# Travel Plan Design Based on Graph Theory and Multiple Linear Regression Equation Model

Yifeng Dou<sup>1, a</sup>, Hailing Xiong<sup>2, b, \*</sup>

<sup>1</sup>College of Computer and Information Science, Southwest University, Chongqing, 400715, China
 <sup>2</sup>College of Computer and Information Science, Southwest University, Chongqing, 400715, China
 <sup>a</sup>email: 1204840733@qq.com, <sup>b</sup>email: xionghl@swu.edu.cn, \*Corresponding author

**Keywords:** Travel Plan Design; Graph Theory; Principal Component Analysis; Goal Programming; Multiple Linear Regression Equation

**Abstract.** Different families have different needs for tours. In order to solve the problem, this paper firstly creates a table which takes all determining varieties into consideration. By adopting principal component for analyzing and filtrating the main factors, the conclusion is as follow: the first principal component is time and the second one is cost; their factor contribution rate is respectively 58.95% and 38.93%. Moreover, the two weighted principal components can make the analysis process easier and reduce error rate. Secondly, in order to provide the best tour plan for different families, this paper selects eight tourist attractions in Xi'an and establishes the goal programming model which covers some factors, such as tour routes, costs and time. Finally, a multivariate linear regression equation is applied to test traveling plans and determine the best one.

# Introduction

With the development of economy and the progress of material civilization, people want to have more time to improve their living standards, but most will choose to travel, so tourism has gradually become a kind of lifestyle. With the summer vacation approaching, many parents will take children to the city tourism, but different families have different needs, such as number, cost limit, time limit and so on. For most families, they require to visit most tourist attractions with the least money in the shortest time. Thus, we choose a tourist city firstly, and consider the travel route, cost, time and other important factors, and design a best travel plan for the different needs of the families.

Different families have different needs. If we want to know what factors have the greatest influence on their choice, we'll need statistics. But too many factors will get the analysis of trip mode into trouble, so it is necessary to identify the most important factor in order to further solve the problem. Refer to the past data table to filter out the appropriate data and create the table. Using principal component analysis method [1] to get a comprehensive index, and then run the main draw of the first component and the second principal component in SAS [2] by data. Besides, considering selected eight spots, whether team tour or self-help tour is chosen, and the tour route is an important part to connect the tourist source places and destinations. Designed tours conducive to select the destination of the tourists have to arrange their travel activities, to avoid "roaming", and it helps to show various tourist spots and help tourists use time rationally, there are also plans to help tourists spend less. Then we use graph theory [3] to tourist attractions as vertices in the graph, traffic routes can be seen as the various attractions on the map edge, the edge weights on behalf of other attractions from a distance to attractions, and finally the use of the shortest path problem solve the problem of routes' design after setting tour spots, using Matlab [4] to solve time and cost required of each plan. At last, through above model established four tour plans, tourists can choose according to their own situation to fit their circumstances plan, but this plan is for the best, and how to build a plan described in this article is the best plan. Multiple linear regression model [5] in this paper carried out a score draw plans created plan, which is the best plan.

## The preparation of the model

(1) Factors selection based on principal component analysis

Different families have different circumstances. Some people need to consider time, and some do cost. Some people travel for shopping and some for sightseeing. Some will consider the child's preferences. Create a table by each of these reasons statistical factors affecting the choice of approach.

Index	Time	Cost	Distance	Family income	Shopping	Tour
	(days)	(RMB)	(km)	(RMB)	(%)	(%)
Variable	X1	X2	X3	X4	X5	X6
Car1	3	7000	400	10000	60	30
Group2	5	5000	1000	6000	20	40
Visit3	4	3000	600	4000	10	30
Other 4	2	1000	100	3000	10	10

Tab.1 Different ways and variable values

There are car, groups, exercise, four other way to travel, time, cost, travel, family income, shopping, sightseeing six indicators, denoted, design principal component relationships:

$$Z = a_1 x_1 + \dots + a_6 x_6 = a X$$

(1)

Find the a(a a = 1), the maximum the Var(Z), Find the characteristic roots and the corresponding unit eigenvectors and cumulative contribution rate. In SAS run the program receive: Tab.2 Eigenvalue of the correlation matrix

Factors	Eigenvalue	Difference	Proportion	Cumulative
1	3.5371	1.2011	0.5895	0.5895
2	2.3360	2.2091	0.3893	0.9789
3	0.1269	0.1269	0.0211	1.0000

As can be seen from Tab.2 the contribution of the first principal component was 58.95%, the cumulative contribution rate before the two principal components reached 97.88%, so just using two principal components can be a good reflection of their needs, so reducing the hassle of has to consider so many factors, but to meet their demands. Another feature of the fourth fifth and sixth value is 0, one can conclude that six indicator variables ( $x_i$ , i = 1,2,3,4,5,6) have a linear relationship, characterized by the maximum of two months the eigenvector corresponding to the value of the first and second derived main components.

Tab.3 Eigenvectors						
	$Z_1$	$Z_2$	$Z_3$	$Z_4$	$Z_5$	$Z_6$
<i>x</i> <sub>1</sub>	0.4016	-0.4265	0.1926	0.1102	0.7795	0.0000
<i>x</i> <sub>2</sub>	0.4788	0.2836	0.1014	-0.8247	0.0000	0.0000
<i>x</i> <sub>3</sub>	0.4128	-0.2942	0.5196	0.1680	-0.5805	0.1977
<i>x</i> <sub>4</sub>	0.3939	0.4387	0.1074	0.3928	-0.0450	-0.6960
<i>x</i> <sub>5</sub>	0.2904	0.5479	-0.0611	0.3495	0.1159	0.6900
<i>x</i> <sub>6</sub>	0.4468	-0.2994	-0.8169	0.0560	-0.2000	-0.1742

Tab.3 shows that the first main component and second main component of the relationship were written:

 $Z_1 = 0.4016x_1 + 0.4788x_2 + 0.4128x_3 + 0.3939x_4 + 0.2904x_5 + 0.4468x_6$ (2)

 $Z_2 = -0.4265x_1 + 0.2836x_2 - 0.2942x_3 + 0.4387x_4 + 0.5479x_5 - 0.2994x_6$ (3)

(2) Analysis of the main factors

Value of each component of the feature vectors of the principal components can be interpreted,

the individual characteristic points corresponding to the largest values of the first feature vectors are positive, it is the number of days the reaction and travel distance of travel, so called first a main component of the time factor. The second largest eigenvalue of the eigenvector corresponding to the first component (time factor), and the third component (stroke coefficients) and sixth component (coefficient of view) is negative, the other positive, it reflect costs and monthly family income situation said second main component of the cost, so the cost and time is the most important factor, then build models from the following discussion and maybe a factor.

## The establishment and solution of the model

(1) Route optimization design of unlimited travel cost and the shortest travel time

Under the conditions of travel costs are not limited to, the requirements of the eight spots are tour again, with a minimum of time. Tourists spent time consisting of two parts, part time on the road, the other part is the time spent in the sights. So T2 is defined as the time visitors spent on the road, T2 attractions for tourists to stay in time, to get the overall objective function is:

$$MinT = T_1 + T_2 \tag{4}$$

Requires visitors to tour again eight attractions, so visitors spend on the road

$$T_1 = \sum_{i=0}^{8} \sum_{j=0}^{8} t_{ij} r_{ij}$$
(5)

(6)

The amount of time visitors stay for eight spots is

$$T_2 = 36$$

Therefore, the ultimate objective function is:

$$MinT = T_1 + T_2 = \sum_{i=0}^{8} \sum_{j=0}^{8} t_{ij} r_{ij} + 36$$
(7)

Because starting from Beilin, then travelling all eight tourist attractions, so

$$\sum_{i=0}^{8} \sum_{j=0}^{8} r_{ij} = 9$$
(8)

To make the time spent exploring the shortest tour route should be surrounded by a closed circle, and each attraction can only go from one side, the other side out, so the constraint conditions are as followed:

$$\sum_{i=0}^{8} r_{ik} = \sum_{j=0}^{8} r_{kj} \le 1(k = 1, 2, ..., 8) \quad and \quad r_{ij} * r_{ji} = 0$$
(9)

When k = 0, the departure from Beilin, so

$$\sum r_{0j} = 1 \tag{10}$$

Similarly:  

$$\sum r_{i0} = 1$$
(11)

According to the objective function and constraints, model building:

$$MinT = T_1 + T_2 = \sum_{i=0}^{8} \sum_{j=0}^{8} t_{ij} r_{ij} + 36$$
(12)

Constraint conditions:

$$\begin{cases} \sum_{j} r_{0j} = 1, \sum_{i} r_{i0} = 1 \\ \sum_{i=0}^{8} \sum_{j=0}^{8} r_{ij} = 9 \\ \sum_{i=0}^{8} r_{ik} = \sum_{j=0}^{8} r_{kj} \le 1 (k = 1, 2, L, 8) \end{cases}$$
(13)

Use of graph theory, can be drawn from the matrix of time. Then using Matlab to solve the

shortest time spending on travelling is 76 hours, then the best route to travel successively Beilin, Yang tomb, Qianling, mausoleum, Zhouzhi, Famen Temple, Museum, Big Wild Goose Pagoda, Cuihuashan.

(2) Route optimization design of the least cost and unlimited tourism time

The whole cost of tourism includes fees of entrance tickets, transportation, eating, and accommodation. The fees are certain, so all design of tourist route must make tourists tour eight spots. And total fees of transportation, accommodation, eating and else must be the least. We express accommodation fees in  $m_1$ , transportation fees in  $m_2$ , eating fees and else in  $m_3$ , fees of entrance tickets in  $m_4$ . So the objective function is:

$$MinM = m_1 + m_2 + m_3 + m_4 \tag{14}$$

We express transportation fees in  $m_2$ , and  $r_{ij}$  is used for deciding 0-1 variables from the i scenic spot to the j scenic spot. Then we express transportation fees from the i scenic spot to the j scenic spot in  $c_{ij}$ . So fees of each scenic spot are:

$$m_2 = \sum_{i=0}^8 \sum_{j=0}^8 r_{ij} c_{ij}$$
(15)

We express fees of entrance tickets in  $m_4$ , and tourists tour 8 scenic spots. So total fees of entrance tickets are:

$$m_{A} = 100 + 122 + 30 + 45 + 70 + 50 + 50 + 81 = 548 \tag{16}$$

Tourists travel ten hours during the daytime, and then need to find hotels. The accommodation fees are average 100 RMB every day. We express total time of tourism in T. Owing to average eating fees every day are 100 RMB, and T is the total time of tourism. So total fees of eating and else (for example: shopping). So total fees of accommodation are:

$$m_1 = m_3 = 100[\frac{T}{10}] = 10T \tag{17}$$

In conclusion, the last objective function is:

$$MinM = m_1 + m_2 + m_3 + m_4 = 20T + \sum_{i=0}^{8} \sum_{j=0}^{8} r_{ij}c_{ij} + 548$$
(18)

Constraint conditions still apply here in IV(A). And constraint condition of total tourism time T is:

$$T = \sum_{i=0}^{8} \sum_{j=0}^{8} t_{ij} r_{ij} + 36$$
(19)

According to the objective function and constraints, we set up model system and the objective function is

$$MinM = m_1 + m_2 + m_3 + m_4 = 20T + \sum_{i=0}^{8} \sum_{j=0}^{8} r_{ij}c_{ij} + 548$$
(20)

On the basis of (13), with the addition of (19) made up the constraint conditions.

Using the method of graph theory and we use Matlab and work out a minimum cost is 1872 RMB. The best tourism route at this time is: The Forest of Steles $\rightarrow$ Yang's Tomb $\rightarrow$ Emperor Qinshihuang $\rightarrow$  QianLing $\rightarrow$ Famen Temple $\rightarrow$ History Museum $\rightarrow$ The Wild Goose Pagoda $\rightarrow$ Cuihua Mountain $\rightarrow$ Building of Great Wild Goose Pagoda.

(3) Travel time is limited, as many attractions travel route optimization design

Because of limited time the problem is given 3 days in the model. Then the total travel time in this 3 days (including the road to spend time and time travel costs) for 30 hours, within the 30 hours as many scenic spots, so reasonable arrangement time, reasonable tourist route is essential. So the objective function of this problem only needs to be improved for the (2) model, other constraints conditions are the same. Then a constraint based on travel time is not more than three days, i.e.

$$T = \sum_{i=0}^{8} \sum_{j=0}^{8} r_{ij} t_{ij} + \frac{1}{2} \times \sum_{i=0}^{8} \sum_{j=0}^{8} r_{ij} \times (t_i + t_j) \le 30$$
(21)

We change the objective function to MaxN. On the basis of (13) and (19), with the addition of (21) made up the constraint conditions. Limited travel time is three days, the expression (13), (19) and (21) can be computed in tourist attractions number is five, the road to spend time of 27 hours, the shortest path between tourism route: The Forest of Steles, Emperor Qinshihuang, History Museum, The Wild Goose Pagoda, Cuihua Mountai, Famen temple.

(4) The cost of travel is limited, as many attractions tourism route optimization design

The problem for travel expenses is limited, in the model given the cost of 1000 RMB, then using this 1000 RMB should be as much as possible of the scenic spots, so reasonable arrangement time, reasonable arrangements for the tour route is crucial. So the objective function of this problem only needs to be improved for the (3) model, other constraints conditions are the same. Then a constraint based on tourism is the cost of not more than 3000 RMB, i.e.

$$M \le 1000$$
 (22)  
Specific expression is:

$$\sum_{i=0}^{8} \sum_{j=0}^{8} r_{ij} c_{ij} + \frac{1}{2} \times \sum_{i=0}^{8} \sum_{j=0}^{8} r_{ij} \times (d_i + d_j) + 20T \le 1000$$
(23)

here we change the objective function to MaxN. Constraint conditions:

$$\sum_{i}^{8} r_{i0} = 1, \sum_{j}^{6} r_{0j} = 1$$

$$N - \sum_{i=0}^{8} \sum_{j=0}^{8} x_{ij} = 0$$
(24)
$$\sum_{i=0}^{8} r_{ik} = \sum_{j=0}^{8} r_{kj} \le 1(k = 1, 2, L_{,8})$$

$$T = \sum_{i=0}^{8} \sum_{j=0}^{8} r_{ij}t_{ij} + \frac{1}{2} \times \sum_{i=0}^{8} \sum_{j=0}^{8} r_{ij} \times (t_{i} + t_{j})$$

$$\sum_{i=0}^{8} \sum_{j=0}^{8} r_{ij}c_{ij} + \frac{1}{2} \times \sum_{i=0}^{8} \sum_{j=0}^{8} r_{ij} \times (d_{i} + d_{j}) + 20T \le 1000$$

Qualified travel expenses for 1000, substitution (24) can be calculated in tourism scenic spots for five least cost 893.8 RMB, the travel route for the forest of steles is obtained by the short circuit: The Forest of Steles, Emperor Qinshihuang, History Museum, The Wild Goose Pagoda, Cuihua Mountain, Famen Temple.

#### Model test through multiple linear regression model

Multiple linear regression equation  $\xi = ax + by + cz + L$ . Verify whether the model is the model for the best travel plans, multiple linear regression analysis were as follows:

$$\begin{cases} y = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m + \varepsilon \\ \varepsilon \sim N(0, \sigma^2) \end{cases}$$
(25)

Where  $\beta_0, \beta_1, \dots, \beta_m, \sigma^2$  are unrelated to the unknown parameters  $x_1, x_2, \dots, x_m$ , which is called the regression coefficients. N independent observations  $(y_i, x_{i1}, \dots, x_{im})$  were obtained,  $i = 1, \dots, n; n > m$ , from (25) to give

$$\begin{cases} y_i = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m + \varepsilon_i \\ \varepsilon_i \sim N(0, \sigma^2), i = 1, 2, \dots, n \end{cases}$$
(26)

The use of Matlab for solving linear regression equation results p = 0.094,  $R^2 = 0.686$ , When the interval is set criteria R> 8 considered equation is linear, the problem is the output value 0.828, so the equation is linear, and  $\beta_0 = 4.6189$ ,  $\beta_1 = 0.2187$ ,  $\beta_2 = -0.0007$ . Here using  $x_1$  to represent the

time variable,  $x_2$  stand for the cost variable, and *y* represents the score value, set the value out of five stars. So

(27)

 $y = 4.6189 + 0.2187x_1 - 0.0007x_2$ 

Table 4 gives the standard of the model test.

Tab.4 Standard of model test						
Standard	Well	Good	General	Poor		
Grading range	≥ 4.5	3.5 ~ 4.5	2.5 ~ 3.5	≤ 2.5		

Test plan (1): Solving the results according to (1), the time variable = 7.6, variable cost = 1894, substituting the formula (27) was:  $4.6189 + 0.2187 \times 7.6 - 0.0007 \times 1894 = 4.95522 > 4.5$ . Comparison Tab.4 scoring criteria, so that a very good plan.

Test plan (2): According to the results of (2) solving, time variable cost variable into the equation (27) was:  $4.6189 + 0.2187 \times 7 - 0.0007 \times 1872 = 4.838 > 4.5$ . Comparison Tab.4 scoring criteria, so the plan two is very good.

Test plan (3): According to the results of (3) solving, time variable cost variable into the equation (27) was:  $4.6189 + 0.2187 \times 2.7 - 0.0007 \times 582 = 4.80199 > 4.5$ . Comparison Tab.4 scoring criteria, so the plan three very well.

Test plan (4): According to the results of (4) solving, time variable cost variable into the equation (27) was:  $4.6189 + 0.2187 \times 4.3 - 0.0007 \times 893.8 = 4.93365 > 4.5$ . Comparison Tab.4 scoring criteria, so the four plans are also very good.

#### **Summary**

This study enriches the research methods to optimize the design of the tourist routes, showing the application potential of the discuss analytical method for the optimization design of tour routes, reduce travel costs of tourism enterprises and tourists. This article discusses how to make tourism travel the shortest distance, least time, the problem of the lowest cost. Model structure is simple, clear, can better solve practical problems; structural model closer to reality; meet market relations; model up easily, without tedious calculations; obtained results of the model closer to reality, in line with our common-sense understanding. Of course, the article also has some shortcomings, due to limited time, so it does not adequately consider the psychological flash prime tourist satisfaction. However, in practical problems, the satisfaction degree of tourists must be considered in the travel route optimization design. Further research can be carried out by APT model [6].

#### Acknowledgement

This work was financially supported by the National Natural Science Foundation of China (41271292, 61303227).

#### References

[1] Jolliffe I T. Principal component analysis[J]. Springer Berlin, 2010, 87(100):41-64.

[2] Institute S. SAS/STAT user's guide[M]. SAS Institute, 1990.

[3] Diestel R. Graph theory[J]. Oberwolfach Reports, 2000, 311(1):67-128.

[4] Hookings A. Numerical Methods with MATLAB(Book Review)[J]. International Journal of Electrical Engineering Education, 2000.

[5] Andrews D F. A Robust Method for Multiple Linear Regession[J]. Technometrics, 2012, 16(4):523-531.

[6] Pieleanu F D. The APT Model and its Applicability in Romania's Case[J]. Romanian Statistical Review Supplement, 2012, 60:103-112.