A principal component analysis using SPSS for Multi-objective Decision Location Allocation Problem

ZIPENG ZHANG
School of Management Science and Engineering
Shandong Normal University
No.88 Culture Rode, Jinan City, Shandong Province
CHINA
zhangzipeng20@126.com

Abstract: - In order to solve the location allocation problem with multi-objective decision (MDLAP), this paper creatively combines a cost-based mathematical optimization model, which transforms the distribution location problem into a two-stage logistics location selection decision one. In the process of solving the bottom model, this paper puts forward many methods such as data standardization, entropy weight, principal component analysis and mathematical expressions by using SPSS soft. In the process of solving the top model, this paper use immune algorithm to apply experimental simulation which based on logistics demand and location data. In the process of analysis, there are many methods to be provided to solve the above model, such as the weighted linear regression method, the similarity analysis system clustering method, the principal component regression method and the Immune algorithm. As a result, the 97 candidate service area in Shandong Province are selected into 9 service areas of Wei Fang, Qingdao, Ping Du, Q Fu and so on, in order to be the better optimal logistics development area. This model avoids the ambiguity of the traditional methods, and we can better solve the optimal number and location problem in the LAP.

Key-Words: - SPSS; location selection model; principal component regression method

1 Introduction

With the rapid development of China's economy, the logistics and distribution business have attracted many researchers from various fields working on it over the past few years. Many significant theories have been proved that the distribution plays a pivotal role in the logistics system. In order to make different requirements to be delivered to the hands of customers accurately and effectively, the main targets of this paper is to make a timely logistics system research, which can gain practical value to solving the above problem. Domestic and foreign scholars have conducted a lot of research and put forward many location models. A large number of studies show: the logistics distribution location problem is a multi-objective optimization problem with complex constraints, and it belongs to the NP-hard problem. These optimization algorithms, such as taboo search algorithm, genetic algorithm and ant colony algorithm, also called heuristic search algorithm, often achieved no better results when the scale is large. we unable to obtain the global optimal location result, because these algorithms' searching speed is slow and easy to fall into local optimum. To quantify and extrapolate a better ideal, we should develop more efficient analytical and computational approaches.

Logistics location allocation problem (LAP) [1] can be traced back to the 1909 issue of Weber, it first treats the LAP from a mathematical point of view. After nearly a century of development, its theory and application have been greatly enriched, the logistics location allocation problem has produced a network location model (DLAP), a single period model (SLAP), capacitated multi stage model (CMLAP) and multi-objective location model (MOLAP) [2-5]. Most of the people research the capacity-limit single stage model to solve LAP up to the present.

At present, the study of logistics function in expressway service area mainly focuses on the feasibility analysis and management research. Ceng Zhaogeng[6] (2008) points out that the expressway service area in China will develop to three directions: first, some will change from rest function to leisure function, second, part of them will become the logistics node, finally, the expressway service area will become an important platform, and it can be used for commerce and trade circulation. Zheng Zhiping[7] (2011), Miu Guosheng[8] (2011)there is a analysis about Fujian, Shanghai, Nanjing expressway service area development advantages of modern logistics base, it clearly pointed out that the highway development of logistics has great potential and puts forward the preliminary plan to introduction of LAP in logistics industry. Liu Ying [9] (2009) made Shandong Province as the research object, his analysis about the third party logistics service area in LAP feasibility is original, and it is based on the development of internal relations between the expressway service area and the development of modern logistics industry, finally he put forward the strategic plan and strategy of LAP in Shandong expressway service area.

In short, solution of the MDLAP yields many insights including the following: one is a discussion about the development direction of expressway service areas. The second is the feasibility and countermeasures for the development of logistics location problem. But the methods above have not become an important issue of long standing interest to economists and transportation scientists, and especially there is no deep research on the expressway service area logistics function network. It should be under the condition of development of key technologies. In summary, the highway logistics function has some academic research papers, but it is mostly feasible research and management to develop its logistics function, the literature of LAP in expressway service area logistics node is rarely, only Xi Jun [10] analysis and discusses comprehensive evaluation method of alternative nodes, it use the principal component in expressway service LAP problem, and also use the qualitative analysis and quantitative analysis method to discuss the level , location and function of the logistics node service areas. Qin Lu [11] (2007) applies the method above to the logistics service area of expressway node partition, he use it to achieve good results in the angel of regional logistics and the expressway service area integration.

The aim of the paper is to solve logistics location problems by using different solution approaches. In this paper, a decision-making mathematical model, with multi-attribute group, is proposed. This paper uses SPSS soft to analyze the relationship between many relevant indicators and logistics demands. In this paper, we put forward four steps to solve the reality network problems of expressway traffic in Shandong province. First it takes the topological structure as a starting point. Secondly, it puts the cost-based and transfer-based LAP problem as the target. In the third step, this paper creatively combines the mathematical optimization model with a two-stage logistics location. In the end, we obtain the solution to cost-based decision problem and the transport-based optimization problem. This paper not only puts forward a method of data standardization processing to solve the bottom model, but also applies the artificial immune algorithm to determine the number and location of the logistical centres. The above two methods remedy the disadvantages of a decision method which is with simple factor. It is easier to get accurate results of location.

2 Description of the LAP Problem

2.1 Definition of the LAP problem

In the basic LAP problem, there is a highway network in the range of a certain area (Shandong province). We can know that the position of the candidate logistics (M) and logistics demand (N) has been fixed. In order to provide finished products for the low cost logistics demand, storage, transfer, processing, management and other services, the system requirement selects one or a plurality of logistics from the candidate logistics nodes. For example, there are 97 expressway service area of expressway in Shandong province, and it is within the scope of their size, location, condition, location and other characteristics are known for each service area, LAP problem demand to determine the number and location of logistics in the service area according to the specific decision method.

2.2 Mathematical model of the LAP problem

The selection of a logistics location among alternative locations is a multi target decision-making problem including both quantitative and qualitative criteria. The mathematical model is used as the logistics node from 97 alternatives one from Shandong province expressway service area, of course it should based on the total demand for logistics and transport costs, thus we can achieve the lowest total cost of logistics and transport in this area. As everyone knows, the goods of transport diversity have various kinds. And different types of goods have different distribution costs, which due to its weight, volume, timeliness and portable degree caused by transportation. The total transportation cost and volume is provided as the target in this paper, we defined a condition that the distribution cost is no related to the types of transport goods. In this paper we builds a LAP optimization model based on this consideration, and the model is close to the reality of the classification in the position of that the candidate logistics and logistics demand nodes is known, of course it is under the situation of not considering the transportation storage fee, management fee and transport cost and freight traffic is proportional to the distance.

The model is as follows：

$$MinC = \sum_{i^{'} \in G_i^{'}} \sum_{i^{''} \in G_i^{''}} \sum_{j^{'} \in G_j^{'}} \sum_{l \in L} \sum_{k \in N} [C_{i^{'}j^{'}}^{l} X_{i^{'}j^{'}k} D_{i^{'}j^{'}}^{l} Q_{i^{'}j^{'}}^{l} + C_{j^{'}i^{''}}^{l} X_{j^{'}i^{''}k} D_{j^{'}i^{''}}^{l} Q_{j^{'}i^{''}}^{l}] + \sum_{j^{'} \in G_j^{'}} H_{j^{'}} (F_{j^{'}c} S_{j^{'}} + W_{j^{'}c} Q_{j^{'}}) \quad (1)$$

Table1 Symbol and Definition of the Algorithm

Symbol	Definition
$G_i^{'} = \{i^{'} / i^{'} = 1,2 \cdots m\}$	Sets of city of vehicle starting nodes
$G_i^{''} = \{i^{''} / i^{''} = 1,2 \cdots m\}$	Sets of city of vehicle finishing nodes
$G_j^{'} = \{j^{'} / j^{'} = m+1 \cdots\}$	Expressway service areas
$N = \{k / k = 1,2 \cdots\}$	The set of vehicles
$L = \{l / l = 1,2,3 \cdots\cdots l\}$	Different types of goods
F_{jc}	Building cost of the logistics
W_{jc}	Operating costs of the logistics
$C_{i^{'}j^{'}}^{l}$	Transportation cost of goods l from starting Node $i^{'}$ to finishing node $j^{'}$
$D_{i^{'}j^{'}}^{l}$	Transportation distance of goods l from starting node $i^{'}$ to logistics node $j^{'}$
$Q_{i^{'}j^{'}}^{l}$	Transportation traffic of goods l from starting node $i^{'}$ to logistics node $j^{'}$
$C_{j^{'}i^{''}}^{l}$	Transportation manage cost of goods l from logistic node $j^{'}$ to finishing node $i^{''}$
$D_{j^{'}i^{''}}^{l}$	Transportation distance of goods l from logistic node $j^{'}$ to finishing node $i^{''}$
$Q_{j^{'}i^{''}}^{l}$	Transportation of goods l from logistic node $j^{'}$ to finishing node $i^{''}$
$S_{j^{'}}$	The scale of expressway service area
$Q_{j^{'}}$	The maximum amount of transit goods of logistics node $j^{'}$

This is the constraints condition about the mathematic model of the LAP problem.

- Decision variables constraints.

$$X_{i^{'}j^{'}k} \le H_{j^{'}} \vee X_{j^{'}i^{''}k} \le H_{j^{'}} \quad (2)$$

We can know from the formula (2) that the service area which was not selected for the logistic. And it is not been provided transit function for any transport service.

- Transport capacity constraints

$$\sum_{i^{'} \in G_i^{'}} \sum_{i^{''} \in G_i^{''}} \sum_{j^{'} \in G_j^{'}} \sum_{l \in L} Q_{i^{'}j^{'}}^{l} \le Q_{i^{'}j^{'}}^{l} \quad (3)$$

$$\sum_{i^{'} \in G_i^{'}} \sum_{i^{''} \in G_i^{''}} \sum_{j^{'} \in G_j^{'}} \sum_{l \in L} Q_{j^{'}i^{''}}^{l} \le Q_{i^{'}j^{'}}^{l} \quad (4)$$

The formula (3) show that the total freight volume of goods from the originating station to logistics less than transit goods from the originating station to the terminal. The formula (4) represents that the total freight volume of goods from the logistics to finishing node less than transit goods from the originating station to the terminal.

- Capacity constraints constraint.

$$\sum_{i^{'} \in G_j^{'}} \sum_{j^{'} \in G_j^{'}} Q_{i^{'}j^{'}} \le Q_{j^{'}} \quad (5)$$

The formula (5) means that the capacity of logistics can meet the demand of passenger transit freight logistics network.

- Cost relations constraints

$$COST = F_{jc} + C_{i^{'}j^{'}}^{l} \; D_{j^{'}i^{''}}^{l} \quad (6)$$

Formula (6) shows the relation about the cost and the transportation cost, which should constitute like the formula (6).

3 Realization of the Algorithm

As for solving the logistics LAP problem, we select some reasonable methods, such as that we not only can save cost, but also speed up circulation efficiency, increase social benefits. About the LAP problem of expressway logistics network[12], we will determine correct logistics centres from the candidate service areas, in the process, various influencing reasons behind the situation are associated together with the LAP problem (including the subjective factors and objective factors) [13]. According to the actual situation and logistics theory, finally we obtained ideal results of our through the comprehensive analysis about the factors.

3.1 Collection for the decision data

The related data of expressway service area in this paper are not only taken from the authority of the statistical yearbook and the official website, they are also obtained through the full market investigation.

By the way, we also use the analytic hierarchy process to consolidate decision-makers' assessments. Through principal component analysis method using SPSS soft, this paper analyzes on 13 indicators shown in Table 2. Through standardized treatment of original matrix such as the economy, politics, population, resources, environment and so on, the correlation matrix will be shown in table 3. All kinds of statistical data required for this paper mainly include the following aspects:

- We gain the dates which contain the number of population, the state of the economy scale, the logistics demand of each candidate centre from the statistical yearbook. They can be used by different theoretic backgrounds and relate differently to the discipline of multi-attribute group decision-making.
- We can gain the expressway service information of Shandong province, which include infrastructure, convenient transportation, toll station spacing, service area scale, function, coverage and the distance information.
- We can also know the policy environment of candidate logistics centre, the degree of public approval, the government support and other information from the government work report.
- The conventional approaches to logistics location problem tend to be less effective in dealing with the imprecise or vagueness nature of the linguistic assessment. In order to ensure data reliability and timeliness, this paper obtains information from the authoritative statistical yearbook. This paper also can make the existing statistical data analysed and processed. It makes the date close to the LAP problem of logistics centre, and the result is more convincing. For example, in the decision-making index of the importance of service route, there is no direct data to display the importance index. According to the shortest path information between each pair of the cities, we need the index only be obtained by statistical analysis of the route information.

To sum up, this paper selects 24 information of expressway service area as the evaluation index from the Shandong Expressway Group, which is shown in the following table 2: logistics centre location decision factors of Shandong Expressway Group:

Table 2 Logistics Centre location Decision Factors of Shandong Expressway Group

Num	Candidate	W_1	W_2	W_3	W_4	W_5	W_6	W_7	W_8	W_9	W_{10}	W_{11}	W_{12}	W_{13}	Score
1	De Zhou	1915	16	2	10	6563	557	150	13000	10000	1	50	5.00	8.00	
2	Xia Jin	1915	5	1	80	6563	557	200	6000	2000	0	40	1.00	7.00	
3	De Nan	1915	16	3	30	6563	557	120	3826	4200	2	60	3.00	8.00	
4	G Tang	1905	5	1	40	7131	579	200	9687	5600	1	80	3.00	7.00	
5	Yu Cheng	1915	16	1	50	6563	557	180	12800	15000	1	50	3.00	8.00	
6	Tan Q	4400	25	4	5	10705	681	120	15691	2625	1	100	7.00	8.00	
7	Tai An	2475	25	4	0	7018	549	130	11042	3625	2	60	7.00	8.00	
8	Nin Yang	2475	30	3	40	7018	549	150	6170	5000	1	50	5.00	8.00	
9	Q Fu	2820	30	3	30	6431	808	350	9639	6350	2	50	5.00	8.00	
10	Z Cheng	2820	16	1	20	6431	808	80	5420	4168	1	50	4.00	7.00	
11	Ten Zhou	1560	16	1	10	4243	373	100	6000	6000	1	60	4.00	8.00	
12	Z Zhang	1560	16	3	0	4243	373	130	11000	8000	1	80	5.00	8.00	
13	X Cheng	1560	16	1	5	4243	373	150	6800	3500	0	20	1.00	8.00	
14	Cao Zhou	1440	16	2	20	4836	829	230	7000	3600	1	30	3.00	8.00	
15	Z Ping	3280	22	3	15	12119	453	90	6392	6500	2	60	7.00	8.00	
16	Z Bo	3280	22	4	0	12119	453	192	12000	9830	2	80	8.00	8.00	
17	Qing Zhou	3600	42	3	20	11862	909	170	6789	6400	2	70	6.00	8.00	
18	Fang Z	3600	42	2	15	11862	909	246	5168	3800	1	80	4.00	8.00	
19	Wei Fang	3600	42	5	0	11862	909	250	12000	12000	3	80	8.00	8.00	
20	Go Mi	3600	15	2	30	11862	909	80	5600	3000	1	30	2.00	8.00	
21	Ping Du	4907	20	4	32	15802	697	278	11000	2000	3	30	7.00	7.00	
22	Lai Xi	4907	15	3	38	15802	697	225	3054	500	2	30	6.00	7.00	
23	Wien Deng	2203	6	2	20	6869	280	170	5000	2700	1	30	3.00	7.00	
24	Qing Dao	6608	21	5	0	25371	872	164	16801	13000	1	100	9.00	8.00	

In the above table of the decision factors shows that: W_1 represents the economy condition of the city which contain candidate service area, W_2 represents the importance of the route in candidate services, W_3 defines a influence index of candidate services area, and W_4 describes the distance between a candidate service area and its adjacent city, W_5 explains a index of logistics demand volume of the regional area which candidate service areas in, W_6 is the population index of city, W_7 is the size of the candidate service area, W_8 gives the gross area of the candidate service area, W_9 represents the turnover of the candidate service area, W_{10} represents the accessibility of the candidate service areas. W_{11} is the cost of transformation and operation of service areas, at the same time the development prospect of the logistics service area is on behalf of W_{12}, W_{13} describes the infrastructure in the service area. The data of W_1, W_5, W_6, are obtained from the economic development situation of Shandong Province Statistical Yearbook and the government work report in 2011. The data of W_6, W_7, W_8, W_{11}, W_{13}, come from the statistical data which belong to Shandong province transportation hall of official website. W_2, W_{10} are collected from the expressway network and the dates of W_3, and W_{12} are obtained through investigation and statistics.

3.2 Data standardization and its formula

In the process of analysing about LAP problem, there is no doubt that, we often encounter a variety of data types, and the difference between the units of measure for various statistical data will lead to the final evaluation results for the convenience of analysis. Under many situations, the values of the qualitative criteria are often imprecisely defined for the decision-makers.

In order to make the perfect decision, we need put the various data model of analysis to be normalized. The standard method used in this thesis is the standard method of maximum and minimum value, the following specific standardization method:

- Positive index (large for optimal index) processing method:

$$X_i^* = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \tag{7}$$

- Negative index (small is better index) processing method:

$$X_i^* = \frac{X_{\max} - X_{\mathrm{i}}}{X_{\max} - X_{\min}} \tag{8}$$

Where X_i describes all kinds of index value for the raw data, X^*_I represents all kinds of data index for the normalized value.

3.3 Determine the weight

The concept of entropy is a function description for the state of system. Value and variation of entropy not only commonly carries to be used on the analysis and comparison, but also be used to calculate a disorder of one system. For the dates in the model of LAP model, the concept of entropy is put forward to measure the difference between the same degree factors in statistics: when a statistical data of each evaluation object is larger, the smaller, entropy, which is said the information provided by the index is larger; when a statistical data of each evaluation object is smaller, the larger, entropy, which describes the effective information index provide is smaller. When the difference between the data of a certain evaluation factor is little, entropy tends to maximize, this can show that the valid information of the index is very low, so we can remove such indicators from the decision model.

- Definition of entropy:

$$H_i = -\frac{1}{\ln q}\sum_{i\in q}\frac{X_i}{\sum_{i\in q} X_i}\ln X_i \tag{9}$$

- Definition of weight:

$$W_i = \frac{1-H_i}{p-\sum_{i\in p} H_i}, (0 \le \mathrm{W}_i \le 1, \sum_{i\in p} W_i = 1) \tag{10}$$

Where p is the number of index data, q is the number of decision-making object decision making problem.

- Principal component analysis method:

$$R_{ij} = \frac{1}{1-q}\sum_{i\in q}\frac{\left(X_i^* - \overline{X}_i^*\right)\left(X_j^* - \overline{X}_j^*\right)}{S_i S_j} \tag{11}$$

$$\overline{X}_i^* = \frac{\sum_{i\in q} X_i^*}{q} \tag{12}$$

$$S_i = \sqrt{\frac{\sum_{i\in q}\left(X_i^* - \overline{X}_i^*\right)^2}{q-1}} \tag{13}$$

This paper got all kinds of data which related to decision model, by using Statistics description function that in SPSS soft. It makes a standard to original data based on mathematical expressions above. Through standardized treatment of original matrix above, the correlation matrix will be shown in table 3, like the following:

Table3 Normalized Location Factors in Decision Model of LAP

N	candidate	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	score
1	De Zhou	0.09	0.30	0.25	0.13	0.11	0.44	0.35	0.72	0.76	0.33	0.38	0.50	1.00	
2	Xia Jin	0.09	0.00	0.00	1.00	0.11	0.44	0.61	0.21	0.12	0.00	0.25	0.00	0.00	
3	De Nan	0.09	0.30	0.50	0.38	0.11	0.44	0.20	0.06	0.30	0.67	0.50	0.25	1.00	
4	Go Tang	0.09	0.00	0.00	0.50	0.14	0.48	0.61	0.48	0.41	0.33	0.75	0.25	0.00	
5	Yu Cheng	0.09	0.30	0.00	0.63	0.11	0.44	0.51	0.71	1.16	0.33	0.38	0.25	1.00	
6	Tan Q	0.57	0.54	0.75	0.06	0.31	0.64	0.20	0.92	0.17	0.33	1.00	0.75	1.00	
7	Tai An	0.20	0.54	0.75	0.00	0.13	0.43	0.25	0.58	0.25	0.67	0.50	0.75	1.00	
8	Nin Yang	0.20	0.68	0.50	0.50	0.13	0.43	0.35	0.23	0.36	0.33	0.38	0.50	1.00	
9	Q Fu	0.27	0.68	0.50	0.38	0.10	0.84	1.36	0.48	0.47	0.67	0.38	0.50	1.00	
10	Z Cheng	0.27	0.30	0.00	0.25	0.10	0.84	0.00	0.17	0.29	0.33	0.38	0.38	0.00	
11	Ten Zhou	0.02	0.30	0.00	0.13	0.00	0.15	0.10	0.21	0.44	0.33	0.50	0.38	1.00	
12	Z Zhang	0.02	0.30	0.50	0.00	0.00	0.15	0.25	0.58	0.60	0.33	0.75	0.50	1.00	
13	Xi Cheng	0.02	0.30	0.00	0.06	0.00	0.15	0.35	0.27	0.24	0.00	0.00	0.00	1.00	
14	Cao Zhou	0.00	0.30	0.25	0.25	0.03	0.87	0.76	0.29	0.25	0.33	0.13	0.25	1.00	
15	Z Ping	0.36	0.46	0.50	0.19	0.37	0.28	0.05	0.24	0.48	0.67	0.50	0.75	1.00	
16	Zibo	0.36	0.46	0.75	0.00	0.37	0.28	0.57	0.65	0.75	0.67	0.75	0.88	1.00	
17	Qing Zhou	0.42	1.00	0.50	0.25	0.36	1.00	0.45	0.27	0.47	0.67	0.63	0.63	1.00	
18	Fang Z	0.42	1.00	0.25	0.19	0.36	1.00	0.84	0.15	0.26	0.33	0.75	0.38	1.00	
19	Wei Fang	0.42	1.00	1.00	0.00	0.36	1.00	0.86	0.65	0.92	1.00	0.75	0.88	1.00	
20	Go Mi	0.42	0.27	0.25	0.38	0.36	1.00	0.00	0.19	0.20	0.33	0.13	0.13	1.00	
21	Ping Du	0.67	0.41	0.75	0.40	0.55	0.66	1.00	0.58	0.12	1.00	0.13	0.75	0.00	
22	Lai Xi	0.67	0.27	0.50	0.48	0.55	0.66	0.73	0.00	0.00	0.67	0.13	0.63	0.00	
23	Wien Deng	0.15	0.03	0.25	0.25	0.12	0.00	0.45	0.14	0.18	0.33	0.13	0.25	0.00	
24	Qing Dao	1.00	0.43	1.00	0.00	1.00	0.94	0.42	1.00	1.00	0.33	1.00	1.00	1.00	

4 Decision-making Model of LAP

In order to meet the target established at the start of the chapter, we firstly use the method of data standardization on the normalized sets to establish the model of the decision-making problem, the result of correlation analysis will guide us to find the correct conclusion that whether the service areas are built into logistics centres. In this section, we use various methods based on thought and layers of SPSS software to mine effective information.

4.1 Linear regression method in SPSS

Take steps of "Analysis"—"Regression" –"Weight estimation" in SPSS software to calculate the weight of influence factors X_1, X_2, X_3......X_{13} by mathematical expressions.

Table 4 Weight of Influence Factor

Index	Weight	Index	Weight
X_1	0.352	X_8	0.119
X_2	0.524	X_9	0.115
X_3	0.129	X_{10}	0.893
X_4	0.236	X_{11}	0.265
X_5	0.484	X_{12}	0.214
X_6	0.087	X_{13}	0.923
X_7	0.032		

According to the concept of entropy weight, we can clearly find that the value of factor X_{13} was significantly higher than other influence factors, so this kind of influence factor index X_{13} from the 24 candidates have no difference, and the global information it contains was significantly less than the rest factors, thus the factors X_{13} can be listed as a weak correlation factor, and deleted, then decision factors W_{13} in table 3 (red line) has been removed.

4.2 Cluster method in SPSS

In the process of analysis of decision factors, we gradually find that: because of the similar factors, the decision contents (each index of candidate service area) between each other have similar results. The index of 24 candidates is divided into several categories, and they have appropriated large-scale system clustering method in SPSS. This can not only reduce the analysis scale, but also makes the differences of the index in the categories as small as possible and the difference between categories as large as possible.

Corresponding analysis of the operation as follows:

- Step 1: Click the "analysis" -- "classification" --

"cluster ".

- Step 2: Make the decision influence factor "X1", "X2"...... "X12" to be selected into "variables" list box.
- Step 3: Select the method: make the conversion value standard for "Z score", select the button below "case", "group".
- Step 4: select 2-12 in clustering scheme s, determine the preservation.

What cluster number we should select is not determined before the above operation, so it requires the calculation of all the results of the 2 to12 class, analysis of the date as follows: it can be seen from the chart that it can display good group and different group if the candidate are divided into 6,7 or 8 categories.

4.3 Comprehensive decision of LAP model

4.3.1 Correlation factor

The method of factor analysis in SPSS software will be used to make the correlation variables divided into a fewer sets of variables. And it has high correlation in the same group, and low correlation in different groups. According to this method, we can calculate the correlation coefficient matrix to solve the original problem. When variance account for 85% of total variances, then we can select the principal components

We can see from the Table 5 that five extracted principal components reflect 89.62% information of primitive variables. It should be proved that there is a strong correlation between all variables before we apply the method of factor analysis to solve LAP problem. For example, the 12 variables of index in table 3 is not independent for each other. In order to analysis of the relationship between the original variables, we should investigate the correlation between each other. This paper uses the correlation analysis of SPSS to discuss the correlation, and gets the following variable correlation table below.

In order to facilitate the layout, this paper only discusses the correlation matrix of the first 8 results, which are visible that there is strong or weak positive correlation between these variables, that is to say the information between the above variables are overlap.

4.3.2 Factor concentrate by SPSS

In a decision problem, there are p decision hypothesis, decision sample in the index data, $X = \left(x_1, x_2, x_3, \cdots\cdots, x_p\right)^T$ is random variables , The common factor which the paper search is $F = \left(f_1, f_2, f_3, \cdots\cdots, f_p\right)^T$, then we can see factor analysis model as:

$$\begin{aligned} X_1 &= a_{11}F1 + a_{12}F_2 + \cdots\cdots + a_{1m}F_m + \varepsilon_1 \\ X_2 &= a_{21}F_1 + a_{22}F_2 + \cdots\cdots + a_{2m}F_m + \varepsilon_2 \\ &\cdots\cdots \\ X_p &= a_{p1}F_1 + a_{p2}F_2 + \cdots\cdots + a_{pm}F_m + \varepsilon_p \end{aligned} \quad (13)$$

$A = (a_{ij})$ is the loading matrix of the factor, a_{ij} describes the load factor ε is the special influence factors outside the factor (the actual analysis is negligible).

We can use regression estimation method to compute the mathematical model of factor scores after calculating the common factor, and then evaluate the case by further calculating the factor scores. The formula of factor scores is:

$$F_i = b_{i1}X_1 + b_{i2}X_2 + \cdots\cdots + b_{in}X_n \left(i = 1, 2, 3\cdots\cdots m\right) \quad (14)$$

According to the step of "analysis", "reduction" and "factor analysis", we can have the following table 5:

Table 5 Total Variance Explained

C	Initial values			Extraction Sums of Squared Loadings		
	Total	Variance%	Cumulative	Total	Variance%	Cumulative
1	4.83	43.98	43.98	4.84	43.98	43.98
2	1.90	17.29	61.27	1.90	17.29	61.27
3	1.15	10.46	71.74	1.15	10.46	71.74
4	1.01	9.23	80.97	1.01	9.23	80.97
5	0.76	6.94	87.91	0.76	6.93	87.91
6	0.43	3.96	91.87			
7	0.37	3.44	95.31			
8	0.20	2.27	97.58			
the principal component analysis method						

Table 6 Component Matrix (Method: Principal Component Analysis)

	Component				
	1	2	3	4	5
X_1	0.781	0.549	-0.274	0.123	-0.003
X_2	0.648	0.170	0.252	-0.589	-0.168
X_3	0.891	0.025	-0.037	-0.232	0.198
X_4	-0.515	0.409	-0.023	0.672	0.181
X_5	0.759	0.283	-0.090	0.202	0.050
X_6	0.490	0.540	0.082	0.298	-0.106
X_7	0.253	0.331	0.554	0.301	0.331
X_8	0.610	-0.213	0.547	0.401	0.226
X_9	0.426	-0.353	0.507	0.341	0.050
X_{10}	0.642	0.315	0.284	-0.694	0.314
X_{11}	0.640	-0.377	0.003	0.151	-0.576
X_{12}	0.922	-0.116	-0.058	-0.191	0.178

Table 7 Component Score Coefficient Matrix

	Component				
	1	2	3	4	5
X_1	-0.088	0.460	-0.087	-0.047	-0.049
X_2	0.109	-0.238	-0.039	0.620	-0.084
X_3	0.310	0.078	-0.029	-0.131	-0.003
X_4	-0.269	0.169	0.010	-0.102	0.427
X_5	-0.122	0.477	-0.028	-0.107	-0.015
X_6	-0.293	0.146	-0.029	0.593	0.050
X_7	0.111	-0.122	0.117	0.009	0.680
X_8	-0.076	0.052	0.445	-0.205	0.154
X_9	-0.130	-0.130	0.498	0.018	0.177
X_{10}	0.500	-0.122	-0.154	-0.101	0.208
X_{11}	-0.129	0.005	0.269	0.215	-0.217
X_{12}	0.275	0.078	0.035	-0.139	-0.046

From the table 5 we can see that the 12 original factors can be summarized into five components (factors), the cumulative percentage of the first component is 43.9% in the total data, second component is 17.3%, third, four or five components of the proportion is respectively 10.5%, 9.2%, 6.9%. That is to say the importance degree of first component in the five components is much larger.

The cover rate of the five components is 87.91% in all data, so we can determine that the principal component analysis method have very ideal effect.

Combined with the component matrix data of table 6, factors of X_3, X_5, X_{12} score higher in the first principal components, X_1, X_6 were higher in the principal component 2, and X_7, X_8, X_9 have high score in the main components of 3, at the same time X_2, X_4, X_{10} have high score in the components of 4, then X_{11} scores higher in the main composition of 5. So it can be classified the higher scores of factors into the corresponding principal component, which make principal components of Z_1 as the logistics influence factors (including X_3, X_5, X_{12}), the principal component Z_2 as the factor of economic population (size X_1, X_6), the principal component Z_3 factor as scale of candidate service area (X_7, X_8, X_9), the main composition of Z_4 as convenience degree of the traffic in candidate service area (including X_2, X_4, X_{10}), the principal component Z_5 as investment and operation cost factors (X_{11}).

4.3.3 To calculate the comprehensive scores

In this paper, from Table 7 we can gain the score coefficient matrix of various components, which can directly gain the main components of each five

Table 8 Principal Component Analysis of the LRP Model

Num	Candidate	Z_1	Z_2	Z_3	Z_4	Z_5	Score
1	De Zhou	-0.19310	-0.23121	0.43171	0.51615	0.35326	-0.18
2	Xia Jin	-1.70152	0.87836	-0.52923	1.77392	0.46605	-0.46
3	De Nan	-0.53884	0.14081	0.03075	-1.06185	0.01839	-0.31
4	Go Tang	-0.74439	-0.39378	0.04199	1.31084	0.39227	-0.24
5	Yu Cheng	-0.56428	-0.89756	1.02267	2.32593	0.63897	-0.04
6	Tan Q	1.00307	-0.90986	-1.10385	-0.03176	-0.67934	0.12
7	Tai An	0.46523	-0.61212	0.17287	-1.60620	0.33403	-0.01
8	Nin Yang	-0.29266	0.21194	0.25602	-0.34314	-0.34767	-0.12
9	Q Fu	0.34766	1.14313	2.17930	0.61989	0.59852	0.68
10	Z Cheng	-0.65233	0.05567	-0.70039	-0.23620	-1.45890	-0.47
11	Ten Zhou	-0.89195	-1.11230	0.01205	-0.83744	-0.33264	-0.68
12	Z Zhang	-0.16508	-1.83325	0.36005	-0.52998	0.24491	-0.38
13	X Cheng	-1.47313	-0.55258	-0.04661	-0.47863	-0.27548	-0.81
14	Cao Zhou	-0.75448	0.63854	0.91603	0.11161	-0.28017	-0.13
15	Z Ping	0.30699	-0.37888	-0.78478	-1.24556	0.30603	-0.11
16	Zibo	1.02575	-1.02600	0.20725	-0.39490	1.11374	0.34
17	Qin Zhou	0.80867	0.74105	0.62835	-0.34144	-1.64334	0.4
18	Fang z	0.38277	1.01734	0.73070	0.26337	-2.34205	0.28
19	Wei Fang	1.89434	0.09325	1.89190	-0.24333	-0.16831	1.01
20	Go Mi	-0.56723	0.93155	-1.48281	0.16544	-1.27883	-0.32
21	Ping Du	0.92539	1.80661	-0.24507	-0.31498	2.19353	0.82
22	Lai Xi	0.14267	2.09787	-1.27411	-0.53466	0.97137	0.31
23	We Deng	-1.11234	-0.07635	-0.57442	-.85497	1.28069	-0.55
24	Qing Dao	2.34878	-0.73223	-2.14036	1.96791	-0.10504	0.86

expressions, the expressions of first composition are as follows:

$$F_1 = 0.150*X_1 + 0.125*X_2 + 0.171*X_3 - 0.099*X_4 + 0.146*X_5 + 0.094*X_6 +$$
$$0.049*X_7 + 0.117*X_8 + 0.082*X_9 + 0.124*X_{10} + 0.123*X_{11} + 0.178*X_{12} \quad (15)$$

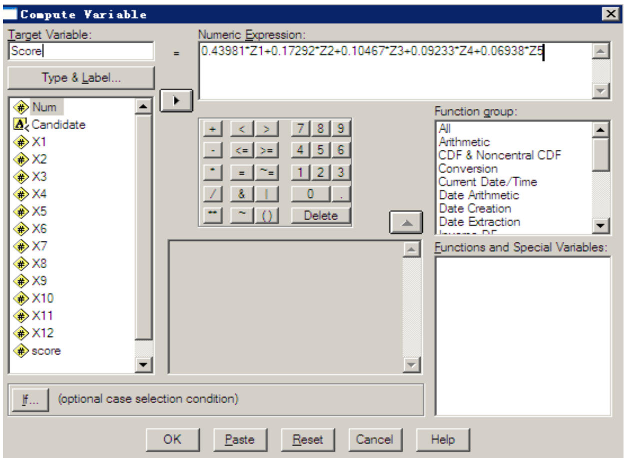

Figure 1 Calculation Window of Component in SPSS

According to the principal component features charts in Table 4 and the component index value in Table 5 we use formula: comprehensive score equal to principal components variance contribution rate*principal component coefficients.

$$R_i = 0.43918*Z_{1i} + 0.17292*Z_{2i} + 0.10467*Z_{3i} + 0.9233*Z_{4i} + 0.6938*Z_{5i}$$

The specific steps as shown in above Figure1, it will generate the comprehensive score of decision-making in raw data generation, see table dotted line.

When we take the scores of the items by order in descending, it is easily been found that the seven service areas such as Wei Fang, Qingdao, Ping Du, Q Fu, Qing Zhou, Zibo, Fang Z, Tan Q can satisfied the demand construction between the 24 candidate service area. We can take this method extending to all 97 service area of Shandong province to make the last results, in addition to the above, the service area of East Lai Wu, Tai An West, Jin Nan, Jin Nin, Yi Sui, Lin Yi North can also be selected as the alternative service area.

4.4 Experimental results of LAP model

In order to find the optimization number of centres in the logistics location problem, this section performs many experimental analyses by using immune algorithm. We can see that, the problem of which service area can be converted into a logistics centre is determined in the previous section 4.3, however, it did not discuss how to determine the number of logistics centre. In this section, there exists much conditional information about such city's date as longitude, latitude and logistics demand. In the condition of this information, we make use of artificial immune algorithm to optimize the solution to the LAP problem. The experimental result is list by figure 1, figure2, figure3, figure4, figure5; and it signs "n" to represent the number of logistics centre among them.

In the section 4.3.3, we select 14 alternative service areas. By comparing the figure1-5, when the number of logistics centre is 8 or 9, the efficiency (such as the optimal service distance and number) of the LAP model is the best.

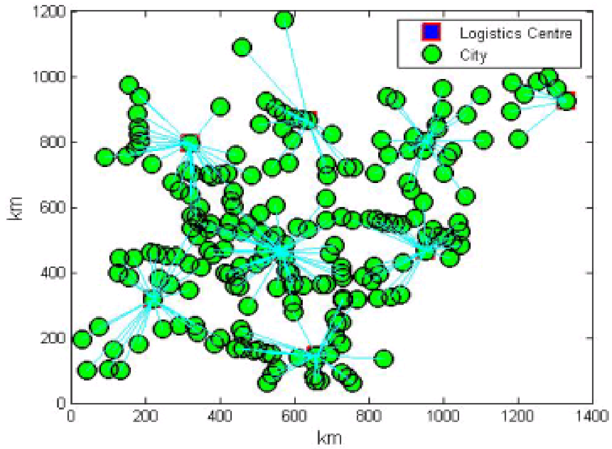

Figure 2 Schematic Diagram of Logistics Centre（n=8）

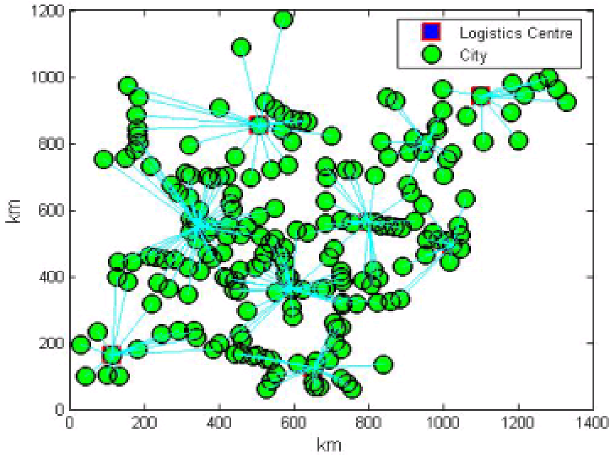

Figure 3 Schematic Diagram of Logistics Centre（n=9）

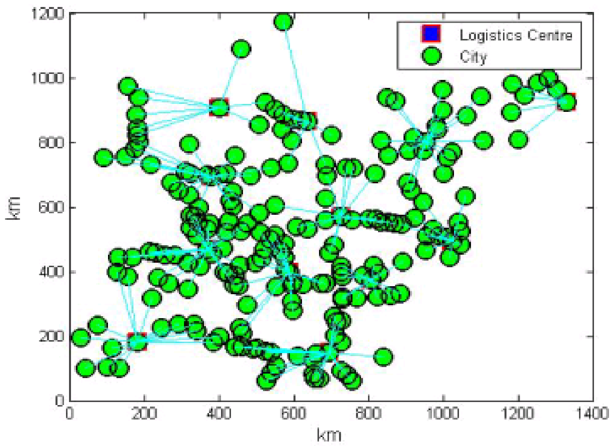

Figure 4 Schematic Diagram of Logistics Centre（n=10）

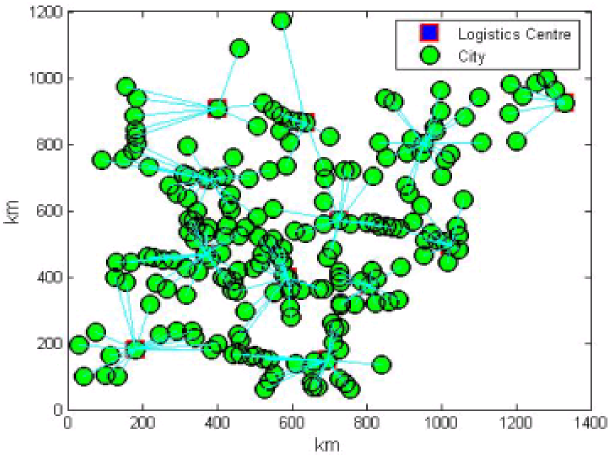

Figure 5 Schematic Diagram of Logistics Centre（n=12）

5 Conclusion

Finally, we have achieved an optimization result by using the method of SPSS and immune algorithm. Taking into account quantitative and qualitative criteria, these approaches are extended to select the best logistics location alternative. In this paper, the real data which collected from the statistical yearbook and the official websites is made as basic research. We firstly make use of the methods in SPSS, which include entropy weight regression, cluster analysis, factor analysis and principal component analysis. Through the optimization of immune algorithm, we finally choose 9 service areas from Shandong expressway group as the most suitable logistic areas. There exists a common feature in such varieties of transport as railway, aviation, shipping, highway and others, that the method above can help them to carry capacity expansion. The methods of logistics location in this paper have more advantages than others. As the transportation of passenger and logistics have many similar characteristics, the methods and ideas of LRP research in highway network can also be applied to other modes of transportation. The single mode of transport in logistics has many cost and efficiency defects, so the future research should focus on the multi-object, multi-mode of transport. It is supposed that the optimized large-scale system in a comprehensive multi-transport network will be used widely.

References:

[1] HU X, ZHANG Y, LI Z. The vehicle route scheduling location problem of multiple distribution s and its solution: An SPSS and genetic algorithm based approach [J]. Logistics Technology, 2010, 1: 029.

[2] Ross A, Jayaraman V. An evaluation of new heuristics for the location of cross-docks distribution s in supply chain network design [J]. Computers & Industrial Engineering, 2008, 55(1): 64-79.

[3] Hu X H, Lu C Z, Li M, et al. Mathematical model for selecting locations for medical and health supplies reserve in Hainan Province [J]. Asian Pacific journal of tropical medicine, 2014, 7(2): 160-163

[4] Yao Z. The Construction of Evaluation Index System of Logistics Service Quality in Online Shopping with Small B2C and C2C as Example [J]. Journal of Anhui Agricultural Sciences, 2013, 1: 177

[5] Gajšek B, Grzybowska K. A cross-county contextual comparison of the understanding of the term logistics platform in practice [J]. Research in Logistics & Production, 2013, 3

[6] Siller A B, Tompkins L. The big four: analyzing complex sample survey data using SAS®, SPSS®, STATA®, and SUDAAN®[C]//Proceedings of the Thirty-first Annual SAS® Users Group International Conference. 2006: 26-29.

[7] Panda S, Padhy N P. Optimal location and controller design of STATCOM for power system stability improvement using PSO [J]. Journal of the Franklin Institute, 2008, 345(2): 166-181.

[8] Holzbeierlein J M, Lopez-Corona E, Bochner B H, et al. Partial cystectomy: a contemporary review of the Memorial Sloan-Kettering Cancer experience and recommendations for patient selection[J]. The Journal of urology, 2004, 172(3): 878-881.

[9] Jiang Z, Wang D. Model and algorithm of location optimization of distribution s for B2C e-commerce [J]. Control and Decision, 2005, 20(10): 1125.

[10] Ji S W, Huang T T, Zhang Y F. Study on Manufacturing Enterprises Distribution Location [J]. Advanced Materials Research, 2014, 834: 1938-1941.

[11] Li Y, Liu X, Chen Y. Selection of logistics location using Axiomatic Fuzzy Set and TOPSIS methodology in logistics management [J]. Expert Systems with Applications, 2011, 38(6): 7901-7908.

[12] Alizadeh B, Burkard R E. Uniform-cost inverse absolute and vertex location problems with edge length variations on trees [J]. Discrete Applied Mathematics, 2011, 159(8): 706-716.

[13] Viti F, Rinaldi M, Corman F, et al. Assessing partial observability in network sensor location problems [J]. Transportation research part B: methodological, 2014, 70: 65-89.