# A Novel Performance Evaluation Method using ISODATA Method

Xueda Li[1, a], Ying Xiao[2, b]

[1] Dalian University of Technology, Panjin, China;

[2] Dalian University of Technology, Panjin, China.

[a]1298359043@qq.com, [b]3600382231@qq.com

## Abstract

**Performance evaluation for employees is very important to a company to keep itself alive and is a major references for employee salaries. Through performance evaluation, the company leaders could know more clearly about the operation of the company. Whereas, problems of high subjectivity, low reasonableness, low effectiveness, etc. always existing in performance evaluation methods and attract interests of many researchers. This paper puts forward a novel performance evaluation method for bank employees using iterative self-organizing data analysis (ISODATA) algorithm. The evaluation factors are selected firstly according to experts' experience to build the data set and model input. A performance clustering method based on ISODATA is designed for obtaining the performance levels. Finally, the performance evaluation is conducted according to the clustering results. The results of experiments demonstrate that the proposed method has better properties in efficiency than the competing methods.**

## Keywords

**Performance evaluation; ISODATA; Clustering method; Evaluation factor.**

## 1. Introduction

The rapid development of China's information technology has provided conditions for the rapid processing and information interconnection of internal affairs of major banks, which has promoted the improvement of bank staff's office efficiency and the efficient processing of bank information [1]-[3]. Therefore, the combination of informationization and banking affairs is applied to a lot of business. In the Chinese banks, informatization has penetrated into all aspects and is also a concrete manifestation of a fast-growing bank. According to the actual situation of the bank, it is very necessary and meaningful to study the bank employee performance evaluation method [4].

Through the research on the current stuation of employee performance evaluation of domestic commercial banks, it is found that domestic commercial banks have not yet implemented a professional employee performance evaluation system, and there is no complete channel for collecting employee performance, which leads to no way for commercial bank leaders and managers. Employees develop more reasonable performance standards. Without understanding the work status of employees, it is impossible to propose a more effective management plan in time, and the concept of "work performance as the center" has not penetrated the hearts of bank leaders and managers. The performance evaluation has been carried out within the bank [5]-[6]. Their performance evaluation system is also in the primary stage. It is not supported by very complete information. It is only a specification for performance evaluation. When the business of the bank employees is divided. It is also a one-sided factor in simply using the deposit amount. This will result in the bank not being able to pay attention to the real needs of employees and not improving the efficiency of employees. Therefore, after paying attention to these problems in commercial banks, many scholars in China are also actively studying the performance evaluation management system for employees in the bank, expecting to obtain the actual situation of system users, but their research still has many limitations. It is not enough to study the functional process and conceptual framework of the performance appraisal system.

This paper first selects the performance evaluation factors of bank employees, and designs the employee evaluation division model based on iterative self-organized data analysis algorithm, so as to achieve a more reasonable division and analysis of employee performance.

## 2. Evaluation Factor Selection

In order to better carry out performance evaluation, the key performance indicator factors of employees are first selected and selected as follows:

### 2.1 State model

### 2.2 Total daily deposits

The total daily deposit, which is the difference between daily income and daily expenditure, is calculated as

$$S_d = W_d - C_d \tag{1}$$

Where $S_d$ is the total daily deposit; $W_d$ is daily income; $C_d$ is the daily expenditure.

### 2.3 Monthly deposits

The total number of monthly deposits, that is, the difference between monthly income and monthly expenditure, which is calculated as

$$S_m = W_m - C_m \tag{2}$$

Where Sm is the total monthly deposit; Wm is the monthly income; and Cm is the monthly expenditure.

### 2.4 Total public deposits

The total amount of existing deposits on the public account, that is, the sum of the number of deposits on the public account at the time of the last liquidation and the income from the last liquidation to the time of the liquidation, minus the time from the second liquidation to the time of the liquidation, the formula is

$$S_{PU1} = S_{PU0} + W_{PU} - C_{PU} \tag{3}$$

Where SPU1 is the total amount of deposits on the public account; SPU0 is the number of deposits on the public account at the time of the last liquidation; WPU is the income from the last liquidation to the time of the liquidation; and CPU is the time from the last liquidation to the current liquidation.

### 2.5 Total private deposits

The total amount of existing deposits on the private account, that is, the sum of the number of deposits on the public account at the time of the last liquidation and the income from the last liquidation to the time of the liquidation, minus the time from the sub-liquidation to the time of the liquidation, the calculation formula is

$$S_{PE1} = S_{PE0} + W_{PE} - C_{PU} \tag{4}$$

Where SPE1 is the total amount of deposits on the public account; SPE0 is the number of deposits on the public account at the time of the last liquidation; WPE is the income from the last liquidation to the time of the liquidation; and CPU is the time from the last liquidation to the current liquidation.

## 3. Performance Evaluation with ISODATA

In the traditional performance appraisal and reward mechanism, the fixed quota method and the fixed ratio are the more common methods. While, these methods will lead to some unreasonable situations. For example, there are 100 employees, and the top 10 employees are fixed monthly for performance rewards. If there is a similar performance between the 11th employee and the 10th employee, there will be no incentives due to the quota problem, which will hurt the enthusiasm of the initiative of employees. similarly, if the performance of the 10th employee and the 9th employee is different, because of the quota problem, the 10th employee, although the performance is not outstanding, still won the award and that will hit the enthusiasm of the top nine award-winning employees [7].

Therefore, on the basis of achieving digital performance appraisal, it is very necessary to combine the effective methods to achieve a dynamic and efficient division of employees' excellent level. Based on the traditional linear fitting algorithm, the company achieves the integrated evaluation of employee performance, and then selects a fixed proportion to classify the superior level [8]. Under the existing competition mechanism, it may lead to unfairness and cannot improve employee enthusiasm more effectively. At the same time, through the clustering of performance appraisal, it can achieve more efficient evaluation of current employee work efficiency. Based on the above, this paper designs an automatic clustering method based on ISODATA to achieve adaptive clustering of employee performance levels. Among them, the normalized employee's various banking data is used as the clustering parameter, and the Euclidean Metric is used as the evaluation index to carry out the adaptive dynamic clustering model design.

The ISODATA and the K-means clustering method are very similar in many places. For example, the cluster center of ISODATA and the clustering center of the mean clustering method are obtained by continuously calculating the mean of the samples. At the same time, in the ISODATA algorithm, exploratory operations and human-computer interaction functions were developed. At the same time, in the process of the loop, ISODATA can merge or split the clusters according to the empirical feedback of the intermediate clustering results, and achieve the result of “self-organization”, which belongs to a heuristic clustering method. For the dynamic clustering requirements of employee performance, that is, the number of unknown clusters, the ISODATA algorithm is more in line with actual needs.

### 3.1 Data normalization

Select the employee's total daily deposit X1, the total monthly deposit X2, the total public deposit X3, and the total private deposit X4 as the parameter indicators for the performance excellence. In order to unify the dimension, eliminate the influence of different units on the clustering results. The parameters of the dimensions are normalized, and the normalized parameter data is input into the clustering model. Specifically, for a dimension parameter Xi, the sample data is assumed to be $x_i^j, i=1,2,...,4, j=1,2,...,n$, then the normalization formula is

$$\begin{cases} \bar{x}_i^j = \dfrac{x_i^j - x_i^{k1}}{x_i^{k2} - x_i^{k1}} \\ x_i^{k1} = \min\{x_i^1, x_i^2, ..., x_i^n\} \\ x_i^{k2} = \max\{x_i^1, x_i^2, ..., x_i^n\} \end{cases} \tag{5}$$

### 3.2 ISODATA algorithm steps

Step 1: Initialization parameter information. In the process of loop calculation, real-time parameter correction is performed according to different situations, and the data is classified according to the given parameter standard;

Step 2: Calculate the distance between each class, that is, the distance between each cluster center. Here, the Euclidean distance is used for clustering;

Step 3-5: According to the set parameter standard information, analyze the current clustering result, and then perform corresponding class processing, such as Step4 split operation, Step5 merge operation, and then obtain new Class clusters and cluster centers;

Step 6: Repeat Step2 to Step5 until the current clustering result satisfies the current parameter condition standard setting.

### 3.3 ISODATA algorithm flow

Step 1: Load n samples $\{\mathbf{X}_i, i=1,2,...,n\}$

Randomly select $m$ initial cluster centers $\{\mathbf{Z}_i, i=1,2,...,m\}$, unlike the K-means clustering method, in the ISODATA algorithm, the number of random initial cluster centers may not necessarily be equal to the given number of categories (considering the ISODATA could "self-organize" data classification, the final number of classifications is also unknown). At the same time, the corresponding parameter setting is made at the beginning: $K$ represents the expected number of cluster centers; $\theta_n$ represents the minimum number of samples in each cluster, and if the number of elements of a cluster is lower than $\theta_n$, then the class A cluster cannot be treated as a separate class; $\theta_s$ represents the standard deviation of the sample distance distribution in a cluster; $\theta_c$ represents the minimum distance between two cluster centers, and the distance between two classes is less than $\theta_c$,then this The two classes are to be merged; $L$ is the maximum number of clusters allowed to be merged for each loop calculation; $I$ is the maximum number of loop calculations.

Step 2: Assign n samples to the nearest cluster $C_j$.

Suppose $d_j = \min\{|\mathbf{X}-\mathbf{Z}_i|, i=1,2,...,n\}$, the distance of $|\mathbf{X}-\mathbf{Z}_j|$ is the smallest, then $\mathbf{X} \in \boldsymbol{C}_j$.

Step 3: If the number of elements in $C_j$ is $N_j < \theta_n$, it means that $C_j$ cannot be a separate class cluster, and the class should be merged with its nearest class and $m = m-1$.

Step 4: Calculate the cluster center result of each current cluster $C_j$.

$$\mathbf{Z}_j = \frac{1}{N_j}\sum_{X \in C_j} \mathbf{X}, \ j=1,2,...,m \tag{6}$$

Step 5: Calculate the average distance between cluster centers of each current cluster $C_j$.

$$\bar{d}_j = \frac{1}{N_j}\sum_{X \in C_j} |\mathbf{X}-\mathbf{Z}_j|, \ j=1,2,...,m \tag{7}$$

Step 6: Calculate the average result of the distance of each sample from the cluster center of the cluster to which it belongs.

$$\bar{d} = \frac{1}{n}\sum_{j=1}^{m} N_j \bar{d}_j \tag{8}$$

Step 7: Analyze the current clustering result according to the parameter standard, and operate:

If $I$ has been cycled, then $\theta_c$=0 goes to Step11 and the calculation ends.

If $m \le \frac{K}{2}$, jump to Step8 to perform the split operation.

If the number of iterations is even, or $m \ge 2K$, then the cluster is no longer classified, and jump to step11 for execution. Otherwise, jump to Step8 to perform the split operation.

Step 8: Perform the calculation of the standard deviation vector $\boldsymbol{\delta}_j = (\delta_{1j}, \delta_{1j}, ..., \delta_{nj})^T$ of the sample distance in each cluster, and the calculation formula is

$$\delta_{ij} = \sqrt{\frac{1}{N_j}\sum_{x \in B_j}(x_{ik} - z_{ij})^2} \tag{9}$$

Among them, $i=1,2,...,n$, $j=1,2,...,m$.

Step 9: Obtain the corresponding maximum component in each $\{\boldsymbol{\delta}_j, j=1,2,...,m\}$ as $\{\delta_{j\max}, j=1,2,...,m\}$.

Step 10: In $\{\delta_{j\max}, j=1,2,...,m\}$, if there is $\delta_{j\max} > \theta_s$, it also meets one of the following two conditions: $\bar{d}_j > \bar{d}$ and $N_j > 2(\theta_n + 1)$, that is, when the total number of samples in $C_j$ exceeds the specified value by more than one time, that is, $m \ge \frac{K}{2}$, $Z_j$ is split into two new cluster centers $\mathbf{Z}_j^+, \mathbf{Z}_j^-$, and $m = m+1$, $\mathbf{Z}_j^+$ is equivalent to the component of $\delta_{j\max}$, which can be added. $k\delta_{j\max}, 0 \le k \le 1$ and $\mathbf{Z}_j^-$ are equivalent to the component of $\delta_{j\max}$ ,and $k\delta_{j\max}$ can be subtracted. If the split operation has been executed in this step, then jump to Step2 is performed; otherwise, continue.

Step 11: Calculate the distance between the two cluster centers of each cluster, and calculate the formula as

$$d_{ij} = |\mathbf{Z}_i - \mathbf{Z}_j|, i=1,2,...,m-1; \ j=i+1,...,m \tag{10}$$

Step 12: Compare the values of $d_{ij}$ and $\theta_c$, and rank the values of $d_{ij} < \theta_c$ in the order of minimum distance, that is, $\{d_{i_1 j_1}, d_{i_2 j_2}, ..., d_{i_L j_L}\}$ and $d_{i_1 j_1} < d_{i_2 j_2} < ... < d_{i_L j_L}$.

Step 13: Combine the two cluster centers $\mathbf{Z}_{i_1}$ and $\mathbf{Z}_{i_2}$ with the distance $d_{i_1 j_1}$, that is, the two classes with the smallest cluster distance to merge, and the new center is

$$\mathbf{Z}_i^* = \frac{1}{N_{i_1} + N_{j_1}} [N_{i_1} \mathbf{Z}_{i_1} + N_{j_1} \mathbf{Z}_{j_1}], i = 1, 2, ..., L \tag{11}$$

Where $N_{i_1}$ and $N_{j_1}$ are the number of elements of the two clusters with $\mathbf{Z}_{i_1}$ and $\mathbf{Z}_{j_1}$ as the cluster centers. With this kind of operation, $Z_i^*$ can be made into a true average vector, and $m = m - L$.

Step 14: If the loop is calculated for $I$ times, the algorithm ends. Otherwise, jump to Step2 to execute. In particular, if there is a user to modify the parameters at this time, jump to Step1 for execution.

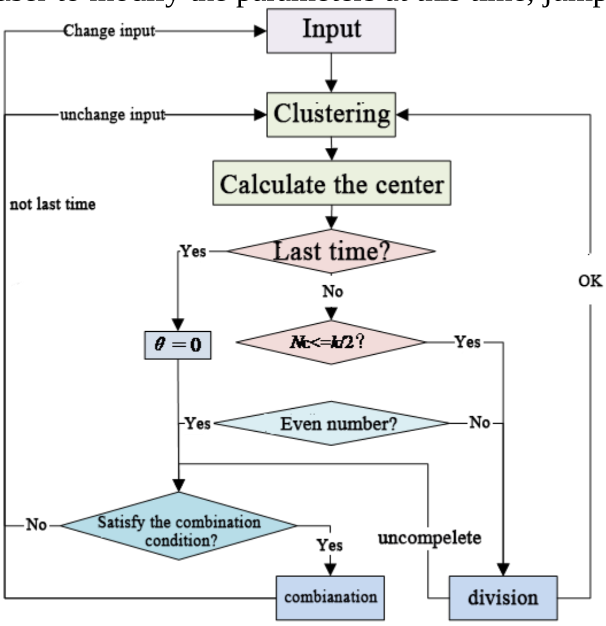


Fig. 1 Flow chart of the ISODATA clustering algorithm.

The algorithm framework of ISODATA is shown in Fig. 1. Through the above analysis, it can be found that by entering the functions of manual parameter setting, self-organization merging, splitting, etc., ISODATA can process the clustering results in real time, merge smaller clusters, and perform larger clusters. Classification and so on. By continuously cycling, the final result can be stabilized and converged, and a satisfactory clustering result is obtained. At the same time, according to the ISODATA algorithm description, this algorithm belongs to an unsupervised heuristic learning method. Compared with the fixed ratio division, ISODATA is more suitable for the fusion evaluation of performance.

## 4. Simulations and Experiments

For the clustering algorithm, the evaluation criteria of the clustering algorithm clustering effect is first established, which is called cluster compactness. For sample $\mathbf{x}_i$, its compactness with its final clustering result category is described as

$$\mathbf{S}_i = \frac{\min(\mathbf{b}_i) - a_i}{\max[a_i, \min(\mathbf{b}_i)]} \tag{12}$$

Where $a_i$ is the average distance between $\mathbf{x}_i$ and other samples of the same type, $\mathbf{b}_i$ is a vector, and its element $\mathbf{X}_i$ is the average distance between points in the class of different classes. For example, the $k$-th element of $\mathbf{b}_i$ is $\mathbf{X}_i$ and $k$ The average distance between elements in category $\mathbf{C}_k$. The larger $\mathbf{S}_i$ is, the more reasonable the $\mathbf{x}_i$ classification is.

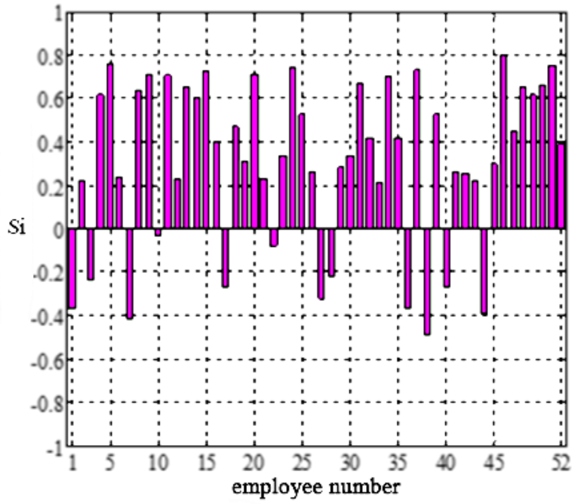


Fig. 2 Clustering experiment results.

The performance data of 52 employees were normalized, and then the European distance was used as a standard. ISODATA clustering, cluster compactness description is shown in Fig. 2.

It can be seen from the figure that from the analysis of clustering results, the compactness description range is -0.4~0.8, and the compactness result is higher. Therefore, the excellent partitioning method based on ISODATA method has better clustering results. Based on the results of the clustering, it is possible to evaluate the excellent and poor conditions of the employees. For example, the clustering results are divided into two categories, indicating that the employees are good or bad; the clustering results are divided into three categories, indicating that the employees are good, medium, and poor; the clustering results are divided into four categories, indicating that the employees. There are four cases of excellent, good, medium and poor. At the same time, combined with the average performance of employees within the various types, the distribution of personnel, and the comparison of multi-month situation, it can provide more adequate decision-making information for employee work status, enterprise development and follow-up planning.

## References

[1] Ling S, Cai Z, Li L, et al. Performance evaluation of deep feature learning for RGB-D image/video classification[J]. Information Sciences, 2017, s 385–386(C):266-283.

[2] Sun J, Wang C, Xiang J, et al. Performance evaluation of heterogeneous bank supply chain systems from the perspective of measurement and decomposition[J]. Computers & Industrial Engineering, 2017.

[3] Jie Y, Dong F, Li D. Management performance evaluation of bank listed companies based on grey integrated clustering[C]. Control & Decision Conference. 2017.

[4] Sardjono W, Pujadi T. Performance evaluation of systems Managed File Transfer in banking industry using IT Balanced Scorecard[C]. International Conference on Information Management & Technology. 2017.

[5] Adeloye A J, Soundharajan B S, Mohammed S. Harmonisation of Reliability Performance Indices for Planning and Operational Evaluation of Water Supply Reservoirs[J]. Water Resources Management, 2017, 31(3):1013-1029.

[6] Chen Y, Chiu Y, Chiu C. The performance evaluation of banks considering risk: an application of undesirable relation network DEA[J]. International Transactions in Operational Research, 2017(3).

[7] Aras G, Tezcan N, Furtuna O K. Multidimensional comprehensive corporate sustainability performance evaluation model: Evidence from an emerging market banking sector[J]. Journal of Cleaner Production.

[8] Baqar M, Zaidi S S H, “Performance evaluation of linear and multi-linear subspace learning techniques for object classification based on underwater acoustics”, in International Bhurban Conference on Applied Sciences & Technology, 2017.