Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2014, 6(5): 442-448

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

Application of Fuzzy Cluster Theory in Logistic Distribution Node Analysis-Use an Agriculture Products Distribution Enterprise as an Example

Xin Lei

Hunan Vocational College of Commerce

ABSTRACT

This paper investigates eight logistic nodes of an agriculture product distribution enterprise with the fuzzy cluster theory applied to the analysis of logistic distribution node. A statistical indicator system of distribution nodes is built based on the condition of these eight nodes of the enterprise. With the fuzzy cluster calculation, the nodes are categorized to economical distribution area and non-economical distribution one. Different strategies are practiced in different areas. This method serves to provide quantitative basis for decision making of agriculture products distribution enterprises.

Key words: Fuzzy cluster analysis, agricultural products, logistic node

---

INTRODUCTION

Nowadays, the consumer's demands for agriculture products have changed quickly for the time being. However, the logistic cost of agriculture products is relatively high because of lack of distribution node and high empty-loaded ratio and low efficiency of backhaul of agriculture products in China. Besides, most of the logistic enterprises of agriculture products in China are weak and small, and only function to transport or transfer agriculture products in the supply chain. Thus, the operational ability of the circulation industries of agriculture products can be improved gradually by optimizing the distribution of agriculture product, and supply chain can be managed effectively to meet customers'' various requirements. This paper aims to analyze the positioning of the logistic distribution node with fuzzy cluster analysis method.

ANALYSIS OF FUZZY CLUSTER THEORY

Fuzzy cluster analysis method refers to the method of identifying the group based on related membership relation with building fuzzy matrix by analyzing attribute of the subject. It can be explained by a characteristic function:

$$A(x)=\begin{cases}1 & x\in A\\ 0 & x\notin A\end{cases} \quad (1)$$

It is inaccurate to use set A to represent the result of distribution or profits of agriculture products circulation enterprises. Therefore, we need to apply fuzzy mathematic method to substitute A in the interval $[0,1]$, and get a value, which is a membership degree, to measure it. Thus, $0 \le A(x) \le 1$, and $A(x)$ is the membership function. It is clear that the concept of characteristic function is widened by the creation of the membership function. The characteristic function describes the space correlation between elements, and the membership degree function further explains low degree of apace correlation. Fuzzy cluster analysis method can be used to group elements in X with the analysis of fuzzy correlations between elements in categorical set X. The first step of fuzzy cluster analysis is to standardize selected indicators and sample data. The selection of categorical indicators determines the quality of categorization, which means statistical indicators, should have high identifiability and, actual meaning and be representative with generality. Selected indicators cannot be used directly. Each element data should be standardized beforehand for further comparison and analysis. The method is as following: to compress data of selected sample $x_1, x_2, \cdots, x_n$ to the closed interval $[0,1]$, we need to use m sample indicators $y_1, y_2, \cdots, y_m$, and x_{ij} represents m indicators of n samples. The standard deviation and mean of the indicator j of n samples are:

$$\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij} \quad (2)$$

And

$$S_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2} \quad (3)$$

Meanwhile, the original sample data can be standardized with the formula:

$$x_{ij}^{`} = \frac{x_{ij} - \bar{x}_j}{S_j} \quad (4)$$

and with formula of maxima standardization:

$$xij = \frac{x_{ij}^{`} - x_{j\min}^{`}}{x_{j\max}^{`} - x_{j\min}^{`}} \quad (5)$$

The standardized data in formula（4）can be compressed to the closed interval $[0,1]$. Following the previous steps, fuzzy similar matrix calibration can be achieved. First, r_{ij}, which represents the similar degree of two samples x_i and x_j, is used to form the fuzzy similar matrix $\underset{\sim}{R} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & \cdots & r_{nm} \end{bmatrix}$, in which r_{ij} represents the variable of similar degree of x_i and x_j. When $r_{ij} \infty 1$, which means the similar degree of x_i and x_j is high, correlation coefficient method, Euclidean distance method and included angle cosine method can be used to determine the value of r_{ij}. Formula of correlation coefficient method is:

$$r_{ij} = \frac{\sum_{k=1}^{m}(x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{\sqrt{\sum_{k=1}^{m}(x_{ik} - \bar{x}_i)^2}\sqrt{\sum_{k=1}^{m}(x_{jk} - \bar{x}_j)^2}} \quad (6)$$

And

$$\bar{x}_i = \frac{1}{m}\sum_{k=1}^{m} x_{ik} \quad (7)$$

$$\bar{x}_j = \frac{1}{m}\sum_{k=1}^{m} x_{jk} \quad (8)$$

Euclidean distance method used the formula (9):

$$d_{ij} = \sqrt{\sum_{k=1}^{m}(x_{ik} - x_{jk})^2} \quad (9)$$

And

$$r_{ij} = 1 - Cd_{ij} \quad (10)$$

In formula (10), C should be given a proper positive value to ensure r_{ij} lies in the interval $[0,1]$. The included angle cosine formula is

$$r_{ij} = \frac{\sum_{k=1}^{m} x_{ik} x_{jk}}{\sqrt{\sum_{k=1}^{m} x_{jk}^2}\sqrt{\sum_{k=1}^{m} x_{ik}^2}} \quad (11)$$

With these three methods, $\underset{\sim}{R}$, as the fuzzy relation degree, only has the characteristics of symmetry and reflexivity, lacking transitivity. Thus, fuzzy similar matrix $\underset{\sim}{R}$ should be transformed to equivalent matrix with Lemma: suppose $\underset{\sim}{R}$ is a fuzzy similar matrix of A, when $|A| = n$, $k \le n$ exists to enable $\underset{\sim}{R}$ become the fuzzy equivalent matrix with the transitive closure $t(\underset{\sim}{R}) = \underset{\sim}{R}$, in which $|A|$ is the element number in the scope A.

$$r_{ij} = \frac{\sum_{k=1}^{m} x_{ik} x_{jk}}{\sqrt{\sum_{k=1}^{m} x_{jk}^2}\sqrt{\sum_{k=1}^{m} x_{ik}^2}} \quad (12)$$

Formula (12) is used with order and in limited times to get the value of k, and fuzzy equivalent matrix. Thus, we can get $t(\underset{\sim}{R})$ as fuzzy equivalent relation to give different values to λ and calculate the horizontal cut set with distinct objects. Fuzzy equivalent relation elements, with values larger than λ, need to be categorized to form a dynamic clustering. Suppose the sample space $X = \{x_1, x_2, \cdots, x_n\}$ with n samples and m characteristics of the categorization, $x_j = \{x_{j1}, x_{j2}, \cdots, x_{jm}\}$, x_{jk} is the number k characteristic of xj. Suppose the class number of λ is r, the sample size of i is n_i, and $x_1^i, x_2^i, \cdots, x_{ni}^i$ belong to cluster i, vector $\bar{x}^i = \{\bar{x}_1^i, \bar{x}_2^i, \cdots, \bar{x}_{ni}^i\}$ is the central vector of cluster i. The corresponding $\bar{x}_k^i$ represents the characteristic mean of number k in cluster i $\bar{x}_k^i = \frac{1}{n_i}\sum_{j=1}^{n} x_{jk}, k = 1,2,\cdots,m$. $\bar{x} = \{\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_m\}$ represents the vector of all sample, and $\bar{x}_k = \frac{1}{n_i}\sum_{j=1}^{n} x_{jk}, k = 1,2,\cdots,m$. Thus statistics can be calculated as:

$$F = \frac{\sum_{i=1}^{r} n_i \left\| \bar{x}^i - \bar{x} \right\|^2 \Big/ (r-1)}{\sum_{i=1}^{r}\sum_{j=1}^{n_i} \left\| x_j^i - \bar{x}^i \right\|^2 \Big/ (n-1)} \quad (13)$$

and the distance between clusters use statistics F to analyze characterization, in which $\left\|\overline{x}^i - \overline{x}\right\| = \sqrt{\sum_{k=1}^{ni}(\overline{x}_k^i - \overline{x}_k)^2}$ is the distance between $\overline{x}^i$ and $\overline{x}$ and $\left\|x_j^i - \overline{x}^i\right\|$ is the distance between x_j^i and $\overline{x}^i$. We can conclude that the more reasonable the grouping is, the larger the value of F will be, which means the value of grouping level λ corresponding to F is the best.

ANALYSIS OF THE APPLICATION OF FUZZY CLUSTERING IN AGRICULTURE PRODUCTS DISTRIBUTION NODE

An agriculture products circulation industry is selected in this paper to analyze the application of the fuzzy clustering method proposed. For this industry, there are eight logistic distribution nodes, and the distribution scheme is analyzed in this part. It is a large professional agriculture products circulation industry specialized in cultivating, processing, purchasing and distributing agriculture and sideline products. This company is professional in providing full service of vegetables, fruit, seafood, meat, fowl, dry condiment and others to airlines, schools, hospitals, mines, governments, companies and middle and high class hotels. It also advocates developing production base of pollution-free agriculture products and strives to achieve standardized cultivation with obeying the standard of food production importing industrial management concept. It acts actively to become the leading company processing pollution-free agriculture products with thousands acres of cultivation land and hundreds of professional companies in cultivating agriculture products. It aims to achieve produce in land, process at once and distribute in time, which satisfies customers with low price but high quality. The production mode is “product→company→customer” and minimized the circulation nodes between production and consumption, which saves the cost of customers and provides all kinds of products from first-processed product to semi-finished one.

This company controls the cost and quality from the origin for its long-time cooperation with long-term suppliers, and ensures to provide dry goods, oil, fruits and seafood with large quantity, good quality and low price. Suppliers are all qualified and can provide inspection report and approved certificate. At the same time, it has complete testing method and equipment and quality control system. Besides the base and suppliers, two levels of inspecting are carries out to ensure the quality. One is the self-inspection of the company, and the other is the final inspection before the products released to the market by professional skilled personnel. Upholding the concept of “service and customers come first”, this company provides professional and comprehensive service. Feedbacks can be sent to the company since the personnel and positions are settled, and managers can arrange the positions to improve service quality. Now, the company has hundreds of employees, and the education level of the people in charge are high, with 90% graduated from junior and 10% got higher education levels. Employees are trained regularly as well to enhance their professional ability and cooperation ability. It has over ten large transporting vehicles and all drivers as well as back-up personnel have the qualification to drive large vehicles, which functions to help the regular trading and distribution.

1 Construction of the Statistical Indicator System of Agriculture Products Distribution Nodes

The characteristics of statistical indicators of agriculture products distribution nodes, including systematic, scientific, concise, quantitative, qualitative and independent, are combined with the situation in operation to build the statistical indicator system of agriculture products distribution nodes (See Table 1).

Tab.1: Statistical Indicator System of Agriculture Products Distribution Nodes

Customer’s Information	Difficulty	Distance between Supplier and Demander A
		Road Condition B
		Limitation Degree of Time Windows C
	Strength	Funding D
		Trading Volume E
		Management Level F
		Network Coverage G
	Others	Loyalty H
		Potential Development Rate I

2 Data Collection

Data collection is the key step to get the statistics of agriculture products distribution nodes. It needs professional agents or specialized departments of agriculture products circulation industry to collect data. In the data, qualitative data that can give results is collected by circulation industry, while quantitative data that can reflect the feature of agriculture products circulation industry are calculated. According to the statistical indicator system in Table 1, there are five quantitative indicators, namely distance between the supplies and demander A which is the distance between two points, limit degree of time windows C, which is a mean and need to collect the trading time of agriculture products distribution nodes, flowing capital D which is the of agriculture products distribution nodes, everyday trading volume E and network coverage area G of agriculture products distribution nodes. The data of the five indicators can be collected from the company, but the other qualitative indicators cannot be calculated. They are investigated by experts(See Table 2).

Tab. 2: Original Data of Statistical Indicator System of Agriculture Products Distribution Nodes

	A	B	C	D	E	F	G	H	I
	Km	1-10 Score	Hour	RMB(10,000)	Ton	1-10 Score	M^2	1-10 Score	Percent
1	3506.41	5	1.5	50	5.5	8	200	7	90
2	1054.38	9	1	60	8.5	5	43.5	9	50
3	3854.87	7	1	100	7.8	9	322	4	60
4	1504.16	6	1	150	4	10	580	10	100
5	1648.91	10	2	70	5	6	429.2	8	120
6	5203.99	8	1	80	3	7	392	6	40
7	1700.37	8	1	30	2.5	6	400	7	80
8	3347.13	3	1	50	2	5	278.3	7	80

3 Clustering Analysis of Agriculture Products Distribution Nodes

First of all, to standardize data and solve the problem of different measurement of statistical indicators of agriculture products distribution nodes, we can use the standard deviation method. The formula is $SD = \sqrt{\frac{\sum X}{N} - (\frac{\sum X}{N})^2}$, in which SD is standard deviation, X is the original data and N represents the total number of data. Standardized value of the original data is as following(See Table 3).

Tab. 3: Standardized Value of Statistical Indicator of Agriculture Products Distribution Nodes

	A	B	C	D	E	F	G	H	I
1	-0.13691	-0.88192	0.84003	-0.63515	0.29721	0.54006	-0.80833	-0.13644	0.47006
2	2.22873	0.88192	-0.50402	-0.36772	1.54863	-1.08012	-1.77677	0.95511	-1.03414
3	-0.01978	0.00000	-0.50402	0.70201	1.25663	1.08012	-0.05337	-1.77378	-0.65809
4	-0.80998	-0.44096	-0.50402	2.03917	-0.32850	1.62019	1.54317	1.50089	0.84611
5	-0.76132	1.32288	2.18407	-0.10029	0.08864	-0.54006	0.61000	0.40933	1.59822
6	0.43373	0.44096	-0.50402	0.16715	-0.74564	0.00000	0.37980	-0.68222	-1.41019
7	0.74402	0.44096	-0.50402	-1.17001	-0.95421	-0.54006	0.42930	-0.13644	0.09402
8	-0.19046	-1.76383	-0.50402	-0.63515	-1.16278	-1.08012	-0.32380	-0.13644	0.09402

After the standardization, we can get the Euclidean Matrix:

$$D = \begin{bmatrix} 0.54842 & 0.76155 & 0.54023 & 0.71409 & 0.73733 & 0.79700 & 0.81813 \\ 0.76155 & 0.47492 & 0 & 0.25133 & 0.58523 & 0.44491 & 0.41454 \\ 0.54023 & 0 & 0.50654 & 0.42368 & 0.81608 & 0.64154 & 0.55785 \\ 0.71409 & 0.25133 & 0.42368 & 0.47264 & 0.53264 & 0.51407 & 0.434460 \\ 0.73733 & 0.58523 & 0.81608 & 0.53264 & 0.51031 & 0.69777 & 0.46626 \\ 0.79700 & 0.44491 & 0.64154 & 0.51407 & 0.69777 & 0.85692 & 0.46626 \\ 0.81813 & 0.41454 & 0.55785 & 0.43446 & 0.46626 & 0.46626 & 0.85026 \end{bmatrix} \tag{14}$$

Thus the fuzzy distance is

$$\underset{\sim}{R} = \begin{bmatrix} 1 & 0.64 & 0.89 & 0.63 & 0.83 & 0.86 & 0.93 & 0.95 \\ 0.64 & 1 & 0.55 & 0 & 0.29 & 0.68 & 0.52 & 0.48 \\ 0.89 & 0.55 & 1 & 0.59 & 0.49 & 0.95 & 0.75 & 0.65 \\ 0.63 & 0 & 0.59 & 1 & 0.55 & 0.62 & 0.6 & 0.51 \\ 0.83 & 0.29 & 0.49 & 0.55 & 1 & 0.6 & 0.81 & 0.54 \\ 0.86 & 0.68 & 0.95 & 0.62 & 0.6 & 1 & 1 & 0.88 \\ 0.93 & 0.52 & 0.75 & 0.6 & 0.81 & 1 & 1 & 0.99 \\ 0.95 & 0.48 & 0.65 & 0.51 & 0.54 & 0.88 & 0.99 & 1 \end{bmatrix} \tag{15}$$

and fuzzy equivalent matrix is

$$t(\underset{\sim}{R}) = \begin{bmatrix} 1 & 0.68 & 0.95 & 0.63 & 0.83 & 0.95 & 0.95 & 0.95 \\ 0.68 & 1 & 0.68 & 0.63 & 0.68 & 0.68 & 0.68 & 0.68 \\ 0.95 & 0.68 & 1 & 0.63 & 0.83 & 0.95 & 0.95 & 0.95 \\ 0.63 & 0.63 & 0.63 & 1 & 0.63 & 0.63 & 0.63 & 0.63 \\ 0.83 & 0.68 & 0.83 & 0.63 & 1 & 0.83 & 0.83 & 0.83 \\ 0.95 & 0.68 & 0.95 & 0.63 & 0.83 & 1 & 1 & 0.99 \\ 0.95 & 0.68 & 0.95 & 0.63 & 0.83 & 1 & 1 & 0.99 \\ 0.95 & 0.68 & 0.95 & 0.63 & 0.83 & 0.99 & 0.99 & 1 \end{bmatrix} \quad (16)$$

which means the different horizontal correlation condition is

$$\begin{aligned} & I=6 \quad J=8 \quad Mx=0.99223 \\ & I=1 \quad J=6 \quad Mx=0.95473 \\ & I=1 \quad J=7 \quad Mx=0.95473 \\ : & I=1 \quad J=3 \quad Mx=0.95234 \\ & I=1 \quad J=5 \quad Mx=0.83332 \\ & I=1 \quad J=2 \quad Mx=0.68294 \\ & I=1 \quad J=4 \quad Mx=0.63043 \end{aligned} \quad (17)$$

Sure, if λ is given different value, the grouping result will be different (See Table 4).

Tab.4: Clustering Results

Value	Cluster Number	Result
0.63	1	（1,2,3,4,5,6,7,8）
0.68	3	（2）、（4）、（1,3,5,6,7,8）
0.83	4	（2）、（4）、（5）（1,3,6,7,8）
0.95	5	（2）、（3）、（4）、（5）（1,6,7,8）
0.99	7	（1）、（2）,（3）、（4）、（5）、（7）（6,8）

We can see from the calculation above that λ lies in the interval $[0,1]$ as it is shown in Table 3 with different corresponding groupings, and forms the dynamic clustering. However, in reality we usually need one grouping method, which means we have to calculate the value of F (specific steps are skipped). When $\lambda = 0.68$, $F = 10.31$ which is the maximum, and the grouping result is the best.

CONCLUSION

From the previous calculation, we can conclude that the result of the clustering depends on the value of λ in the closed interval $[0,1]$. If it has to be grouped precisely, larger λ value is needed, which means the precision of grouping is influenced by λ. Similarly, with the results from running situation of agriculture products circulation industry and clustering analysis, when $\lambda = 0.68$, $F = 10.31$ which is the maximum, therefore the range is selected as $0.68 \le \lambda \le 0.83$ to proceed the clustering and divide distribution nodes into three areas, (2), (4) and (1,3,5,6,7,8) to reflect the similarity and difference of distribution nodes. Considering the indicators in Table 1, the better reasonable economic area is （1,3,5,6,7,8）and（4）for the distribution distance is suitable, while area（2）is a non-economic area with low development potential because the distance is relatively long and the cost is too high. To conclude, the traditional administrative division is different from the geography division. The area division of agriculture products distribution nodes using fuzzy clustering method is more reasonable and can be improved with situations of the agriculture products circulation industry. However, the traditional method is simplistic in analyzing the factors and can only group roughly without considering other factors influencing the routine and method. The fuzzy clustering analysis method has comprehensively taken into consideration of the factors that may influence these aspects in trading, and provides a way to organize the whole distribution network. Thus, companies can outsource the non-economic distribution area and cut the distribution cost to have a larger network.

REFERENCES

[1] Li Kai; *Study on the Construction of a Agricultural Logistic System in Liaoning Province*[D]; Shenyang

University of Technology; **2010**.
[2] Zhang Zhanyi; *Optimization and Simulation of Order Picking in Agricultural Products Logistics Center Based on eM-Plant* [D]; Jilin University; **2007**.
[3] Liu Xiaofei; *Xinjiang Shihezi Agricultural Logistics Development Pattern Reclamation* [D]; Shihezi University; **2011**.
[4] Yu Xiaodong; *Study on Logistic Distribution Model and Equipment Strategy of Agricultural Products* [D]; Inner Mongolia University of Technology; **2010**.
[5] Zhang Tao, Wang Mengguang; *Solving the VRP by a Hybrid Algorithm of Genetic Algorithm and 3-opt Algorithm* [J]; Journal of Northeastern University (Natural Science); **1999** (03).
[6] DuanMeihua; *Problems in Constructing Chongqing Agricultural Products Logistics Center* [D]; Southwest University; **2011**.
[7] Zhou Peide, Zhou Zhongping, Zhang Huan; *Journal of Beijing Institute of Technology*; **2000** (02).
[8] Paskaleva K. *International Journal of Innovation and Regional Development*, v.1, n.4, pp.405-422, 25 January, **2009**.
[9] Arora A..*Traffic Engineering and Control*, v 53, n 10, p 375-378, November **2012**.
[10] B. Chen and H. H. Cheng. *IEEE Trans. Intelligent Transportation Systems*, vol.11, no.2, pp.485–497, **2010**.