A Novel Application of Fuzzy Statistical Methods in Economic Decisions Making

Yue Zhang*

Shenyang Li Gong university
Shenyang, China

Abstract — To analyze the agricultural economy of L province and improve the competitiveness of regional economy, an efficient way is needed to categorize the cities in L province in order to make specific economic policies. After introducing the basic theories of regional agricultural development, this paper uses the weighted fuzzy c-means clustering algorithm to analyze the economy data of L province. Choosing of 9 indicators of L province's agricultural economy, the fuzzy c-means clustering algorithm divided 14 cities into 5 clusters. The results of the clustering analysis provides a foundation to the further study of disparities of economic development and adopted policies in different clusters in assisting the economic decisions of the government.

Keywords — weighted fuzzy c-means clustering; regional agricultural economy; fuzzy analysis

I. INTRODUCTION

With the development of global economic integration, China's agricultural industry is facing unprecedented impact and challenges. The current agricultural production infrastructure is still weak, and the lack of capital and technology investment, the low level of productivity, suboptimal design of industrial structure, the poor quality of agricultural products, low economic, make it important to enhance the domestic and international competitiveness of agricultural development [1]. In fact, considering the status of the public service and strategic position, the agricultural industry has always been the focal point of the Chinese government. Since 2004, the "three rural issues" has been the center of debate for nine consecutive years. It is an important task to enhance the domestic and international competition strength of various regions, and to consolidate the achievements of agricultural development. The development of modern agriculture, efforts to improve the comprehensive agricultural production capacity, and enhance the competitiveness of modern agricultural industry is the focus of the government policies [2]. In addition, in 2012 the government document, the agricultural science and technology has been the major theme for the first time, including the emphasis on technology and driven innovation, leading the development of modern agriculture, agricultural production, methods to increase farmers' income [3].

An important category of fuzzy statistical analysis is c-means clustering, which is known as non supervised classification [4]. It is based on the internal structure of the sample set. The sample set is divided into a series of subsets. The sample set is divided into a series of subsets. The sample sets are classified into different categories. The traditional clustering method is bipartition. But with the fact that the nature of the object is hard to defines as simple black and white. The fuzzy clustering analysis method can accurately describe the sample and all kinds of classes [5]. The uncertainty relationship between the two can be more objective to reflect the actual situation, as the fuzzy c-means clustering method become the mainstream of clustering analysis.

So far, many fuzzy clustering algorithms have been proposed, among which the fuzzy c-means clustering algorithm based on B standard function is the most perfect and applied method [6]. In practical application, it has many factors that affect a given sample set, and these factors are different in all situations.

II. REGIONAL AGRICULTURAL DEVELOPMENT THEORIES

Agricultural competitive power is the comparison of the efficiency of agricultural production in a certain market economy environment, and the specific performance in the unit of agricultural production efficiency, scale efficiency and resource costs, in order to ensure the sustainable and stable development of agriculture, to make sure the profit continuing increase [7]. Provincial agricultural competitiveness belongs to the category of regional agricultural competitiveness and the theoretical foundation of its formation is the agricultural location theory, resource endowment theory [8]. The regional agricultural competitiveness is the important factor of agricultural production structure and regional agricultural production. The important factors of agricultural production structure difference between provinces includes the innovation of agricultural science and technology and the active level of agricultural science and technology [9].

The development and competitiveness of agricultural development and the competitiveness of the province is determined by many factors, which can achieve the effective integration of the various factors in order to maximize the development of regional agriculture [10]. Specifically, the definition of the agricultural competitiveness of the province can be summarized as the following four aspects:

1) Integrated agricultural production capacity. The agricultural comprehensive production capacity is the premise of the sustainable development of agriculture and the improvement of the efficiency of agriculture. Only by improving the comprehensive production capacity of
agriculture and it can we realize the function the rural areas to enhance the agricultural efficiency and increasing farmers’ income.

(2) The degree of agricultural marketization, namely, the competition of the market of the province's agriculture. Modern agriculture is a market oriented agriculture, which is facing the dual pressures of the domestic market and the international market. On the one hand, the development of agriculture in the province should face the competition of the domestic related industries, on the other hand, based on the international background of the trend of economic globalization, it should actively integrate into the world agricultural market. Therefore, the development of the province's agriculture should make full use of the two markets and two kinds of resources, and promote the sustainable development of agriculture in the region [11].

(3) Agricultural sustainable development ability. Agricultural sustainable development depends on the support of agricultural science and technology. The potential power of the long-term development of agriculture is an important indicator of the evaluation of agricultural competitiveness, but also an important source of improving the competitiveness of the industry. The sustainable development of agriculture in the province is based on the background of agricultural marketization and internationalization, and the sustainable development of agriculture promotes the improvement of the competitiveness of agriculture in the province. At the same time, the sustainable development of agriculture needs the support of science and technology, and the development of the sustainable and stable development of agriculture depends on science and technology investment and innovation.

(4) Long-term agricultural efficiency. The purpose of any industrial development is to profit, and then promote the expansion of the industry, agricultural production is no exception. As our country is gradually entering the stage of expansion of the industry, agricultural production is no exception. As our country is gradually entering the stage of expansion of the industry, agricultural production is no exception. As our country is gradually entering the stage of expansion of the industry, agricultural production is no exception. As our country is gradually entering the stage of expansion of the industry, agricultural production is no exception. As our country is gradually entering the stage of expansion of the industry, agricultural production is no exception. As our country is gradually entering the stage of expansion of the industry, agricultural production is no exception. As our country is gradually entering the stage of expansion of the industry, agricultural production is no exception.

III. WEIGHTED FUZZY C-MEANS CLUSTERING ALGORITHM

A. Data set’s fuzzy c partition

Set up classification data set \( X = \{x_1, x_2, \ldots, x_n\} \) as a limited sample set, \( x_k = (x_{k1}, x_{k2}, \ldots, x_{kn}) \) as the index vector of sample \( x_k (k = 1, 2, \ldots, n) \), \( x_{ik} (l = 1, 2, \ldots, r) \) is the index value of dimension \( l \) in sample \( x_k \). To conduct the fuzzy clustering analysis, the fuzzy partition \( c \) of \( X \) must satisfy:

\[
X_1 \cup X_2 \cup \ldots \cup X_c = X
\]

\[
X_i \cap X_j \neq \Phi, 1 \leq i \neq j < c
\]

\[
X_i \neq \Phi, X_i \neq X, 1 \leq i \leq c
\]

Function \( u_{ik} = u_{X_i}(x_k) \) represents the relationship of sample \( x_k \) and fuzzy set \( X_i (1 \leq i \leq c) \) , \( u_{ik} \) is the subordinate relationship of fuzzy c partition’s fuzzy set \( X_i \), thus \( u_{ik} \in [0,1] \). The fuzzy partition \( c \) is represented by the subordinate relationship matrix \( U_f = (u_{ik})_{c \times n} \), which also called fuzzy partition matrix. Therefore, the space of fuzzy partition \( c \) in sample \( X \) is:

\[
M_f = \{U_f \in R^{cm} | u_{ik} \in [0,1], \forall i, k; \sum_{j=1}^{c} u_{ik} = 1, \forall k; 0 < \sum_{k=1}^{c} u_{ik} < n, \forall i \} \quad (4)
\]

B. Weights determination

In weighted fuzzy c-means clustering algorithm, the weight value is a core parameter. The methods of determining the weight value include coefficient method, entropy weight method, Gauss function weighting method and principal component analysis. This paper uses the principal component analysis method to determine the weights, and study has verified by the principal component analysis method’s validity in c-means clustering algorithm [13].

Assume the principal component matrix of the sample is:

\[
U = \begin{bmatrix}
u_{i1} & u_{i2} & \cdots & u_{ir} \\
u_{i1} & u_{i2} & \cdots & u_{ir} \\
\vdots & \vdots & \ddots & \vdots \\
u_{i1} & u_{i2} & \cdots & u_{ir}
\end{bmatrix} \quad (5)
\]

In (5), \( r \) is the index number.

\( \lambda = (\lambda_1, \lambda_2, \ldots, \lambda_r) \) is the row vector composed by \( r \) characteristic roots, \( \lambda(I=1, 2, \ldots, r) \) is the characteristic root of \( u_i \), and \( \lambda_1 > \lambda_2 > \cdots > \lambda_r \).

\( a = (a_1, a_2, \ldots, a_r) \) is square contribution vector, then:

\[
a_j = \frac{\lambda_j}{\sum_{i=1}^{r} \lambda_i} \quad (6)
\]

Index coefficient vector \( b = (b_1, b_2, \cdots, b_r) \) is a set composed by corresponding index of \( l \), then:

\[
b = U \ast a \quad (7)
\]

After standardization of index vector \( b \), the weights vector \( W = (w_1, w_2, \ldots, w_r) \), then:

\[
w_j = \frac{|b_j|}{\sum_{k=1}^{r} |b_k|} \quad (8)
\]

C. Weighted fuzzy c-means clustering algorithm

The weighted fuzzy c-means clustering algorithm proposed in this paper is based on the traditional c-means clustering algorithm, the Euclidean distance in target function is defined by weighted Euclidean distance, which satisfy various conditions in the process of fuzzy clustering analysis. Therefore, the basic idea behind the weighted fuzzy c-means clustering algorithm is same as the traditional fuzzy c-means clustering algorithm.
Set target function of weighted fuzzy c-means clustering algorithm as:
\[ J^*_w(U^*_f, P^*) = \sum_{k=1}^{n} \sum_{i=1}^{w} (u^*_{ik})^m (d^*_{ik})^2, m \in [1, \infty) \] (9)

In (9):
\[ (d^*_{ik})^2 = \|x_i - p^*_k\|^2 = (x_i - p^*_k)^T A (x_i - p^*_k) \] (10)
\[ A = \begin{bmatrix} w_1 & & \\ & \ddots & \\ & & w_n \end{bmatrix} \]
\[ P^* = (p^*_1, p^*_2, \ldots, p^*_c) \] (11)

The minimum value of the objective function is \( \min \{ J^*_w(U^*_f, P^*) \} \), and each row of division matrix \( U^*_f \) is independent, therefore:
\[ \min \{ J^*_w(U^*_f, P^*) \} = \min \{ \sum_{i=1}^{n} (u^*_{ik})^m (d^*_{ik})^2 \} = \sum_{i=1}^{n} \min \{ \sum_{k=1}^{c} (u^*_{ik})^m (d^*_{ik})^2 \} \] (12)

The constraining condition is \( \sum_{i=1}^{c} u^*_{ik} = 1 \), with Lagrange multiplication, the solution is:
\[ F = \sum_{i=1}^{n} (u^*_{ik})^m (d^*_{ik})^2 + \lambda \left( 1 - \sum_{i=1}^{c} u^*_{ik} \right) \] (13)

The optimal first order necessary condition is:
\[ \frac{\partial F}{\partial u^*_{ik}} = \left( 1 - \sum_{i=1}^{c} u^*_{ik} \right) \lambda m (u^*_{ik})^{m-1} (d^*_{ik})^2 = 0 \] (14)

From (13) and (14), there is:
\[ u^*_{ik} = \frac{1}{\lambda m} \sum_{i=1}^{c} \left( \frac{d^*_{ik}}{(d^*_{ik})^2} \right) \frac{1}{\lambda m} \] (15)

When \( d^*_{ik} = 0 \), then \( u^*_{ik} = 1 \).

In the same way, \( J^*_w(U^*_f, P^*) \) is the minimum value of \( p^*_i \), the partial derivative of (9) is:
\[ \frac{\partial}{\partial p^*_i} (x_i - p^*_k)^T A (x_i - p^*_k) = 0 \] (16)

The solution of (16) is:
\[ p^*_i = \frac{\sum_{k=1}^{n} (u^*_{ik})^m x_k}{\sum_{k=1}^{n} (u^*_{ik})^m} \] (17)

The whole process of weighted fuzzy c-means clustering algorithm is the iterative operations of (15) and (17), until the terminal condition of the iterative is met to obtaining the results of fuzzy c-means clustering analysis.

IV. RESULTS

Based on the L Province’s agricultural economy development in actual situation, this paper chose 9 important figures to representing the agricultural economy after screening as the indexes of the fuzzy c-means clustering analysis:

X1: rural per caplital cultivated land area (hectare/person)
X2: grain yield per unit area (kg/hectare)
X3: the increase of the index value of forest and pasture ( pant)area accounts for the proportion of the area of the crops sown (%)
X4: vegetable planting area accounts for the proportion of the area of the crop sown area (%)
X5: effective irrigation area accounts for the proportion of cultivated land area (%)
X6: total power of agricultural machinery (KW-hr)
X7: labor productivity
X8: the increase of the index value of forest and pasture (pant)area accounts for the proportion of cultivated land area (%)
X9: farmers’ capita net income (yuan)

The relevant indicators of agricultural economy in L province are shown in table 1.

<table>
<thead>
<tr>
<th>City</th>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>X4</th>
<th>X5</th>
<th>X6</th>
<th>X7</th>
<th>X8</th>
<th>X9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lanzhou</td>
<td>0.1285</td>
<td>3017</td>
<td>61.09</td>
<td>22.86</td>
<td>0.3970</td>
<td>151</td>
<td>13843</td>
<td>105.01</td>
<td>4588</td>
</tr>
<tr>
<td>Jiaju</td>
<td>0.1293</td>
<td>8661</td>
<td>22.35</td>
<td>10.07</td>
<td>1.0000</td>
<td>12</td>
<td>39559</td>
<td>107.13</td>
<td>7865</td>
</tr>
<tr>
<td>Jinchang</td>
<td>0.2478</td>
<td>6604</td>
<td>69.38</td>
<td>13.94</td>
<td>0.8920</td>
<td>89</td>
<td>21464</td>
<td>104.75</td>
<td>5935</td>
</tr>
<tr>
<td>Baiyin</td>
<td>0.2422</td>
<td>2078</td>
<td>80.87</td>
<td>5.74</td>
<td>0.3102</td>
<td>175</td>
<td>11993</td>
<td>106.02</td>
<td>3387</td>
</tr>
<tr>
<td>Tianshui</td>
<td>0.1932</td>
<td>3048</td>
<td>71.39</td>
<td>11.64</td>
<td>0.0059</td>
<td>116</td>
<td>9760</td>
<td>107.50</td>
<td>2825</td>
</tr>
<tr>
<td>Wuwei</td>
<td>0.1016</td>
<td>6723</td>
<td>65.18</td>
<td>14.76</td>
<td>0.7112</td>
<td>342</td>
<td>17354</td>
<td>106.13</td>
<td>4451</td>
</tr>
<tr>
<td>Zhangwei</td>
<td>0.2533</td>
<td>6729</td>
<td>69.28</td>
<td>7.39</td>
<td>0.6879</td>
<td>200</td>
<td>79025</td>
<td>106.00</td>
<td>5575</td>
</tr>
<tr>
<td>Pingliang</td>
<td>0.1692</td>
<td>2019</td>
<td>72.19</td>
<td>10.53</td>
<td>0.1191</td>
<td>94</td>
<td>13621</td>
<td>107.21</td>
<td>3137</td>
</tr>
<tr>
<td>Jiuquan</td>
<td>0.2287</td>
<td>7547</td>
<td>28.64</td>
<td>14.99</td>
<td>0.9890</td>
<td>194</td>
<td>41720</td>
<td>104.23</td>
<td>7234</td>
</tr>
<tr>
<td>Qingyang</td>
<td>0.1437</td>
<td>2759</td>
<td>68.76</td>
<td>12.49</td>
<td>0.1040</td>
<td>311</td>
<td>12661</td>
<td>105.75</td>
<td>3254</td>
</tr>
<tr>
<td>Dingxi</td>
<td>0.1792</td>
<td>2477</td>
<td>72.91</td>
<td>4.02</td>
<td>0.1810</td>
<td>211</td>
<td>8465</td>
<td>105.75</td>
<td>2602</td>
</tr>
<tr>
<td>Longnan</td>
<td>0.1071</td>
<td>3013</td>
<td>75.96</td>
<td>7.49</td>
<td>0.2170</td>
<td>153</td>
<td>7656</td>
<td>102.05</td>
<td>2130</td>
</tr>
<tr>
<td>Linfu</td>
<td>0.0138</td>
<td>4923</td>
<td>79.46</td>
<td>4.79</td>
<td>0.3921</td>
<td>75</td>
<td>6191</td>
<td>106.26</td>
<td>2375</td>
</tr>
<tr>
<td>Gannan</td>
<td>0.1028</td>
<td>2493</td>
<td>51.24</td>
<td>1.07</td>
<td>0.0926</td>
<td>33</td>
<td>8508</td>
<td>107.53</td>
<td>2689</td>
</tr>
</tbody>
</table>
When the fuzzy clustering number c=5, the fuzzy partition matrix is shown in Table 2.

<table>
<thead>
<tr>
<th></th>
<th>0.9345</th>
<th>0.1701</th>
<th>0.1463</th>
<th>0.0803</th>
<th>0.0434</th>
<th>0.2376</th>
<th>0.0253</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0156</td>
<td>0.1994</td>
<td>0.5195</td>
<td>0.0538</td>
<td>0.0204</td>
<td>0.3246</td>
<td>0.9074</td>
<td></td>
</tr>
<tr>
<td>0.0071</td>
<td>0.4011</td>
<td>0.1589</td>
<td>0.0157</td>
<td>0.0091</td>
<td>0.1171</td>
<td>0.0268</td>
<td></td>
</tr>
<tr>
<td>0.0173</td>
<td>0.1257</td>
<td>0.0709</td>
<td>0.0828</td>
<td>0.8632</td>
<td>0.1286</td>
<td>0.0161</td>
<td></td>
</tr>
<tr>
<td>0.0252</td>
<td>0.1036</td>
<td>0.1044</td>
<td>0.7674</td>
<td>0.0640</td>
<td>0.1922</td>
<td>0.0244</td>
<td></td>
</tr>
<tr>
<td>0.1234</td>
<td>0.0208</td>
<td>0.2031</td>
<td>0.0338</td>
<td>0.3187</td>
<td>0.1720</td>
<td>0.1048</td>
<td></td>
</tr>
<tr>
<td>0.1234</td>
<td>0.0208</td>
<td>0.2031</td>
<td>0.0338</td>
<td>0.3187</td>
<td>0.1720</td>
<td>0.1048</td>
<td></td>
</tr>
<tr>
<td>0.0663</td>
<td>0.0499</td>
<td>0.0906</td>
<td>0.0167</td>
<td>0.0993</td>
<td>0.0906</td>
<td>0.0600</td>
<td></td>
</tr>
<tr>
<td>0.0252</td>
<td>0.9034</td>
<td>0.0357</td>
<td>0.0060</td>
<td>0.0475</td>
<td>0.0395</td>
<td>0.0333</td>
<td></td>
</tr>
<tr>
<td>0.5204</td>
<td>0.0118</td>
<td>0.1255</td>
<td>0.0408</td>
<td>0.1982</td>
<td>0.4868</td>
<td>0.6621</td>
<td></td>
</tr>
<tr>
<td>0.2647</td>
<td>0.0141</td>
<td>0.5452</td>
<td>0.9027</td>
<td>0.3363</td>
<td>0.2111</td>
<td>0.1398</td>
<td></td>
</tr>
</tbody>
</table>

The whole process of weighted fuzzy c-means clustering algorithm is to use (15) and (17) to carry out the iterative operations of fuzzy partition matrix until the terminal condition of fuzzy clustering is satisfied. The fuzzy cluster dendrogram is shown in Figure 1.

From Figure 1, L Province agricultural economic can be divided into five classes of fuzzy cluster. The first class only includes Lanzhou; second class includes Jinchang, Wuwei, Zhangye; third class includes Jiayuguan Jiuquan; fourth class includes Tianshui, Pingliang, Linxia, Gannan; fifth class includes Baiyin, Qingyang, Dingxi, Longnan.

When the fuzzy clustering number c=6, the fuzzy partition matrix is shown in Table 3. The fuzzy cluster dendrogram is shown in Figure 2.

From Figure 2, L Province agricultural economic can be divided into six classes of fuzzy cluster. The first class includes Longnan, Linxia; second class includes Baiyin, Dingxi, Qingyang; third class includes Lanzhou; fourth class includes Jiayuguan, Jiuquan; fifth class includes Tianshui, Pingliang, Gannan; sixth class includes Jinchang, Wuling, Zhangye.

Figure 3 The Vnew index value with changing of fuzzy clustering number C

From Figure 3, it can be seen that the Vnew index reached the minimum value at the c=5, that is, the optimal fuzzy clustering number is 5. Thus, the agricultural economy of L province is divided into five categories of comparison which close to the actual situation.

V. CONCLUSION

In this paper, a weighted fuzzy c-means clustering algorithm is used to cluster analysis the regional agricultural economy in L province. Through the general analysis of the regional agricultural economy in L Province, combined with the theory of regional agricultural economy, the agricultural economy of 14 cities in L province is divided into 5 categories according to 9 indicators.

The 5 clusters are:

- Lanzhou
- Jinchang, Wuwei, Zhangye
- Jiayuguan Jiuquan
- Tianshui, Pingliang, Linxia, Gannan
Baiyin, Qingyang, Dingxi, Longnan
The clusters of the L province provides a reference for the policies of further development of agricultural in the region. The limitations of this paper is that the algorithm still has the possibility to improve to achieve better simulation to the reality.

CONFLICT OF INTEREST
The author confirms that this article content has no conflict of interest.

REFERENCES