Research on Identification of Consumer Product Quality Safety Factors Based on Improved FCM Algorithm

Yong Kuai¹, Haiyan Wang², Yanfeng Wang³,⁷ and Yingcheng Xu³

¹Wuhu Institute of Standardization, Wuhu, 230001, China
²School of Management and E-Business, Zhejiang Gongshang University, Hangzhou 310018, China
³China National Institute of Standardization, Beijing 100191, China

*Corresponding author e-mail: wangyf@cnis.ac.cn

Abstract. In view of the shortcomings of FCM algorithm, the membership degree and the cluster category number are improved to perfect the FCM algorithm. The performance of the improved algorithm is verified by a case of consumer product quality as a data source, and the results show that with the improved algorithm, both clustering accuracy rate and F value are improved.

Keywords: Algorithm, Consumer product, Quality safety, Factor

1. Introduction
The consumer product quality safety has become a hot issue in the society, which is related to the people's physical and mental health and social harmony and stability. At present, the research in the field of consumer product quality safety in China is still in its initial stage, and the systematic method for identifying and analyzing the impact factors of consumer product quality safety has yet to be established. Based on the improved FCM clustering algorithm, the impact factors of consumer product quality safety can be effectively identified and extracted, and the scientificity and effectiveness of the supervision on consumer product quality safety can be further improved.

2. Another Section of Your Paper
FCM clustering algorithm is a fuzzy clustering algorithm proposed by Dunn and developed by Bezdek. To some extent, the reliability of FCM algorithm on clustering effect and stability depends on the initial setting of parameters. Reasonable selection of the cluster number and the initial value of fuzzy weighted index will affect the clustering results. In addition, FCM algorithm has some shortcomings[1-3]:

(1) FCM algorithm has certain limitations in the reasonable identification of outlier data;
(2) Whether in the FCM algorithm can be reasonably selected will directly affect the clustering results.

In order to overcome the above two shortcomings, this paper will improve the reasonable selection of cluster number and weighted index and the optimization of data identification of outliers in the following two aspects: firstly, FCM algorithm is sensitive to the outlier data, so it is necessary to
modify the membership degree of the data, i.e. modify the membership degree of the outlier data which deviates from the cluster center, so that it will have less impact on the clustering effect in the process of continuous update of the cluster center; secondly, the cluster number c should be reasonably selected. It is necessary to introduce the fuzzy division validity function. By this theory, we can compare the advantages and disadvantages of different cluster numbers, so that the number of categories can be determined automatically.

3. Improved FCM Algorithm

3.1. Weighted Membership Degree

In order to reduce the impact of such outlier $X$ on the clustering effect, an improved membership method is proposed in this paper. The basic idea of this method is as follows[3-6]: For a given dataset $D = \{d_1, d_2, d_3, \ldots, d_n\}$, add a weighted value $\alpha$ to the fuzzy membership degree $u_{ik}$ of the dataset $D$ studied, so as to reduce the impact of the outlier on the cluster center $P_i$ and improve the clustering analysis. The purpose of this method is to increase the impact of data point $d_i$ with high membership degree $u_{ik}$ on the cluster center position $P_i$ by introducing the weighted membership value $\alpha$, and to reduce the impact of data point $d_i$ with small membership degree $u_{ik}$ on the cluster center position $P_i$. The improved formula is shown below:

$$U_{ik}^{(b)} = \frac{1}{1 + \alpha} u_{ik}^2 + \frac{\alpha}{1 + \alpha} u_{ik}, \alpha \in [0, 1]$$

(1)

As can be seen from the above formula, the value range of $U_{ik}^{(b)}$ is $[0, 1]$, and when the membership degree $u_{ik} = 0$, $U_{ik}^{(b)} = 0$; when the membership degree $u_{ik} = 1$, $U_{ik}^{(b)} = 1$. Since the above function is a concave function, we can see that after the improvement of the membership degree in the interval $[0, 1]$, $U_{ik}^{(b)}$ has been reduced to a certain extent compared with the original value of $U_{ik}^{(a)}$ (the smaller the degree of membership, the more obvious the corresponding reduction). It can be clearly seen from the above formula that the impact of data point $d_i$ with small membership degree on the cluster center $P_i$ is reduced. $\alpha$ is the weighting coefficient, which is generally in the range of $[0, 1]$. The larger $\alpha$ is, the smaller the adjustment range of membership degree $u_{ik}$ will be, and the result is relatively accurate; the smaller $\alpha$ is, the greater the adjustment range of membership degree $u_{ik}$ will be, and the result is better. Its advantage lies in speeding up the adjustment of membership degree $u_{ik}$ and improving the efficiency of the whole algorithm.

3.2. Determination of the cluster category number c

Fuzzy division coefficient is a very important factor in determining the cluster category number c. The main disadvantage of fuzzy division coefficient $Z(u_{ik}, c)$ is that with the increase of the cluster category number $c$, the function tends to decrease monotonously. When the value of $c$ is 2, $Z(u_{ik}, c)$ will always get the maximum value, so it is not ideal to determine the best cluster number $c$ by using $Z(u_{ik}, c)$. Meanwhile, the function $Z_i(u_{ik}, c)$ tends to decrease monotonously with the
increase of the cluster category number \( c \), so it is not ideal to take \( Z_1(u_{ik}, c) \) as a clustering validity function alone. Since \( \frac{1}{c} \leq Z(u_{ik}, c) \leq 1 \) and \( \frac{1}{c} \leq Z_1(u_{ik}, c) \leq 1 \), the validity function for determining the cluster category number \( c \) is defined as \( |Z(u_{ik}, c) - Z_1(u_{ik}, c)| \), and hence the following clustering validity function is obtained[7-8].

\[
Z_2(u_{ik}, c) = Z(u_{ik}, c) - Z_1(u_{ik}, c)
\]

(2)

For the "optimal" finite set of \( U \), if \((u^*, c^*)\) satisfies:

\[
Z_2(u^*_{ik}, c^*) = \min_c \left\{ \min_{\alpha_k} Z(u_{ik}, c) \right\}
\]

(3)

Then \( Z_2(u^*_{ik}, c^*) \) is the optimal validity cluster, and \( c^* \) is the optimal category number. After summarizing the above, a new validity cluster function is proposed as follows:

\[
\min Z_j(u_{ik}, c) = Z_2^{(i+1)}(u_{ik}, c) / Z_2^{(i)}(u_{ik}, c)
\]

(4)

Solve for the minimum value of this function. When \( c = c^* \), the function has a minimum value, and \( c = c - 1 \); otherwise, \( c = c + 1 \), and the iteration needs to continue.

3.3. Steps of algorithm
The steps of improved FCM algorithm are as follows:

1. Given the cluster category number \( c \) (1 < \( c \leq n \), \( n \) is the number of samples) and the fuzzy index \( m(1 \leq m < +\infty) \), the cluster center matrix is \( P = [P_1, P_2, \ldots, P_i, \ldots, P_c] \). Set the iteration stop threshold \( \omega^{(i)} \), the cluster prototype \( P_i \) and the iteration counter \( a^{(i)} = 0 \).

2. Update the division matrix \( U^{(a)} \) with the iterative formula below. The iterative formula of membership degree is a mapping from a point to a set. In practical calculation, the following membership degree updating formula is usually adopted[6]:

\[
u_{ik}^{(a)} = \left\{ \sum_{k=1}^{c} \left[ \frac{j_{ik}^{(a)}}{j_{ik}^{(a)}} \right]^{2} \right\}^{-1}, d_{ik}^{(a)} > 0
\]

(5)

When \( d_{ik}^{(a)} = 0 \), \( u_{ik}^{(a)} = 1 \). When \( d_{ik}^{(a)} = 0 \quad \forall k \neq r \), \( u_{ik}^{(a)} = 0 \).

We can improve the membership degree to reduce the impact of the outlier on the cluster center:

\[
U_{ik}^{(b)} = \frac{1}{1+\alpha} u_{ik}^{2} + \frac{\alpha}{1+\alpha} u_{ik}, \alpha \in [0, 1]
\]

(6)

3. The cluster prototype \( P_i^{(m+1)} \) can be updated by using the following formula:
\[ P_i^{(v+1)} = \frac{\sum_{t=1}^{n} (u_{ik}^{(v+1)})^m d_t}{\sum_{t=1}^{n} (u_{ik}^{(v+1)})^m} , k = 1, 2, \ldots, c \] (7)

(4) if \( \| P_i^{(v)} - P_i^{(v+1)} \| < \omega \), the algorithm stops and outputs the division matrix \( U \) and the cluster prototype \( P_i \); otherwise, let \( v = v + 1 \), execute (2).

(5) For the following formula, determine whether \( Z_f(u_{ik}, c) \) is greater than or equal to 1.

\[ Z_f(u_{ik}, c) = \frac{Z_2^{(v+1)}(u_{ik}, c)}{Z_2^{(v)}(u_{ik}, c)} \] (8)

When \( c = c^* \), \( Z_f(u_{ik}, c) \) is greater than 1, and \( c = c - 1 \); otherwise, \( c = c + 1 \), and the iteration needs to continue. By this method, we do not need to determine the cluster number \( c \) in advance, and instead it can be determined automatically in an iterative manner in the process of clustering. The algorithm flow is shown in Figure 1:[8-10]:

**Figure 1.** Improved FCM Algorithm Flow Chart
4. Experimental Analysis

4.1. Data sources
It uses IRIS dataset as a test set to verify the improved FCM algorithm, and applies the “consumer product quality safety accident cases" as the data source to complete the identification and analysis of consumer product quality safety factors. The IRIS dataset uses the characteristics of iris as the data source, and the dataset contains 150 datasets. There are 200 cases of consumer product quality safety accidents, the data of which comes from consumer complaints, online public opinion and injury surveillance.

4.2. Result analysis

![Clustering results based on traditional FCM algorithm](image1)

![Clustering results based on improved FCM algorithm](image2)

**Figure 2.** Clustering Results Before and After FCM Algorithm Improvement Based on the IRIS Dataset

![Clustering results based on traditional FCM algorithm](image3)

![Clustering results based on improved FCM algorithm](image4)

**Figure 3.** Clustering Results Before and After FCM Algorithm Improvement Based on the Consumer Product Quality Safety Accident Cases
5. Conclusion
In view of the defects of the existing FCM algorithm, this paper suggests improvement in the following two aspects: firstly the weighted membership degree; secondly the optimal selection of cluster category number. After introducing new weight variables into the membership degree and putting forward a new membership degree update function, the impact of the outlier on the cluster center begins to decrease. According to the clustering validity function, this paper proposes a new method to discriminate the clustering validity function. In the experimental process, the IRIS dataset is used to verify the performance of the improved algorithm, and factor clustering analysis is carried out in combination with the dataset of typical consumer product quality safety accident cases. The results show that the clustering results based on the improved FCM algorithm are obviously better than those based on the unimproved FCM algorithm.

Acknowledgments
We would like to acknowledge that this research is supported and funded by the the Basic Scientific Research Business Projects 552020Y-6770-1, and Major Projects of the National Social Science Fund of China under Grant No. 18ZDA079.

References
[1] T Kalaiselvi, P Sriramakrishnan. Rapid brain tissue segmentation process by modified FCM algorithm with CUDA enabled GPU machine. International Journal of Imaging Systems and Technology, 2018, 28(3).
[2] Karan Sikka, Nitesh Sinha, Pankaj K. Singh, Amit K. Mishra. A fully automated algorithm under modified FCM framework for improved brain MR image segmentation. Magnetic Resonance Imaging 2009, 27:994-1004.
[3] S. Rajeshbabu, B.V. Manikandan. Detection and classification of power quality events by expert system using analytic hierarchy method. Cognitive Systems Research, 2018, 52:729-740
[4] Jyoti Prakash Singh, Yogesh K. Dwivedi, Nripendra P. Rana. Event classification and location prediction from tweets during disasters.. Annals of Operations Research, 2019, 283 (3):737-757.
[5] Xue Ying, Sha Xiuyan. Grey prediction model based on improved fuzzy clustering algorithm. Statistics and Decision, 2017(09): 27-30. Jorge Enrique Rodriguez, Jenny P. Ortiz. Bayesian methods for classification inappropriate web pages.. Vision electric, 2017, 1:10-10.
Du A N, Fang B X. Novel Approaches for Web Filtering Based on User Interest Focusing Degree. . International Journal of Innovative Computing Information and Control, 2008. 4(4): 1325-1334

[6] Liang Chao, Su Jinsheng, Zhao Hui, Zhao Dewei, Yang Jing, Liu Dunnan, Wang Lingxiang. Aggregation of Typical Diurnal Curves of Regional Distribution Based on FCM Algorithm. IOP Conference Series: Earth and Environmental Science, 2021, 651(2).

[7] Chen Rui, Zhang Fengbin, Xi Liang. Anomaly Detection Algorithm Based on FCM with improved Krill Herd. Journal of Physics: Conference Series, 2019, 1187(4).

[8] Li Xiang, Li Ke, Li Zhongxue. A method for determining the number of manufacturing units based on fuzzy C-means algorithm. Modern Manufacturing Engineering, 2017(05): 132-138.

[9] Xu Qin, Tang Dunbing, Cai Qixiang. Improved fast fuzzy C-means clustering image segmentation algorithm. Journal of Nanjing University of Science and Technology, 2016, 40(03): 309-314.