Hindawi
Complexity
Volume 2021, Article ID 5571683, 12 pages
https://doi.org/10.1155/2021/5571683

Research Article

Early Warning of Financial Risk Based on K-Means Clustering Algorithm

Zhangyao Zhu[1] **and Na Liu**[2]

[1]*School of Accounting, Southwestern University of Finance and Economics, Chengdu, Sichuan 611130, China*
[2]*International Business School, Southwestern University of Finance and Economics, Chengdu, Sichuan 611130, China*

Correspondence should be addressed to Na Liu; yona@smail.swufe.edu.cn

Received 10 February 2021; Revised 24 February 2021; Accepted 27 February 2021; Published 5 March 2021

Academic Editor: Wei Wang

Copyright © 2021 Zhangyao Zhu and Na Liu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The early warning of financial risk is to identify and analyze existing financial risk factors, determine the possibility and severity of occurring risks, and provide scientific basis for risk prevention and management. The fragility of financial system and the destructiveness of financial crisis make it extremely important to build a good financial risk early-warning mechanism. The main idea of the K-means clustering algorithm is to gradually optimize clustering results and constantly redistribute target dataset to each clustering center to obtain optimal solution; its biggest advantage lies in its simplicity, speed, and objectivity, being widely used in many research fields such as data processing, image recognition, market analysis, and risk evaluation. On the basis of summarizing and analyzing previous research works, this paper expounded the current research status and significance of financial risk early-warning, elaborated the development background, current status and future challenges of the K-means clustering algorithm, introduced the related works of similarity measure and item clustering, proposed a financial risk indicator system based on the K-means clustering algorithm, performed indicator selection and data processing, constructed a financial risk early-warning model based on the K-means clustering algorithm, conducted the classification of financial risk types and optimization of financial risk control, and finally carried out an empirical experiments and its result analysis. The study results show that the K-means clustering method can effectively avoid the subjective negative impact caused by artificial division thresholds, continuously optimize the prediction process of financial risk and redistribute target dataset to each cluster center for obtaining optimized solution, so the algorithm can more accurately and objectively distinguish the state interval of different financial risks, determine risk occurrence possibility and its severity, and provide a scientific basis for risk prevention and management. The study results of this paper provide a reference for further researches on financial risk early-warning based on K-means clustering algorithm.

1. Introduction

Financial risk is the possibility of potential losses in direct investment caused by company loans, fiscal finance, and other economic factors and corresponding consequences of the economic shocks. The outbreak of financial crises often leads to currency depreciation, exchange rate fluctuations, market downturns, economic recession, and sometimes even the decline of neighboring businesses, countries, and even the world economy because of the contagiousness of financial risk or crisis [1]. Therefore, the fragility of financial system and the destructiveness of financial crisis make it extremely important to build a good financial risk early-warning mechanism [2]. The expansions of transaction scale promote the development of the market, increase competition, and encourage enterprises to innovate. The early warning of financial risk is to identify and analyze existing financial risk factors, determine the possibility and severity of occurring risks, and provide scientific basis for risk prevention and management [3]. The objects in the same cluster are similar to each other and different from objects in different clusters, which is an unsupervised learning process. The content of financial risk early-warning can be divided into financial risk organization form, indicator system, and prediction method in detail; the functions of financial risk early-warning include timely grasp of trends, effective evaluation, and timely adoption of relevant regulatory measures, thereby reducing the harm of financial risks [4].

Cluster analysis is, when the category of studied object is not known in advance, to group the similarities into one category based on the degree of affinity, so that the same category can achieve the maximum homogeneity and minimize the heterogeneity, while the different categories achieve the maximum homogeneity and minimum heterogeneity [5]. Clustering analysis algorithms can be summarized into three different types: trying to find an optimal partition to divide the data into a specified number of clusters; trying to find a method of clustering structure hierarchy; and trying to find a method based on probability model for potential cluster modeling [6]. The K-means cluster analysis method can effectively avoid the subjective negative impact caused by the artificial threshold value, so it can more accurately and objectively distinguish the state intervals of different financial risks. K-means is the most widely used clustering method so far, whose main idea is to gradually optimize clustering results and constantly redistribute target dataset to each clustering center to obtain optimal solution; and biggest advantage lies in its simplicity, speed, and objectivity, being widely used in many research fields such as data processing, image recognition, market analysis, and risk evaluation [7]. K-means clustering algorithm is to select K data as the initial centroid of each category and divide them into K categories according to the principle of one category with the smallest distance, and then the divided mean values are judged according to the square error criterion function for determining whether the division is converged: if convergence, the algorithm is over; otherwise, continue to redivide and update the value of each cluster center successively until the optimal clustering result is obtained [8].

On the basis of summarizing and analyzing previous research works, this paper expounded the current research status and significance of financial risk early-warning, elaborated the development background, current status and future challenges of the K-means clustering algorithm, introduced the related works of similarity measure and item clustering, proposed a financial risk indicator system based on the K-means clustering algorithm, performed indicator selection and data processing, constructed a financial risk early-warning model based on the K-means clustering algorithm, conducted the classification of financial risk types and optimization of financial risk control, and finally carried out an empirical experiments and its result analysis. The study results of this paper provide a reference for further researches on financial risk early-warning based on K-means clustering algorithm. The detailed chapters are arranged as follows: Section 2 introduces the related works of similarity measure and item clustering; Section 3 proposes a K-mean-clustering-algorithm-based financial risk indicator system, including indicator selection and data processing; Section 4 constructs a financial risk early-warning model based on the K-means clustering algorithm; Section 4 carries out an empirical experiments and its result analysis; Section 6 is the conclusion.

2. Related Works

2.1. Similarity Measure. The K-means clustering does not need to store the distance matrix, occupies a small memory and has the advantages of large processing data volume and fast running speed. However, it should be noted that K-means clustering can only achieve local optimization and can only handle continuous data variables, the K-means clustering method needs to specify the number of clusters before clustering, and the clustering results are easily affected by the initial effect of clustering can only get a local optimal solution [9]. Model clustering requires a lot of prior knowledge to be able to give a suitable clustering model; although density clustering is universal, it lacks generality; while dividing clustering requires a given number of clusters before clustering, but the clustering effect and clustering speed are better, for example, consider stock price data [10]. They are typical time series data, but stock prices are independent of each other; future prices will only be affected by the closing price of the previous day. The content of monitoring can be divided into financial risk organization forms, indicator systems, and forecasting methods. Since the attribute values of the processed data objects often differ greatly in units and value ranges, it is considered to perform certain data processing to make the value range above the same benchmark [11].

Commercial banks and other financial institutions form a complex network relationship through balance sheets, credit business, and other channels. Financial risk early-warning work mainly includes risk identification, risk assessment, risk early warning, and risk treatment, which can be divided into financial risk organization form, indicator system, and prediction method in detail. When banks are subject to internal and external shocks to trigger debt defaults, liquidity risks are spread through the credit channels associated with interbank business, and the activities of other banks and financial institutions will be affected by risk spillovers, causing systemic risks [12]. At the same time, driven by factors such as information asymmetry and investor irrationality, the risk contagion process in the capital market tends to accelerate [13]. In order to reduce the crisis caused by systemic risks, it is necessary to manage systemic risks reasonably and the frequency of financial crises has made the prevention of systemic risks more and more important [14]. With the cross-development of econometric methods and system engineering methods, the analysis of financial risk contagion effects from the perspective of complex network relevance provides a new research perspective in this field. Microindividual risk spillovers and network effects formed by risk contagion have become decoding the focus of the tail event. When systemic risks accumulate to a certain extent and are released, it will cause a large number of financial institutions to close down, which will spread to the entire financial system and trigger a systemic financial crisis [15].

2.2. Item Clustering. Risk management activities should involve three elements: price, preference, and probability. Price is used to determine the cost that must be paid to prevent various risks: probability is used to estimate the likelihood of these risks occurring; preference is used to determine the ability and willingness to bear risks and confidence. Risk management must integrate the three elements to make systematic and dynamic rational decision-making, so as to achieve a balance between financial risk and risk preference, so that investors can bear the risks they are willing to take and obtain the greatest risk reward [16]. K-means is the most widely used clustering method so far, whose main idea is to reveal the industry clustering and sector transfer characteristics of the stock market, while the clustering based on the linear trend characteristics of the time series mainly reflects the similarity between the fluctuations of individual stocks and stock indicator fluctuations. It is especially important that it can make the institutional body composed of several individual decision makers optimally control risks in risk management, so that the entire organization will not suffer excessive risk losses due to the behavior of a certain decision maker [17]. It emphasizes the consistent, accurate, and timely measurement of risks faced by financial institutions and tries to establish a rigorous procedure to analyze the distribution of total risks in the transaction process, asset portfolio, and other business activities and how to price and rationally allocate capital for different types of risks. At the same time, a dedicated risk management department has been established within financial institutions, dedicated to preventing and deflating risks and digesting the resulting costs [18].

The bank's credit department evaluates the potential and development prospects of the borrower and its unit and integrates the borrower's income and the ability to return the loan conditions and enters the evaluation system in a unified manner [19]. The existing records and information will be the applicant's risk evaluation result and perform evaluation and submit it to credit granting personnel to provide a basis for making decisions. Using the K-means algorithm is actually a process of solving optimization, and the objective function exceeds the minimum value in some partial intervals, but there is only one global minimum [2]. The objective function is required to be consistent with the error sum of the square sum of the search direction during the search operation. K-means clustering algorithm is a new scientific and effective bank loan risk management analysis algorithm, which mainly performs quantitative analysis [20]. When dealing with cubes, the size reduction technology can be used as the dimension of the two-dimensional transformation, and the dimensionality reduction actually uses some means to process the higher-dimensional data into the lower-dimensional data, and at the same time, the similarity between the data and the data before processing is consistent with the most basic data. In this way, it can perform cluster analysis on the lower-dimensional data obtained after processing [21–24].

3. Indicator System of Financial Risk Based on K-Means Clustering Algorithm

3.1. Indicator Selection. The establishment of a financial risk early-warning system can be roughly based on the three variables of macroeconomics, financial system, and foreign trade and economics and the macroeconomic level mainly considers factors closely related to the economy and the financial system. Its indicators include output value growth rate, fiscal deficit/output value, unemployment rate, consumer price indicator, and fixed asset investment growth rate; the financial system level mainly considers the currency market; the steady development indicators of the capital market include growth rate, stock financing, commodity housing price volatility, and stock market value volatility. After the selection of early-warning indicators is completed, the operation of the financial system needs to be divided into four states according to the magnitude of the risk, namely, the safe state, the basic safe state, the light risk state, and the severe risk state. In view of this, many scholars have invested in this research work and have made much major progress in selecting financial risk early-warning indicators and establishing financial risk early-warning models. The fragility of financial system and the destructiveness of financial crisis make it extremely important to build a good financial risk early-warning mechanism. Therefore, the main advantages of K-means clustering algorithm are simple, fast, and efficient and scalable for big data and the clustering quality of the standard K-means clustering algorithm is highly dependent on the initial clustering center. Figure 1 shows the framework of financial risk early-warning based on k-means clustering algorithm.

When the mean and variance of the portfolio are the same, investors often choose a portfolio with a larger third-order moment and sometimes even put the third-order moment in a more important position. It is worth pointing out that whether it is variance, absolute deviation, or skewness, lower-than-average returns and higher-than-average returns are placed in the same position. It is the income that is lower than the expected value, because this is the real loss or risk, usually called the lower risk.

A sample point data set is given as $X = \{x_1, x_2, \cdots, x_n\}$ and the number of clusters is set as k, and $\{y_1, y_2, \cdots, y_k\}$ is a division of the set X; if the center of the class y_i is z, then the cluster problem is to find k classification centers $z_1, z_2, \ldots, z_k$ so that the sum of the distances from all sample points x_i to a nearest center point z_i is extremely small:

$$Y_i = \sum_{i=1}^{k} \frac{y_i - x_i}{z_i - x_i}. \quad (1)$$

After the early-warning indicator system is established, the abovementioned early-warning indicator data is affected by different directions and magnitudes, which will affect the mining of data rules and the accuracy of subsequent prediction results. Therefore, the original data needs to be normalized before the early-warning model is established as follows:

$$Q_i = a \cdot q_i e^{x_i} - b \cdot w_i e^{y_i}, \tag{2}$$

where q_i is a financial risk early-warning indicator; w_i is a different year; a is the value of different indicators in different years; and b is an indicator obtained after standardization. Given the number of samples to be clustered and the number of clusters, the samples are divided into corresponding classes according to the similarity of the data, which is helpful in the financial risk early-warning based on K-means clustering algorithm.

After establishing the hierarchical structure, aiming at a certain criterion of the k-th layer, all n factors related to the k-1 layer are analyzed by pairwise comparison, and the extension interval number is used to quantitatively express the relative importance between the two and construct the extension judgment matrix W_{ij}:

$$W_{ij} = \frac{\left(1/r_{ij}\right) - \left(1/t_{ij}\right)}{\Delta r_{ij} - \Delta t_{ij}}, \tag{3}$$

where r_{ij} is an extension interval number; t_{ij} is the endpoint of the extension interval element in the i-th row and j-th column of the matrix W_{ij}; Δr_{ij} is the single ordering of the i-th factor on a certain layer to a certain factor on the j layer; and Δt_{ij} is the extension interval weight of the j-th factor of the i-th level to a certain factor of the previous level.

The advantage of this type of measurement is that its statistical and financial meanings are simple, clear, and easy to understand and can be directly used to establish corresponding portfolio selection models. However, in reality, due to the variability of the joint distribution of the return on the set of securities to be invested, to fully reflect its random characteristics, in theory, all the moment information of its order should be considered at the same time. If the risk is only measured based on a few low-order moments of the return distribution, it will cause model errors due to the failure to describe a certain characteristic of the distribution. Because of its advantages in decision theory and actual investment, the research and application of various forms of lower half-moment risk measures are still popular recently. The distribution of center point in different state intervals for similarity measure and item clustering are shown in Figure 2.

3.2. Data Processing. Economic and financial risks are relatively complicated, and they are determined by the risks of all aspects; therefore, the economic and financial risk evaluation indicator system is a complete system, and the indicators are interrelated to form an organic whole. The higher the level of financial development, the more it can improve the efficiency of enterprises and promote technological progress, mainly because the improvement of the level of financial development can reduce transaction costs, expand transaction scale, and increase the level of division of labor and specialization. Financial innovation has caused a large amount of financial capital in the market to stay in the financial market, and various financial derivatives continue to appear and this kind of speculation has increased the risk of the bubble economy. The influencing factors of risk evaluation of listed companies include scale factor, profit factor, liquidity factor, and operation factor. That is, listed companies aim to make profits, and their operations are commercialized and investors and creditors pay more attention to the scale and profitability of enterprises. In the upward phase of the economic cycle, the increase in the profit of financing platforms makes the actual financing leverage ratio lower than the bank's expected financing leverage ratio. Banks increase their risk appetite, increase the expected financing leverage ratio, and provide more loans to financing platforms, resulting in the actual leverage ratio gradually rise. Figure 3 shows the flowchart of financial risk early-warning based on K-means clustering algorithm.

The ultimate line of defense against risks is to maintain sufficient own capital. Furthermore, because the free capital of commercial banks is limited after all, they must maintain special reserves, capital loss reserves, and bad loan reserves to compensate for the loss of loan principal and interest according to specific business conditions. In order to avoid dependence on the choice of measurement unit, the data should be standardized first and standardized measures try to give equal weight to all variables, which is useful when there is no prior knowledge about the data. However, users may want to give some variables greater weight and the data service aggregation layer exists to solve the performance of the computing service of the higher-level organization.

In the collaborative filtering recommendation algorithm, there are mainly similarity measurement methods such as cosine similarity, correlation similarity, and adjusted cosine similarity; cosine similarity is selected as the similarity measurement method, and the formula is

$$E_{ij}\left(u_{ij}, o_{ij}\right) = \int_{i,j=1}^{k} \frac{u_{ij} \cdot o_{ij}}{\left\|u_{ij}\right\| \cdot \left\|o_{ij}\right\|}, \tag{4}$$

where u_{ij} is the rating of user u on item i and item j; o_{ij} is the number of users who rate item i and item j at the same time; $\|u_{ij}\|$ is the set of rated items of user u; $\|o_{ij}\|$ is calculated from the target item collection of average offset items; and k is the number of elements in the collection.

The error sum of squares criterion calculates the sum of squares of distances from all observation points to their class centroids, and it is a criterion function p (x_{ij}) to measure the variability within a class:

$$p_{ij} = \left(\frac{1}{a_{ij}} - \frac{1}{s_{ij}}\right)\left(\frac{1}{d_{ij}} - \frac{1}{g_{ij}}\right), \tag{5}$$

where a_{ij} is the variable value of the j-th observation of the i-th type; s_{ij} is the mean value of the j-th variable of the i-th type; d_{ij} is the value of the j-th observation variable of the i-th type; and g_{ij} is the j-th variable of the i-th square of the mean difference.

In addition, it can buffer the results of each query and analysis without each time and can copy, update, and summarize the required data from the node according to the task. All start calculations from the bottom node again and users can perform query, report, analysis, and other services

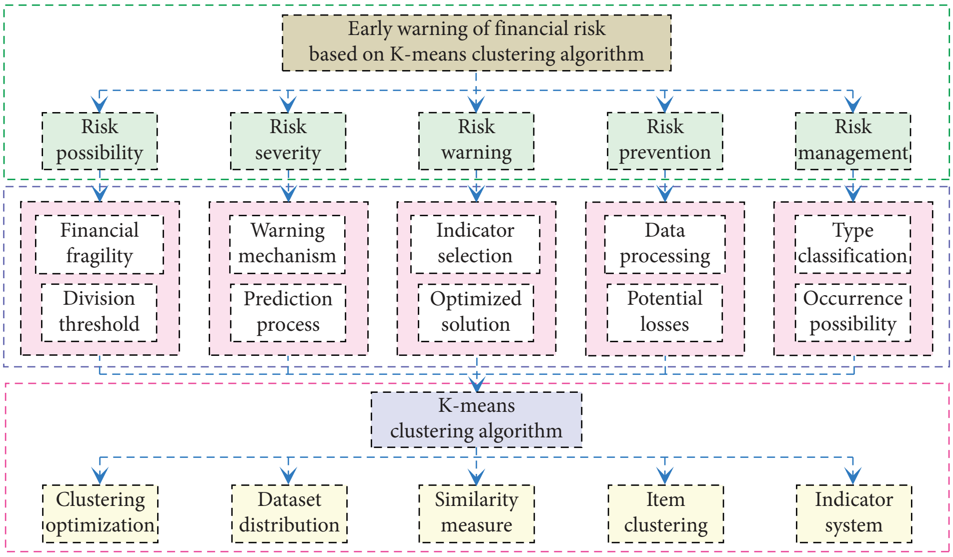


Figure 1: Framework of financial risk early-warning based on K-means clustering algorithm.

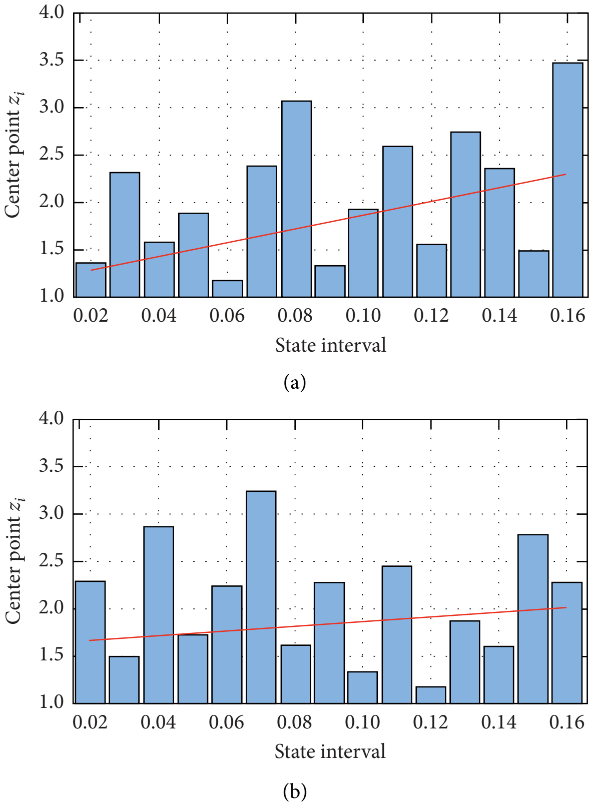


Figure 2: Distribution of center point in different state intervals for similarity measure (a) and item clustering (b).

through a unified access interface. The construction of the data center of the financial supervision information system includes the data center and the bureau data center; the data center includes the data processing center, and the data collection and information release center includes the data processing center and data center.

4. Early-Warning Model of Financial Risk Based on K-Means Clustering Algorithm

4.1. Type Classification of Financial Risk. Because many financial early-warning indicators are complex and related, it is necessary to reduce the dimensionality through principal

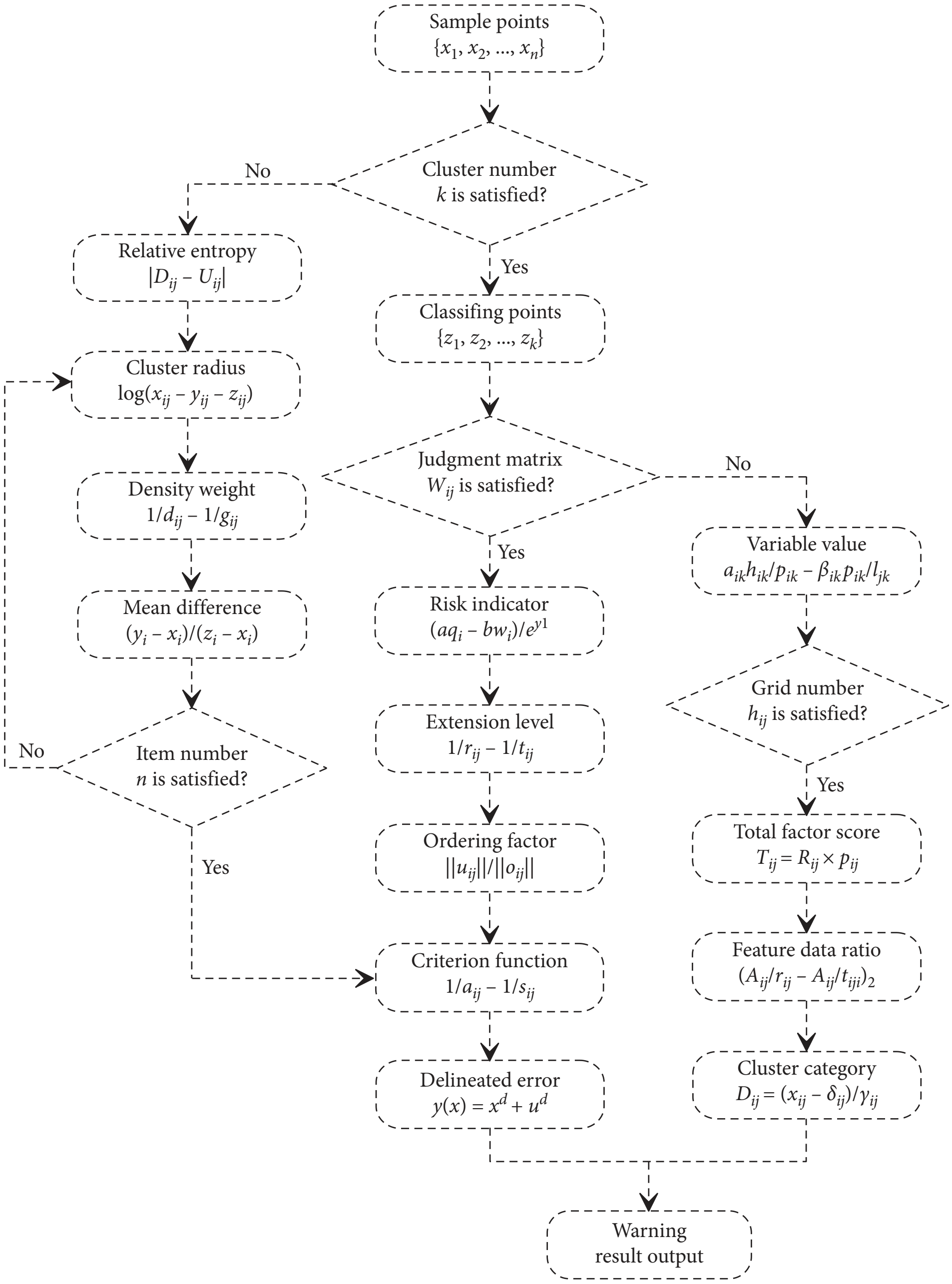


Figure 3: Flowchart of financial risk early-warning based on K-means clustering algorithm.

component analysis first, so that multiple indicators can be transformed into a few uncorrelated comprehensive indicators, and the information loss is small. The K-means cluster analysis method can effectively avoid the subjective negative impact caused by the artificial threshold value, so it can more accurately and objectively distinguish the state intervals of different financial risks. When the financial time series feature is read in, the feature is compared with all subnodes of the root node of the core tree, and the corresponding cluster center point with the closest distance is found, and the feature is added to the corresponding subtree (Figure 4). When financial risk early-warning indicator q_i is 0.2, 0.4, 0.6, and 0.8, respectively, both of the financial risk levels and early-warning error rates show increasing trends. When q_i is equal to 0.4, the risk level is generally high; when q_i is equal to 0.2, the risk level is generally low; when q_i is equal to 0.2, the error rate is generally high; when q_i is equal to 0.8, the error rate is generally low. The data are used as the category center point, and a new category node is added to the core tree. When the feature number is searched to the category node, compare the distance between the feature data and the center node.

Loan repayment forecast and customer credit policy analysis is very important to banking business. There are many factors that have varying degrees of impact on loan repayment performance and customer credit rating calculations. Feature selection and attribute correlation calculation help to identify important factors, eliminating

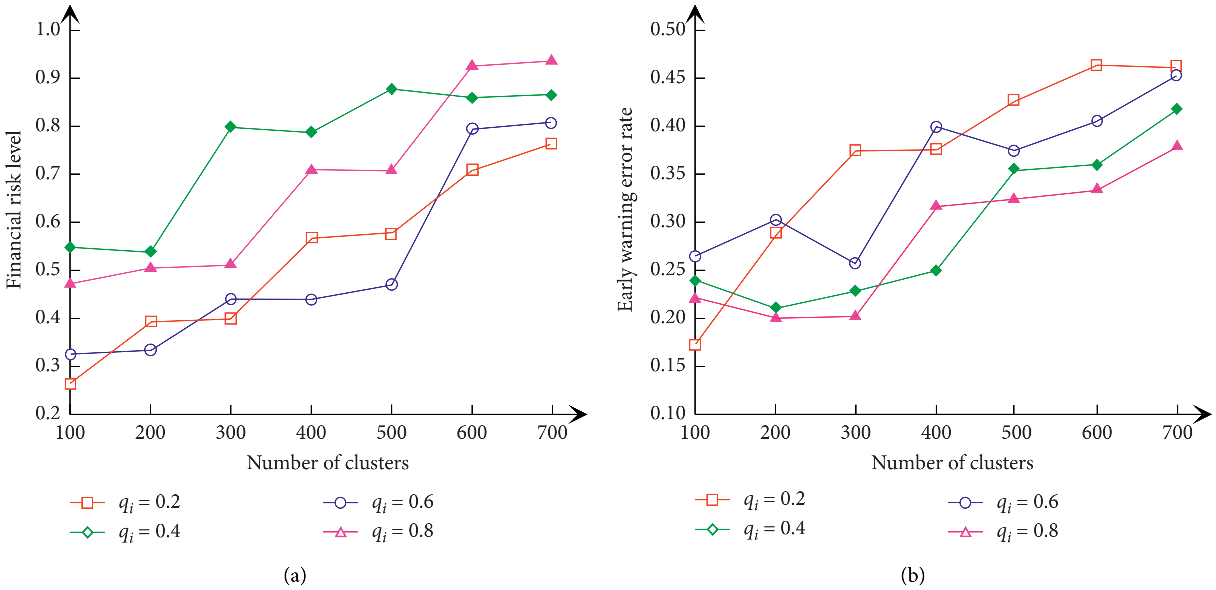


FIGURE 4: Financial risk levels (a) and early-warning error rate (b) at different number of clusters when financial risk early-warning indicator qi is 0.2, 0.4, 0.6, and 0.8, respectively.

nonrelated factors, for example, factors related to loan repayment risk include loan rate, loan maturity, debt ratio, repayment to income ratio, customer income level, education level, area of residence, and credit history. The bank can then adjust the loan issuance policy accordingly to grant loans to those who have been rejected before, but the basic information shows that they are relatively low-risk applications.

It is assumed that a total of k samples are added in the j time period in the i-th grid; it can be considered that the density weights of these k samples are all 1, and then the sum of the density weights in the i-th grid R_{ij} is

$$R_{ij} = \sum_{i,j=1}^{k} \alpha_{ij} \frac{h_{ij}}{p_{ij}} - \beta_{ij} \frac{p_{ij}}{l_{ij}}, \tag{6}$$

where h_{ij} is the number of grids entered by the j-th new sample in time i; l_{ij} is the number of dense grids added by the j-th new sample in time i; α_{ij} is the distance between sample point i and point j; and β_{ij} is the similarity between sample i and sample j.

The ratio of the contribution rate of the variance of each factor to the contribution rate of the total variance of the factor is calculated, and this is used as a weight for weighting, and the total factor score is summarized as

$$T_{ij} = R_{ij} \cdot p_{ij}. \tag{7}$$

As the core feature data in each category increases, the center point of each category needs to be recalculated; the core feature ratio is calculated as whether to update the cluster center point. It is supposed the number of core feature data is n, and the number of total feature data is m, and the core feature data ratio U_{ij} is calculated as follows:

$$U_{ij} = \left(\frac{A_{ij}}{T_{ij}} - \frac{A_{ij}}{R_{ij}} \right)^2, \tag{8}$$

where A_{ij} is the category divergence of the cluster; the smaller the A_{ij}, the more compact the data of the cluster center, the smaller the cluster radius, and vice versa.

The objects in the same cluster are similar to each other and different from objects in different clusters, which is an unsupervised learning process. The clustering algorithm automatically divides the big data set into several different clusters and similar subsets within the clusters according to the attributes of the data itself, so as to solve the problems of boundary value, mean square error, and cross-validation of credit scoring. Clustering analysis algorithms can be summarized into three different types: trying to find an optimal partition to divide the data into a specified number of clusters; trying to find a hierarchical method of clustering structure; and modeling potential clusters based on model probability method.

4.2. Warning Optimization of Financial Risk. The asset portfolio obtained by selecting stocks in different categories is better than the result of randomly selecting stocks from both the perspective of return and risk, that is, clustering is effective. The clustering results show that the clustering based on the scale curve can reveal the industry clustering and sector transfer characteristics of the stock market, while the clustering based on the linear trend characteristics of the time series mainly reflects the similarity between the fluctuations of individual stocks and stock indicator fluctuations. The difference between financial risk and general risk is the risk arising from financial activities such as capital borrowing and capital management (Figure 5). When q_i is equal to 0.2, the cluster purity firstly decreases from ~92% to ~86% (time window ranging 0 to 30), then remain nearly constant around ~86% (time window ranging 30 to 110); and then increase from ~86% to ~96% (time window ranging 110 to 150). When q_i is equal to 0.2, the cluster purity firstly increase from ~86% to ~95% (time window ranging 0 to 30), then decrease from ~86% to ~88% (time window ranging 30 to 110); and then increase from ~88% to ~94% (time window ranging 110 to 150). The functions of financial risk early-warning include timely grasp of trends, effective evaluation, and timely adoption of relevant regulatory measures, thereby reducing the harm of financial risks.

Financial risk early-warning work mainly includes risk identification, risk assessment, risk early warning, and risk treatment, which is similar to the process of the immune system from identifying antigens to eliminating them. Financial risk identification and assessment is to understand financial risks and their sources through investigations and using some qualitative and quantitative methods to measure the size of the risk is the first step in financial risk early-warning. For risks exceeding the level of the police, relevant mechanisms are adopted for risk warning. The former requires certain analysis and calculations to obtain, while the latter can be directly extracted from the data set. The construction of the data center of the financial supervision information system includes the data center and the bureau data center. Compared with the former, the determination steps are simpler, but the representativeness of the selected initial condensation points is usually lower. When financial risk early-warning decision makers encounter a risk event for the first time, they will be slow to deal with it, and may cause losses, but through the accumulation of experience, they can react quickly and improve risk prevention when similar risks strike and the efficiency of control.

Because the selected data has large differences in interval span and dimension, a standardized method will be used to preprocess the data to eliminate the influence of dimension:

$$D_{ij} = \frac{x_{ij} - \delta_{ij}}{\gamma_{ij}}, \tag{9}$$

where γ_{ij} is the overall mean and δ_{ij} is the overall standard deviation. As an expression of a measure of uncertainty, such a concept can also be used as a measure of the degree to which a certain probability distribution density $p(x_i)$ deviates from a given standard distribution $u(x_i)$, which is called relative entropy:

$$\left|D_{ij} - U_{ij}\right| = \log\left(x_{ij} - y_{ij} - z_{ij}\right). \tag{10}$$

The sum should be performed on all possible values of the feature; the smaller the relative entropy, the greater the difference between the two types of probability distributions, but when the two types of probability distributions is exactly the same, the relative entropy reaches the maximum.

When the user encounters similar risk events again, user can refer to the accumulated professional knowledge and experience in financial risk early-warning management in a timely manner. The process by which decision makers make decisions and solve problems reflects the continuous learning of risk problems, repeated memory of similar problems, and ultimately decision-making and feedback. The generation of financial risk identifiers depends on the characteristics of the risk factors in the corresponding gene bank; each financial risk early-warning factor group corresponds to a corresponding risk identifier.

5. Empirical Analysis

5.1. Empirical Experiment. This paper selects 60 multinational companies in the A-share market as the research objects. Given that these companies are all listed companies and are basically in a mature period, the comprehensive consideration of the solvency of the companies, operation ability, profitability, and risk control ability is analyzed based on the five-year corporate financial data from 2015 to 2019. After screening the indicators, this paper selects two evaluation indicators for the financial system and foreign economic and trade: the financial system mainly considers the stable development of the money market and the capital market. The indicators include growth rate, credit to output ratio, loan interest rate and deposits, interest rate ratio, financial institution deposit-loan ratio, stock financing, commodity housing price volatility, and stock market value volatility. Foreign economy and trade mainly considers import and export, foreign exchange transactions, and exchange rate fluctuations. Its main indicators include actual exchange rate volatility, foreign exchange reserve growth rate, foreign debt growth rate, and import and export growth rate. This paper first uses 2018 data as the basic sample for training and testing; then, in order to determine the consistency of the risk assessment standard, the final cluster center in 2018 is used as the evaluation standard for another three years, and the 2016, 2017, and 2019 samples are calculated point to the distance between the training sample in 2018 and the final cluster center, and then financial risk warning was conducted for 60 multinational companies in the past three years.

The profitability indicator mainly analyzes the ability of a company to make profits; the better the company's profitability means that it is likely to obtain enough cash to repay its debts, and its credit status will be better. Debt solvency is of great significance to creditors and insufficient solvency of an enterprise may result in creditors not being able to recover the principal and interest of bonds in full and in time. The growth ability of an enterprise indicates the long-term expansion ability of the enterprise and the future production and operation strength of the enterprise. Data cleaning routines usually include: filling in missing data values, smoothing noisy data, and identifying or removing outliers to resolve data inconsistencies. Problematic data often lead to distortion of the mining results and also causes the mining process to fall into chaos, resulting in unreliable output.

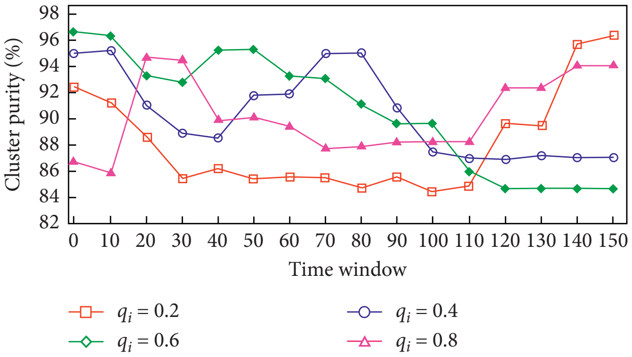


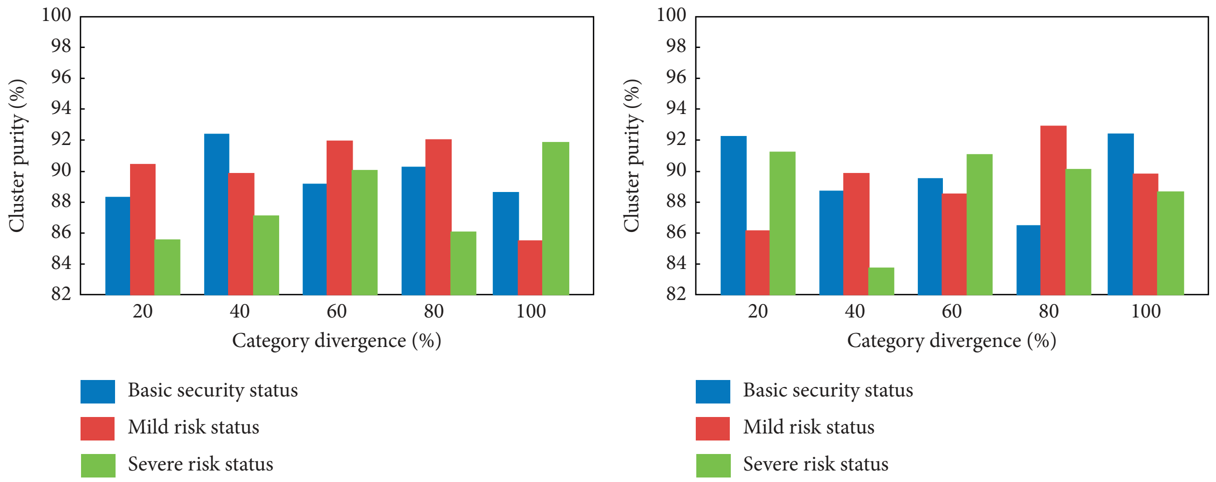
Figure 5: Cluster purities at different time windows when financial risk early-warning indicator qi is 0.2, 0.4, 0.6, and 0.8, respectively.

Although most data mining processes involve the processing of incomplete or noisy data, they are not robust, and often focus on how to avoid excessively accurate descriptions of the data by the mined patterns.

5.2. Result Analysis. The establishment of various mechanisms for predictive finance is mainly based on a variety of financial activities, this is the main content, and the object of establishment is the process of realization of all finance and is based on various basic theories related to finance. For reference, the system established by various advanced predictive financial-related technologies, various-system related indicators, various predictive models, and different signals are used to supervise the process of financial operation and obtain a series of supervision results and make financial decisions in various situations caused by these results. Principal component analysis is a statistical process that uses orthogonal transformation to convert a set of observations of possible related variables into a set of linear unrelated variable values called principal components. The definition of this conversion is that the first principal component has the limit possibly large variance, and each subsequent component in turn has the highest variance under constraints (Figure 6). There are two commonly used methods to determine the initial condensation point, one is to synthesize the initial condensation point, and the other is to use actual observations as the initial condensation point. The former requires certain analysis and calculations to obtain, while the latter can be directly extracted from the data set. Compared with the former, the determination steps are simpler, but the representativeness of the selected initial condensation points is usually lower.

Cluster analysis is a kind of unguided observational learning. Its basic principle is based on the properties of the sample itself, without any model for reference or following, that is, without prior knowledge, mathematical methods are used to follow a certain similarity or difference indicator, calculate the similarity between samples, and cluster the samples according to this similarity. The main advantages of K-means clustering algorithm are simple, fast, and efficient and scalable for big data and the clustering quality of the standard K-means clustering algorithm is highly dependent on the initial clustering center. Using a random initial clustering center on the one hand may get very poor clustering results; on the other hand, it will make the clustering results of the algorithm unstable. A good clustering method produces high-quality clusters: high intra-cluster similarity and low intercluster similarity. There are two criteria for evaluating clustering quality: internal quality evaluation criteria and external quality evaluation criteria; the internal quality evaluation criteria evaluate the clustering effect by calculating the average similarity within the cluster, the average similarity between the clusters or the overall similarity (Figure 7). When dealing with cubes, the size reduction technology can be used as the dimension of the two-dimensional transformation, and the dimensionality reduction actually uses some means to process higher-dimensional data into lower-dimensional data.

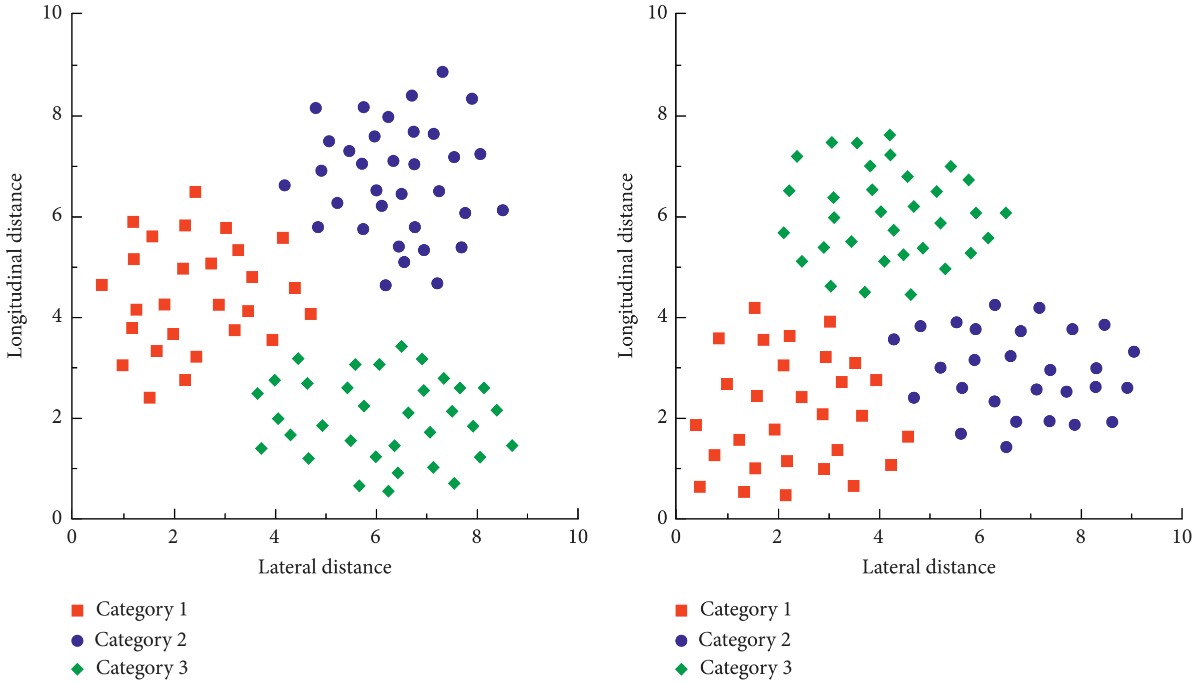
Companies in the start-up period are generally small in scale, simple in corporate governance, more concentrated in policy formulation and management, and more focused on innovation. At the same time, such companies also face a lot of uncertainty, so the book-to-market ratio is higher, and the company's characteristics are more risky. Because each group will make the centroid step smaller and smaller and gradually converge, the program usually presets a threshold to avoid excessive computing time. The last position of the centroid is the basis for us to delimit the cluster, which determines the cluster to which each sample belongs. The larger the cluster size, the more samples that are predicted to be true, and the easier the true samples are to be included, so the recall rate will increase. But if the cluster size becomes larger, there will be many false samples included in the cluster, so the accuracy rate will be reduced, and vice versa. Therefore, recall rate and precision rate are a pair of contradictory indicators; compared with other companies, the probability of financial institutions losing money or even going bankrupt is very low. Even if similar negative events occur, the rights and interests of investors can be protected



(a)

(b)

Figure 6: Cluster purities of three different risk statuses for similarity measure (a)and item clustering (b).



(a)

(b)

Figure 7: K-means clustering results of financial risk warning for similarity measure (a) and item clustering (b).

to a certain extent. The higher the tangible assets, the better the quality of the company's assets; the current debt ratio reflects the short-term liabilities of the listed company; the lower the current liabilities, the lower the company's short-term debt burden, and the better the company's prospects.

6. Conclusions

This paper proposed a financial risk indicator system based on the K-means clustering algorithm, performed indicator selection and data processing, constructed a financial risk early-warning model based on the K-means clustering algorithm, conducted the classification of financial risk types and optimization of financial risk control, and finally carried out an empirical experiments and its result analysis. Using some qualitative and quantitative methods to measure the magnitude of risk is the first step in financial risk warning, for risks that exceed the level of the police, relevant mechanisms are adopted to carry out risk warning. The use of K-means algorithm is actually the process of solving optimization and the objective function exceeds the minimum value in some parts of the interval, but only the global minimum exists. The standard function requires the sum of the squares of the search direction during the search operation error is consistent, and model clustering requires a lot of prior knowledge to be able to give a suitable clustering model. The larger the cluster size, the more samples that are predicted to be true, and the easier the true samples are to be included, so the recall rate will increase. The main advantages of K-means clustering algorithm are simple, fast, efficient, and scalable for big data and the clustering quality of the standard K-means clustering algorithm is highly dependent on the initial clustering center. Although density clustering is universal, it lacks generality; while dividing clustering requires a given number of clusters before clustering, the clustering effect and clustering speed are better. Although most data mining processes involve the processing of incomplete or noisy data, they are not robust and often focus on how to avoid excessively accurate descriptions of the data by the mined patterns. The study results show that the K-means clustering method can effectively avoid the subjective negative impact caused by artificial division thresholds, continuously optimize the prediction process of financial risk, and redistribute target dataset to each cluster center for obtaining optimized solution, so the algorithm can more accurately and objectively distinguish the state interval of different financial risks, determine risk occurrence possibility and its severity, and provide a scientific basis for risk prevention and management. The study results of this paper provide a reference for further researches on financial risk early-warning based on K-means clustering algorithm.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

References

[1] C. Diks, C. Hommes, and J. Wang, "Critical slowing down as an early warning signal for financial crises?" *Empirical Economics*, vol. 57, no. 4, pp. 1201–1228, 2019.

[2] J. F. Kölbel, T. Busch, and L. M. Jancso, "How media coverage of corporate social irresponsibility increases financial risk," *Strategic Management Journal*, vol. 38, no. 11, pp. 2266–2284, 2017.

[3] K. Bouslah, L. Kryzanowski, and B. M'Zali, "Social performance and firm risk: impact of the financial crisis," *Journal of Business Ethics*, vol. 149, no. 3, pp. 643–669, 2018.

[4] S. Srinivasan and T. Kamalakannan, "Multi criteria decision making in financial risk management with a multi-objective genetic algorithm," *Computational Economics*, vol. 52, no. 2, pp. 443–457, 2018.

[5] M. Z. Hossain, M. N. Akhtar, R. B. Ahmad, and M. Rahman, "A dynamic K-means clustering for data mining," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 13, no. 2, pp. 521–526, 2019.

[6] R. Jothi, S. K. Mohanty, and A. Ojha, "DK-means: a deterministic k-means clustering algorithm for gene expression analysis," *Pattern Analysis and Applications*, vol. 22, no. 2, pp. 649–667, 2019.

[7] P. M. Shakeel, S. Baskar, V. S. Dhulipala, and M. M. Jaber, "Cloud based framework for diagnosis of diabetes mellitus using K-means clustering," *Health Information Science and Systems*, vol. 6, no. 1, pp. 1–7, 2018.

[8] C. Slamet, A. Rahman, M. A. Ramdhani, and W. Darmalaksana, "Clustering the verses of the Holy Qur'an using K-means algorithm," *Asian Journal of Information Technology*, vol. 15, no. 24, pp. 5159–5162, 2016.

[9] S. Bekiros, D. K. Nguyen, L. Sandoval Junior, and G. S. Uddin, "Information diffusion, cluster formation and entropy-based network dynamics in equity and commodity markets," *European Journal of Operational Research*, vol. 256, no. 3, pp. 945–961, 2017.

[10] K. Balakrishnan, R. Watts, and L. Zuo, "The effect of accounting conservatism on corporate investment during the global financial crisis," *Journal of Business Finance and Accounting*, vol. 43, no. 5-6, pp. 513–542, 2016.

[11] N. Coombs, "What is an algorithm? Financial regulation in the era of high-frequency trading," *Economy and Society*, vol. 45, no. 2, pp. 278–302, 2016.

[12] D. K. Sharma, S. K. Dhurandher, D. Agarwal, and K. Arora, "kROp: k-Means clustering based routing protocol for opportunistic networks," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 4, pp. 1289–1306, 2019.

[13] B. Becker and V. Ivashina, "Financial repression in the European sovereign debt crisis," *Review of Finance*, vol. 22, no. 1, pp. 83–115, 2018.

[14] P. Fränti and S. Sieranoja, "K-means properties on six clustering benchmark datasets," *Applied Intelligence*, vol. 48, no. 12, pp. 4743–4759, 2018.

[15] D. Musto, G. Nini, and K. Schwarz, "Notes on bonds: illiquidity feedback during the financial crisis," *The Review of Financial Studies*, vol. 31, no. 8, pp. 2983–3018, 2018.

[16] K. Valaskova, T. Kliestik, and M. Kovacova, "Management of financial risks in Slovak enterprises using regression analysis," *Oeconomia Copernicana*, vol. 9, no. 1, pp. 105–121, 2018.

[17] D. Abuaiadah, "Using bisect k-means clustering technique in the analysis of Arabic documents," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 15, no. 3, pp. 1–13, 2016.

[18] D. Fernández-Arias, M. López-Martín, T. Montero-Romero, F. Martínez-Estudillo, and F. Fernández-Navarro, "Financial soundness prediction using a multi-classification model: evidence from current financial crisis in OECD banks," *Computational Economics*, vol. 52, no. 1, pp. 275–297, 2018.

[19] I. Korol and A. Poltorak, "Financial risk management as a strategic direction for improving the level of economic security of the state," *Baltic Journal of Economic Studies*, vol. 4, no. 1, pp. 235–241, 2018.

[20] I.-D. Borlea, R.-E. Precup, F. Dragan, and A.-B. Borlea, "Centroid update approach to K-means clustering," *Advances in Electrical and Computer Engineering*, vol. 17, no. 4, pp. 3–10, 2017.

[21] S. Sankhwar, D. Gupta, K. C. Ramya, S. Sheeba Rani, K. Shankar, and S. K. Lakshmanaprabu, "Improved grey wolf optimization-based feature subset selection with fuzzy neural classifier for financial crisis prediction," *Soft Computing*, vol. 24, no. 1, pp. 101–110, 2020.

[22] W. Wang, F. Xia, H. Nie et al., "Vehicle trajectory clustering based on dynamic representation learning of internet of vehicles," *IEEE Transactions on Intelligent Transportation Systems*, p. 1, 2020.

[23] Z. Liu, W. Wei, H. Wang, Y. Zhang, Q. Zhang, and S. Li, "Intrusion detection based on parallel intelligent optimization feature extraction and distributed fuzzy clustering in WSNs," *IEEE Access*, vol. 6, pp. 72201–72211, 2018.

[24] S. Xia, D. Peng, D. Meng et al., "A fast adaptive k-means with No bounds," in *Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence*, p. 1p. 1, November 2020.