brand is one of the most important determinants of consumer decision-making. At the same time, if the package goods and services cannot be apart, brands are considered as the credit guarantee of retailers.

Domestically, Jie Zhang and Jianan Zhong [3] conducted research to analyze how sale promotion influences customers' minds and predict the purchase choices of customers. Gang Du and Zhenyu Huang [4] employed the Teradata platform to build decision-making tree model to predict customers' purchasing behaviors, further improving the efficiency and accuracy of prediction. Zhanbo Zhao, Luping Sun, and Meng Sun [5] discovered that factors influencing page view and sales volume are

substantially different. To be more specific, price, scale, reputation, and insurance have a significant influence on page view and sales volume. Zhihai Hu, Dandan Zhao and Yi Zhang [6] employed sales of skin care products on Taobao as an example to analyze the influence of online comments on sales volume. The aforementioned researches mainly explored certain factors influencing sales volume but lacked generality. Therefore, online sellers were unable to determine the influential order of all these factors.

With respect to the research methods, current researches mainly employed three methods: Grey Relational Analysis, C2C Model, and BP Neural Network Fitting. As for Grey Relational Analysis, Fatao Wang employed Grey Relational Analysis to determine the main factors for the development of online shopping. Naicong Hou, Xu Zhang, Enjun Zhang [7] presented reputation as the most influential factor of purchase. Xiao Shi [8] conducted a quantitative research of the interrelation of sales and price, comment rate, popularity with the utilization of Grey Relational Analysis. As for the C2C model, Youzhi Xue and Yongfeng Guo [9] employed a Tobit model to discover that customers valued more on price and delivery fee. Jingsha Fu [10] created a quantify model of influential factors. As for BP Neural Network Fitting, Yanli Ma built an evaluating system including refund rate, descriptions and online comments. All these aforementioned methods are theoretically capable of analyzing the influence of certain factors on sales volume but are lack of practicality.

In conclusion, current researches have failed to analyze influential factors in a systematic and comprehensive way, and they have failed to reveal specific characteristics that achieve higher sales volume. Therefore, our research results improve the current research methods by offering a clear view into the characteristics that cellphones with high sales volume have and applying our results to predicting sales volume.

1.3 Research purpose and significance

Since online sellers constantly worry about ways to promote sales volume, we conduct research in the hope of offering a practical solution by determining which characteristics contribute to improving sales volume. Our research purposes can be summarized as below:

i. To conduct qualitative research to have a general understanding of the characteristics that contribute to high sales volume.

ii. To conduct quantitative research to rank factors that are considered to have an influence on sales volume.

iii. To determine specific characteristics within each factor that contribute to the highest sales volume.

iv. To predict the sales volume of cellphones with a given characteristic.

Our research results will be of great reference and help to online cellphone sellers by offering a clear explanation of what kinds of cell phones have the highest sales

volume. Online cellphones sellers can consequently adjust their products according to our research results to achieve higher sales volume.

1.4 Research method and train of thinking

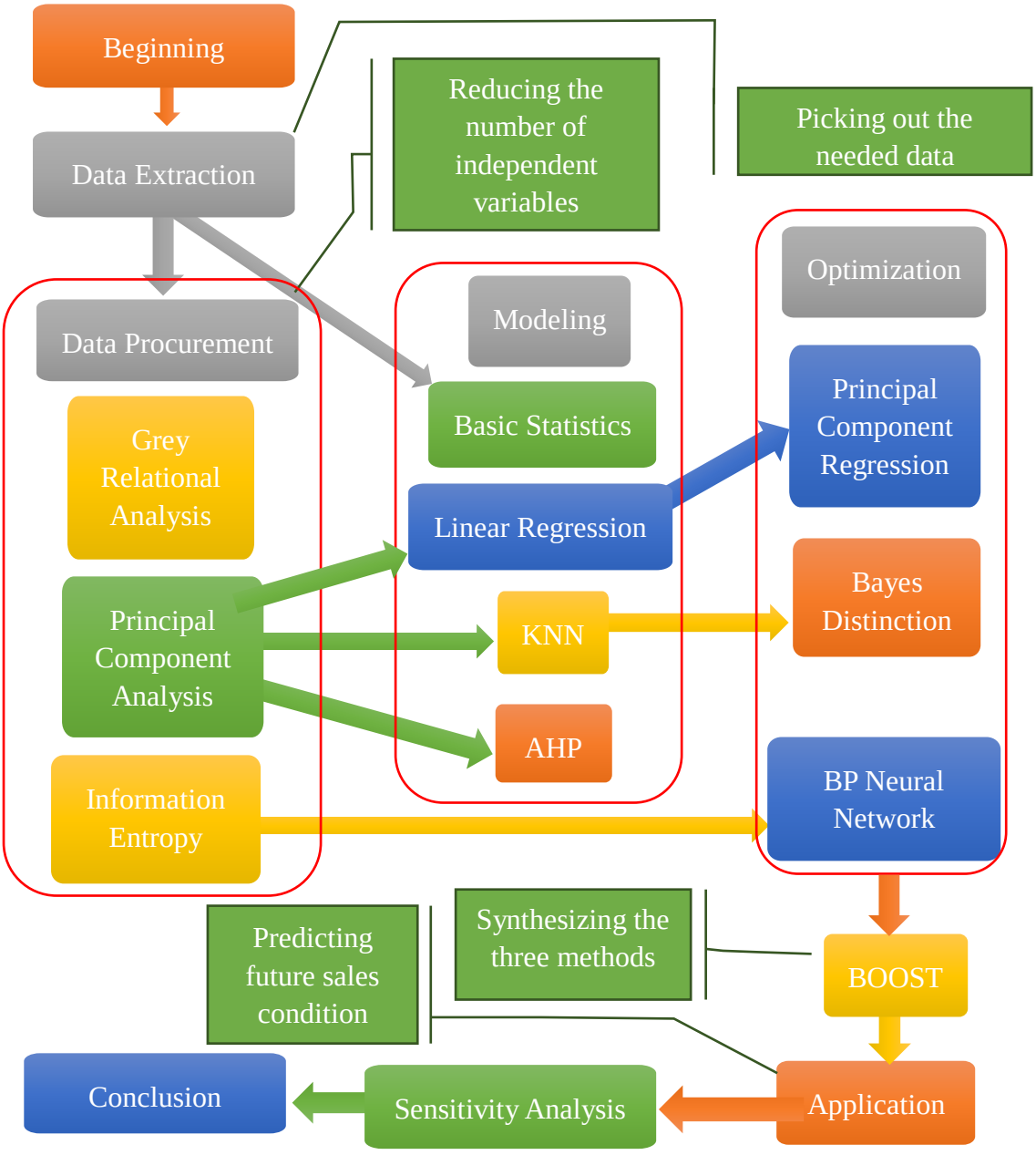

Figure 1: The flow chart of the whole modeling process

Figure 1 below presents the whole modeling process. After gathering data of information about product selling in AliExpress, we extract useful and relevant data concerning different influential factors and conduct basic statistics for further research. Next, we come to the data procurement to reduce the number of independent factors. In this process, we apply three different method——Grey Relational Analysis, Principal Component Analysis, and Information Entropy. The Grey Relational Analysis fails to reduce the number of influential factors, while the other two methods

effectively complete the goal. Then we apply the results of data procurement for modeling. In the modeling process, we apply results from Principal Component Analysis to Analytic Hierarchy Process, KNN, and Linear Regression. At this point, we have reached the conclusion of the rank of different independent factors. Furthermore, we conduct optimization to each model. We optimize KNN by Bayes Distinction and Linear Regression by Principal Component Regression, while we optimize Entropy of Information to BP Neural Network Fitting.

Afterward, we employ the BOOST algorithm to synthesize the three methods and reach the conclusion that which characteristics contribute to the highest sale volume. Finally, we practice the application of our research results by predicting future sales conditions.

2 Assumptions

2.1 Assumptions

- Category Click Rate represents the ratio of the number of people who buy that certain type of cell phone to the number of people who click on the picture online for more detail.
- Category Convert Rate represents the ratio of the number of people who click on the picture for more detail to the number of people who browse the internet and see the picture of that certain type of cell phone.
- We assume that considering these two sets of data as the bases for the Information Gain provide authentic information and reflect the ratio of people who are interested in and actually buy the cell phone. In this way, the data are also in a more consistent and standardized form which is convenient for later grouping and processing.

2.2 Definitions

Table 1: the definition of notations

Notation	Definition
A_{ij}	The element in i^{th} Row and j^{th} Column in matrix A
X	The independent variables matrix
x	Row vector of independent variables
y	Row vector of dependent variables
$\overline{x}$	The algebra average of several data
Y_1	Click rate sequence in Grey Relational Analysis or click rate matrix in other parts
Y_2	Convert rate sequence in Grey Relational Analysis or convert rate matrix in other parts
X_k	The k^{th} independent variable sequence in Grey Relational

	Analysis
Y_k^n	The n^{th} number in the k^{th} dependent variable sequence in Grey Relational Analysis
X_k^n	The n^{th} number in the k^{th} independent variable sequence in Grey Relational Analysis
ΔX_k^n	The difference between every two adjacent terms in independent variable sequences in Grey Relational Analysis
ΔY_k^n	The difference between every two adjacent terms in dependent variable sequences in Grey Relational Analysis
$CC(Y_k)$	The correlation coefficient of k^{th} dependent variable sequence
$CC(Y_k, X_l)$	The correlation coefficient between k^{th} dependent variable sequence and l^{th} independent variable sequence
$\gamma(Y_k, X_l)$	The correlation degree between the k^{th} dependent variable sequence and l^{th} independent variable sequence
$E(X)$	The information entropy regarding the set of incidence X
P_i	The possibility that incident numbered i will happen in the set X
$E(global)$	The information entropy of Category Click and Convert Rate
$IGain$	The information gain of individual variables related to the Category Click and Convert Rate.
$A_{i,j}$	The data in the i^{th} line and j^{th}column in the table of data processing concerning information entropy
x_p	The p^{th} original variable
z_q	The q^{th} New variable
m	The number of samples
l	The number of variables in each sample
$x_{ij}{}^*$	The standardized data at row i and column j
x_{ij}	The data at row i and column j before standardization
R	The correlation coefficient matrix in principal component analysis
λ_q	The q^{th} characteristic roots or eigenvalues in AHP
$a_q(A_q)$	The q^{th} characteristic vectors
a_{pq}	The pth value of the q^{th} characteristic vectors
$(w_1 \quad \dots \quad w_n)$	Weight vector in Weight Determination Method
n	The number of choices of target layer in Weight Determination Method

$\boldsymbol{w}$	The eigenvector in Weight Determination Method
$\boldsymbol{\beta}$	Coefficient matrixes of the original data
$\boldsymbol{\beta'}$	Coefficient matrixes of Principal Component Regression
$\boldsymbol{P\{X\}}$	The probability that satisfies condition X
$\boldsymbol{\alpha}$	Reliability in Regression
$\boldsymbol{\theta}$	Parameters to be estimated of the ensemble in Regression
$\boldsymbol{\hat{\theta}_1}$	The confidence upper limit in Regression
$\boldsymbol{\hat{\theta}_2}$	The confidence lower limit in Regression
$\boldsymbol{d(X,Y)}$	The Mahalanobis distance of the data
$\boldsymbol{\Sigma}$	The covariance matrix
$\boldsymbol{P(B_i\|A)}$	Posteriori probability in Bayes Distinction
$\boldsymbol{P(A\|B_i)}$	Priori probability in Bayes Distinction
$\boldsymbol{P(B_i)}$	The frequency at which the sample appears in Bayes Distinction
$\boldsymbol{G_i}$	The ensemble in Bayes Distinction
$\boldsymbol{f(x)}$	Probability density function of G_i in Bayes Distinction
$\boldsymbol{p_i}$	The priori probability of G_i In Bayes Distinction
$\boldsymbol{k}$	The number of G_i in Bayes Distinction
$\boldsymbol{P\left(j/i\right)}$	The conditional probability of wrongly categorizing the sample of G_i to the ensemble G_j
$\boldsymbol{C\left(j/i\right)}$	The loss caused by the wrong categorization
$\boldsymbol{D_k}$	A division of a set of distinction samples
$\boldsymbol{ECM}$	The average wrong distinction loss

3 Data Procurement and Process

3.1 Data extraction

We have obtained information about sale records on AliExpress, which is under the control of Alibaba. The original data is in the appendix. With the algorithm and formula given by AliExpress, we convert the original data into the readable and understandable data, which can also be seen in the appendix. [11][12]

We utilize PYTHON to extract the parameter cells, which contain several standardized descriptions of the phones. With the help of XLRD module and XLWR module, we search for cells with the assigned field one after another. We divide the searching process into two stages. The first stage is to separate the entire parameters into several fields that contain only one property each; The second stage is to check what each field denotes and use numerical data to characterize the words. For instance, when we search for the battery property, which is detachable, not detachable, or unknown, we first split the cell by "
" which stands for breaks to obtain strings that merely possess one property in lieu of many. Then we use the "if" function to determine whether the obtained string includes target string, which is "yes" or "no" standing for detachable or not detachable. If it includes the prior one, we define the corresponding value in the new Excel table as 1. If it includes the latter one, we define

the corresponding value in the new Excel table as 2. If it includes neither one, we define the corresponding value in the new Excel table as 0, which stands for unknown.

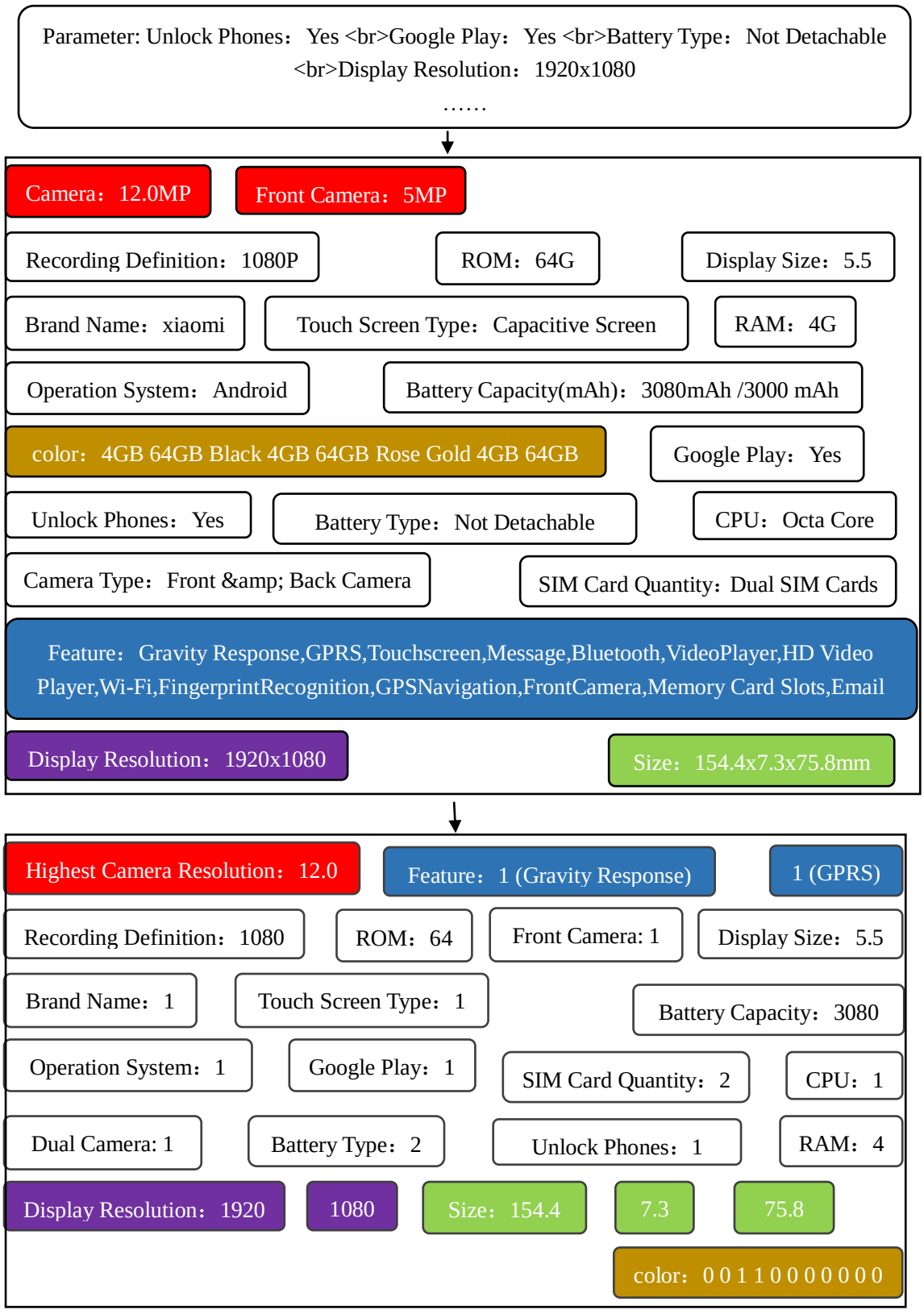

Figure 2: Data Extraction diagram

We set Unlock Phones, Google Play, Battery Type, Display Resolution, Operation

System, Gravity Response, GPRS, SIM Card Quantity, Size, Battery Capacity, Camera, Recording Definition, Display Size, Brand Name, CPU, Touch Screen Type, RAM, and ROM as the keywords for the first stage; we set “yes” and “no” as the keywords for the second stage.

In the second stage, there are some individual cases for us to pay attention to. When we extract the color parameters, we search the name of the colors individually, for the reason that a page may contain phones with various colors. We use the binary combinations to express the colors of the phones. We set White, Blue, Rose, Gold, Silver, Grey, Pink, Brown, Orange, Yellow, and Red as the detection keywords, which allows us to obtain eleven-dimensional binary array to demonstrate the colors. The following figure 2 illustrates the process.

When we are extracting the highest camera resolution fields, we search all the fields with “camera: ” and comparing the numerical part of all the fields featuring above, retaining the largest one and disposing of the rest.

As for extracting the size, we come up with a problem that some of the dimensions are expressed in inches, while others are in centimeters or millimeters, triggering inconsistency in units. To solve this issue, we first use “x” or “*” to split the value of three dimensions, before we multiply the three parameters, get the volume of the phones, and use a method to determine the critical value that decides the unit of the phones. We select a phone that we regard as normal, calculating the volume among in inches, millimeters, and centimeters. We then obtain the square roots of the products of the size in inches and centimeters, as well as in centimeters and in millimeters, which are regarded as the critical value. We obtain the essential values of volume, which are 36.86334 and 4712.451. If the product is less than 36.86334, we appreciate the unit as inches. Then we multiply the length, width, and height of the phone with 25.4 to obtain the corresponding value in millimeters. If the product is more than 36.86334 but less than 4712.451, we regard the unit as centimeters. Then we multiply the length, width, and height of the phone with 10 to obtain the corresponding value in millimeters. If the product is more than 4712.451, we regard the unit as millimeters. Then we straight write the length, width, and height of the phone into the tables.

Finally, we write the value into the Excel table and obtain the data that we use, which can be seen in the appendix.

3.2 Grey Relational Analysis

In the real world, it is commonly seen that what influences a system tends to be multi-factors instead of a single counterpart, while the relationship between the factors is complicated, which gives rise to the fact that it is easy to cover up its essence with mere regards of its appearance, which makes it difficult to get accurate information and distinguish the primary and secondary factors. The grey system analysis method is essentially an analytic method that replaces discrete data with linked concepts. [8]

The grey system theory holds that, although the appearance of the objective system seems to be complicated, and the data is irrelevant, it always functions as a whole,

which means it is not random but proves to contain some inherent laws that can be discovered and explored, and the key is how to choose the proper way to figure out the rules of the data and utilize them.

The gray correlation degree is calculated as following in general: First, we standardize the collected evaluation data to ensure that it is treated without dimension; we obtain the sequence of difference and compute the maximum and minimum variance of the series of difference; we calculate the correlation coefficient and the calculation correlation degree.

Specifically, we consider the dependent variables, which are click rate and conversion rate, as the reference sequence. As shown in the appendix, we let the following sequence 1 denotes the click rate sequence:

$$Y_1 = Y_1^1, Y_1^2, Y_1^3, Y_1^4, \dots, Y_1^{1324} \tag{1}$$

And we let the following sequence 2 as the convert rate sequence:

$$Y_2 = Y_2^1, Y_2^2, Y_2^3, Y_2^4, \dots, Y_2^{1324} \tag{2}$$

We consider the 26 series of independent variables as comparing sequence. As shown in the appendix, we let the following sequence 3 denotes the Google play sequence:

$$X_1 = X_1^1, X_1^2, X_1^3, X_1^4, \dots, X_1^{1324} \tag{3}$$

And so on, we let the following sequence 4 as the can-design-product sequence:

$$X_{26} = X_{26}^1, X_{26}^2, X_{26}^3, X_{26}^4, \dots, X_{26}^{1324} \tag{4}$$

Then we standardize the data, making the variance of each sequence change into 1 and the mean into 0. We compute the difference between every two adjacent terms, which can be shown as following formula 5-6:

$$\Delta X_k^n = X_k^{n+1} - X_k^n (k\epsilon\{k\epsilon\mathbb{N}^*|k \le 26\}, n\epsilon\{n\epsilon\mathbb{N}^*|n \le 1323\}) \tag{5}$$

$$\Delta Y_k^n = Y_k^{n+1} - Y_k^n (k\epsilon\{k\epsilon\mathbb{N}^*|k \le 2\}, n\epsilon\{n\epsilon\mathbb{N}^*|n \le 1323\}) \tag{6}$$

We finally calculate correlation coefficients and the correlation degree, of which the formula is as following formula 7-9.

$$CC(Y_k) = \left| \sum_{i=1}^{1323} \frac{\Delta Y_k^i}{n} \right| (k\epsilon\{k\epsilon\mathbb{N}^*|k \le 2\}) \tag{7}$$

$$CC(Y_k, X_l) = \left| \sum_{i=1}^{1323} \frac{\Delta Y_k^i - \Delta X_l^i}{n} \right| (k\epsilon\{k\epsilon\mathbb{N}^*|k \le 2\}, l\epsilon\{l\epsilon\mathbb{N}^*|l \le 26\}) \tag{8}$$

$$\gamma(Y_k, X_l) = \frac{1 + CC(Y_k)}{1 + CC(Y_k) + CC(Y_k, X_l)} (k\epsilon\{k\epsilon\mathbb{N}^*|k \le 2\}, l\epsilon\{l\epsilon\mathbb{N}^*|l \le 26\}) \tag{9}$$

From the obtained correlation degree, we find that the independent variables which have less value in them are apt to have higher correlation values, symbolizing that a closer connection with the dependent variables. Moreover, the independent variables which have the same number of value possess the same correlation degree, rendering it impossible for us to distinguish how close the connections are between these

independent variables and the target dependent variables. We can conclude that the Grey Relational Analysis suits for continuous variables rather than discrete variables, indicating that it is not an ideal technique for us to determine how tight the relationship is under this situation.

Table 2: Grey Relational Analysis Result

	Google Play	Battery Type	Battery Capacity(mAh)	Display Resolution	Operation System	SIM Card Quantity	
Click Rate	0.958033	0.958033	0.54663	0.958033	0.958033	0.958033	
Convert Rate	0.989033	0.989033	0.563529	0.989033	0.989033	0.989033	
	Recording Definition (P)	Touch Screen Type	RAM(G)	ROM(G)	CPU	Display Size (inches)	Size
Click Rate	0.958033	0.958033	0.588991	0.337011	0.958033	0.958033	0.771116
Convert Rate	0.989033	0.989033	0.570534	0.330881	0.989033	0.989033	0.805193
	Highest camera resolution(MB)	Dual Camera	Front Camera	Brand	Color	Feature	Price
Click Rate	0.771525	0.958033	0.958033	0.958033	0.486192	0.958033	0.555845
Convert Rate	0.805639	0.989033	0.989033	0.989033	0.499513	0.989033	0.539377
	SearchCnt	GoodCommentCount	Score	IsGalleryFeatured	IsHighQuality	CanDesignProduct	
Click Rate	0.752451	0.550346	0.958033	0.958033	0.958033	0.958033	
Convert Rate	0.722596	0.56748	0.989033	0.989033	0.989033	0.989033	

Information Entropy

Information entropy is used to reflect the complexity of the information being processed. Higher information entropy value indicates a higher degree of information complexity. Thus, information entropy can be applied to analyze the information in a quantitative way. Information entropy is defined by the formula 10 below:

$$E(X) = -\sum_{i=1}^{n} p_i \log_2(p_i) \tag{10}$$

where $E(X)$ represents the Information Entropy of X, the set of incidents taken into consideration (in the formula 10 the total number of incidents is n) and $p_i$represents the possibility that the incident numbered i will happen in the set X. The information entropy is calculated in the form of the sum of each individual incident.

However, in determining which factor is more important for us to take into consideration among 26 individual variables related to the cell phone as extracted, what is needed should be the amount of information that can be acquired from analyzing on factor instead of its complexity as reflected by the information entropy. Therefore, we utilize information gain to consider which factors are the top ones that should be taken into account as the most crucial. In other words, what kinds of factors contribute more or promote the sale of the smartphones in general. The calculation of information gain of each factor involves its information entropy and is a deliberate and complex process. In the next part of this section, we will mainly discuss the data processing related to the information gain.

First, we identify 26 individual variables as the potentially influential factors for sale volume of cell phones including Google play, battery type, brand, RAM, ROM, dual camera, front camera, display size, etc. Then, types of data representing the actual sale volume of cellphones are regarded as the bases for calculating the information gain. Instead of choosing the actual sales volume, we consider the Category Click Rate and Category Convert Rate. Reasons are illustrated in the assumption.

We then divide the Category Click Rate and Category Convert Rate into five groups respectively and reasonably, according to the individual value of the data, from high to low, categorized from 1 to 5. After categorizing the data related to Category Click and Convert Rate, we use formula 10 to calculate the global information entropy of those two sets respectively. As applying the formula to the Category Click Rate, $E(X)$ now represents the information entropy of the Category Click Rate, and p_i represents the possibility of category numbered i will happen. Especially, since there are 5 categories, the number n equals 5. The same can be applied to the Category Convert Rate, and the final results are shown in table 3 below:

Table 3: The Global information entropy

	Category Click Rate	Category Convert Rate
Global information entropy $E(global)$	2.200779	2.081891

The information we can get from each individual variable is calculated respectively, and the individual variables can be generally classified into two groups: group one with relevant data presenting in inconsistent ways, including factors like Is Gallery Featured and Dual Camera, in which the data only consist of 1, 0, or -1(in other words, the data are expressed in simple forms and can be calculated artificially); group two with relevant data presenting in consistent forms, including factors like Display Size and Display Resolution, in which the data are in various forms and need grouping for further calculation.

As for group one, we take ROM as an example to illustrate how the information entropy is calculated based on the grouping of Category Click Rate. First, we do the grouping and data processing. The data of ROM are presented as discrete variables, including 2, 4, 8, 16, 32, 64, 128, and 256. The grouping of data in ROM should also be related to the grouping of Category Click Rate, so accordingly, there are in total 40 groups, which are presented in table 4 below:

Table 4: The grouping of ROM related to Category Click Rate

ROM	Group number in Category Click Rate	1	2	3	4	5
	2	0	3	2	1	0
	4	5	6	13	7	0
	8	64	38	63	54	8
	16	110	89	132	143	36
	32	64	44	82	63	29
	64	83	38	62	49	16
	128	3	4	5	7	0
	256	0	0	1	0	0

Let i represents the i^{th} line in the table of the forty groups (as distinguished by double cross lines), j represents the j^{th} column in the table, and A_{ij}represents the number in the unit of the i^{th} line and j^{th} column. Thus, in the unit $A_{4,1}$, the number 110 represents that there are in total 110 data in ROM that are 16 and also in the group 1 as categorized according to the Category Click Rate. Notice that the sum of all the forty groups should equal to the total number of data (and in our data processing, the total number of data available is 1324).

After the grouping of ROM data related to the Category Click Rate, we further calculate the information entropy of the data in each line using formula 10. The information entropy of ROM in each line is shown in table 5:

Table 5: Information entropy of ROM

					Information entropy $E(ROM)$
0	3	2	1	0	1.459148
5	6	13	7	0	1.89366
64	38	63	54	8	2.122787
110	89	132	143	36	2.205866
64	44	82	63	29	2.242444
83	38	62	49	16	2.160525
3	4	5	7	0	1.931295
0	0	1	0	0	0

In order to acquire the total amount of information we can gain from the independent variable ROM, we need to further calculate the possibility that each line will happen. As for the first line $A_{1,j}$, we calculate the times at which data 2 appears and then

divide the total number of data, 1324. Then, we multiply the possibility to the information entropy of each line, the results are shown in table 6:

Table 6: The Information entropy, possibility and their products of ROM

Information entropy	Possibility	Product
1.459148	0.004531722	0.006612
1.89366	0.023413897	0.044338
2.122787	0.171450151	0.363952
2.205866	0.385196375	0.849692
2.242444	0.212990937	0.47762
2.160525	0.187311178	0.40469
1.931295	0.014350453	0.027715
0	0.000755287	0

The sum of all eight products is the total information entropy we can get from the individual variable ROM. However, for the information gain as related to the Category Click Rate, we need to use the global information entropy of Category Click Rate to subtract the sum of the product above, as the following formula 11 presents:

$$IGain(Category\ Click\ Rate, Rom) = E(global) - \sum Information\ entropy \times Possibility \quad (11)$$

where $E(global)$ here represents the global information entropy of the Category Click Rate, since the gain is related to the Category Click Rate. The final gain is presented in table 7 below:

Table 7: The final information gain of ROM

	Sum of the products	$IGain$
ROM	2.174619842	0.026159369

Similarly, the information gain of ROM related to the Category Convert Rate can also be calculated using the method above, and the only difference will be the data in the 40 groups and in the final formula, $E(global)$ should represent the global information entropy of the Category Convert Rate.

As for the group two, we consider the Search Count (the number of time that a certain type of phone is exposed to the customer) as related to the Category Click Rate in order to illustrate the difference of data processing from group one. From the data we have extracted, it is obvious that the data in the Search Count are not discrete and the majority of the data of this independent variable are different. However, for the calculation of the information entropy, the number of data in the group should reach a substantial amount, or the final result will be meaningless. Thus, we divide the 1324 data into 5 groups reasonably in order to ensure the number of data in each group for an effective final result.

We divide the data into 5 groups, which are: [0,3000], [3000,30000], [30000,100000], [100000,500000] and [500000, max value]. The later data processing parts are similar

to that for the independent variables in group one. The following table 8 presents the grouping of Search Count after the data division:

Table 8: Grouping of Search Count

Search Cnt	Group number in Category Click Rate	1	2	3	4	5
	[0,3000)	220	2	17	13	2
	[3000,30000)	31	106	107	94	19
	[30000,100000)	34	43	77	66	8
	[100000,500000)	29	61	96	81	8
	[500000, max value)	15	10	63	70	52

The data can later be processed as the same way above, and the final information gain of the Search Count related to the Category Click Rate is as shown in table 9:

Table 9: The final information gain of Search Count

	Sum of the products	$IGain$
Search Count	1.808392333	0.392386753

As for other individual variables whose data are not discrete numbers, the same data processing method can be applied. Thus, the information gain of each 26 individual variables as related to the Category Click Rate and Category Covert Rate can thus be calculated. The final information gain is presented in table10 and 11:

Table10: Information gain of each individual variables related to the Category Click Rate. The higher information gain indicates the factor is more important. The table below ranks the independent variables and produces the final result.

Comment Count	0.732792417
GoodCommentCount	0.680453664
Search Count	0.392386753
Score	0.173242295
Brand	0.124112475
Is Gallery Featured	0.060358001
Battery Capacity(mAh)	0.050232189
RAM(G)	0.031072544
Size	0.028079662
Highest camera resolution	0.026589357
ROM(G)	0.026159369
Price	0.02284587
Color	0.021268354
Display Resolution	0.019607145
Feature(gravity and GPRS)	0.016959237
Is High Quality	0.016942427
CPU	0.016022113
Recording Definition (P)	0.012770426

Display Size	0.011465334
Battery Type	0.010742922
Touch Screen Type	0.008087216
Operation System	0.006372009
SIM Card Quantity	0.006106585
Front Camera	0.001569914
Dual Camera	0.00153786
Google Play	0.000108253

Table 11: Information Gain of each individual variables related to the Category Convert Rate

Comment Count	0.950131659
GoodCommentCount	0.910616696
Search Count	0.631528548
Score	0.288394004
Brand	0.261397755
IsGalleryFeatured	0.220147548
Battery Capacity(mAh)	0.102065066
Highest camera resolution	0.067310002
Color	0.052728838
Size	0.052070786
Price	0.040606491
RAM(G)	0.040286271
ROM(G)	0.039561052
Recording Definition (P)	0.038203566
CPU	0.037314954
Display Resolution	0.032897632
Battery Type	0.0300117
Feature(gravity and GPRS)	0.024420792
Display Size	0.020633742
IsHighQuality	0.014778733
SIM Card Quantity	0.010565193
Operation System	0.008369328
Front Camera	0.006603513
Touch Screen Type	0.005934561
Google Play	0.005319893
Dual Camera	0.002904891

Higher information gain of the individual variable indicates greater importance of that factor contributing to the sale volume of the product. From tables 10 and 11 above, we can conclude that Comment Count, Good Comment and Search Count contribute more to the sales volume of the products as a whole, while Score, Brand and Is gallery Featured are also significant factors promoting the phone's sales. One thing particularly noticeable is that Category Click Rate and Category Convert Rate are considered separately in the data processing, but they yield a similar final result, the

fact of which lend the method credibility. Thus, when deciding which variables are more crucial and can be taken into account for the modeling further, the method of information entropy is a relatively clear and reliable way.

3.4 Principal Component Analysis

We regard the sales condition as dependent variables and the properties of phones as independent variables. We try to reduce the dimensionality, diminishing the vast amount of the original data and variables into fewer data and variables, while the new variables can retain the information in the original data by and large. [13]

We utilize the 26 original variables mentioned in 3.4 as the original data. We still use X to denote independent variables matrixes, $Y_n(n \in \{n \in N^* | n \leq 2\})$. The original variables are $x_p(p \in \{p \in N^* | p \leq l\})$; the new variables are $z_q(q \in \{q \in N^* | q \leq p\})$. We use m to denote the number of samples; we use l to denote the number of variables in each sample. Thus, the data matrix is as matrix 12

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1l} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{ml} \end{bmatrix} \tag{12}$$

Since the data vary in dimensions and ranges, we need to standardize the data. We adopt the variance standardization technique to operate the data so that the variance of the standardized data is 1, while we conduct the central translation so that the mean of the data is 0. The formula is as formula 13

$$\overline{x_j} = \sum_{t=1}^{i} \frac{x_{tj}}{i}, \sigma_j = \sqrt{\sum_{i=1}^{n} \frac{\left(\overline{x_j} - x_{ij}\right)^2}{n-1}}, {x_{ij}}^* = \frac{x_{ij} - \overline{x_j}}{\sigma_j} \tag{13}$$

${x_{ij}}^*$ denotes the standardized data at row i and column j; x_{ij}denotes the data at row $i$and column j before standardization. i denotes total column number and j denotes total row number.

Then we establish the correlation coefficient matrix R. The formulas are shown in formula 14-15.

$$r_{ij} = \frac{\sum_{k=1}^{n}(x_{ki} - \overline{x}_i)(x_{kj} - \overline{x}_j)}{\sqrt{\sum_{k=1}^{n}(x_{ki} - \overline{x}_i)^2 \sum_{k=1}^{n}(x_{kj} - \overline{x}_j)^2}} \tag{14}$$

$$R = \left(r_{ij}\right)_{l \times l} \tag{15}$$

Then we obtain the characteristic roots $\lambda_q(q \in \{q \in N^* | q \leq l\})$ which satisfy$\lambda_x > \lambda_y$ for $\forall 1 \leq x < y \leq q$ and characteristic vectors $a_q(q \in \{q \in N^* | q \leq l\})$to determine the load a_{pq}on each new principal component variables $z_q$of the original variables x_p, which are equal to the q^{th}larger characteristic values of the correlation matrix corresponding to the eigenvectors. $a_{pq}$is the p^{th} value of the q^{th}characteristic vectors. The formula is as formula 16:

$$RA = \lambda A \tag{16}$$

In the formula, A denotes each characteristic vector, λ denotes each characteristic value. The characteristic roots are shown in table 12. Characteristic vector matrix is in the appendix.

Table 12: Principal Component Regression Characteristic Value

5.830569	2.108099	2.023791	1.548141	1.270862	1.142889	1.079039
0.984911	0.973456	0.896728	0.878234	0.848337	0.823186	
0.770015	0.68834	0.646611	0.571062	0.49684	0.465049	0.413485
0.364857	0.324523	0.276729	0.245753	0.200672	0.127823	

The contribution rate formula and the total contribution rate formula is as 17-18.

$$\frac{\lambda_i}{\sum_{k=1}^{q}\lambda_k}\ (i=1,2,\ldots,p) \quad \text{and} \quad \frac{\sum_{k=1}^{i}\lambda_k}{\sum_{k=1}^{q}\lambda_k}\ (i=1,2,\ldots,p) \tag{17-18}$$

We obtain the total contribution rate until the fourteenth principal component is 81.45%, which is larger than 80%. Therefore, we take the first fourteenth eigenvalue as the principal component. Suppose the principal component is formula set 19

$$\begin{aligned} z_1 &= a_{11}x_1 + a_{21}x_2 + a_{31}x_3 + a_{41}x_4 + a_{51}x_5 + \cdots + a_{26\,1}x_{26} \\ z_2 &= a_{12}x_1 + a_{22}x_2 + a_{32}x_3 + a_{42}x_4 + a_{52}x_5 + \cdots + a_{26\,2}x_{26} \end{aligned} \tag{19}$$

...

$$z_{14} = a_{1\,14}x_1 + a_{2\,14}x_2 + a_{3\,14}x_3 + a_{4\,14}x_4 + a_{5\,14}x_5 + \cdots + a_{26\,14}x_{26}$$

Since the data are standardized before the analysis, each coefficient is equally likely. We can use the independent variables of which the principal component coefficients are relatively larger in the first several principal components. For instance, there are several original variables in the first principal components, of which the coefficients are relatively larger among all the coefficients of the original variables in the first principal components. We choose two deputies of them to denote the resembling original variables and repeat the process for the second and third principal components. Thus we obtain 6 properties of the phones for further analysis, which are Display Resolution, Recording Definition, RAM, ROM, CPU, Highest camera resolution, and Price.

4 Modeling

4.1 Basic Statistics

After obtaining the original data, we do the basic statistics process. We set the click rate and the convert rate as the dependent variables, while other variables as independent variables. On the one hand, we make pie charts, as well as line charts, reveal the proportions of the phones with each characteristic over the ensemble, as shown in figure 3-4. On the other hand, to show the cross relationship between the independent variables and dependent variables, we draw the bivariate tables to reveal the proportions of the phones with each characteristic over a certain type of phones. We first categorize the continuous variables into several ranges to discretize the variables. Table 13 is the statistic table of Battery Capacity. We divide the click rate

into 5 categories, which are 0-0.1, 0.1-0.2, 0.2-0.225, 0.225-0.3, 0.3-0.464. We divide the convert rate into 5 categories, which are 0-0.1, 0.1-0.2, 0.20-0.22, 0.22-0.23, 0.23-0.468.

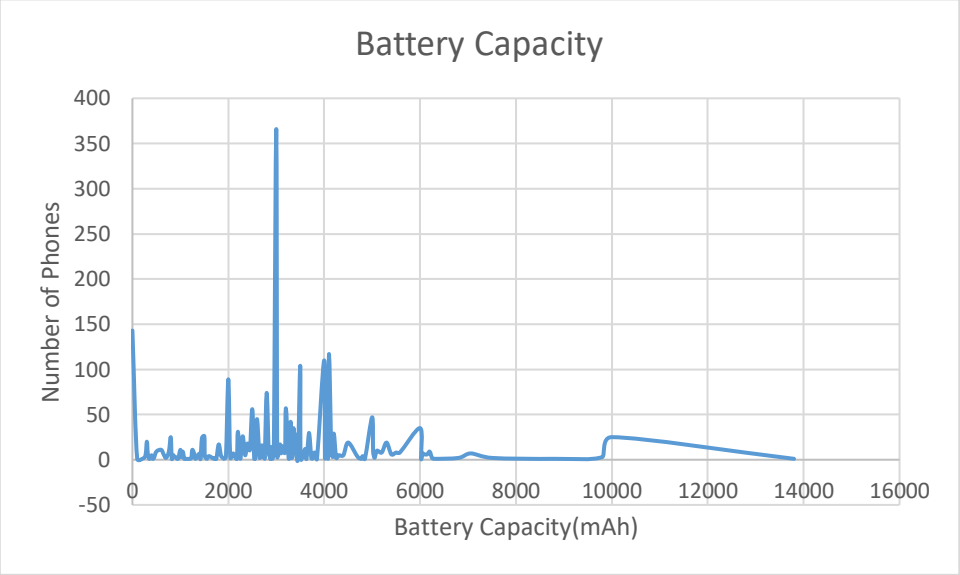

Figure 3: Line Chart of Battery Capacity. The Capacity focuses on 3000, 4000, and 4100 mAh.

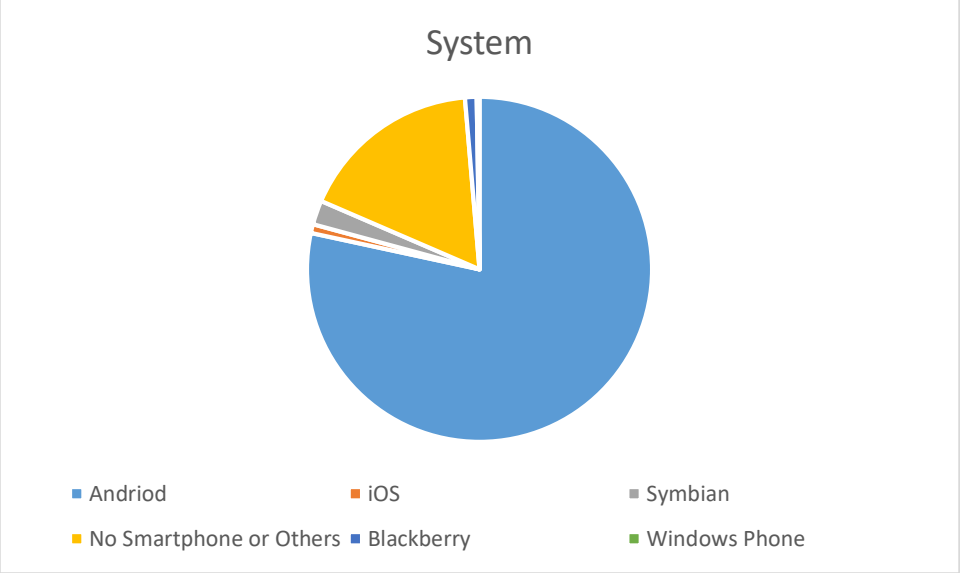

Figure 4: Pie Chart of System. The Android system takes a major proportion.

Table 13: Statistics from Battery Capacity to Click Rate Category. The upper-middle Battery Capacity reveals a better sales.

ClickrateCatagory / mAh	1	2	3	4	5
Lower than 3000	93	67	92	88	14
3000	59	40	66	60	8
More than 3000 but less than 4000	80	57	85	79	16
4000mAh to 4100	24	19	53	60	42
More than 4100	73	39	64	37	9

The previous charts demonstrate, for instance: most of the phones possess 3000 mAh, 4000 mAh, or 4100 mAh battery. The phones with Android systems lead the ranking of systems, while Apple system is the second one. For the phones which achieve

higher click rate category, they are more likely to have the upper-middle battery capacity.

4.2 Weight Determination Technique

In order to choose by diverse factors and judge the sales of certain types of phones, we create a new Weight Determination Technique, which imitates the Analytic Hierarchy Process (AHP), to achieve the goal which is to determine the weight of each option in complicated and uncertain problems. We define the properties of the phones, which are display resolution, recording definition, RAM, ROM, CPU core, highest camera resolution, and price, obtained from the Principal Component Analysis, as the scheme layer, while defining the click rate and the convert rate as the target layer, to build up the weight determining model with one mere layer but several groups.

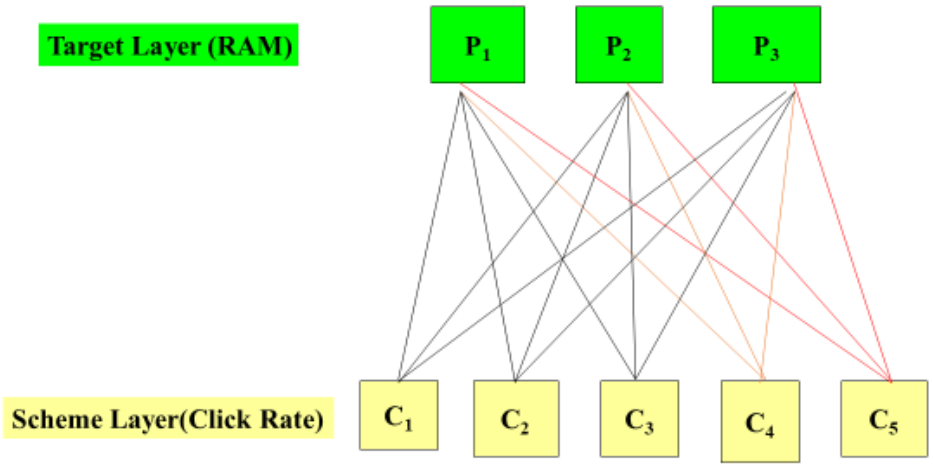

Figure 5: Structure diagram. We use only two layers but with several groups.

We divide RAM into three groups, less than 1 GB, no less than 1 GB but less than 4 GB, and more than 4 GB, of which are groups 1, 2, and 3 respectively. We divide ROM into three groups, less than 8 GB, no less than 8 GB but less than 64 GB, and more than 64 GB, of which are groups 1, 2, and 3 respectively. We also divide display resolution, recording definition, highest camera resolution, and price into several categories, of which the standard is the same as what we do in the Information Entropy part. Figure 5 shows the diagram from RAM to click rate. [14]

First, we define the amounts of phones that possess certain properties under certain types of sales conditions, which refers to the amount of a certain target choice under a certain scheme layer condition, as w. In accordance with the target choice, we obtain a weight vector $(w_1 \quad \cdots \quad w_n)(n$ stands for the number of choices of target layer). We compute the ratio between the number, $w_i(1 \leq i \leq n)$, of each scheme layer choice under a common target layer choice and regard it as the weight of paired comparison matrix. As they are consistent matrixes, we do not need to apply consistency tests to the matrixes, for they are automatically consistent, which means

that the eigenvalues are all identical. With the help of the formula of the eigenvalue and eigenvectors shown in formula 20,

$$Aw = \lambda w \tag{20}$$

we can obtain the eigenvectors, w. Composing the eigenvalues of each scheme layer, we obtain the eigenvector matrixes as well as weight vector matrixes from the target layer to the scheme layer.

Then we repeat the process from each scheme layer, which is the sales condition, to each target layer, which is the properties of the phones, to achieve the goal that for each scheme the sum of the weight vector is 1 to transversely compare which option is more welcomed under the same sales condition. Comparing the weight of each scheme to one single target vertically, we obtain which kinds of phones are more welcomed under the same standard.

Finally, we draw the statistical chart with each weight vector, such as stacked column charts, to clearly express the interference of the properties of the phones to the result. The charts are shown in the appendix, one of which is shown in figure 6.

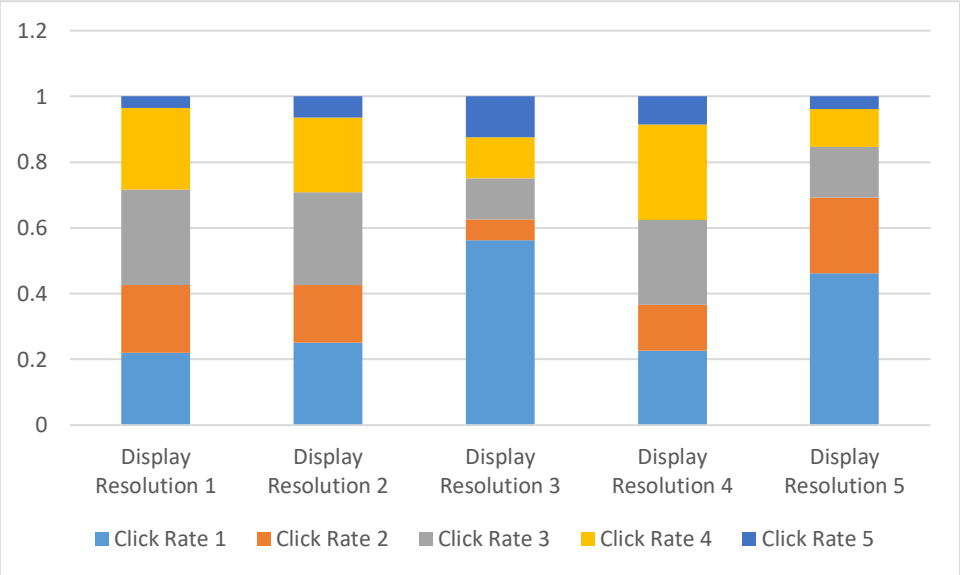

Figure 6: Result analysis. The middle Display Resolution experiences both a better sale and a worse sale.

We can clearly see that phones with middle display resolution tend to attract more customer to click in and purchase. Phones with lower and higher recording definition are more welcomed, while phones with medium counterpart are less intriguing. Phones with lower RAM, ROM, and CPU involve in more click rate, whereas phones with higher equivalents involve in more convert rate. For both highest camera resolution and price, the medium ones are both attractive.

4.3 Linear Regression

The third modeling method we use is Linear Regression. We can regard the properties of phones as independent variables, and the sales as dependent variables. Based on the samples, each data can be viewed as a mapping from the independent variables, which

are the properties, to the dependent variables, which are sales. As each information is expressed numerical, we can find the function from the independent variables to the dependent variables through linear regression from the data. [15]

Let x_1 to x_7 respectively denote display resolution, recording definition, RAM, ROM, CPU core, highest camera resolution, and price. Let $y_1$denotes click rate and $y_2$denote convert rate. The value of the independent variables and dependent variables is the numbers of each option. We utilize regression formula 21.

$$y_n = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 (n \in \{1,2\}) \quad (21)$$

Let X denotes the independent variables matrix; $Y_n (n \in \{n \in N^* | n \leq 2\})$ denote dependent variables matrixes;β denotes coefficient matrixes. We apply Least Square Regression Method to the issue, of which the formula is shown in formula 22:

$$\beta' = (X^T X)^{-1} X^T Y = \left(\sum x_i \, {x_i}^T\right)^{-1} \left(\sum x_i \, y_i\right) (i \in \{i \in N^* | i \leq n\}) \quad (22)$$

The formula is set to solve out the value of the coefficient matrixes of point estimation. With MATLAB giving solution, we obtain the coefficient matrixes which are presented in table 14:

Table 14: Linear Regression Coefficient

	Click Rate	Convert Rate
β_0	0.017671	0.018321
β_1	7.71E-10	6.24E-10
β_2	1.83E-06	1.68E-06
β_3	-0.00055	-0.00053
β_4	8.85E-06	8.64E-06
β_5	-0.00062	-0.00081
β_6	2.61E-06	-9.50E-06

Point estimation possesses a drawback that it cannot express the accuracy of the data obtained. Thus we utilize interval estimation to reuse the Least Square Regression Method, the formula as in formula 23:

$$P\{\hat{\theta}_1 < \theta < \hat{\theta}_2\} = 1 - \alpha \quad (23)$$

θ denotes the parameters to be estimated of the ensemble; P denotes probability; $\hat{\theta}_1$denotes Confidence upper limit; $\hat{\theta}_2$denotes Confidence lower limit; αdenotes reliability which satisfies $0 < \alpha < 1$. In this way, we obtain formula 24

$$P\{\hat{\beta}_{n\,1} < \beta < \hat{\beta}_{n\,2}\} = 1 - \alpha \; (n \in \{n \in N^* | n \leq 2\}) \quad (24)$$

With the MATLAB program, we set α as 0.95, under which the regression coefficient bound is shown in table 15.

Residual graphs are in the appendix, one of which is shown in figure 7. When examining correlation coefficients, we find the correlation coefficients are as presented in table 16:

Table 15: Linear Regression Coefficient Bound

	Click Rate	Convert Rate	Click Rate	Convert Rate

	Lower Bound	Lower Bound	Upper Bound	Upper Bound
β_0	0.013262	0.013966	0.02208	0.022677
β_1	-5.27E-10	-6.58E-10	2.07E-09	1.91E-09
β_2	-2.08E-06	-2.20E-06	5.75E-06	5.55E-06
β_3	-0.0014	-0.00136	0.0003	0.000311
β_4	-3.08E-05	-3.06E-05	4.85E-05	4.78E-05
β_5	-0.00195	-0.00212	0.000702	0.000499
β_6	-0.00019	-0.00019	0.00019	0.000185

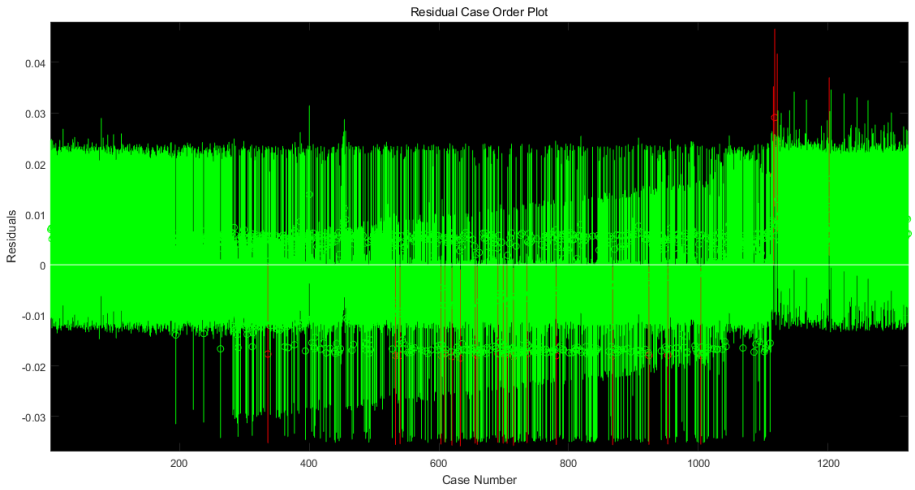

Figure 7: Residual Case Order Plot of Linear Regression

The comparison between the distinction result and the real result is shown in table 17, which shows the accuracy of the distinction. In light of the fact that the accuracy is relative low, which is insufficient to reveal the features of each variable precisely, we consider taking the advantage of other methods.

Table 17: Linear Regression Correlation Coefficient, which is not high enough for further analysis.

Click Rate	Convert Rate
0.115124	0.162528

4.4 KNN Algorithm

In accordance with the given data, we try to randomly sample two-thirds of the data as learning samples and one-third of the data as the test data to highly merge the vast amount of the data and find the shared features and characteristics of each sample to obtain the common properties of the phones under similar sales condition to determine the relationship. [16]

We utilize Mahalanobis distance distinction to operate these data, which is processed after principal component analysis and features eradicating the dimension of each independent variables. The formula is as the following formula 25.

$$d(x,y)=\sqrt{(x-y)\Sigma^{-1}(x-y)^T} \tag{25}$$

Among the formula, x and y denote two row vectors; Σdenotes the covariance matrix; $d(x,y)$denotes the obtained Mahalanobis distance of the data.

For the click rate, we correctly categorized 51 samples out of 444, achieving an accuracy of 11%; for the convert rate, we correctly categorized 72 samples out of 444, achieving an accuracy of 16%, which is too low for further application. Thus, we made an optimization in 5.2.

5 Optimization

5.1 Principal Component Regression

Principal Component Regression suits explicitly for the problems that have a vast amount of independent data types, not all of which are tightly connected to the dependent data, which means some of the data are loosely related to the data. In view of considering that our problem has 26 independent variables, the method is highly compatible with our research.

We can still do as part 4.3, regarding the sales condition as dependent variables and the properties of phones as independent variables. We try to reduce the dimensionality, diminishing the vast amount of the original data and variables into fewer data and variables, while the new variables can retain the information in the original data by and large. [17]

We utilize the 26 original variables mentioned in 3.4 as the original data. We still use $X$to denote independent variables matrixes, $Y_n(n \in \{n \in N^* | n \leq 2\})$. The original variables are $x_p(p \in \{p \in N^* | p \leq l\})$; the new variables are $z_q(q \in \{q \in N^* | q \leq p\})$. We use m to denote the number of samples and use $l$to denote the number of variables in each sample.

Applying Least squares regression, point estimation and interval estimation method which has previously been mentioned, we obtain the principal coefficient matrixβ'as shown in table 18 with formula 25.

$$y_n{}^* = \beta_1{}'z_1 + \beta_2{}'z_2 + \beta_3{}'z_3 + \cdots + \beta_{14}{}'z_{14}(n \in \{n \in N^* | n \leq 2\}) \quad (26)$$

Table 18: Coefficient Matrix of principal component

Click rate	Convert rate	Click rate bond		Convert rate bond	
0.006462	0.0062	0.004059	0.008866	0.003961	0.008439
-0.00041	-0.00037	-0.00051	-0.00032	-0.00046	-0.00028
0.0002	2.03E-05	-0.00024	0.000636	-0.00039	0.000427
0.002007	0.001957	0.001797	0.002217	0.001762	0.002153
0.000336	0.000307	0.000162	0.00051	0.000145	0.000468
0.000999	0.0011	0.000672	0.001325	0.000796	0.001405
-0.00419	-0.00407	-0.00458	-0.00379	-0.00444	-0.0037
0.001884	0.001854	0.001271	0.002498	0.001283	0.002426
-0.00422	-0.0042	-0.00502	-0.00343	-0.00494	-0.00347
-0.0012	-0.00111	-0.00209	-0.00031	-0.00194	-0.00029
0.000897	0.000965	0.000346	0.001447	0.000452	0.001478

-0.00077	-0.00071	-0.00107	-0.00047	-0.00099	-0.00043
-0.00069	-0.00067	-0.00091	-0.00048	-0.00087	-0.00047
-0.00056	-0.00059	-0.00098	-0.00014	-0.00098	-0.0002
-0.00091	-0.0009	-0.00106	-0.00076	-0.00104	-0.00076

The correlation coefficients of this method are 0.805032 and 0.826614, which are satisfactory for further calculation. Ultimately, we conduct the inverse standardization process and obtain the equation interpreted in the original data, which is formula 27, and the final coefficient matrix, as shown in table 19.

$$y_n = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_{26} x_{26} (n \in \{1,2\}) \quad (27)$$

Table 19: Final Coefficient Matrix of original variables

Click rate	Convert rate	Click rate Bound		Convert Rate Bound	
0.0064624	0.0062001	0.0040588	0.008866	0.0039614	0.0084389
-0.000786	-0.000838	-0.001242	-0.00033	-0.001262	-0.000413
-0.000207	-0.000214	-0.000476	0.000063	-0.000465	0.0000374
1.42E-08	4.98E-08	-0.000465	0.0004646	-0.000433	0.0004328
4.31E-10	-3.61E-09	-0.000131	0.0001314	-0.000122	0.0001224
0.0006882	0.0005926	0.0003245	0.0010518	0.0002539	0.0009314
0.000785	0.0007886	0.0012951	0.0002749	0.0012637	0.0003134
1.08E-07	-1.16E-07	-0.000158	0.0001582	-0.000147	0.0001471
0.0009	0.0008054	0.0007563	0.0010438	0.0006715	0.0009393
0.0000249	0.0000324	0.0000949	-0.000045	0.0000976	-3.27E-05
0.0000167	0.0000174	0.0001232	-8.98E-05	0.0001166	-8.17E-05
-8.94E-05	-9.18E-05	-0.000135	-4.39E-05	-0.000134	-4.94E-05
-0.000236	-0.000209	-0.000139	-0.000334	-0.000118	-0.0003
-5.25E-05	-4.31E-05	-0.000661	0.0005556	-0.00061	0.0005232
-4.03E-05	-3.61E-05	-9.11E-05	0.0000104	-8.34E-05	0.0000112
-0.004318	-0.004261	-0.004665	-0.003971	-0.004584	-0.003938
-0.00228	-0.002102	-0.002718	-0.001841	-0.00251	-0.001693
-9.58E-06	-3.22E-06	-0.000632	0.0006129	-0.000583	0.0005766
-0.000208	-0.000244	-0.0003	-0.000117	-0.000329	-0.000158
0.0009217	0.001011	0.0011259	0.0007175	0.0012013	0.0008208
-4.81E-06	-4.41E-06	-3.56E-06	-6.05E-06	-3.24E-06	-5.56E-06
2.92E-10	9.89E-09	-6.03E-05	0.0000603	-5.62E-05	0.0000562
2.09E-06	1.99E-06	-8.34E-05	0.0000876	-7.77E-05	0.0000816
0.0033081	0.0033025	0.0036388	0.0029774	0.0036105	0.0029945
0.0009639	0.0010324	0.0003938	0.001534	0.0005014	0.0015634
-0.002635	-0.002544	-0.003126	-0.002145	-0.003001	-0.002087
0.0016046	0.0016475	0.0011791	0.00203	0.0012513	0.0020438

5.2 Bayes Distinction

Bayes Distinction ideally satisfies the requirements of such issue that each individual of the ensemble exists at different frequencies, which indicates that we need to take

into consideration that the different possibilities that each individual exists. As for our research, each phone is obviously impossible to appear at identical frequencies, so we apply Bayes Distinction to our study.

In the distance distinction method above, it does not take into account the frequency of each sample in the whole and does not take into account the loss caused by the wrong distinction. The Bayes distinction method modifies on the basis of distance distinction, and the formula is defined as in formula 28: [18]

$$P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{\Sigma P(A \mid B_i)P(B_i)} \tag{28}$$

Among which $P(B_i|A)$represents a posteriori probability; $P(A|B_i)$ represents a prior probability; $P(B_i)$represents the frequency at which the sample appears; Σ represents the total covariance matrixes. The distinction rule is that the posterior probability is the highest and the average wrong distinction loss is the lowest, which brings out the rule is as follows: If the condition meets the following formula 29:

$$P(G_l \mid x_0) = \frac{p_l f_l(x_0)}{\Sigma p_j f_j(x_0)} = \max_{1 \le i \le k} \frac{p_i f_i(x_0)}{\Sigma p_j f_j(x_0)} \tag{29}$$

Then we categorize x_0 into G_l, among which G_i is the ensemble, $f(x)$ is the probability density function of G_i, p_i is prior probability of G_i, which is the probability that it belongs a certain category when sample x_0 occurs, and k is the number of G_i. The solution formula for distinction analysis is as the following formulas 30-31:

$$ECM = \sum_{i=1}^{k} p_i \sum_{j \neq i} C(j/i) P(j/i) \tag{30}$$

$$p(j/i) = P(X \in D_j / G_i) = \int_{D_j} f_i(x)dx \qquad i \neq j \tag{31}$$

In this case, $P\left(j/_i\right)$ represents the conditional probability of wrongly categorizing the sample of $G_i$to the ensembleG_j. $C\left(j/_i\right)$is the loss caused by this categorization.$D_k$is a division of a set of distinction samples. ECM is the average wrong distinction loss. The solution to a Bayes distinction analysis is to make the smallest set of solutions.

Using the MATLAB program, we still randomly sample $\frac{2}{3}$of the ensemble as a learning sample and$\frac{1}{3}$as a test set to carry out Bayes distinction solution. We utilize the data after principal component analysis to study the condition of the distinction.

The result is shown in the appendix, part of which is as following figure 8-9 and table 20. For instance, the number “91” shows that there are 91 samples with 2G RAM are judged as click rate category 1.

For the click rate, we correctly categorized 89 samples out of 444, achieving an accuracy of 20%; for the convert rate, we correctly categorized 176 samples out of 444, achieving an accuracy of 40%, which is relatively higher than the accuracy obtained from KNN algorithm.

Table 20: RAM result in click rate. Both better sales and worse sales focus on higher RAM.

	Category 1	Category 2	Category 3	Category 4
0.125	2	0	0	0
0.5	7	0	0	1
1	49	0	1	7
1.5	1	0	0	0
2	91	0	15	25
3	41	1	64	19
4	1	13	68	2
6	0	11	22	0
8	0	0	1	0

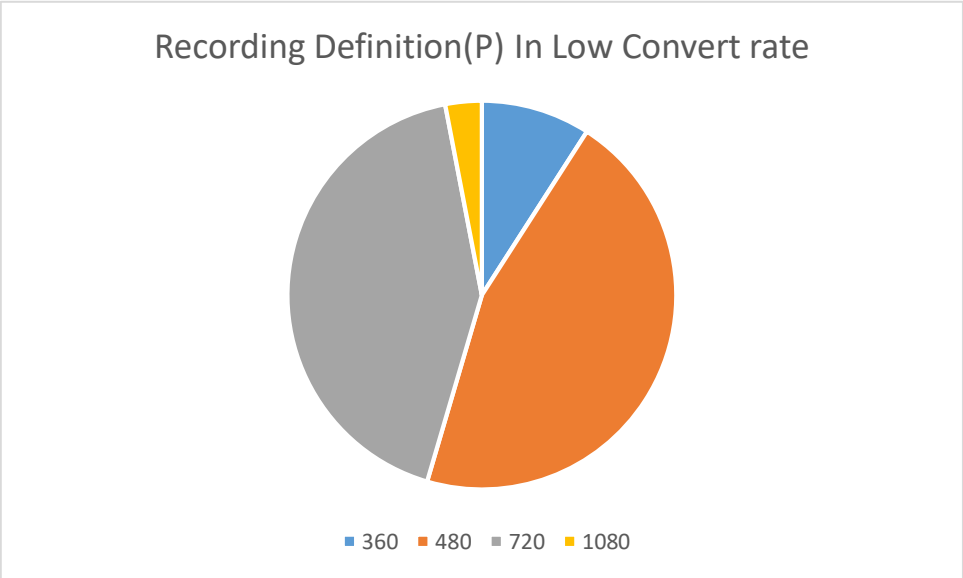

Figure 8: Bayes Result of Recording Definition in Low Convert Rate. Lower Recording Definition has a relatively lower Convert Rate.

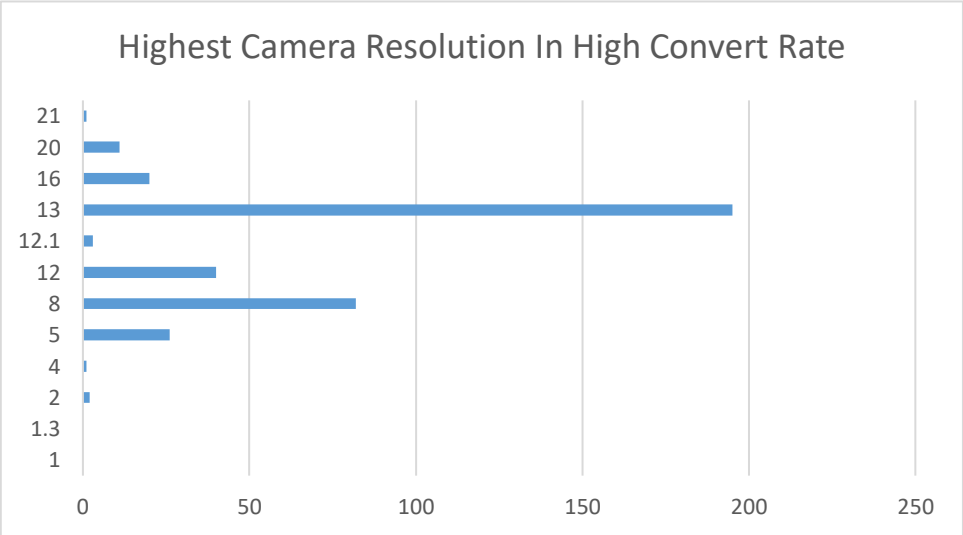

Figure 9: Bayes Result of Highest Camera Resolution in High Convert Rate. Phones with Higher Camera Resolution demonstrates better sales.

From the results given, we can clearly figure out the trend that the higher the mobile configuration is, the high click rate and convert rate the sample has. To be specific, phones with higher display resolution, higher recording definition, higher camera resolution, more CPU cores, larger RAM, and more spacious ROM are apt to reveal more satisfactory sales condition. The phones that display weaker sales performance tend to possess lower counterparts of the features listed above.

The comparison between the distinction result and the real result is shown in table 17, which shows the accuracy of the regression. We can see that it is much higher than linear regression.

Table 17: Principal Regression Correlation Coefficient, which is higher enough for the Regression.

Click Rate	Convert Rate
0.805032	0.826614

5.3 BP Neural Network Fitting

BP Neural Network is a kind of multilayer feed-forward network, which highly fits for the problem that there are data with a certain scale, the relationship between which is not too complicated to identify. When it comes to our target, we have a middle-sized database, while the process we want is fitting, which is not too intricate, which shows that the model can be applied to our goal.

We utilize BP neural network fitting as another method to promote the accuracy of the regression. BP neural network works to encode itself with its high-dimensional features and to carry out dimension reduction processing towards high-dimensional data. It is marked by a feature extraction model with unsupervised learning, which can also combine a few basic features to obtain higher-layer abstract features. [19]

We utilize Tangent Sigmoid function as the transfer function; we use Levenberg Marquardt algorithm (trainlm) as the training algorithm; we use the Gradient descent with momentum weight and bias learning function (learngdm) as the learning algorithm; we use the mean square error (MSE) method as the learning function. The structure of the network and the performance plot are shown in figure 10 and 11.

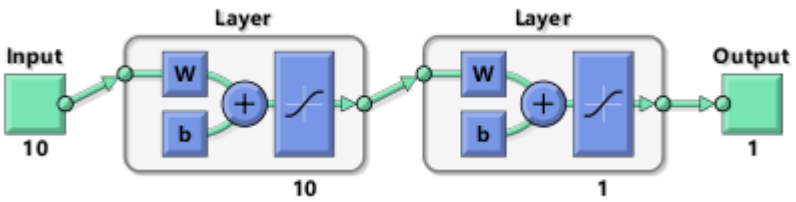

Figure 10: BP Neural Network Structure. The layer number, which is 10, does not consumes too much time while the result is satisfactory.

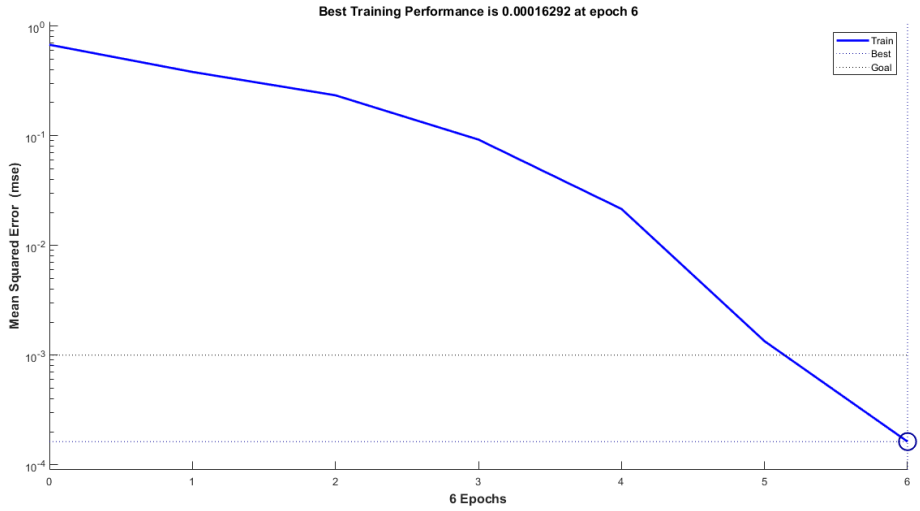

Figure 11: the performance plot of BP Neural Network. The training performance is enhancing rapidly.

We utilize the properties after the Information Entropy analysis to conduct the process. Using the MATLAB program, we still randomly samples $\frac{2}{3}$of the ensemble as a learning sample and$\frac{1}{3}$as a test set to carry out the BP neural network fitting.

Table 21: BP Neural Network Result. The error of some numbers is lower than 1%.

O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
					ClickRate mean							ConvertRate	
0.02078	0.011384	0.008857	0.004261	0.021461	0.013348		0.014479	-0.00362	0.02533	0.022221	0.007262	0.013133	
0.020471	0.002053	0.005953	0.002532	-0.00187	0.005829		0.016294	0.001955	0.009609	0.013991	0.007917	0.009953	
0.021792	0.020835	0.013656	0.017043	0.010632	0.016791		0.016819	0.009979	0.020041	0.002321	0.021112	0.014054	
0.015907	0.013295	0.005368	0.002473	0.001955	0.0078		0.017374	0.007013	0.009318	0.010874	0.007137	0.010343	
0.012225	0.020101	0.01652	0.019162	0.021755	0.017953		0.016328	0.008709	0.013901	0.014945	0.017033	0.014183	
0.020453	0.010045	0.006676	0.002474	-0.00278	0.007373		0.013149	0.007797	0.007339	0.013637	0.008797	0.010144	
0.015864	0.010883	0.006219	0.002389	-0.00382	0.006307		0.015677	0.009632	0.007558	0.010519	0.007801	0.010238	
0.016053	0.003646	0.006422	0.0025	-0.00062	0.0056		0.016442	0.00814	0.008473	0.011153	0.007686	0.010379	
0.021603	0.021945	0.016347	0.019545	0.016197	0.019127		0.015357	0.019486	0.012833	0.01432	0.012956	0.01499	
0.018423	0.012065	0.007084	0.002474	-0.00151	0.007707		0.013997	0.007761	0.007401	0.013494	0.008178	0.010166	
0.01248	0.0216	0.012888	0.017519	0.017937	0.016485		0.015002	0.012329	0.016478	0.005335	0.012801	0.012389	
0.016136	0.010158	0.006237	0.002339	0.000183	0.007011		0.015353	0.009618	0.007716	0.011093	0.008007	0.010357	
0.015526	0.008238	0.006689	0.002239	0.002612	0.007061		0.014566	0.011249	0.006636	0.006922	0.004731	0.008821	
0.017688	0.010685	0.008449	0.005342	0.007715	0.009976		0.013769	0.008002	0.002354	0.021218	0.006657	0.0104	
0.024372	0.022045	0.016973	0.019874	0.018377	0.020328		0.014417	0.027674	0.015893	0.014049	0.012057	0.016818	
0.005814	0.020225	0.00924	0.00513	0.011588	0.010399		0.01433	0.010461	0.012536	0.009471	0.006633	0.010686	
0.023505	0.012874	0.016141	0.019691	0.01486	0.017414		0.01492	0.024921	0.017553	0.01721	0.012591	0.017439	
0.021031	0.003062	0.006779	0.002463	-0.00594	0.005479		0.016598	-0.00305	0.011454	0.017018	0.008545	0.010112	
0.007895	0.019178	0.005459	0.002725	0.006349	0.008321		0.014353	-0.00967	-0.00093	0.019134	0.006623	0.005903	
0.004709	0.011095	0.006211	0.002544	0.004508	0.005813		0.018337	0.007168	6.45E-05	0.010817	0.004856	0.008249	
0.020242	0.001131	0.006219	0.002481	-0.00251	0.005513		0.015004	0.007746	0.008241	0.012762	0.008476	0.010446	
0.015613	0.014638	0.006331	0.002231	0.00345	0.008452		0.014434	0.011731	0.006555	0.007063	0.005558	0.009068	
0.022037	0.009821	0.006919	0.001946	0.000596	0.008264		0.02166	0.003533	0.007823	0.010596	0.007032	0.010129	
0.016142	0.010266	0.017648	0.019628	0.022345	0.017206		0.015378	0.016143	0.022779	-0.00053	0.019856	0.014726	
-0.00863	0.017637	0.019993	0.017025	0.030315	0.015269		0.016931	0.016732	0.012842	-0.00257	0.009733	0.010734	

We divide the learning samples into five groups, each time using four of the groups to carry out a model and then test the test set. Therefore we can obtain five identical models, and then we use the BOOST algorithm to get the means of the five result. The result is in the appendix, part of which is as the following table 21.

It can be seen that some of the predicted data run an accuracy that is higher than 99%.

We also utilize a formula to measure the error of our estimation, reaping an average score of 9.45 of click rate and 9.46 of convert rate out of 10, which shows that this model can successfully reflect the trend. The formula is as the following formula 32.

$$S_k = max\left(0{,}10 - 10 \times \left|\frac{log_{10}\left|\frac{x_{predict}}{x_{real}}\right|}{5}\right|\right) \tag{32}$$

5.4 BOOST Algorithm

We utilize BOOST algorithm to obtain the average value of each method of the samples. The formula is as the following formula 33

$$\bar{x} = \sum_{i=1}^{3} \frac{x_i}{3} \tag{33}$$

The theory of BOOST algorithm is as follows. For a complicated issue, it is a better judgment when synthesizing the judgment of each expert than that of a sole expert. For each step, we generate a model accumulate each model to a whole model, which enables us to analyze the problems.

In the formula, x_1denotes the original result of Principal Component Analysis. x_2 denotes the result of Bayes distinction. x_3 denotes the original result of BP neural network fitting. For each category in Bayes distinction, we utilize the mid-value of each interval to numerate each category. We divide the click rate into 5 categories, which are 0-0.1, 0.1-0.2, 0.2-0.225, 0.225-0.3, 0.3-0.464, as well as the convert rate into 5 categories, which are 0-0.1, 0.1-0.2, 0.20-0.22, 0.22-0.23, 0.23-0.468. Therefore, we use 0.05, 0.15, 0.2125, 0.2625, and 0.382 to denote the 5 result of the categories. We use 0.05, 0.15, 0.21, 0.225, and 0.349 to denote the 5 result of the categories.

6 Application

Table 22: Final Result

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	ID	Unlock PH	Google Pl	Battery T	Battery C	Display R	Display R	Operatior	SIM Card	Recordin	Touch Scr	RAM(G)	ROM(G)	CPU	Display S	Size_X(mm	Size_Y(mm	Size_Z(mm	Hi
2																			
3	9083.0	1.0	1.0	2.0	4100.0	1280	720	1.0	2.0	720.0	1.0	2.0	0.0	1.0	5.0	139.24	69.96	8.65	13
4	9072.0	1.0	1.0	2.0	4100.0	1280	720	1.0	2.0	1080.0	1.0	3.0	0.0	1.0	5.0	139.24	69.96	8.65	13
5	8160.0	1.0	1.0	2.0	3120.0	1280	720	1.0	2.0	720.0	1.0	2.0	0.0	2.0	5.0	139.5	70.4	8.5	13
6	9273.0	1.0	1.0	2.0	4100.0	1920	1080	1.0	2.0	1080.0	1.0	4.0	64.0	0	5.5	151.0	76.0	8.45	13
7	8585.0	1.0	1.0	2.0	4100.0	1920	1080	1.0	2.0	1080.0	1.0	4.0	64.0	0	5.5	151.0	76.0	8.45	13
8	8574.0	1.0	1.0	2.0	4100.0	1280	720	1.0	2.0	720.0	1.0	3.0	0.0	1.0	5.0	139.24	69.96	8.65	13
9	9835.0	1.0	1.0	2.0	4100.0	1920	1080	1.0	2.0	1080.0	1.0	3.0	64.0	0	5.5	151.0	76.0	8.35	13
10	9078.0	1.0	1.0	2.0	4100.0	1280	720	1.0	2.0	720.0	1.0	2.0	0.0	1.0	5.0	139.24	69.96	8.65	13
11	9863.0	1.0	1.0	2.0	4100.0	1920	1080	1.0	2.0	1080.0	1.0	4.0	64.0	0	5.5	151.0	76.0	8.45	13
12	6724.0	1.0	1.0	2.0	4100.0	1920	1080	1.0	2.0	1080.0	1.0	3.0	32.0	0	5.5	151.0	8.35	76.0	13
13	9089.0	1.0	1.0	2.0	3060.0	1920	1080	1.0	2.0	1080.0	1.0	3.0	32.0	0	5.5	153.6	75.2	7.25	12
14	9856.0	1.0	1.0	2.0	4100.0	1920	1080	1.0	2.0	1080.0	1.0	4.0	64.0	0	5.5	151.0	76.0	8.45	13
15	9097.0	1.0	1.0	2.0	4100.0	1920	1080	1.0	2.0	1080.0	1.0	3.0	32.0	0	5.5	151.0	76.0	8.35	13
16	36.0	1.0	2.0	1.0	1500.0	128	160	4.0	2.0	360.0	3.0	0.125	2.0	0	1.77	115.0	49.0	13.5	1.(
17	8592.0	1.0	1.0	2.0	3060.0	1920	1080	1.0	2.0	1080.0	1.0	3.0	32.0	0	5.5	153.6	75.2	7.25	12
18	8362.0	1.0	1.0	2.0	3060.0	1920	1080	1.0	2.0	1080.0	1.0	4.0	32.0	0	5.5	153.6	75.2	7.25	12
19	8579.0	1.0	1.0	2.0	3060.0	1920	1080	1.0	2.0	1080.0	1.0	4.0	32.0	0	5.5	153.6	75.2	7.25	12
20	8188.0	1.0	1.0	2.0	3000.0	1920	1080	1.0	2.0	1080.0	1.0	4.0	32.0	0	5.5	151.1	74.2	7.5	21
21	1799.0	1.0	2.0	1.0	3000.0	320	240	4.0	3.0	360.0	3.0	0.125	2.0	0	2.4	132.0	62.0	22.0	2.(
22	503.0	1.0	2.0	1.0	1450.0	240	320	4.0	2.0	360.0	3.0	0.125	2.0	0	2.4	55.0	126.5	15.5	2.(
23	7531.0	1.0	0	1.0	2600.0	1920	1080	1.0	1.0	1080.0	1.0	2.0	16.0	2.0	5.0	136.6	69.8	7.9	13
24	240.0	1.0	2.0	1.0	1500.0	320	240	4.0	2.0	360.0	3.0	0.125	2.0	0	1.8	100.0	43.0	18.0	0.:
25	9612.0	1.0	1.0	2.0	3060.0	1920	1080	1.0	2.0	1080.0	1.0	3.0	32.0	0	5.5	153.6	75.2	7.25	12

We use the data which have exactly one zero of each data as the test sets and conduct

the BP neural network illustrated in part 5 to obtain the final result to show that our models and methods can be applied to a broader range. We use the data after Principal Component Analysis for Principal Component Regression and Bayes Distinction and data after Information Entropy for BP neural network Fitting. The following table 22 is a part of the final result.

7 Sensitivity Analysis

Sensitivity analysis is a method of studying and analyzing the sensitivity of the model to changes in system parameters or surrounding conditions. In our team's optimization methods, it can detect the stability of our model, especially when the given data is not accurate.

In this part, we will mainly discuss the sensitivity of the application part. If we give the test set of the data an increase or a decrease of 1%, by changing the value of the original data matrix on the program, we discover that the output data of the principal component regression changes precisely 1%; almost all the results in the Bayes Distinction part have no difference in categories; the majority of the output of BP neural network model fluctuates 1% approximately. The output after the change is small enough for us to make a further adjustment. Therefore, it is acceptable in the modeling. This sensitivity analysis also indicates that our model has universality and can be applied to more situations. For instance, if there is some error in the data, out final result does not vary rapidly correspondingly. Therefore, our model is relatively stable. The data of Sensitivity Analysis can be referred to the appendix part.

8 Conclusion

8.1 Strength and Weakness

The method we propose in the paper has effectively made up the vacancy and deficiency of the previous evaluating process regarding the sale volume of cell phones, and several main advantages are as the following. For a start, it presents the ranking of the most important individual variables within the cell phone market, the results of which are seldom considered by manufacturers but actually of great significance. Manufacturers can take specific traits of cellphones into consideration, deciding which types or combinations of traits are more profitable to produce and fit the need of their target customers. Furthermore, as the application section in the paper indicates, the process we propose can also be applied to pragmatic purposes. By using the method linked with BP neural network, the process can successfully predict the outcome of the sale volume of cellphones before they are released into the market, and the margin of error is within an acceptable level. Besides the application of the evaluating process in real life, the method itself is also more advanced and comprehensive than that in the previous thesis. For the method of information entropy in data processing and BP neural network in the optimization, they not only fit in the exact needs of the data being processed and the expected outcome, but they are also

more precise and reliable, ensuring the credibility of the model as a whole.

Admittedly, there are several shortcomings concerning the whole paper, like for some particular methods including Grey Relational Analysis, the results are not very desirable, and they are not entirely useful for the later optimization. However, considering the techniques being applied as a whole, the advantages obviously outweigh the deficiencies, thus making the modeling reliable for reference and have high practical value.

8.2 Conclusion

We have discovered some intriguing and unexpected conclusion throughout the modeling process.

- It is the most amusing that phones sold either the best or the worst focus on the highest RAM, ROM, and CPU, which means that these three factors attract the customers a lot. If the mobile manufacturer is willing to enhance the phone specs without adding price, they should consider promoting the RAM, ROM, and CPU for the most priority.
- Phones with upper-middle display resolution sell better, while phones with middle and the highest counterparts sell worse. Phones with the lowest and highest recording definition gain better sales. The sellers and manufacturers should not pay much attention to these factors because these factors are less concerned by customers.
- For the Highest Camera Resolution and Price, the ones with middle and lower-middle condition sell well, and the ones with upper-middle or the highest equivalents sell experience a tough sell. Maybe the prices are too high for ordinary users to purchase, while the users do not need such high specs on phones. Compared with adding the versatile specs, the manufactures ought to choose lowering down the prices rather.

Despite the specific conclusion and some reasonable explanation, our research also yields significant results in the following four aspects. First, the method concerning AHP produces a qualitative analysis of which specific traits in the individual variables promote the sale of the phones the best way. For example, regarding the display resolution, target readers can clearly make out the third category as bringing more profit and contributing more to the sale. The results can be compiled into graphs and thus providing the whole picture in a straightforward way. Besides, AHP is an easily accessible method and produces a relatively reliable result.

Second, the quantitative research can be used to rank the factors and determine which elements are the most crucial ones that the manufacturers should take into consideration. The results reflect the customers' tendency towards different types of cell phones, and their preference is carefully studied using information entropy in the data processing. The ranking of individual variables gives the target readers a broader view of which ones are the keys to promoting the sale and lays the stepping stone for

the further optimization relating to the different traits in individual variables.

Third, the optimization process allows the determination of specific characteristics that contribute to the highest sales volume. The optimization using Bayes distinction and BP neural network further explores the result in particular details. For example, as for the individual variable like color, it can be analyzed that gold contributes to the highest sale, the fact of which will definitely give the manufacturers more detailed references when making decisions about the production of certain cell phones. For other variables, the same method can be applied, either, yielding valuable insight into specific traits.

Last but not least, the sales volume of the cell phones can be successfully predicted by applying the method in the optimization process, as mentioned in the application section. These methods enable the manufacturers to predict the sale with given characteristics, and according to the sensitivity analysis and data testing, the model is reliable and can be applied for other practical uses.

9 Reference

[1] J. Chevalier, D. Mayzlin. The Effect of Word Of Mouth on Sales: Online Book Review [J]. Working Paper, 2003.12.

[2] Michael D. Smith, Erik Brynjolfsson. Consumer Decision-Making at an Internet Shop bot: Brand Still Matters [J]. The Journal of Industrial Economics, 2001.12(4):541-558.

[3] Jie Z., Jianan Z. Research of promotion's influence to customers' purchasing behaviors [A]. In The 11th National Conference on Psychology [C]. Kaifeng, China, 2007: 278.

[4] Gang D., Zhenyu H. Prediction of customers' purchasing behaviors in the Big Data environment [J]. Modernization of Management, 2015, 1(14): 40-42.

[5] Zhanbo Z., Luping S. and Meng S., Research of comparison between factors in C2C influencing page view and sales volume[J].Journal of Management Science,2013,26(1): 58-67.

[6] Zhihai H., Dandan Z. and Yi Z. An Empirical Study on the Effect of Online Reviews on Product Sales [J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2015, 12(11): 52-55.

[7] Naicong H., Xu Z., Enjun Z. Grey Relational Analysis of online reputation and sales volume——movie data as an example [J], Modernization of Management, 2015, 2(10):28-30.

[8] Xiao S. Research of influential factors of online sales based on Grey Relational Analysis [D], Yunnan University of Finance and Economics, Yunnan, 2017.

[9] Youzhi X., Yongfeng G. Competitive Strategy of E-business Sellers on Consumer-to-Consumer Platform: Based on Data from Taobao.com [J], Nankai Business Review, 2012, 15(1): 129-140.

[10] Jingsha F. The Study of Influencing Factors and Index System of C2C Online Shop Sales Volume Based on the Soft Set Theory [D], Chongqing Jiaotong University, Chongqing, 2016.

[11] Wenxuan H. Study on the factors influencing the purchase behavior of Wechat business customers [D], Nanchang University, Nanchang, 2016.

[12] Jiao L. Research on customer purchase behavior analysis system based on Data Mining [J], Time Finance, 2015, 2(2): 320-321.

[13] http://blog.csdn.net/MATLAB_matlab/article/details/59483185?locationNum=10&fps=1, MATLAB principal component analysis.

[14] Mingbei C., Gang H., Guoufu Z. Comprehensive evaluation of takeaway website based on AHP method——Eleme website as an example [J]. Modern Business, 2015, 12: 57-58.

[15] https://wenku.baidu.com/view/99c8408e6529647d272852cd.html, Interval estimation and linear regression analysis with MATLAB.

[16] Jiang W. Comparative Study of Fisher Discriminant and Mahalanobis Distance Discriminant [J]. Journal of Ningbo Polytechnic, 2017, 21(5); 91-94.

[17] Yimeng F. Analysis of influencing factors of customer purchase behavior in E-commerce [J], Industrial & Science Tribune, 2014, 13(8): 138-139.

[18] Haiwei W. Yu X., Yalin W.A bivariate hierarchical Bayesian approach to predicting customer purchase behavior [J], Journal of Harbin Engineering University, 2007, 28(8): 949-954.

[19] Wu P. Application of Cigarette Sales Forecasting Based on Neural Network [J], Computer Simulation, 2012, 29(3): 227-230.

10 Acknowledgement

小组分工如下。

在本文写作过程中，钱成完成了摘要、信息熵、优缺点分析及结论部分的制作及写作；曹凌微完成了研究背景、目前研究状况、研究意义、研究方法的制作和写作；田肇阳完成了数据处理、灰色关联度、主成分分析和回归、权重确定方法、线性回归、距离和贝叶斯判别、BP 神经网络的制作和写作。

以下为指导老师与参赛成员简历。

吴昊，副教授，工作于清华大学数学科学系。 教育背景：应用数学专业博士，2009 年 7 月，清华大学，导师：金石教授。数学专业学士，2004 年 7 月，清华大学。 工作经历：长聘副教授，2016 年 12 月至今，数学科学系，清华大学。准聘副教授，2013 年 12 月至 2016 年 11 月，数学科学系，清华大学。博士后，2009 年 11 月至 2010 年 10 月，数学学院，保罗 •萨巴蒂大学(图卢兹三大)，合作导师：Prof. Naoufel Ben Abdallah。系主任助理，2017 年 9 月至今，数学科学系，清华大学。教育委员会成员，2016 年 12 月至今，中国工业与应用数学学会。 主要奖励：曾获全国优秀博士学位论文提名，国务院学位委员会，2012。优秀青年论文一等奖，中国计算数学学会，2011。青年教师教学优秀奖，清华大学，2017。北京高校第八届青年教师教学基本功比赛一等奖，北京市教育委员会，2013。

王殿军，男，汉族，1960 年 9 月生于陕西。1982 年 1 月在陕西师范大学数学系获得理学学士学位。1997 年 7 月在北京大学数学学院获得博士学位。1997 年 8 至 1999 年 7 月为清华大学数学系博士后。1999 年 8 月至 2006 年 12 月为清华大学数学系副教授、教授，先后担任过数学系研究生工作组组长、党委副书记、党委书记。2007 年 1 月起任清华大学附属中学校长。王殿军长期在大学的教学科研一线工作，主讲过十余门课程，其中北京市和清华大学的精品课程各一门，近五年所讲授的主要课程教学评估均居清华大学前 5%。完成了国家自然基金等各类科研项目近十项，发表学术论文 30 余篇，改编和编著出版书籍各两部，指导博士后 2 名、博士生 1 名、硕士生 5 名。曾荣获清华大学优秀辅导员 “林枫奖”、 清华大学优秀教学成果奖、清华大学青年教师教学优秀奖、北京市优秀教学成果奖以及“北京市教育创新标兵”、“北京市优秀教师”等荣誉称号。

钱成，男，现就读于清华大学附属中学，品学兼优，全面发展。2017 年中考总分 574 分（北京市海淀区裸分并列第二名），清华附中 2017 学年启迪奖学金的获得者。在学校成绩多次名列年级总分前三，其中高一第一学期期末考试总成绩排名第一。初中担任班级体育委员，高中担任学习委员，并多次被评为海淀区三好学生，优秀学生干部。同时，他积极组织和参加校内外的各项活动，担任清华附中上地学校课外辅导员及清华附中模拟联合国社团学术委员会成员，并曾参与 2018 清华大学钱学森力学班 oricplus 科创营。此外，竞赛方面曾获 2018 中年全国中学生数理化学科能力竞赛高一年级组数学一等奖、物理一等奖；北京市初中数学联赛二等奖；北京市高一数学竞赛一等奖；北京市应用数学竞赛一等奖（建模论文二等奖）；ASDAN 美式数学竞赛几何学个人赛第二名，团队力量赛第二名，团队车轮战第 11 名，团队总分第 9 名；2018 年 AMC12 总成绩 124（全球前 1%，AIME 折合成绩 204）等多项国内国际优异成绩。建模方面，曾获美国

高中生数学建模比赛荣誉提名，清华大学登峰杯数学建模比赛省级一等奖，复赛二等奖及全国总决赛三等奖。

田肇阳，男，现就读于清华大学附属中学。本人成绩每学期各科都是优，身体素质和体质也均为优。10 年级上期中考试全年级排名第 20 名，10 年级上期末考试全年级排名第 30 名，10 年级下期末考试全年级排名第 40 名（约 600 名学生）。2017 年 11 月参加美国高中数学建模比赛（HiMCM）获得 Honorable Mentioned 奖项。2018 年全国中学生数理化学科能力竞赛高一年级组数学一等奖、物理一等奖、化学一等奖。2018 年 AMC12 国际数学竞赛 118 分，进入全球前 1%；AIME 折算成绩 178 分。2018 年 3 月获得清华大学主办的“登峰杯”数学建模比赛省级赛区一等奖。2018 年 5 月获得清华大学主办的“登峰杯”数学建模比赛复赛二等奖及全国总决赛三等奖。2018 年 8 月参加 AMT 比赛获得个人代数前 25%，几何前 40%，微积分第三名，团队力量赛第二名，团队车轮战第 11 名，团队总分第 9 名成绩。2017 年 9 月开始参加清华附中计算机科学实验室，利用 Python 及 Arduino 完成《给盲人带上眼睛——识别红绿灯的提醒系统》项目制作；利用 Urllib，Tensorflow，Jieba 及 Wordcloud 模块完成《关键词搜索及二分类》项目制作 2018 年 1 月至今入选英才计划，完成《近似新闻合并及正负面评价》项目。

曹凌微，女，现就读于清华大学附属中学。本人成绩优异，年级排名前 3%（前 20 名），英语成绩尤为突出：2018 年 8 月托福 115 分；2018 年 3 月 SAT 1540 分；2018 年 5 月 AP 微积分 5 分。曾参与的学术活动：2018 年 7 月参加清华大学钱班和清华附中共同开展的首届 ORIC+创新营；2017 年 7 月参加斯坦福大学夏校数学 Logic and Problem-solving 课程三周。曾获奖项：2018 年 3 月获第三届登峰杯数学建模竞赛北京市一等奖；2018 年 5 月获第三届登峰杯数学建模竞赛复赛二等奖；2018 年 8 月获第三届登峰杯数学建模竞赛全国总决赛三等奖；2018 年 1 月获第十届全国中学生数理化学科能力展示活动北京赛区高一年级数学一等奖；2017 年 12 月获全国中学生英语能力竞赛高一年级全国一等奖；2018 年 8 月获 ASDAN 美式数学竞赛个人代数第八名，几何前 10%，团队力量赛第二名，团队车轮战第 11 名，团队总分第 9 名；2018 年 1 月获美国高中生数学建模大赛（HiMCM） Honorable Mention 奖；2018 年 AMC12 国际数学竞赛 105 分，进入全球前 1%；AIME 折算成绩 178 分。

11 Declaration

本参赛团队声明所提交的论文是在指导老师指导下进行的研究工作和取得的研究成果。尽本团队所知，除了文中特别加以标注和致谢中所罗列的内容以外，论文中不包含其他人已经发表或撰写过的研究成果。若有不实之处，本人愿意承担一切相关责任。

参赛队员：

指导老师：

2018 年 9 月 28 日

12 Appendix

12.1 PYTHON Code

```
import xlrd
import xlwt
ExcelFile=xlrd.open_workbook(r'C:\Users\tianzhy\Desktop\sumai.xlsx')
sheet=ExcelFile.sheet_by_name('速卖')
workbook = xlwt.Workbook(encoding = 'ascii')
worksheet = workbook.add_sheet('My Worksheet')
row=0
for i in range (1,2291):
    count=0
    for j in range (0,19):
        temp = str(sheet.cell(i,j).value)
        if float(temp)==0.0:
            count=count+1
        #print(j)
    if count<=3:
        row=row+1
        for j in range (0,56):
            temp = str(sheet.cell(i,j).value)
            worksheet.write(row, j, label = str(temp))
        worksheet.write(row, 56, label = str(i+1))
    #print(i)
workbook.save('Excel_Workbook.xls')

import xlrd
import xlwt
ExcelFile=xlrd.open_workbook(r'C:\Users\tianzhy\Desktop\detail_numerical.xls')
sheet=ExcelFile.sheet_by_name('My Worksheet')
workbook = xlwt.Workbook(encoding = 'ascii')
worksheet = workbook.add_sheet('My Worksheet')
for i in range (1,2291):
    temp = str(sheet.cell(i,7).value)
    if ('Gravity Response' in temp) :
        worksheet.write(i, 0, label = str(1))
    else:
        worksheet.write(i, 0, label = str(0))
    if ('GPRS' in temp):
        worksheet.write(i, 1, label = str(1))
    else:
        worksheet.write(i, 1, label = str(0))
workbook.save('Excel_Workbook.xls')
```

```
import xlrd
import xlwt
ExcelFile=xlrd.open_workbook(r'C:\Users\tianzhy\Desktop\工作簿 1.xlsx')
sheet=ExcelFile.sheet_by_name('Sheet1')
workbook = xlwt.Workbook(encoding = 'ascii')
worksheet = workbook.add_sheet('My Worksheet')
for i in range (1,2291):
    temp = str(sheet.cell(i,0).value)
    if 'x' in temp:
        temp = str.split(temp,'x')
        worksheet.write(i, 0, label = str(temp[0]))
        worksheet.write(i, 1, label = str(temp[1]))
        worksheet.write(i, 2, label = str(temp[2]))
    elif '*' in temp:
        temp = str.split(temp,'*')
        worksheet.write(i, 0, label = str(temp[0]))
        worksheet.write(i, 1, label = str(temp[1]))
        worksheet.write(i, 2, label = str(temp[2]))
    elif 'X' in temp:
        temp = str.split(temp,'X')
        worksheet.write(i, 0, label = str(temp[0]))
        worksheet.write(i, 1, label = str(temp[1]))
        worksheet.write(i, 2, label = str(temp[2]))
    else:
        worksheet.write(i, 0, label = str(temp[0]))
        worksheet.write(i, 1, label = str(temp[1]))
        worksheet.write(i, 2, label = str(temp[2]))
        temp=[0,0,0]
    temp[0]=float(temp[0])
    temp[1]=float(temp[1])
    temp[2]=float(temp[2])
    judge=temp[0]*temp[1]*temp[2]
    if judge<36.8633431902425:
        temp[0]=temp[0]*25.4
        temp[1]=temp[1]*25.4
        temp[2]=temp[2]*25.4
    elif judge<4712.4514674042:
        temp[0]=temp[0]*10
        temp[1]=temp[1]*10
        temp[2]=temp[2]*10
    worksheet.write(i, 3, label = str(temp[0]))
    worksheet.write(i, 4, label = str(temp[1]))
    worksheet.write(i, 5, label = str(temp[2]))
```

```
        #worksheet.write(i, 0, label = str(count))
workbook.save('Excel_Workbook.xls')

import xlrd
import xlwt
ExcelFile=xlrd.open_workbook(r'C:\Users\tianzhy\Desktop\sumai.xlsx')
sheet=ExcelFile.sheet_by_name('速卖')
workbook = xlwt.Workbook(encoding = 'ascii')
worksheet = workbook.add_sheet('My Worksheet')
properties1=['Unlock Phones','Google Play','Battery Type','Battery Capacity','Display Resolution','Operation System','Feature','SIM Card Quantity','Recording Definition','Touch Screen Type','RAM','ROM','color']

properties2=['Size','Display Size']
properties3=['Camera：','Camera Type','Front Camera：']
properties4=['CPU：Octa Core','CPU：Quad Core','CPU：Dual Core']
for i in range (1,2291):
    temp = sheet.cell(i,6).value
    temp = str.split(temp,'<br>')
    length=len(temp)
    for j in range (0,13):
        vari=0
        for k in range (0,length):
            if properties1[j] in temp[k]:
                vari=temp[k]
        worksheet.write(i, j, label = str(vari))
    vari=0
    for k in range (0,length):
        if properties4[0] in temp[k]:
                vari=properties4[0]
        if properties4[1] in temp[k]:
                vari=properties4[1]
        if properties4[2] in temp[k]:
                vari=properties4[2]
    worksheet.write(i, 13, label = str(vari))
    vari=0
    for k in range (0,length):
        if 'Size' in temp[k]:
            vari=vari+1
            if vari < 3:
                worksheet.write(i, (13+vari), label = str(temp[k]))
    vari=0
    for k in range (0,length):
        if 'Camera：' in temp[k]:
```

```
                vari=vari+1
                worksheet.write(i, (15+vari), label = str(temp[k]))

workbook.save('Excel_Workbook.xls')

import xlrd
import xlwt
ExcelFile=xlrd.open_workbook(r'C:\Users\tianzhy\Desktop\detail_numerical.xls')
sheet=ExcelFile.sheet_by_name('My Worksheet')
workbook = xlwt.Workbook(encoding = 'ascii')
worksheet = workbook.add_sheet('My Worksheet')
for i in range (1,2291):
    temp = str(sheet.cell(i,13).value)
    if ('Black' in temp) or ('black' in temp):
        worksheet.write(i, 0, label = str(1))
    else:
        worksheet.write(i, 0, label = str(0))
    if ('White' in temp) or ('white' in temp):
        worksheet.write(i, 1, label = str(1))
    else:
        worksheet.write(i, 1, label = str(0))
    if ('Blue' in temp) or ('blue' in temp):
        worksheet.write(i, 2, label = str(1))
    else:
        worksheet.write(i, 2, label = str(0))
    if ('Rose' in temp) or ('rose' in temp):
        worksheet.write(i, 3, label = str(1))
    else:
        worksheet.write(i, 3, label = str(0))
    if ('Gold' in temp) or ('gold' in temp) or ('champange' in temp) or ('Champange' in temp):
        worksheet.write(i, 4, label = str(1))
    else:
        worksheet.write(i, 4, label = str(0))
    if ('Silver' in temp) or ('silver' in temp):
        worksheet.write(i, 5, label = str(1))
    else:
        worksheet.write(i, 5, label = str(0))
    if ('Grey' in temp) or ('grey' in temp) or ('titanium' in temp) or ('Titanium' in temp):
        worksheet.write(i, 6, label = str(1))
    else:
        worksheet.write(i, 6, label = str(0))
    if ('Pink' in temp) or ('pink' in temp):
        worksheet.write(i, 7, label = str(1))
    else:
```

```
            worksheet.write(i, 7, label = str(0))
        if ('Brown' in temp) or ('brown' in temp):
            worksheet.write(i, 8, label = str(1))
        else:
            worksheet.write(i, 8, label = str(0))
        if ('Orange' in temp) or ('orange' in temp):
            worksheet.write(i, 9, label = str(1))
        else:
            worksheet.write(i, 9, label = str(0))
        if ('Yellow' in temp) or ('yellow' in temp):
            worksheet.write(i, 10, label = str(1))
        else:
            worksheet.write(i, 10, label = str(0))
        if ('Red' in temp) or ('red' in temp):
            worksheet.write(i, 11, label = str(1))
        else:
            worksheet.write(i, 11, label = str(0))

workbook.save('Excel_Workbook.xls')

import xlrd
import xlwt
ExcelFile=xlrd.open_workbook(r'C:\Users\tianzhy\Desktop\sumai.xlsx')
sheet=ExcelFile.sheet_by_name('速卖')
workbook = xlwt.Workbook(encoding = 'ascii')
worksheet = workbook.add_sheet('My Worksheet')
for i in range (1,2291):
    vari=0
    temp = sheet.cell(i,6).value
    if 'Dual Camera' in temp:
        vari=1
    if 'Dual camera' in temp:
        vari=1
    if 'Dual Front Camera' in temp:
        vari=1
    if 'Dual Back Camera' in temp:
        vari=1
    if 'Dual front Camera' in temp:
        vari=1
    if 'Dual Rear Camera' in temp:
        vari=1
    if 'Dual rear Camera' in temp:
        vari=1
    if 'Dual back Camera' in temp:
```

```
            vari=1
        worksheet.write(i, 0, label = vari)
        vari=0
        if 'Front Camera' in temp:
            vari=1
        if 'front Camera' in temp:
            vari=1
        worksheet.write(i, 1, label = vari)
workbook.save('Excel_Workbook.xls')

import xlrd
import xlwt
ExcelFile=xlrd.open_workbook(r'C:\Users\tianzhy\Desktop\sumai.xlsx')
sheet=ExcelFile.sheet_by_name('速卖')
workbook = xlwt.Workbook(encoding = 'ascii')
worksheet = workbook.add_sheet('My Worksheet')
for i in range (1,2291):
    temp = sheet.cell(i,5).value
    temp = str.split(temp)
    length=len(temp)
    count=0
    for j in range (0,length):
        if temp[j]=='Xiaomi'or temp[j]=='xiaomi'or temp[j]=='XIAOMI':
            count=1
        elif temp[j]=='Huawei'or temp[j]=='HUAWEI'or temp[j]=='huawei':
            count=2
        elif temp[j]=='MEIZU'or temp[j]=='meizu'or temp[j]=='Meizu':
            count=3
        elif temp[j]=='LENOVO'or temp[j]=='Lenovo'or temp[j]=='lenovo':
            count=4
        elif temp[j]=='IPHONE'or temp[j]=='iphone'or temp[j]=='iPhone':
            count=5
        elif temp[j]=='OPPO'or temp[j]=='Oppo'or temp[j]=='oppo':
            count=6
        elif temp[j]=='Vivo'or temp[j]=='vivo'or temp[j]=='VIVO':
            count=7
        elif temp[j]=='Nubia'or temp[j]=='NUBIA'or temp[j]=='nubia':
            count=8
        elif temp[j]=='samsung'or temp[j]=='Samsung'or temp[j]=='SAMSUNG':
            count=9
        elif temp[j]=='ZTE'or temp[j]=='zte':
            count=10
        elif temp[j]=='HOMTOM'or temp[j]=='homtom'or temp[j]=='Homtom':
            count=11
```

```
            elif temp[j]=='DOOGEE'or temp[j]=='Doogee'or temp[j]=='doogee':
                count=12
            elif temp[j]=='Letv'or temp[j]=='LeTv'or temp[j]=='letv' or temp[j]=='LETV':
                count=13
            elif temp[j]=='Blackview'or temp[j]=='BLACKVIEW'or temp[j]=='blackview':
                count=14
            elif temp[j]=='NOKIA'or temp[j]=='Nokia'or temp[j]=='nokia':
                count=15
        worksheet.write(i, 0, label = str(count))
workbook.save('Excel_Workbook.xls')
```

12.2 MATLAB Code

```
%clear B
jitiaoshuju=size(A);
jitiaoshuju=jitiaoshuju(1,1);
R=A(:,1);
xingbie=max(R);
R=A(:,2);
jigecanhe=max(R);
%AÊÇÔ-Ê¼ÊýÝ
for b=1:xingbie%ÐÔ±ð
B{b}=[];
end
for c=1:jitiaoshuju%¼¸ÌõÊýÝ
for d=1:xingbie%ÐÔ±ð
if A(c,1)==d
B{d}=[B{d};A(c,:)];
end
end
end
for e=1:xingbie%ÐÔ±ð
for f=1:jigecanhe%¼¸¸ö²Í°Ð
T=B{e}(:,2);
Q=find(T(T==f));
U(e,f)=max(Q);
end
end
%UÊÇÔ-Ê¼¸öÊý£¬±ÈÈçµÚ¶þÐÐµÚÒ»ÁÐ¾ÍÊÇÅ®µÄ²Í°ÐÑ¡Ïî¹1µÄÓÐ¶àÉÙ¸öÈË
for l=1:xingbie%ÓÐ¼¸¸öÐÔ±ð
for j=1:jigecanhe%ÓÐ¼¸¸ö²Í°Ð
for k=1:jigecanhe%ÓÐ¼¸¸ö²Í°Ð
C{l}(j,k)=U(l,j)/U(l,k);
end
```

```
end
end
for i=1:xingbie%ÓÐ¼¸¸öÐÔ±ð
t=C{i};
[x,lumda]=eig(t);
r=abs(sum(lumda));
n=find(r==max(r));
max_lumda_A(1,i)=lumda(n,n);
max_x_A{i}=x(:,n); %ÌØÕ÷Öµ
max_x_A{i}=max_x_A{i}./sum(max_x_A{i});
end
for p=1:xingbie%ÓÐ¼¸¸öÐÔ±ð
for q=1:jigecanhe%ÓÐ¼¸¸ö²Í°Ð
max_x_AB(p,q)=max_x_A{p}(q,1);%ÌØÕ÷ÏòÁ¿£¬µÚÒ»ÐÐÊÇÐµÄµÄ¸÷ÏîÈ¨ÖØ£¬µÚ¶þ
ÐÐÊÇÅ®µÄµÄ¸÷ÏîÈ¨ÖØ
end
end

for i =1:1324
if A(i)==2
       A(i)=1;
elseif A(i)==4
       A(i)=1;
elseif A(i)==8
       A(i)=2;
elseif A(i)==16
       A(i)=2;
elseif A(i)==32
       A(i)=2;
elseif A(i)==64
       A(i)=3;
elseif A(i)==128
       A(i)=3;
elseif A(i)==256
       A(i)=3;
%%else
      %% A(i)=5;
end
end

p=p';
t=t';
net=newff(minmax(p),[10 1]);
net.trainParam.epochs=1000;
```

```
net.trainParam.goal=0.001;
net.trainParam.show=50;
net.trainParam.lr=0.05;
net.trainParam.mc=0.9;
net=train(net,p,t);
A=sim(net,test);
A=A';

[m,n]=size(A);
[p,q]=size(Z);
for j =1:n
    B{j}=A(:,j);
    B{j}=B{j}/std(B{j});

for i=1:(m-1)
       B_{j}(i)=B{j}(i)-B{j}(i+1);
end
end
for j =1:q
    Y{j}=Z(:,j);
    Y{j}=Y{j}/std(Y{j});

for i=1:(p-1)
       Y_{j}(i)=Y{j}(i)-Y{j}(i+1);
end
end
for i=1:q
    aver(i,1)=mean(Y_{i});
for j = 1:n
       aver(i,(j+1))=abs(sum(Y_{i})-sum(B_{j}));
       final(i,j)=(1+aver(i,1))/(1+aver(i,1)+aver(i,(j+1)));
end
end
%averv0=mean(x0_);
%averv1=averv1/1323;
%final=(1+averv0)/(1+averv0+averv);

B=0:1:15;
B=B';
B_=zeros(16,1);
B=[B B_];
for i=1:2290
    B((A(i,1)+1),2)=B((A(i,1)+1),2)+1;
end
```

```
count=0;
for i = 1:16
if B(i,2)~=0
       count=count+1;
       C(count,1)=B(i,1);
       C(count,2)=B(i,2);
end
end

%yangbenµÚÒ»ÁÐÊÇ·ÖÀàÇÃ½øÈ¥
%bÊÇ´ýÅÐµÄÇÃ½øÈ¥£¬gÇÃ½øÈ¥
%iiiÊÇ¸ÅÂÊ£¬½á¹û
%HÊÇ°óÑé¸ÅÂÊ£¬½á¹û
%g-group·ÖÀàÊý£¬°óÀ´Ð´ÁË¸ö×Ô¶¯¼ì²â·ÖÀàÊýµÄ£¬²»»¹ýÃ»ÔÚmatlabÏÂÐ©£¬°Ç°Ç
[m,n]=size(yangben);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
for i=1:g
groupNum(i)=0;
group(i)=0;
for j=1:m
if yangben(j,1)==i
group(i)=group(i)+1;
end
end
if i==1
groupNum(i)=group(i);
else
groupNum(i)=groupNum(i-1)+group(i);
end
end
group;
groupNum; %¼ÆËã·ÖÀà¸öÊýÊý×é
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%¼ÆËã×ÜÆ½¾ùÖµ
% for j=1:n-1
% TotalMean(j)=0;
% for i=1:m
% TotalMean(j)=TotalMean(j)+yangben(i,j+1);
% end
% TotalMean(j)=TotalMean(j)/m;
% end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%
```

```
GroupMean=[];
for i=1:g
if i==1
low=1;
up=groupNum(i);
else
low=groupNum(i-1)+1;
up=groupNum(i);
end
matrix=yangben(low:up,:);
MatrixMean=mean(matrix); %¸÷·ÖÀà×éÆ½¾ùÖµ
GroupMean=[GroupMean;MatrixMean];

for u=low:up
for v=2:n
C(u,v-1)=yangben(u,v)-MatrixMean(v);
end
end
end

C;
GroupMean;
V=C'*C/(m-g);
V_inv=inv(V); %¶Ô¾ØÕóVÇóÄæ
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%
GroupMean=GroupMean(:,2:n);
Q1=GroupMean*V_inv;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%
for i=1:g
lnqi(i)=log(group(i)/m);
mat=GroupMean(i,:);
Q2(i)=lnqi(i)-0.5*mat*V_inv*mat';
end
lnqi;
Q2;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%

[u,v]=size(b);
result=[];
for i=1:u
x=b(i,:);
```

```
yy=Q1*x'+Q2';
result=[result yy];
end
res=result'; %¼ÆËãµÄ´ýÅÐÊý¾Ý¶Ô¸÷±ê×¼Êý¾ÝµÄÏßÐÔ¼ÆËãÖµ

[rows,cols]=size(result);
for i=1:cols
iljj=0;
mlljj=result(:,i);
for j=1:rows
iljj=iljj+exp(result(j,i)-max(mlljj));
end
for j=1:rows
houyangailv(j,i)=exp(result(j,i)-max(mlljj))/iljj;
end
end
H=houyangailv'; %°óÑé¸ÅÂÊ

iii=[];
for a=1:u
k=max(H(a,:));
for ii=1:g
if k==H(a,ii)
iii=[iii;ii];
end
end
end

clear ccatagorydetectionijkmn
for i=1:7
    c{i}=[];
end
for i=1:442
    catagory=b(i,1);
for j =1:7
       [m,n]=size(c{j});
if n~=0
          detection=0;
for k =1:n

if c{j}(5,k)==a(i,j)
                c{j}(catagory,k)=c{j}(catagory,k)+1;
                detection=1;
end
```

```
end
if detection==0
                c{j}(5,(n+1))=a(i,j);
                c{j}(catagory,(n+1))=(c{j}(catagory,(n+1)))+1;
end

end
if n==0
           c{j}(5,1)=a(i,j);
           c{j}(catagory,1)=(c{j}(catagory,1))+1;
end
end
end
for i=1:7
    c{i}=c{i}';
end
```

10.3 Application Result

Bayes Click	Bayes Convert	Principal Click	Principal Convert	BP Click	BP Convert	Average Click	Average Convert
0.2625	0.225	0.301081	0.295534	0.047775	0.024434	0.203786	0.181656
0.05	0.05	0.140721	0.143245	0.021181	0.021401	0.070634	0.071549
0.2625	0.225	0.121122	0.11836	0.009838	-0.05456	0.131153	0.096266
0.05	0.225	0.099311	0.099699	0.015475	-0.045	0.054929	0.093231
0.05	0.05	1.166903	1.174594	0.013802	-0.07697	0.410235	0.382542
0.05	0.225	0.058086	0.055002	0.012924	-0.05778	0.040336	0.074074
0.05	0.225	0.248493	0.250987	0.015712	-0.04218	0.104735	0.144603
0.05	0.05	0.706112	0.71033	0.021142	0.023265	0.259085	0.261198
0.05	0.225	0.238837	0.239042	0.016657	-0.03991	0.101832	0.141379
0.05	0.05	0.284154	0.284908	0.021321	0.016654	0.118492	0.117188
0.05	0.05	0.153205	0.151119	0.023743	0.016128	0.075649	0.072416
0.05	0.05	0.838394	0.842787	0.013709	-0.06495	0.300701	0.275945
0.2625	0.225	0.050065	0.036009	0.016702	0.008986	0.109756	0.089998
0.05	0.05	0.03951	0.033058	0.020839	0.018061	0.036783	0.033706
0.05	0.05	0.427411	0.431128	0.019589	0.019448	0.165667	0.166859
0.05	0.225	0.061471	0.06089	0.015967	-0.04453	0.042479	0.080453
0.05	0.225	0.061778	0.059916	0.02975	-0.05035	0.047176	0.078188
0.05	0.05	0.106117	0.109004	0.029872	-0.03933	0.061996	0.03989
0.05	0.05	0.058695	0.056355	0.022316	0.018414	0.043671	0.041589
0.05	0.05	0.409615	0.413317	0.01951	0.021002	0.159708	0.16144
0.2625	0.225	0.054849	0.047701	0.019698	0.011567	0.112349	0.094756
0.05	0.225	0.45968	0.459352	0.018527	0.016732	0.176069	0.233695

0.05	0.05	0.387575	0.390328	0.019166	0.019707	0.152247	0.153345
0.05	0.225	0.044472	0.041542	0.015017	-0.05861	0.036496	0.06931
0.05	0.05	0.919941	0.925382	0.020089	0.020919	0.33001	0.3321
0.05	0.05	0.710686	0.713729	0.012175	-0.07251	0.25762	0.230405
0.05	0.225	0.029344	0.02205	0.01429	-0.0593	0.031211	0.062583
0.05	0.05	0.121712	0.121506	0.017301	0.013074	0.063004	0.061527
0.05	0.05	0.159957	0.15993	0.021647	0.014211	0.077201	0.074713
0.05	0.225	0.057494	0.057361	0.046611	-0.04982	0.051368	0.077515
0.2625	0.225	0.128441	0.129147	0.049592	0.020074	0.146844	0.12474
0.05	0.05	0.0211	0.018417	0.02409	0.011114	0.03173	0.02651
0.05	0.225	0.267234	0.267913	0.008311	-0.06043	0.108515	0.144161
0.05	0.05	0.11542	0.11341	0.022578	0.016628	0.062666	0.060012
0.05	0.05	0.08274	0.079886	0.019898	0.009123	0.050879	0.046336
0.05	0.05	0.035513	0.035484	0.017282	-0.04392	0.034265	0.013855
0.05	0.225	0.272611	0.272038	0.016109	-0.05984	0.112907	0.145733
0.05	0.05	0.070123	0.06746	0.022535	0.012018	0.047553	0.043159
0.05	0.225	0.056996	0.056187	0.016619	-0.0509	0.041205	0.076762
0.05	0.05	0.311108	0.313147	0.017928	-0.04026	0.126345	0.107629
0.05	0.05	0.046446	0.046241	0.049272	0.012992	0.048573	0.036411
0.05	0.05	0.096495	0.096192	0.019702	0.010709	0.055399	0.0523
0.05	0.225	0.023168	0.021322	0.016456	-0.05952	0.029875	0.062269
0.2625	0.225	0.023472	0.015893	0.01577	-0.05274	0.10058	0.062719
0.05	0.225	0.063036	0.06313	0.041782	-0.05208	0.051606	0.078683
0.05	0.05	0.130617	0.132127	0.020563	0.013422	0.06706	0.065183
0.05	0.05	0.056659	0.055431	0.019311	0.009797	0.04199	0.038409
0.05	0.05	0.153791	0.154382	0.021909	0.015791	0.075233	0.073391
0.05	0.225	0.050062	0.047132	0.014266	-0.05669	0.03811	0.071813
0.05	0.05	0.401322	0.396777	0.019591	0.020679	0.156971	0.155819
0.05	0.05	0.042529	0.039504	0.018259	0.006759	0.03693	0.032088
0.05	0.225	0.046048	0.042978	0.013266	-0.05885	0.036438	0.06971
0.05	0.05	0.072228	0.071846	0.019819	0.012318	0.047349	0.044721
0.2625	0.225	0.008559	0.004133	0.018996	0.019257	0.096685	0.082796
0.05	0.05	0.05594	0.055802	0.020961	0.014394	0.0423	0.040065
0.2625	0.225	0.071697	0.064676	0.02177	0.022783	0.118656	0.104153
0.05	0.225	0.032809	0.029794	0.015003	-0.05685	0.032604	0.065982
0.05	0.05	0.303993	0.303849	-0.00484	0.0218	0.116386	0.125216
0.05	0.225	0.02684	0.023795	0.022054	0.015454	0.032965	0.088083
0.2125	0.225	0.010919	0.014113	0.010625	0.014996	0.078015	0.084703
0.05	0.05	0.063761	0.058907	0.037174	0.012365	0.050312	0.040424
0.05	0.05	0.08451	0.084139	0.020973	0.013408	0.051828	0.049182
0.05	0.05	0.957088	0.966342	0.020351	0.026999	0.34248	0.347781
0.05	0.05	0.079281	0.079612	0.022356	0.014837	0.050545	0.04815
0.2125	0.225	0.007936	-0.00086	0.01403	0.015025	0.078156	0.079722

0.2625	0.225	0.065479	0.051174	0.011993	-0.05492	0.113324	0.073753
0.05	0.05	0.081789	0.082327	0.023207	0.012109	0.051666	0.048146
0.05	0.05	0.211924	0.213195	0.023184	-0.02773	0.095036	0.078487
0.05	0.225	0.037688	0.035776	0.021638	0.018533	0.036442	0.093103
0.05	0.225	0.046328	0.045606	0.02291	0.01285	0.039746	0.094485
0.05	0.225	0.046882	0.043287	0.014918	-0.05514	0.037266	0.07105
0.05	0.225	0.023604	0.020592	0.014003	-0.0574	0.029202	0.06273
0.2125	0.225	0.006932	0.003512	0.008363	0.005225	0.075932	0.077912
0.05	0.225	0.107906	0.101205	0.023197	0.016882	0.060368	0.114362
0.2625	0.349	0.046636	0.031028	0.02256	-0.04799	0.110565	0.110679
0.2625	0.225	0.159246	0.146043	0.021014	0.011773	0.147586	0.127605
0.05	0.225	0.017755	0.016308	0.017763	-0.05388	0.028506	0.062477
0.05	0.05	0.172193	0.173618	0.021146	0.016857	0.081113	0.080158
0.2125	0.225	0.006541	-0.00187	0.010914	0.015761	0.076652	0.079629
0.05	0.225	0.019879	0.016972	0.015368	-0.05581	0.028416	0.062055
0.05	0.05	0.012198	0.009536	0.034172	-0.00937	0.032123	0.016721
0.05	0.225	0.01658	0.013583	0.014871	-0.05724	0.02715	0.060449
0.05	0.225	0.043114	0.035716	0.016119	-0.05591	0.036411	0.068267
0.2125	0.225	0.010337	0.001918	0.013426	0.013825	0.078755	0.080247
0.2125	0.225	0.010963	0.01043	0.010482	0.007622	0.077981	0.081017
0.05	0.05	0.043176	0.042843	0.019313	0.009654	0.037496	0.034166
0.2125	0.225	0.007149	0.003611	0.008852	0.008215	0.076167	0.078942
0.05	0.225	0.023931	0.021039	0.01376	-0.06077	0.02923	0.061758
0.05	0.05	0.100603	0.10202	0.023325	0.015937	0.057976	0.055986
0.2625	0.349	0.011219	0.003559	0.015098	-0.0486	0.096272	0.10132
0.2125	0.225	0.003645	0.003773	0.001698	-0.03455	0.072614	0.06474
0.05	0.05	0.066361	0.066042	0.020988	0.013059	0.045783	0.043033
0.2125	0.225	0.005905	-0.00243	0.009273	0.00535	0.075892	0.075973
0.05	0.225	0.03728	0.037025	0.025724	-0.04938	0.037668	0.070883
0.2125	0.225	0.00517	0.003057	0.003035	-0.03324	0.073568	0.064938
0.2125	0.225	0.003833	-0.00425	0.001923	-0.03413	0.072752	0.062209
0.05	0.05	0.020376	0.020014	0.020601	-0.05363	0.030325	0.005462
0.2125	0.225	0.010629	0.009626	0.003222	0.008785	0.07545	0.081137
0.2125	0.225	0.005152	0.003957	0.002019	-0.03304	0.073224	0.065307
0.05	0.05	0.013796	0.01065	0.021336	0.016562	0.028377	0.025737
0.2125	0.225	0.007317	0.003736	0.009564	0.006794	0.07646	0.07851
0.05	0.05	0.285532	0.287084	0.020283	0.017855	0.118605	0.118313
0.05	0.05	0.029818	0.029556	0.022125	-0.03342	0.033981	0.01538
0.2125	0.225	0.006412	0.008398	0.012103	0.019711	0.077005	0.08437
0.05	0.225	0.056872	0.057289	0.041746	-0.05331	0.049539	0.076327
0.05	0.225	0.02168	0.013711	0.008839	-0.05388	0.02684	0.061609
0.05	0.225	0.27461	0.268199	0.019176	-0.0553	0.114595	0.145968
0.2125	0.225	0.008016	0.004362	0.005632	0.006494	0.075383	0.078619

0.2125	0.225	0.009297	0.008249	0.002735	0.007893	0.074844	0.08038
0.05	0.05	0.026959	0.026361	0.018757	-0.04092	0.031905	0.011814
0.05	0.05	0.028931	0.025484	0.018529	0.007957	0.032487	0.027813
0.05	0.225	0.048058	0.045066	0.022431	0.016809	0.040163	0.095625
0.2125	0.225	0.011357	0.010473	0.003062	0.008683	0.07564	0.081385
0.2125	0.225	0.008931	0.007993	0.001995	0.00784	0.074476	0.080277
0.05	0.225	0.020309	0.017448	0.013675	-0.0585	0.027995	0.061316
0.2125	0.225	0.007707	0.003987	0.033202	0.011352	0.08447	0.080113
0.05	0.225	0.008246	0.007104	0.018907	-0.05212	0.025718	0.059995
0.05	0.05	0.021423	0.021151	0.023556	-0.03263	0.03166	0.01284
0.2625	0.225	0.029511	0.021986	0.020884	0.012993	0.104298	0.08666
0.2125	0.225	0.005237	0.001894	0.000943	-0.03714	0.072893	0.06325
0.05	0.225	0.041358	0.040852	0.01934	-0.0451	0.036899	0.073584
0.05	0.05	0.029407	0.028856	0.02031	0.012507	0.033239	0.030454
0.05	0.225	0.023048	0.02133	0.016487	-0.05954	0.029845	0.062264
0.2125	0.225	0.007227	-0.00093	0.009995	0.006268	0.076574	0.076779
0.382	0.349	0.000474	-0.00617	0.029938	-0.03327	0.137471	0.103187
0.2125	0.225	0.003404	-0.00492	0.001145	-0.03817	0.07235	0.060637
0.2125	0.225	0.005929	-0.00214	0.002944	-0.0385	0.073791	0.061452
0.2125	0.225	0.007107	0.006384	-0.00476	-0.02763	0.071616	0.067918
0.2125	0.225	0.007211	0.006589	0.001433	-0.03111	0.073715	0.066825
0.2125	0.225	0.008844	0.008095	-0.00581	-0.02622	0.071844	0.068958
0.2125	0.225	0.008742	0.007927	-0.00664	-0.0261	0.071535	0.068941
0.2125	0.225	0.004588	-0.00359	0.002976	-0.03823	0.073355	0.06106
0.2125	0.225	0.009737	0.008954	-0.00538	-0.03678	0.072285	0.065725
0.2125	0.225	0.005842	0.002583	0.00356	-0.03568	0.073967	0.063968
0.2125	0.225	0.009736	0.008953	-0.00537	-0.03649	0.072289	0.06582
0.2125	0.225	0.006041	-0.00238	0.00819	0.005465	0.075577	0.07603
0.05	0.225	0.051646	0.049489	0.021027	0.012499	0.040891	0.095663
0.2125	0.225	0.010266	0.00932	-0.00707	-0.03118	0.071899	0.067712
0.2125	0.225	0.003554	-0.00456	0.003841	-0.03775	0.073298	0.060896
0.2125	0.225	0.005441	0.00203	0.00256	-0.03863	0.0735	0.0628
0.2125	0.225	0.007262	0.005939	-0.00623	-0.0318	0.071177	0.066379
0.05	0.05	0.020607	0.020204	0.012958	0.010416	0.027855	0.026873
0.05	0.05	0.022177	0.020814	0.018936	0.00952	0.030371	0.026778
0.2125	0.225	0.009875	0.001768	0.009749	0.005862	0.077375	0.077543
0.05	0.05	0.114214	0.115638	0.023639	0.015651	0.062618	0.06043
0.2125	0.225	0.01193	0.010972	0.002416	0.006727	0.075615	0.0809
0.2125	0.225	0.009059	0.007769	-0.00621	-0.06008	0.071783	0.057561
0.05	0.05	0.069129	0.068707	0.019155	0.005347	0.046095	0.041351
0.05	0.05	0.212041	0.213253	0.020847	0.017145	0.094296	0.093466
0.2125	0.225	0.00864	0.00916	0.008708	0.019435	0.076616	0.084532
0.2125	0.225	0.006305	-0.00132	0.006431	-0.03179	0.075079	0.063963

0.2625	0.225	0.015216	0.011977	0.018604	0.020035	0.098773	0.085671
0.05	0.05	0.022746	0.021435	0.021955	0.010881	0.031567	0.027439
0.2625	0.225	0.004393	-0.00268	0.017402	0.016233	0.094765	0.079518
0.2125	0.225	0.008666	0.000453	0.008176	0.005651	0.076447	0.077035
0.05	0.225	0.014505	0.006801	0.014164	-0.05842	0.026223	0.057795
0.05	0.225	0.029027	0.02591	0.014523	-0.05599	0.031183	0.064974
0.2125	0.225	0.007448	0.003982	0.00731	0.006281	0.075753	0.078421
0.2625	0.225	0.018527	0.017602	0.020393	0.020422	0.100473	0.087675
0.05	0.05	0.042026	0.039118	0.017737	0.007654	0.036587	0.032257
0.05	0.225	0.331541	0.333914	0.014941	-0.06625	0.132161	0.164223
0.2125	0.225	0.005455	-0.00266	0.00119	-0.03827	0.073048	0.061357
0.2125	0.225	0.012391	0.011502	0.010113	0.008919	0.078335	0.081807
0.05	0.225	0.014426	0.00058	0.012753	-0.05659	0.025726	0.056331
0.05	0.225	0.011038	0.007793	0.015099	-0.05155	0.025379	0.060414
0.2125	0.225	0.007116	0.006546	0.004305	-0.02703	0.07464	0.068172
0.2125	0.225	0.006364	-0.00174	0.000389	-0.03959	0.073085	0.061223
0.05	0.05	0.030969	0.03046	0.016133	-0.09231	0.032368	-0.00395
0.05	0.05	0.041989	0.041225	0.024206	0.015106	0.038732	0.035443
0.2125	0.225	0.006068	0.002641	0.001369	-0.04098	0.073312	0.06222
0.2125	0.225	0.010283	0.008887	0.002416	0.006727	0.075066	0.080204
0.05	0.225	0.014355	0.013712	0.016826	-0.04724	0.02706	0.063822
0.05	0.225	0.009614	0.006398	0.013621	-0.05867	0.024412	0.057578
0.05	0.05	0.011085	0.010683	0.011712	-0.04207	0.024266	0.006205
0.05	0.05	0.024787	0.024539	0.019065	0.01	0.031284	0.02818
0.2625	0.225	0.01189	0.004122	0.018492	0.018244	0.097627	0.082456
0.2125	0.225	0.01139	0.010615	0.00267	0.008315	0.07552	0.08131
0.2625	0.225	0.01581	0.008043	0.021035	0.014318	0.099782	0.082454
0.05	0.05	0.027528	0.026768	0.018935	0.01077	0.032154	0.029179
0.05	0.05	0.146195	0.147953	0.02167	0.014909	0.072622	0.070954
0.05	0.225	0.012646	0.009517	0.01619	-0.05372	0.026279	0.060267
0.05	0.225	0.013892	0.0128	0.016021	-0.05959	0.026638	0.059404
0.2125	0.225	0.007266	0.006005	0.002201	-0.03094	0.073989	0.066689
0.05	0.05	0.021308	0.020514	0.01933	0.008859	0.030213	0.026458
0.05	0.225	0.189112	0.192334	0.015699	-0.06461	0.084937	0.117574
0.05	0.05	0.142136	0.142049	0.015453	-0.04105	0.069196	0.050334
0.05	0.05	0.148518	0.148921	0.021764	0.014453	0.073427	0.071125
0.05	0.05	0.276366	0.264585	0.052117	0.023249	0.126161	0.112611
0.2625	0.225	0.010303	0.002507	0.009749	-0.05589	0.094184	0.057206
0.2125	0.225	0.006215	-0.00817	-0.00032	-0.03466	0.072799	0.060724
0.2125	0.225	0.007926	########	0.012253	0.014051	0.07756	0.079663
0.2125	0.225	0.009719	0.009138	0.000374	-0.08663	0.074198	0.049171
0.2125	0.225	0.010238	0.009419	0.000391	-0.02953	0.074376	0.068297
0.2125	0.225	0.006739	0.005964	-0.00528	-0.02343	0.071321	0.069177

0.05	0.05	0.022889	0.022175	0.021974	0.011524	0.031621	0.0279
0.05	0.05	0.021889	0.020806	0.018825	0.009917	0.030238	0.026908
0.05	0.05	0.102282	0.101981	0.019771	0.011515	0.057351	0.054499
0.2125	0.225	0.008776	0.007424	0.001543	-0.02451	0.074273	0.069306
0.2125	0.225	0.012034	0.010843	0.006964	0.011446	0.077166	0.08243
0.2125	0.225	0.012645	0.011759	0.032645	0.014862	0.08593	0.083873
0.05	0.225	0.01525	0.013787	0.020646	0.011275	0.028632	0.083354
0.2125	0.225	0.006153	0.00279	0.003873	-0.04123	0.074175	0.062186
0.2125	0.225	0.005238	-0.00272	0.001981	-0.03653	0.07324	0.061917
0.2125	0.225	0.009704	0.001139	0.006595	0.004754	0.076266	0.076964
0.2125	0.225	0.007505	0.004021	0.011899	0.011975	0.077301	0.080332
0.2125	0.225	0.004383	0.002958	0.00147	-0.03563	0.072784	0.064108
0.05	0.05	0.016457	0.013194	0.016896	0.019142	0.027784	0.027445
0.05	0.225	0.06849	0.061434	0.021317	0.013423	0.046602	0.099952
0.2125	0.225	0.007851	-0.00034	0.008375	0.005534	0.076242	0.076732
0.05	0.225	0.019538	0.016346	0.014395	-0.0515	0.027978	0.063282
0.2125	0.225	0.008866	0.008007	0.002504	0.007833	0.074623	0.08028
0.2125	0.225	0.008521	0.000354	0.008581	0.005284	0.076534	0.07688
0.2125	0.225	0.00776	0.004291	0.008908	0.010604	0.076389	0.079965
0.2125	0.225	0.008879	0.005321	0.009061	0.007696	0.076813	0.079339
0.05	0.225	0.01688	0.013538	0.013647	-0.05888	0.026843	0.059885
0.2125	0.225	0.004865	0.001278	0.001244	-0.0387	0.07287	0.062524
0.05	0.05	0.073389	0.073277	0.02152	0.013877	0.048303	0.045718
0.2625	0.225	0.069466	0.061968	0.020799	0.019072	0.117588	0.102013
0.2125	0.225	0.005393	-0.00247	0.004086	-0.04003	0.073993	0.060834
0.2125	0.225	0.01215	0.011037	0.006082	0.008364	0.076911	0.081467
0.05	0.225	0.02346	0.019896	0.015425	-0.05364	0.029629	0.063752
0.05	0.05	0.01633	0.013126	0.018571	0.009475	0.0283	0.0242
0.05	0.225	0.021668	0.020445	0.017369	0.004842	0.029679	0.083429
0.2625	0.225	0.008397	0.005027	0.01419	0.01422	0.095029	0.081416
0.05	0.225	0.022944	0.015037	0.013643	-0.0581	0.028862	0.060645
0.05	0.225	0.018746	0.019121	0.013199	-0.05789	0.027315	0.062076
0.05	0.05	0.042536	0.041844	0.01938	0.011102	0.037305	0.034316
0.2125	0.225	0.007773	0.007257	0.000346	-0.01889	0.07354	0.071122
0.2125	0.225	0.012191	0.011324	0.005658	0.012083	0.076783	0.082802
0.05	0.225	0.045833	0.042801	0.022026	0.016382	0.039287	0.094728
0.05	0.225	0.014539	0.013379	0.03326	-0.05816	0.032599	0.060072
0.05	0.05	1.842754	1.851109	0.02723	-0.01368	0.639994	0.629142
0.2625	0.225	0.01866	0.010826	0.010803	-0.05841	0.097321	0.05914
0.2125	0.225	0.005478	0.002011	-0.00463	-0.03717	0.071115	0.063279
0.2125	0.225	0.005205	0.001767	0.000531	-0.03617	0.072745	0.063532
0.2125	0.225	0.008186	3.88E-07	0.01224	0.013073	0.077642	0.079358
0.05	0.225	0.062721	0.060188	0.021876	0.015631	0.044866	0.100273

0.2125	0.225	0.006413	0.002731	0.000125	-0.00961	0.073013	0.072705
0.05	0.05	0.023908	0.023306	0.018539	0.004969	0.030816	0.026092
0.2125	0.225	0.012638	0.011516	0.007637	0.014536	0.077591	0.083684
0.05	0.225	0.052156	0.050381	0.020638	0.008122	0.040931	0.094501
0.2125	0.225	0.008049	-0.00015	0.008539	0.005569	0.076363	0.076807
0.2125	0.225	0.010283	0.008887	0.002416	0.006727	0.075066	0.080204
0.05	0.05	0.033099	0.033057	0.017914	-0.04353	0.033671	0.013174
0.05	0.225	0.012346	-0.00199	0.013945	-0.05816	0.02543	0.054948
0.05	0.225	0.021146	0.020853	0.04742	0.017029	0.039522	0.087627
0.05	0.225	0.02212	0.018965	0.01802	0.008293	0.030046	0.084086
0.05	0.05	0.125575	0.122975	0.02215	0.010401	0.065908	0.061126
0.2625	0.225	0.049448	0.036176	0.020347	0.021365	0.110765	0.094181
0.05	0.05	0.356942	0.359282	0.023274	0.020159	0.143405	0.143147
0.05	0.05	0.168361	0.169295	0.021903	0.015031	0.080088	0.078109
0.05	0.05	0.151001	0.152199	0.02252	0.015042	0.074507	0.072413
0.05	0.05	0.103999	0.105048	0.020955	0.012843	0.058318	0.055964
0.05	0.05	0.844454	0.844337	0.025795	0.000313	0.30675	0.298217
0.2125	0.225	0.005866	-0.00886	0.011111	0.016296	0.076492	0.077478
0.05	0.05	0.10245	0.102947	0.020947	0.010108	0.057799	0.054352
0.05	0.225	0.059872	0.052483	0.011717	-0.05989	0.04053	0.072531
0.05	0.05	0.204978	0.206755	0.023622	0.020508	0.092867	0.092421
0.05	0.05	0.053262	0.052931	0.020776	0.012347	0.041346	0.038426
0.05	0.225	0.025138	0.025226	0.022422	0.017912	0.03252	0.089379
0.05	0.225	0.026865	0.026295	0.017613	-0.04441	0.031493	0.068963
0.05	0.225	0.033728	0.032893	0.021447	-0.05059	0.035058	0.069101
0.05	0.05	0.03963	0.039072	0.022987	-0.06509	0.037539	0.007995
0.2125	0.225	0.005303	-0.00274	0.004478	-0.03539	0.074094	0.062291
0.2625	0.349	9.09616	9.118142	0.034491	-0.1052	3.13105	3.120648
0.2625	0.05	0.785604	0.816868	0.035319	-0.04666	0.361141	0.273403
0.05	0.05	0.671339	0.675279	0.019997	0.024817	0.247112	0.250032
0.2625	0.225	0.883821	0.880353	0.010231	-0.06141	0.385517	0.34798
0.2625	0.349	6.919327	6.945776	0.00938	-0.08818	2.397069	2.402198
0.2625	0.225	0.846089	0.842904	0.010162	-0.0601	0.372917	0.335933
0.382	0.349	7.16185	7.179173	0.012326	-0.09859	2.518725	2.476527
0.2625	0.225	0.945303	0.94349	0.009658	-0.06012	0.40582	0.369456
0.2625	0.05	0.373528	0.383608	0.009096	-0.04901	0.215041	0.1282
0.2625	0.225	0.063014	0.05562	0.014447	-0.05349	0.11332	0.075709
0.382	0.349	0.07983	0.06655	0.018882	-0.04924	0.160237	0.122102
0.2625	0.225	0.054785	0.04724	0.014298	-0.05379	0.110528	0.072816
0.05	0.05	0.460305	0.46433	0.021374	0.020999	0.177226	0.178443
0.382	0.349	0.041972	0.0348	0.029855	0.02318	0.151276	0.13566
0.382	0.349	11.4282	11.4552	0.0365	-0.11111	3.948901	3.897695
0.2625	0.225	3.048716	3.057682	0.005921	-0.09288	1.105713	1.063266

0.2625	0.349	12.524	12.5511	0.035871	-0.11862	4.274122	4.260496
0.2625	0.225	1.612981	1.610383	0.001919	-0.05607	0.6258	0.593106
0.05	0.05	0.294202	0.288847	0.046674	0.017109	0.130292	0.118652
0.2625	0.05	1.144992	1.144997	0.018258	0.025127	0.47525	0.406708
0.2625	0.225	0.127119	0.11991	0.015408	-0.05705	0.135009	0.095954
0.05	0.225	0.042112	0.039941	0.015921	-0.05273	0.036011	0.070738
0.05	0.225	0.116847	0.113293	0.022334	0.021449	0.06306	0.119914
0.382	0.349	7.549541	7.578432	0.010822	-0.08613	2.647454	2.613769
0.2625	0.05	0.583085	0.595301	0.00789	-0.05132	0.284492	0.197994
0.2625	0.225	0.10881	0.101731	0.008454	-0.05752	0.126588	0.089736
0.2625	0.225	0.101932	0.094958	0.020866	0.020915	0.128433	0.113624
0.2625	0.225	0.113682	0.106418	0.021067	0.013923	0.132416	0.115114
0.2625	0.225	0.609268	0.603864	0.010462	-0.05297	0.294077	0.258632
0.05	0.05	0.784498	0.792061	0.013634	-0.05421	0.282711	0.262616
0.05	0.05	5.877811	5.903869	0.010725	-0.04756	1.979512	1.968771
0.2625	0.225	1.346343	1.347062	0.007878	-0.06269	0.538907	0.503125
0.2625	0.349	3.003834	3.015559	0.006	-0.07736	1.090778	1.095734
0.2625	0.225	0.081122	0.067548	0.020647	0.022424	0.121423	0.10499