

Research Article

Research on Artificial Intelligence Classification and Statistical Methods of Financial Data in Smart Cities

Xuezhong Fu 

School of Economics, Hainan University, Haikou 570228, China

Correspondence should be addressed to Xuezhong Fu; xuezhongfu2572021@126.com

Received 3 November 2021; Revised 7 December 2021; Accepted 15 December 2021; Published 5 January 2022

Academic Editor: Huihua Chen

Copyright © 2022 Xuezhong Fu. 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.

In order to improve the effect of financial data classification and extract effective information from financial data, this paper improves the data mining algorithm, uses linear combination of principal components to represent missing variables, and performs dimensionality reduction processing on multidimensional data. In order to achieve the standardization of sample data, this paper standardizes the data and combines statistical methods to build an intelligent financial data processing model. In addition, starting from the actual situation, this paper proposes the artificial intelligence classification and statistical methods of financial data in smart cities and designs data simulation experiments to conduct experimental analysis on the methods proposed in this paper. From the experimental results, the artificial intelligence classification and statistical method of financial data in smart cities proposed in this paper can play an important role in the statistical analysis of financial data.

1. Introduction

With the acceleration of economic globalization, financial information has become more transparent and authentic. In this context, accounting standards are converging faster and faster globally. According to incomplete statistics, there are nearly 120 countries and regions that have adopted the International Financial Reporting Standards (IFRS) in the global plan or have adopted the International Financial Reporting Standards (IFRS). The international financial crisis that broke out in 2008 made people realize that improving the transparency of accounting information and formulating a set of globally unified and high-quality accounting standards are vital to the stability and healthy development of the global financial system and capital markets. In this context, countries and regions have accelerated discussions and research on the international convergence of accounting standards [1].

Because keyword retrieval technology cannot meet the needs of some occasions, in recent years, some new researches have begun to focus on improving the efficiency and accuracy of information retrieval technology. Information extraction technology is one of the technologies for

obtaining specific events or the relationship between events and events. It is a process of extracting structured and unambiguous information from unstructured free text or other information resources. Information extraction technology can not only filter out information that is not useful to users, but also generate specific information that users are interested in. Information extraction technology integrates artificial intelligence and natural language processing technology and plays an indispensable role in the field of information retrieval. To evaluate the market value of a company, such as the evaluation of intangible assets, all financial data of the company and the financial data of the industry are required, including the company's financial statements and notes to the statements, relevant industry financial news, and macro- and microeconomic data [2]. Most of these data are stored on the Internet in formats such as web pages (HTML) and PDF and are distributed on various websites that provide financial information. At present, people can only manually collect and sort relevant financial data from different data sources, carefully sort and analyze and filter out the data they want, but there is no place to directly get all the data they want. This takes a lot of manpower and time, and most of the work is repetitive work.

According to the preestablished evaluation model, the financial data evaluation results are obtained, which will greatly improve the efficiency of decision-making and provide support for financial-related decisions such as stock investment, mergers and acquisitions, and financial risk assessment [3].

This article combines big data technology to conduct financial data classification research and explores intelligent methods suitable for contemporary financial data classification.

2. Related Work

The literature [4] put forward some suggestions on the classification of financial assets. It includes two parts: one is the understanding of the different levels of financial transactions, and the other is how to deal with the various levels of relationships in the classification of financial instruments during the transaction process. The literature [5] conducted a comparative analysis on the four categories of financial assets and believed that financial assets should also be included in the foreclosed assets, and the financial assets in this part should be handled in accordance with the relevant regulations in the four categories. The literature [6] specifically analyzed and explained the revision of the classification of nonfinancial assets and financial assets in the national economic accounting system. The literature [7] studied the classification of financial instruments and the measurement model of financial instruments, explored the standard and root causes of the classification of financial instruments, and studied the future direction of the classification standard. According to the current classification standards of financial instruments, the literature [8] elaborated the classification standards of held-to-maturity investments in detail. The literature [9] believed that, under the current standards, the main reason for the strong subjectivity of the classification of financial assets lies in the intention of the management. Therefore, there may be errors in the classification of financial assets in actual work, and the relevant discrimination methods are given in the article.

With the changes in the economic environment, follow-up measurement issues and other related issues derived from the classification of financial assets have gradually emerged, and scholars have launched relevant discussions on these issues. Since the promulgation of the new accounting standards and the full implementation of the financial instrument standards, many scholars have begun to pay attention to whether the problem of financial asset classification will bring earnings management opportunities to enterprises. Therefore, most of the research and discussion on this issue have been concentrated in recent years. Literature [10] believes that the current research does not clearly define the criteria for the classification of financial assets, especially for transactional financial assets and available-for-sale financial assets, which only require classification according to the holding intention and ability of the enterprise. Therefore, companies can make full use of the subjectivity of financial asset classification to manipulate corporate profits. Literature [11] made a specific analysis on

the impact of the classification of equity financial assets on the basis of a full analysis of corporate equity financial assets. Literature [12] conducted an empirical study on the classification of listed companies' financial assets under the new standards. The research perspective is management, accounting policies, and earnings management. Studies have shown that when the initial classification of financial assets is performed, listed companies have obvious intentions to hold financial assets and can be clearly classified; however, during the holding period of financial assets, listed companies have more room for profit operation. The accounting treatment of available-for-sale financial assets can be managed according to the management's wishes. Literature [13] analyzed the classification standards of financial instruments of listed companies and concluded that because the classification of financial instruments is subjective and selective, holders will choose according to their own interests to maximize the reliability of financial information. Literature [14] investigated and studied the classification of financial instruments of listed companies and concluded that the classification of financial instruments of listed companies has a certain tendency according to the wishes of management. It can be found that many scholars have discovered in their research that the classification of financial assets will indeed bring opportunities for companies to manage earnings. The classification of financial assets is closely related to fair value measurement, and some scholars have discussed this. Literature [15] puts forward new insights in the classification, reclassification, and measurement of financial assets to prevent corporate management from adjusting corporate profits at will. Research in this area also exists in the field of practice. Literature [16], in studying the classification of financial instruments of securities companies and the fair value of financial instruments, concluded that the accounting information disclosure system of enterprises on financial instruments is still not perfect. Commercial banks hold the largest scale of financial assets, so some scholars have carried out research on the impact of financial asset classification on commercial banks' profits. Literature [17] analyzed the classification of financial instruments of commercial banks and concluded that the management of the enterprise has a risk appetite for the classification of financial instruments, and it replaces liquidity with robustness. Literature [18] uses the data of Liangmianzhen, Minmetals Development, and Qianjiang Biochemical Company as examples to discuss the earnings management of listed companies based on the perspective of financial asset classification and accounting confirmation differences. Literature [19] compares the differences between IFRS9 and the current accounting standards in the classification of financial assets and analyzes the differences between IAS39 and IFRS9 financial asset classification of 16 listed banks, showing the possible impact of commercial banks in the future. Financial asset standards are closely related to the financial environment. Literature [20] puts forward suggestions for strengthening supervision and protection. Literature [21] discusses the problems that financial asset reclassification may bring to corporate profits and losses, including the recognition of corporate profits and losses in

financial asset holdings, the impact of financial asset impairment on profits and losses, and the possible problems of financial asset reclassification.

3. Financial Data Intelligent Mining Classification Method

We assume that there are p indicators in the actual problem under discussion, and we regard these p indicators as p random variables, denoted as X_1, X_2, \dots, X_p . Principal component analysis is to transform the problem of p indicators into a problem of discussing the linear combination of p indicators. These new indicators F_1, F_2, \dots, F_k ($k \leq p$) fully reflect the information of the original indicators in accordance with the principle of preserving the amount of main information and are independent of each other.

This process of reducing multiple indicators to a few comprehensive indicators is mathematically called dimensionality reduction. The usual method of principal component analysis is to seek the linear combination F_i of the original indicators.

$$\begin{aligned} F_1 &= \mu_{11}X_1 + \mu_{21}X_2 + \dots + \mu_{p1}X_p, \\ F_2 &= \mu_{12}X_1 + \mu_{22}X_2 + \dots + \mu_{p2}X_p, \\ &\dots \\ F_p &= \mu_{1p}X_1 + \mu_{2p}X_2 + \dots + \mu_{pp}X_p. \end{aligned} \quad (1)$$

It satisfies the following conditions:

- (1) The sum of the squares of the coefficients of each principal component is 1, that is,

$$\mu_{1i}^2 + \mu_{2i}^2 + \dots + \mu_{pi}^2 = 1. \quad (2)$$

- (2) The principal components are independent of each other; that is, there is no overlapping information, that is,

$$\text{Cov}(F_i, F_j) = 0, \quad i \neq j, i, j = 1, 2, \dots, p. \quad (3)$$

- (3) The variance of the principal components decreases successively, and the importance decreases successively, namely,

$$\text{Var}(F_1) \geq \text{Var}(F_2) \geq \dots \geq \text{Var}(F_p). \quad (4)$$

The conclusion of the two linear algebras is as follows:

- (1) If A is a real symmetric matrix of order p , then an orthogonal matrix U must be found to make

$$U^{-1}AU = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_p \end{bmatrix}_{p \times p}. \quad (5)$$

Among them, λ_i ($i = 1, 2, \dots, p$) is the characteristic root of A .

- (2) If the unit eigenvector corresponding to the characteristic root of the above matrix is μ_1, \dots, μ_p ,

$$U = (\mu_1, \mu_2, \dots, \mu_p) = \begin{bmatrix} \mu_{11} & \mu_{12} & \dots & \mu_{1p} \\ \mu_{21} & \mu_{22} & \dots & \mu_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{p1} & \mu_{p2} & \dots & \mu_{pp} \end{bmatrix}. \quad (6)$$

Then, the eigenvectors corresponding to different eigenvalues of the real symmetric matrix U are orthogonal, that is, $U^T U = U U^T = I$ [22].

Conclusion: $\sum x$ is the covariance matrix of the random vector $X = X_1 + X_2 + X_3 + \dots + X_p$. It has eigenvalue $\lambda_1, \lambda_2, \dots, \lambda_p$ and eigenvector $\mu_1, \mu_2, \dots, \mu_p$, where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$. Then, the principal components are

$$\begin{aligned} F_1 &= \mu_{11}X_1 + \mu_{21}X_2 + \dots + \mu_{p1}X_p, \\ F_2 &= \mu_{12}X_1 + \mu_{22}X_2 + \dots + \mu_{p2}X_p, \\ &\dots \\ F_p &= \mu_{1p}X_1 + \mu_{2p}X_2 + \dots + \mu_{pp}X_p. \end{aligned} \quad (7)$$

At this time, $\text{Var}(F_i) = U^T \sum x U = \lambda_i$, $i = 1, 2, \dots, p$. $\text{Cov}(F_i, F_j) = 0$, $i \neq j, i, j = 1, 2, \dots, p$.

It is written in matrix form: $F = U^T X$.

$$U = (\mu_1, \dots, \mu_p) = \begin{bmatrix} \mu_{11} & \mu_{12} & \dots & \mu_{1p} \\ \mu_{21} & \mu_{22} & \dots & \mu_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{p1} & \mu_{p2} & \dots & \mu_{pp} \end{bmatrix}, \quad (8)$$

$$X = (X_1, X_2, \dots, X_p).$$

The two basic concepts are as follows:

- (1) Contribution rate: the proportion of the variance of the i -th principal component in the total variance, $\lambda_i / \sum_{i=1}^p \lambda_i$, is called the contribution rate, which reflects how much information the original P indicators have and how comprehensive they are.
- (2) Cumulative contribution rate: the comprehensive ability of the first k_i principal components is described by the variance of these k principal components and the proportion $\sum_{i=1}^k \lambda_i / \sum_{i=1}^p \lambda_i$ in the total variance, which is called the cumulative contribution rate.

One of the purposes of our principal component bond analysis is to replace the original p indicators with as few principal components F_1, F_2, \dots, F_k ($k \leq p$) as possible. In actual work, the number of principal components depends on the amount of information that can reflect the original variable, that is, the cumulative contribution rate.

Principal component analysis is a method that converts multiple indicators into a few indicators and can maintain the correlation of the largest original data. The more important variance contribution β_i ($i = 1, 2, \dots, k$) in principal component analysis represents the maximum value of the

variance contribution of the i -th common factor after eliminating the influence of $(i-1)$ common factors. It is mainly used to measure the importance of the i -th common factor. Therefore, the corresponding evaluation model can be established with β as the weight: $F = \beta_1 F_1 + \beta_2 F_2 + \dots + \beta_k F_k$. Among them, F_1, F_2, \dots, F_k is the corresponding k common factors used to comprehensively describe the original indicators, and the comprehensive score is calculated and sorted.

In summary, the calculation process of finding the principal components has the following steps. We set n samples, where each sample has m data, denoted as

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix}. \quad (9)$$

- (1) Standardization of sample data: in order to achieve the standardization of sample data, the mean and variance of the sample data need to be required. The standardization of sample data is based on the mean and variance of the data. The essence of standardization is to transform the sample into standardized data with a mean value of 0 and a variance of 1. That is, the normalized transformation of the column is

$$x_{ij}^* = \frac{(x_{ij} - \bar{x}_j)}{\delta_j}. \quad (10)$$

Among them,

$$i = 1, 2, \dots, n,$$

$$j = 1, 2, \dots, m,$$

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}, \quad (11)$$

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

The resulting standardized matrix X^* is written as

$$X^* = \begin{bmatrix} x_{11}^* & \cdots & x_{1m}^* \\ \vdots & \ddots & \vdots \\ x_{n1}^* & \cdots & x_{nm}^* \end{bmatrix}. \quad (12)$$

- (2) Calculation of the correlation matrix: for given n samples, we use a computer to calculate the correlation coefficient matrix of the indicator variables:

$$R = \begin{bmatrix} r_{11} & \cdots & r_{1m} \\ \vdots & \ddots & \vdots \\ r_{m1} & \cdots & r_{mm} \end{bmatrix} = \frac{1}{n} X'^* X. \quad (13)$$

Among them,

$$r_{ij} = \frac{1}{n} \sum_{i=1}^n X_{ij} X_{ik} = \frac{1}{n} x_j' x_k, \quad j, k = 1, 2, \dots, m. \quad (14)$$

- (3) Find eigenvalues and eigenvectors. The obtained correlation matrix is R to solve the characteristic equation:

$$|R - \lambda f| = 0. \quad (15)$$

By solving the characteristic equation, k eigenvalues ($i = 1 \sim m$) and the eigenvector corresponding to each eigenvalue can be obtained:

$$Q_i = (a_{i1}, a_{i2}, \dots, a_{ip}), \quad i = 1 \sim k. \quad (16)$$

And the eigenvectors corresponding to $\lambda_1 > \lambda_2 > \lambda_3 > \lambda_k > 0$ are orthogonal to each other.

By the above method, k ($k \leq p$) principal components can be obtained. $\lambda_i / \sum_{i=1}^k \lambda_i$ is the contribution rate of the i -th principal component, denoted as β_i , which is

$$\beta_i = \frac{\lambda_i}{\sum_{i=1}^k \lambda_i}. \quad (17)$$

Among the k principal components, the sum of the contribution rates of the first q principal components is the cumulative contribution rate of the first q principal components, denoted as α :

$$\alpha = \frac{\sum_{i=1}^q \lambda_i}{\sum_{i=1}^k \lambda_i}. \quad (18)$$

The number of principal components can be determined by the cumulative contribution rate. Usually the cumulative contribution rate $\alpha \geq 0.85$ as the standard. For the selected q principal components, if the cumulative contribution rate reaches 85%, that is, $\alpha \geq 0.85$, the principal components can be determined as q . It represents the selected q principal components and basically retains the information of the original p variables. When determining the number of principal components, the number of principal components should be minimized under the condition of the cumulative contribution rate constraint.

- (4) Find the factor load α_i :

$$\alpha_i = \sqrt{\lambda_i} \alpha_i. \quad (19)$$

According to formula (19), we calculate the factor loading matrix and then calculate the score of each factor:

$$F_i = \alpha_i x, \quad i = 1, 2, \dots, k. \quad (20)$$

- (5) According to the numerator of the factor and the size of the contribution rate, we calculate the comprehensive score:

$$F = \beta_1 F_1 + \beta_2 F_2 + \cdots + \beta_k F_k. \quad (21)$$

Finally, it is sorted according to the comprehensive score. PCA analyzes the system with fewer q indicators instead of the original p indicators, which brings great convenience to the comprehensive evaluation of the system.

A large amount of original detection value data flow information is used to detect lag correlation using the BRAID method, and a series of time series with lag correlation can be found, and the lag time of each time series relative to the original reference series can be calculated.

We assume that there are n time series with a lagging correlation after the detection of the BRAID method. Each time series has a piece of data. Since the original detection data are data values obtained at equal intervals of time, a time variable t is introduced in the representation of the time series, and the data value corresponding to each time point is $x(t)$. Then the original multidimensional data flow is denoted as

$$X = (X_1(t), X_2(t), \dots, X_n(t)), \quad t = 0, 1, \dots, m-1. \quad (22)$$

This can be written in matrix form, namely,

$$X = \begin{bmatrix} x_1(0) & x_1(1) & \cdots & x_1(m-1) \\ x_2(0) & x_2(1) & \cdots & x_2(m-1) \\ \vdots & \vdots & \ddots & \vdots \\ x_n(0) & x_n(1) & \cdots & x_n(m-1) \end{bmatrix}. \quad (23)$$

The lag time is recorded as $\Delta t_1, \Delta t_2, \dots, \Delta t_n$.

Among them, $\Delta t_1 = 0$ (that is, x is the original reference sequence), and $\Delta t_2, \dots, \Delta t_n$ is the lag time of the sequence x_2, \dots, x_n relative to the sequence x_1 .

The first principal component analysis method is applied to "aligned" data streams; that is, the data is synchronized. This article discusses the lag-related multidimensional data flow. The data that needs principal component analysis is not synchronized, but has a certain lag time relative to the original time series. Therefore, it is necessary to perform principal component analysis on the detected multidimensional data stream with lagging correlation. The first task is to remove the lagging part of each time series, and "align" the data flow, that is, temporarily ignore the time parameters, and only keep the parts with similar changing trends in each time series, in order to achieve the purpose of multidimensional data flow synchronization. Figure 1 shows the unsynchronized time series, and Figure 2 shows the synchronized time series.

If the data corresponding to w time points in each data stream are obtained, the synchronized multidimensional data stream is recorded as

$$X = \begin{bmatrix} x_1(0) & x_1(1) & \cdots & x_1(w-1) \\ x_2(\Delta t_2) & x_2(\Delta t_2+1) & \cdots & x_2(\Delta t_2+w-1) \\ \vdots & \vdots & \ddots & \vdots \\ x_i(\Delta t_i) & x_i(\Delta t_i+i) & \cdots & x_i(\Delta t_i+w-1) \\ \vdots & \vdots & \ddots & \vdots \\ x_n(\Delta t_n) & x_n(\Delta t_n+i) & \cdots & x_n(\Delta t_n+w-1) \end{bmatrix}. \quad (24)$$

That is, $X = (X_i(\Delta t_i), X_i(\Delta t_i+1), \dots, X_i(\Delta t_i+w-1))$, where $i = 1, 2, \dots, n$.

For the above synchronized multidimensional data stream with delayed correlation, we apply the principal component analysis method to find the principal components. The steps of calculating the principal components include standardizing the sample data, calculating the correlation matrix, finding the eigenvalues and eigenvectors, and finding the principal components.

- (1) Standardize multidimensional data streams. In order to achieve the standardization of multidimensional data streams, we should find the mean and variance of each data stream. The standardization of the data stream is based on the mean and variance of each data in the data stream. The essence of standardization is to transform the data stream into standardized data with a mean value of 0 and a variance of 1, which is to perform a standardized transformation on the column of X :

$$x_i^*(j) = \frac{(x_i(j) - \overline{x(j)})}{\delta_j}. \quad (25)$$

Among them,

$$i = 1, 2, \dots, n,$$

$$j = 1, 2, \dots, m,$$

$$\overline{x(j)} = \frac{1}{n} \sum_{i=1}^n x_i(j), \quad (26)$$

$$\delta_j^2 = \frac{1}{n} \sum_{i=1}^n (x_i(j) - \overline{x(j)})^2.$$

Then we get the standardized matrix, denoted as

$$X = \begin{bmatrix} x_1(0) & x_1(1) & \cdots & x_1(w-1) \\ x_2(0) & x_2(1) & \cdots & x_2(w-1) \\ \vdots & \vdots & \ddots & \vdots \\ x_n(0) & x_n(1) & \cdots & x_n(w-1) \end{bmatrix}. \quad (27)$$

- (2) Calculate the correlation matrix. For the w data streams with lagging correlations that have been detected, we use a computer to calculate the correlation coefficient matrix of the indicator variables:

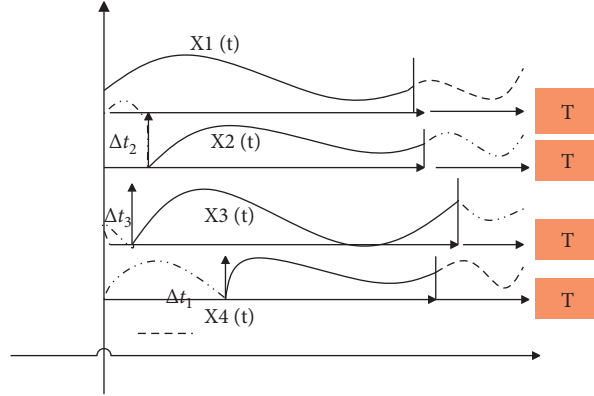


FIGURE 1: Unsynchronized time series.

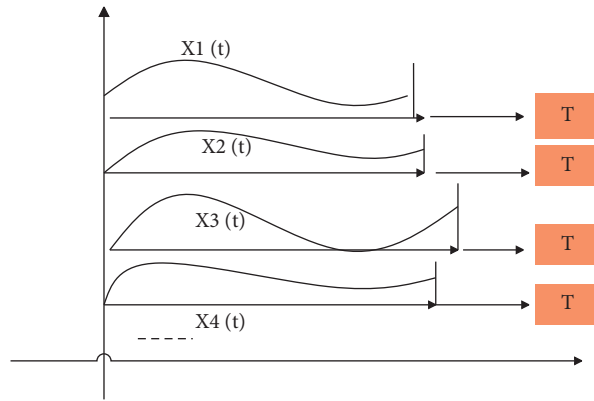


FIGURE 2: Synchronized time series.

$$R = \begin{bmatrix} r_{11} & \cdots & r_{1m} \\ \vdots & \ddots & \vdots \\ r_{m1} & \cdots & r_{mm} \end{bmatrix} = \frac{1}{n} X'X. \quad (28)$$

Among them,

$$r_{ij} = \frac{1}{n} \sum_{i=1}^n X_{ij}X_{ik} = \frac{1}{n} x_j'x_k, \quad j, k = 1, 2, \dots, m. \quad (29)$$

- (3) Find eigenvalues and eigenvectors. The obtained correlation matrix is R to solve the characteristic equation: $|R - \lambda f| = 0$.

By solving the characteristic equation, w eigenvalues can be obtained, and the eigenvector corresponding to each eigenvalue: $U_i = (u_{i1}, u_{i2}, \dots, u_{iw})$, $i = 1 \sim w$, and $\lambda_1 > \lambda_2 > \dots > \lambda_w > 0$, and its corresponding eigenvector are orthogonal to each other.

- (4) Find the principal components. The algorithm can obtain w principal components by the above method. $\lambda_i / \sum_{i=1}^w \lambda_i$ is the contribution rate of the i -th principal component, denoted as β_i , namely,

$$\beta_i = \frac{\lambda_i}{\sum_{i=1}^w \lambda_i}. \quad (30)$$

Among the w principal components, the sum of the contribution rates of the first k principal components is called the cumulative contribution rate of the first k principal components, denoted as α :

$$\alpha = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^w \lambda_i}. \quad (31)$$

The number of principal components can be determined by the cumulative contribution rate. This paper takes the cumulative contribution rate $\alpha \geq 0.85$ as the standard. For the selected k principal components, if the cumulative contribution rate reaches 85%, that is, $\alpha \geq 0.85$, then the principal components can be determined as k . It means that the selected k principal components basically retain the information of the original w variables. When determining the number of principal components, the number of principal components should be reduced as much as possible under the condition of $\alpha \geq 0.85$.

Using the obtained k principal components to linearly combine them can express the original multidimensional data flow of the lag correlation, namely,

$$\begin{aligned}
F_1(t) &= u_{11}x_1(t) + u_{12}x_2(t + \Delta t_2) + \cdots + u_{1i}x_i(t + \Delta t_i) + \cdots + u_{1k}x_k(t + \Delta t_k), \\
F_2(t) &= u_{21}x_1(t - \Delta t_2) + u_{22}x_2(t) + \cdots + u_{2i}x_i(t + \Delta t_i - \Delta t_2) + \cdots + u_{2k}x_k(t + \Delta t_k - \Delta t_2), \\
&\dots \\
F_i(t) &= u_{i1}x_1(t - \Delta t_i) + u_{i2}x_2(t - \Delta t_i + \Delta t_2) + \cdots + u_{ii}x_i(t) + \cdots + u_{ik}x_k(t - \Delta t_i - \Delta t_k), \\
&\dots \\
F_n(t) &= u_{n1}x_1(t - \Delta t_n) + u_{n2}x_2(t - \Delta t_n + \Delta t_2) + \cdots + u_{ni}x_i(t - \Delta t_n + \Delta t_i) + \cdots + u_{nk}x_k(t + \Delta t_n - \Delta t_k).
\end{aligned} \tag{32}$$

Furthermore, the algorithm obtains a comprehensive evaluation function:

$$F = \beta_1 F_1 + \beta_2 F_2 + \cdots + \beta_k F_k. \tag{33}$$

It can be seen that, by performing principal component analysis (PCA) on the lag-related multidimensional data stream, the original multidimensional data stream represented by the data corresponding to the larger m time points can be converted into the smaller k principal components to represent the original multidimensional data stream, which brings great convenience to the storage and use of data stream and also provides help for the analysis and reconstruction of lag-related multidimensional data stream.

4. Classification and Statistics of Financial Data in Smart Cities Based on Artificial Intelligence

The overall framework of financial asset classification is shown in Figure 3.

There are three major factors in sequence analysis. (1) Sequence duration: it refers to the length of the analyzed sequence in the time dimension, such as the time period during which user behavior occurs. (2) Time folding window: it means that a series of actions occurring within a certain period of time will be regarded as occurring simultaneously. (3) Time interval: it refers to the time interval of the discovered time series model. An example of the data mining process is shown in Figure 4.

The fast hierarchical clustering algorithm based on root finding is based on the following assumption: a data set can be divided into several clusters, and there is a core point in each cluster, and this point can represent other points in the cluster. This core point is not necessarily the center of mass of the cluster; it exists in an area where the data is densely distributed. The area with dense data distribution here refers to the closest distance between the data points in this area relative to the surrounding neighborhood of this area. We call the core point of the cluster the root node, as shown in Figure 5.

The structure design of the financial data classification system is shown in Figure 6. The system obtains sample files

from the sample database after data extraction and pre-processes the sample files into training samples and test samples. The training samples are used to train the classifier, and the test samples are used to evaluate the classification performance of the classifier, and the training process of the classifier is adjusted according to the evaluation results to obtain the final classifier.

The evaluation process of the classification effect is shown in Figure 7. The source of the test sample is the same as the training sample. It is taken from the sample database and preprocessed. The test sample also carries the actual category label.

The simple understanding of financial data classification task refers to the process of categorizing financial data of an unknown category based on its content when the classification system is known. Financial data classification is essentially a process of identifying the characteristics of financial data patterns. After the financial data is preprocessed and expressed as a feature vector, it is input into the classifier, and the model is trained. For the financial data to be classified, it is also expressed as a feature vector and then input into the classifier for classification. Figure 8 shows the principle of financial data classification. Financial data content processing will be introduced in detail later.

As shown in Figure 9, we have given a basic model framework based on rough set and KNN text classification. In this model framework, we can clearly see that this model is composed of three parts: rough set-related preprocessing module, KNN classification algorithm module, and quality evaluation module. First, in the text collection stage, the relevant web text data is obtained from the internet, and the relevant data is used as the training set and test set of the web text for subsequent operations, and then the next step is entered.

After constructing the above intelligent model, this paper conducts statistical analysis of the intelligent model and evaluates the effect of financial data mining and financial data classification, and the results are shown in Figure 10.

It can be seen from Figure 10 that the intelligent algorithm proposed in this paper can play a certain role in financial data processing. After evaluating the intelligent data decision in this paper, the results are shown in Figure 11.

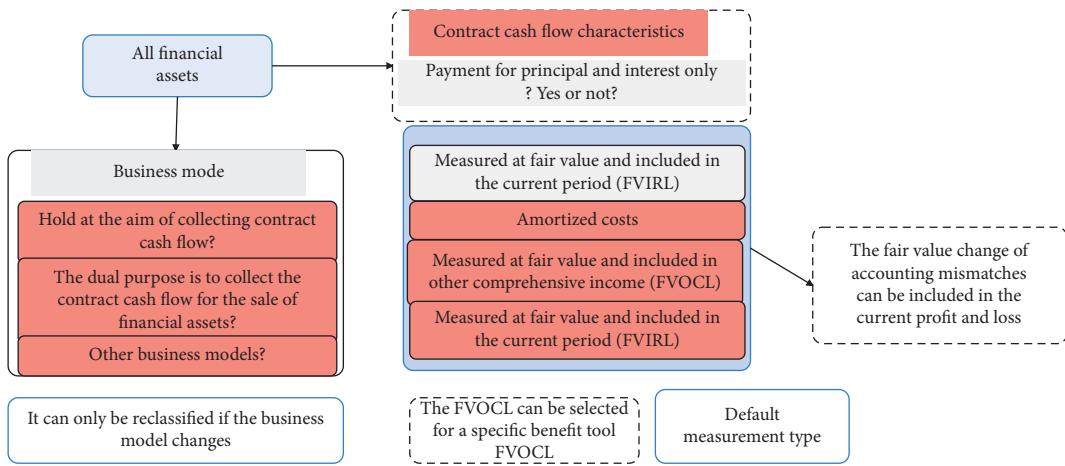


FIGURE 3: The overall framework of financial asset classification.

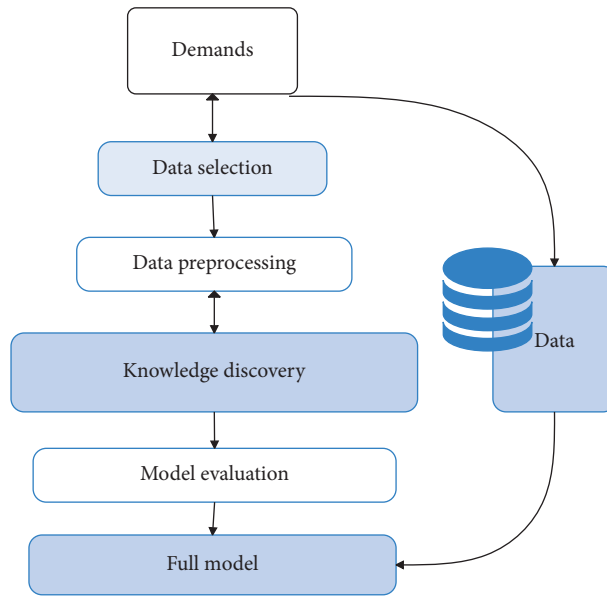


FIGURE 4: The data mining process.

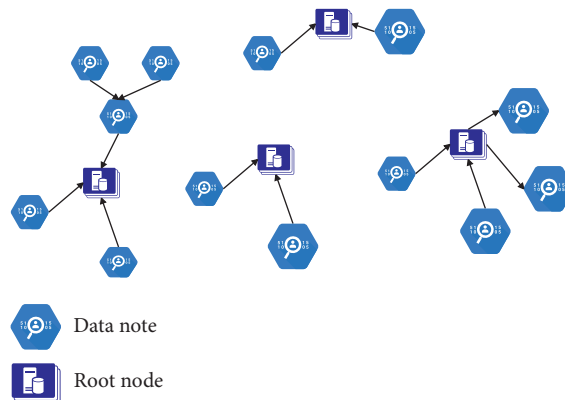


FIGURE 5: A possible result of finding the root node on the sample data set.

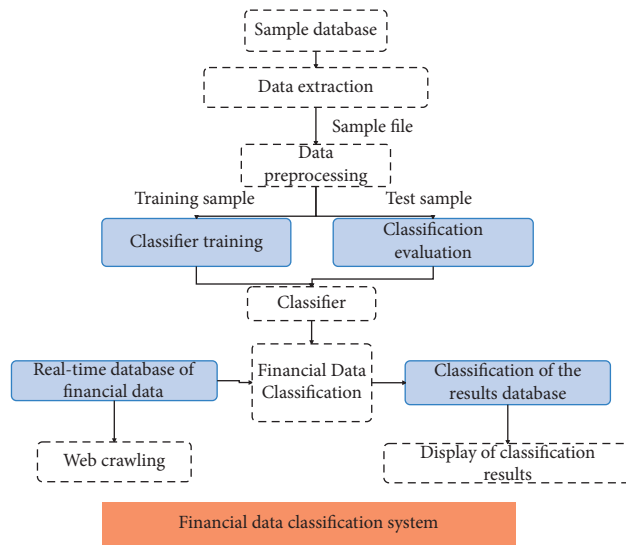


FIGURE 6: Financial data classification system structure.

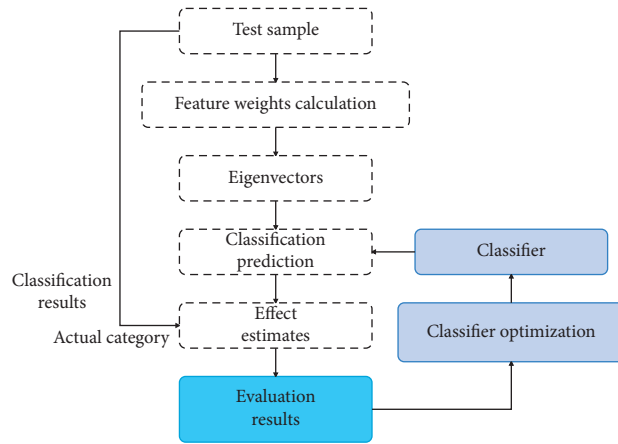


FIGURE 7: Classification and evaluation process diagram.

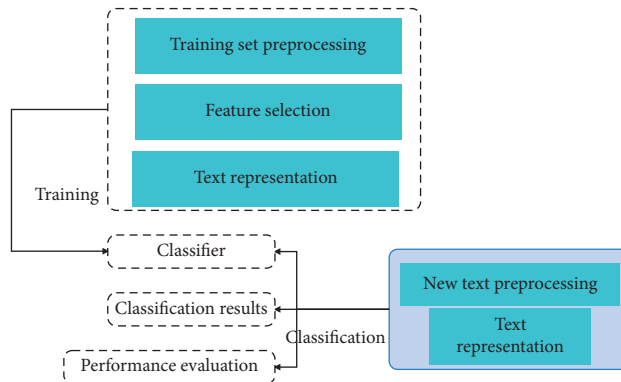


FIGURE 8: Principles of financial data classification.

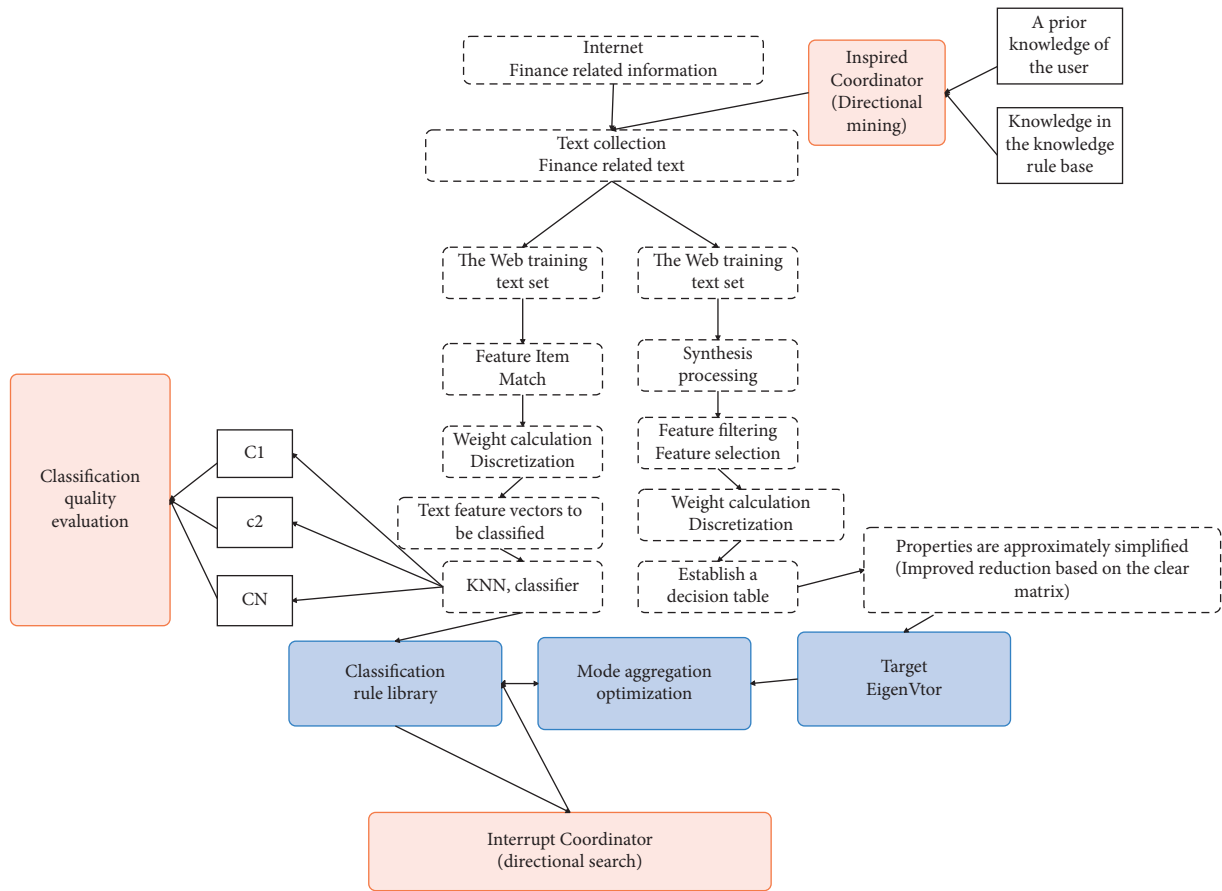


FIGURE 9: Financial data subsystem model framework.

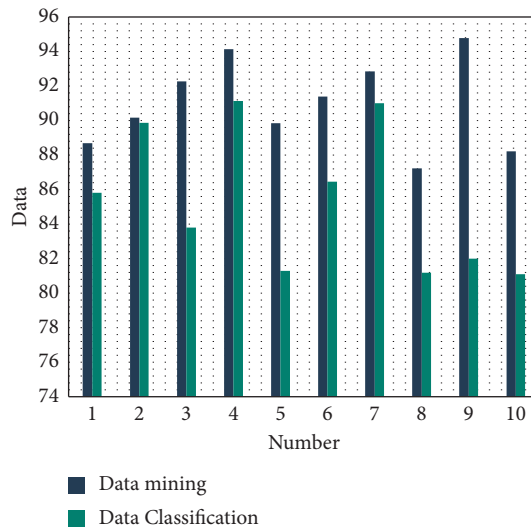


FIGURE 10: Evaluation of financial data processing effect.

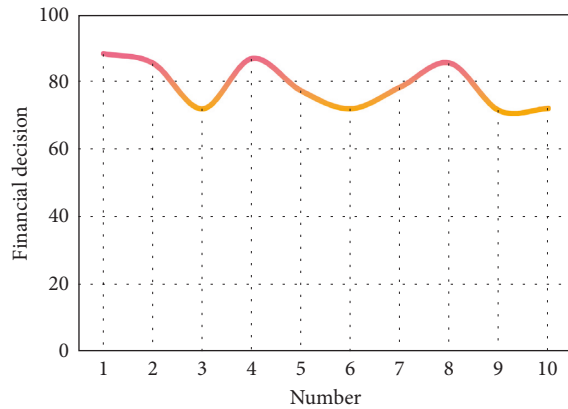


FIGURE 11: Intelligent data decision statistics.

From the above research, we can see that the artificial intelligence classification and statistical method of financial data in smart cities proposed in this paper can play a certain auxiliary role in financial decision-making.

5. Conclusion

The prediction problem of financial data processing is mainly the prediction of financial time series. Financial time series can be regarded as a special time series, which has the following three characteristics. (1) The generation process of financial time series is more complicated, and there are many influencing factors. (2) Most financial time series contain a large number of unpredictable influence factors. (3) The composition of the data in the financial time series is relatively complex and usually exhibits nonlinearity. Artificial intelligence information processing methods such as neural networks, chaos theory, and genetic algorithms can well adapt to these three characteristics, which have become advanced methods to solve financial data processing problems. This paper combines big data technology to carry out financial data classification research and extracts financial data from different data sources in an automatic and intelligent way. Users can not only get the data they want in the shortest time, but also improve the accuracy and effectiveness of the data. Through experimental research, we know that the artificial intelligence classification and statistical method of financial data in smart cities proposed in this paper can play a certain auxiliary role in financial decision-making.

Data Availability

The labeled dataset used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The author declares no conflicts of interest.

Acknowledgments

This study was sponsored by Hainan University.

References

- [1] P. Rajsic, A. Weersink, A. Navabi, and K. Peter Pauls, "Economics of genomic selection: the role of prediction accuracy and relative genotyping costs," *Euphytica*, vol. 210, no. 2, pp. 1–18, 2016.
- [2] F. Jahedpari, T. Rahwan, S. Hashemi et al., "Online prediction via continuous artificial prediction markets," *IEEE Intelligent Systems*, vol. 32, no. 1, pp. 61–68, 2017.
- [3] V. Daksiya, H. T. Su, Y. H. Chang, and E. Y. M. Lo, "Incorporating socio-economic effects and uncertain rainfall in flood mitigation decision using MCDA," *Natural Hazards*, vol. 87, no. 1, pp. 515–531, 2017.
- [4] S. Lahmiri, "A variational mode decomposition approach for analysis and forecasting of economic and financial time series," *Expert Systems with Applications*, vol. 55, no. 8, pp. 268–273, 2016.
- [5] N. Gordini, "A genetic algorithm approach for SMEs bankruptcy prediction: empirical evidence from Italy," *Expert Systems with Applications*, vol. 41, no. 14, pp. 6433–6445, 2014.
- [6] A. Ferramosca, A. H. González, and D. Limon, "Offset-free multi-model economic model predictive control for changing economic criterion," *Journal of Process Control*, vol. 54, no. 3, pp. 1–13, 2017.
- [7] C. J. Jane, "A hybrid model combined grey prediction and autoregressive integrated moving average model for talent prediction," *Journal of Grey System*, vol. 21, no. 2, pp. 91–102, 2018.
- [8] A. Khadjeh Nassirtoussi, S. Aghabozorgi, T. Ying Wah, and D. C. L. Ngo, "Text mining for market prediction: a systematic review," *Expert Systems with Applications*, vol. 41, no. 16, pp. 7653–7670, 2014.
- [9] M. Ellis and P. D. Christofides, "Integrating dynamic economic optimization and model predictive control for optimal operation of nonlinear process systems," *Control Engineering Practice*, vol. 22, no. 1, pp. 242–251, 2014.
- [10] W. Schultz, W. R. Stauffer, and A. Lak, "The phasic dopamine signal maturing: from reward via behavioural activation to formal economic utility," *Current Opinion in Neurobiology*, vol. 43, no. 5, pp. 139–148, 2017.
- [11] C. J. Lee and E. B. Andrade, "Fear, excitement, and financial risk-taking," *Cognition and Emotion*, vol. 29, no. 1, pp. 178–187, 2015.
- [12] S. M. Bartram, G. W. Brown, and W. Waller, "How important is financial risk?" *Journal of Financial and Quantitative Analysis*, vol. 50, no. 4, pp. 801–824, 2015.
- [13] W. W. Cooper, A. T. Kingyens, and J. C. Paradi, "Two-stage financial risk tolerance assessment using data envelopment analysis," *European Journal of Operational Research*, vol. 233, no. 1, pp. 273–280, 2014.
- [14] W. Montford and R. E. Goldsmith, "How gender and financial self-efficacy influence investment risk taking," *International Journal of Consumer Studies*, vol. 40, no. 1, pp. 101–106, 2016.
- [15] S. Muthii Wanjohi, J. G. Wanjohi, and J. M. Ndambiri, "The effect of financial risk management on the financial performance of commercial banks in Kenya," *International Journal of Finance and Banking Research*, vol. 3, no. 5, pp. 70–81, 2017.
- [16] J. Magendans, J. M. Gutteling, and S. Zebel, "Psychological determinants of financial buffer saving: the influence of financial risk tolerance and regulatory focus," *Journal of Risk Research*, vol. 20, no. 8, pp. 1076–1093, 2017.
- [17] E. S. Cole, D. Walker, A. Mora, and M. L. Diana, "Identifying hospitals that may be at most financial risk from medicaid

- disproportionate-share hospital payment cuts,” *Health Affairs*, vol. 33, no. 11, pp. 2025–2033, 2014.
- [18] A. Zalik, “Resource sterilization: reserve replacement, financial risk, and environmental review in Canada’s tar sands,” *Environment and Planning A: Economy and Space*, vol. 47, no. 12, pp. 2446–2464, 2015.
- [19] C. Lucarelli, P. Uberti, and G. Brighetti, “Misclassifications in financial risk tolerance,” *Journal of Risk Research*, vol. 18, no. 4, pp. 467–482, 2015.
- [20] E. Y. Chan, “Physically-attractive males increase men’s financial risk-taking,” *Evolution and Human Behavior*, vol. 36, no. 5, pp. 407–413, 2015.
- [21] M. Banam and A. Mehrazeen, “The relationship of information asymmetry, institutional ownership and stock liquidity with income smoothing in tehran stock exchange,” *Journal of Management and Accounting Studies*, vol. 4, no. 3, pp. 6–11, 2016.
- [22] B. F. Jones, “The human capital stock: a generalized approach,” *The American Economic Review*, vol. 104, no. 11, pp. 3752–3777, 2014.