Scholar Publishing Group
International Journal of Engineering Technology and Construction
https://doi.org/10.38007/IJETC.2020.010103
ISSN 2789-7176 Vol. 1, Issue 1: 29-42

Machine Learning Algorithm based on Spatiotemporal Kriging Model in The Construction of National Stock Market Network

Xianzhen Xu*

Qingdao University, Qingdao, China

xuxianzhen@qdu.edu.cn

**corresponding author*

Keywords: Network Construction, Spatiotemporal Kriging Model, Machine Learning Algorithm, National Stock Market, Spatiotemporal Variation

Abstract: Kriging in space-time has obvious advantages in analyzing the continuous change of natural characteristic quantity. This method can use the time correlation to improve the analysis accuracy. In the analysis of stock market network construction, Kriging is still not applied. The purpose of this paper is to construct the national stock market network based on the machine learning algorithm of spatiotemporal Kriging model. Firstly, the basic theory of the basic properties of the network is introduced. Secondly, the principle of machine learning algorithm is further studied. After distinguishing the classical Kriging model and the spatiotemporal Kriging model, the data of stock market network is constructed. Some problems related to the interpolation with spatiotemporal Kriging method are further studied, such as the reasons for the improvement of precision, the selection of database capacity in the process of spatiotemporal interpolation, etc. The experimental results show that the spatiotemporal Kriging method is used to calculate the spatiotemporal variation function, Cst(0,0) is estimated to be 0.062, and the spatiotemporal variation function is constructed. The average path of stock network is 2.0476, which is close to 2.6095 of random network. At the same time, the clustering coefficient of stock market network is 7619, which is much larger than that of random network. For any real network, if it satisfies the following two conditions at the same time, it can be called a small world effect.

1. Introduction

Using network method to analyze stock market is a new method and perspective to study stock market. It will be one of the research directions in the future to apply robots to the field of quantitative investment. Robot technology is in the ascendant, and has made remarkable achievements in image recognition, search recommendation and many other fields; compared with time series analysis, machine learning model can quickly process and analyze massive data, and often has better generalization ability. It is a hot topic to apply robots to financial data mining. The stock market is a barometer of the economic situation of various countries. The research on the fluctuation of securities market is an important subject of the research on the operation of securities market. The research on the dynamic characteristics of the complex network of the national securities market provides a new perspective for the analysis and research of the operation of the securities market, a new idea for the management of the securities market regulators, and a new analysis method for investors. The ordinary Kriging method interpolates fixed points according to the position relationship between variables. The spatiotemporal Kriging method extends the general Kriging method to the spatiotemporal dimension considering the time-varying characteristics of the stock market.

Ibrahim Fayad analyzed the comprehensive correlation between the canopy height estimated by the lidar sensor and the geological, meteorological, phenological and other auxiliary data, and proposed a canopy height map with high accuracy and spatial resolution, which is of great value. In this study, the canopy height extracted from airborne and satellite lidar is first inferred from the existing environmental data. The canopy height map estimated from air or glas calibration data sets using random forest (RF) regression shows similar accuracy (RMSE is better than 6.5m). In order to improve the accuracy of canopy height estimation, the regression Kriging method was used. The results show that RMSE of regression Kriging from glas data set (from 6.5m to 4.2m) and from airborne lidar data set (from 5.8m to 1.8m [1]). Nussa gabah B. Raja proposed an effective method to fit the time series of spatial functions to the observed data. This method is closely related to Kriging. In this paper, Kriging renewal model is represented by stochastic state equation, and time correlation is introduced into Kriging. This allows a recursive solution similar to Kalman filter to estimate the time series of the function, thus avoiding the growing data problem associated with Kriging's other spatiotemporal expansion [2]. In the rotating motor, according to the step equal time interval of the rotating angle, the finite element method is used for electromagnetic analysis. If the uncertainty of time variable needs to be handled, the time interval should be less than the uncertainty value. Even if the agent model is applied, the cost will increase. Therefore, Junyong Jang proposed a spatiotemporal Kriging agent model considering time and space variables in the design phase. The model processes time variables and predicts the response between time intervals to obtain the output of small uncertainty of time variables. The predicted performance is verified by simulation data, and the estimated reliability of the model is compared with the measured data of the manufacturing prototype [3].

The innovation of this paper is as follows: in the research content, the specific application direction of machine learning in stock market data mining is explored. Machine learning is a hot research direction nowadays, and quantitative trading will also be an important investment way in China's stock market in the future. The combination of the two has broad prospects for development. Secondly, in terms of research methods, according to the network theory, taking the national stock market as the research object, we use a newly proposed and efficient machine learning algorithm, the spatiotemporal Kriging model to build the network model about the stock market of various countries. This paper verifies the effectiveness of the algorithm in the stock market data mining.

2. Proposed Method

2.1. Basic Nature of Network

(1) Graph representation of network

For the same network, due to different definitions of points and edges, it may be represented as different figures [4-5]. Given a network $G = (V, e)$, if any node pair (I, J) and (J, I) in point set V correspond to the same connection edge in the network, then the network is an undirected network, otherwise the network is a directed network; if each edge in the network has a weight to measure the strength of the connection between the node pairs, then the network is a weighted network, otherwise the network is a weighted network unauthorized network [6-7].

(2) Centrality

The degree centrality of nodes is the degree value of nodes, which is the simplest but very important attribute of nodes [8]. In the undirected network, the number of neighbor nodes of node I is recorded as K_i,.

(3) Centrality of eigenvectors

A is the adjacency matrix of the network. If there is a connection relationship between node i and node j, the corresponding element $A_{ij} = 1$ in the adjacency matrix, otherwise $A_{ij} = 0$.

If the above formula is written in vector form, then $ACe = \lambda\ Ce$ is the characteristic equation of adjacency matrix A[9].

(4) Average path length and clustering coefficient

The average path length and clustering coefficient of the network are used to describe the compactness between nodes [10]. That is, the average distance between any two nodes in the network, where n is the node size in the network. However, for the actual large-scale network, its connectivity is often not guaranteed, that is, there may not be a path between a pair of nodes [11-12]. In order to avoid the divergence of the calculation results, the harmonic average of the distance between nodes in the network is usually defined as the average path length L:

$$L = \frac{1}{\frac{1}{2}N(N-1)} \sum_{i \geq j} d_{ij} \tag{1}$$

In this way, the average path length according to the above formula is always a finite value.

For a node i in an undirected network with no right, its degree value is k_i, that is, it has ki neighbor nodes with direct edges connected, and its k_i neighbors have the possibility of being neighbors to each other. In network science, clustering coefficient is used to describe the probability of neighbor relationship between any two neighbors of a node [13-14]. The clustering coefficient Ci of a node i whose degree value is ki is defined as

Where E_i is the number of edges between the ki neighbors of node i, and $ki\ (ki\text{-}1)\ /\ 2$ is the maximum possible number of connected edges between the ki neighbors of node I. Obviously, the value of clustering coefficient C_i satisfies $0 \leq C_i \leq 1$. For a network, the average clustering coefficient of nodes is defined as the clustering coefficient C of the network, namely:

$$C = \frac{1}{N} \sum_{i=1}^{N} C_i \tag{2}$$

2.2. Machine Learning Algorithm

Machine learning, as one of the most popular directions in recent years, is a new subject which uses computer technology, studies the existing data laws and forecasts and analyzes the unknown data, also known as statistical learning [15]. From the perspective of predictive output, including supervised learning: Based on the known dependent variable label, building the mapping relationship between the input and output of data; and unsupervised learning, i.e. without knowing the output label of data, transforming and analyzing it based on the existing input data [16-17].

(1) Single machine learning model

1) Decision tree

Decision tree is a tree transformation structure that classifies and regresses data samples, including internal nodes, leaf nodes and directed edges [18]. The internal node represents a split feature, which determines the split and growth of the tree. The leaf node represents the output of the model, while the directed edge connects the nodes in the tree [19]. When building a decision tree, starting from the root, the sample data is divided into different sub nodes by setting criteria for a certain feature in the multi-dimensional features of the input data, and the process is repeated until it reaches the leaf node and outputs the classification [20].

2) K-nearest neighbor algorithm

The KNN algorithm is more intuitive. For a sample in the set to be predicted, calculate the distance from all the samples to this sample in the training set, sort according to the distance, select the closest sample points, and determine the label of the prediction sample according to the number of votes in the category [21-22]. Compared with other machine learning algorithms, the under neighbor algorithm has no explicit modeling process and is a kind of “inert learning”. In the training stage, it only stores and sorts the samples, obtains the prediction data, and then calculates [23].

3) Logical regression

The linear model constructed under the mean square error loss function can estimate the conditional mean value of the dependent variable, but when the dependent variable is a binary variable in {0,1}, there will be deviation in the fitting by using the linear regression model, which cannot guarantee that the estimated dependent variable is in this range. Therefore, it is necessary to transform the general linear model and use the generalized linear model to estimate [24].

4) Neural network

The most basic component of artificial neural network (ANN) is neuron. By receiving the input of n-dimensional data, the input value and weight w_i vector are combined linearly. Finally, the activation function $f(w_i, x_i)$ is used for nonlinear transformation, that is, the construction process of a single neuron is completed.

(2) Integrated machine learning model

1) Gradient lifting tree

The basic learner in GBDT is the cart decision tree. Considering that the gradient describes the direction of the maximum value in the direction derivative of the function, through the linear superposition of the decision tree model, each decision tree fits the previous negative gradient, and multiple trees can gradually approach the minimum value of the loss function.

2) Random forest

In random forest, two methods are used to introduce random factors and avoid over fitting: disturbance of samples and disturbance of attribute characteristics. In the selection of training data, the bootstrap method is used for repeated sampling with put back, so that for each tree in the training set (1 / E) samples are not extracted, called out of bag data (OOB), which can be used for estimation and verification on the verification set of the model; Breiman has proved that the test error of out of bag data in the random forest is an unbiased estimation of the prediction error of unknown data. It can be used to evaluate the accuracy of classifier. In the random disturbance of attribute features, the selection of split variables on internal nodes is not in all features, but in the first place, a subset of all features is selected randomly, and then split variables are selected from this subset to achieve the purpose of disturbance of attribute features of each tree.

2.3. Kriging Method

The linear regression part provides the global approximation of the simulation, that is, the expectation of the response value. In the classical Kriging model, there are three kinds of regression function: constant type, linear type and quadratic polynomial type. Generally, the form of regression function does not play a decisive role in fitting accuracy. Linear type and quadratic polynomial type are often selected, and random error provides the approximation of local deviation of simulation, that is, the local change of response value.

(1) Kriging method flow

Kriging model combined with response surface method forms the classical Kriging method. The idea is similar to the general response surface method, the specific process and the specific iteration steps are as follows:

1) Select the mean point μx as the initial sample center point;

2) According to the Bucher design, the sample points are determined and the corresponding structural response values are calculated;

3) According to the sample points and their response values, Kriging model parameters are selected to build Kriging model;

4) Use form to calculate the reliability index β and the most likely failure point X*;

5) According to formula 3, a new sample center point x is obtained;

$$X = \mu_X + \frac{G(\mu_X)}{G(\mu_X) - G(X^*)}(X^* - \mu_X) \tag{3}$$

6) Repeat steps 2 to 5 until the convergence criteria are met as shown in Figure 1.

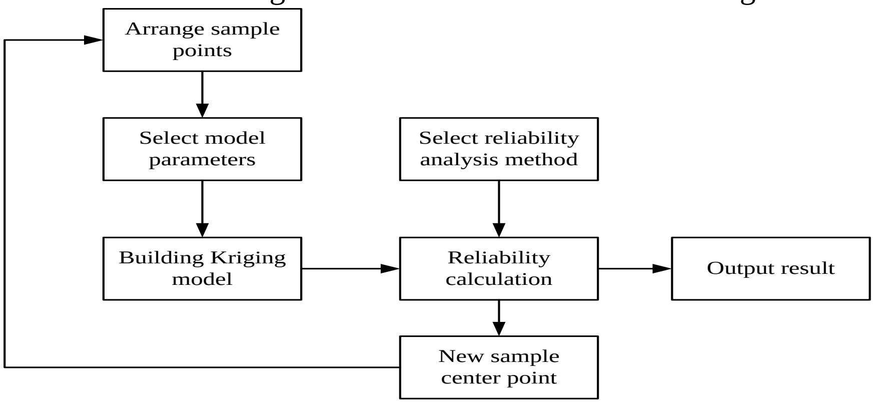

Figure 1. Classical Kriging method

(2) Regionalization variable

Regionalization variable is the theoretical basis of spatial statistics. It is a random function related to location, which usually shows different quantitative characteristics with the change of spatial location. It also has two very important characteristics, one of which is randomness, and the other is structural, that is, it has certain autocorrelation between any two adjacent points in the spatial area.

(3) Covariance function and variation function

Covariance function and variation function are the most basic two functions in spatial statistics. They are based on the theory of regionalized variables and are the main tools to describe regionalized variables. The covariance function of stochastic process refers to the second-order mixed center distance of two random variables Z (T1), Z (T2) at time T1 and T2 of stochastic process Z (T)

As cov (T1, T2) or C (T1, T2), we call it the covariance function of the random process Z (T). From the above definition, we can see that covariance function is a function related to time t, that is, its value changes with the changes of time T1 and T2, so we can use it to describe the autocorrelation of random process between these two times. Similarly, when Z (s) is a regionalized variable, its covariance function can be defined as the second-order mixed center distance of Z (s) and Z (s + H), where Z (s) and Z (s + H) respectively represent the values of random variables at two points s (s is the position of space points) and S + H (H is the vector) in space

Recorded as cov (s, S + H) or C (s, S + H), we call it the covariance function of the regionalized variable Z (s). The covariance function of Z (s) is a function related to position point s and vector H. When s and H are the determined values, two points in the space are determined, corresponding to the random variables Z (s) and Z (s + H), respectively.

Variogram, also known as structure function and variogram, is a function related to the distance between two points in space. It is not only a special basic tool for spatial data analysis, but also plays a very important role in spatial statistics. The most important thing for us to do spatial analysis is to find a suitable variogram, because it can not only reflect the structural changes of regional variables, but also describe their random changes, and it is the basis of many spatial statistical calculations.

Under one-dimensional condition, when the spatial point s changes on the one-dimensional x-axis, the variation function of the regionalized variable Z (s) in the x-axis direction is defined as the variance of the difference between Z (s) and Z (s + H), which is recorded as s, h, where Z (s) and Z (s + H) respectively represent the values of the regionalized variable at points s and S + H. It can be seen from the definition that the variation function s and H are related to s and h, where h is a vector rather than a scalar.

In practical application, the regionalized variable Z (s) we encounter is not only one-dimensional, but more two-dimensional or three-dimensional, so the latter is the focus of our research. Similarly, under the condition of two-dimensional or three-dimensional, we can define the variogram as the variance of the increment of Z (s) and Z (s + h) in any direction with a distance of $|h|$ (the module representing vector h).

Because the variogram plays a very important role in the existence of autocorrelation in the random process, the most important thing is to fit the variogram when using Kriging method for interpolation or prediction, but in most cases, we still need to determine the value of parameters to get the accurate variogram, that is, the nugget value, the base value and the range. In most practice, we do not know the specific parameter value of the theoretical variogram model, so we need to analyze the sample data to obtain the model. Due to the lack of spatial sample data and the lack of data, only through the sample data cannot get the accurate variogram model, so we need to select the best one from the existing theoretical variogram model to fit the empirical variogram, and then constantly adjust the parameter value, so that the existing theoretical variogram can fit the empirical variogram to a great extent.

In spatial statistics, in order to better show the change characteristics of regional random variables, we need to constantly optimize the image of empirical variogram function to better describe its change law. In the process of optimization, the key is to select the model matching the empirical variogram, such as exponential, spherical and Gaussian models. In the process of selecting the appropriate model, we need to select the appropriate variogram model by comparing the RMS difference of sample points.

2.4. Kriging Method in Time and Space

(1) Spatiotemporal analysis

Spatiotemporal data are often encountered in many scientific disciplines, especially in the environment, meteorology or geophysics. Spatiotemporal data analysis is an effective way to extract information and knowledge from massive geographic spatiotemporal data, such as precipitation, daily average temperature and air pollution concentration. Therefore, it is very important to construct spatiotemporal distribution model for the dynamic process of spatial and temporal evolution.

(2) Kriging algorithm in time and space

Kriging interpolation is first used to find gold deposits. Based on the theory of variational function and structural analysis, the method makes linear unbiased optimal estimation of the regionalized variables in a finite area. The spatiotemporal Kriging method is based on the ordinary Kriging method, which extends from the spatial domain to the spatiotemporal domain, provides structural information by establishing spatiotemporal variation function, and interpolates the estimated points.

(3) Spatiotemporal Variogram

As the basic tool of Kriging interpolation, variogram can reflect the spatial variation characteristics of regionalized variables, especially the structure of regionalized variables through randomness. Under the condition that the regionalized variable Z (x) ∈ s is second-order stationary, the formula for defining the spatial variation function is as follows:

Where: X is the coordinate of the sampling point in the spatial domain; h is the spatial distance.

After extending to the space-time domain, let the spatiotemporal regionalized variable Z (s, t) ∈ s × t satisfy the second-order stationary hypothesis, s represents the space domain, T represents the time domain, and the spatiotemporal change function can be defined as:

Where: (s, t) represents the coordinates of sampling points in the space-time domain; HS and HT represent the space distance and time distance respectively.

In order to quantitatively describe the characteristics of the whole regionalized variables, it is necessary to calculate the experimental variogram value and select the appropriate theoretical model to fit the optimal spatiotemporal variogram. Least square method and weighted regression method are commonly used in fitting. The spatiotemporal variogram is constructed by using the separation model and the non separation model, and the spatiotemporal variogram model is obtained by fitting. Therefore, how to obtain a reasonable spatiotemporal variogram is the key to solve the Kriging interpolation problem of spatiotemporal discrete data.

(4) Kriging interpolation in time and space

By introducing the correlation and randomness of space-time dimension, an effective space-time variation function is constructed, and the Kriging interpolation formula of space-time is extended to the space-time domain

$$Z^{*}(s,t)=\sum_{i=1}^{n}\lambda_i Z(s_i,t_i) \tag{4}$$

Where: (Si, Ti) represents the coordinates of sampling point I in the space-time domain. The weight coefficient λ I (I = 1,2,..., n) must satisfy the unbiased estimation that Z * (s, t) is Z (Si, Ti) and the estimation variance is the smallest. Under the second-order stationary assumption, Kriging space-time equation is used to calculate:

$$\begin{cases} \sum_{j=1}^{n} \lambda_j \gamma\big((s_i,t_i),(s_j,t_j)\big) - \mu = \gamma\big((s_i,t_i),(s,t)\big) \\ \sum_{i=1}^{n} \lambda_i = 1 \end{cases} \tag{5}$$

(5) Time variogram fitting

Due to the small number of fixed stations and the influence of space distance, the location of evaluation points changes with time and the location of each fixed station is different. Therefore, according to the calculation results, the time-varying functions γ_t (SJ, HT) (J = 1,2,..., K) of K fixed observation stations are fitted respectively. The nugget index C0, arch height C and change a of each time varying function are estimated by inverse distance weighting of time varying function γ t (HT). The calculation formula is as follows:

$$\gamma_t(h_t) = \frac{1}{k}\sum_{j=1}^{k} \lambda_j \gamma_t(s_j, h_t) \tag{6}$$

Where: λ_j is the distance weight coefficient from the fixed station to the evaluation point.

(6) Fitting of spatial variogram with multi period superposition

In fact, the number of sample point pairs cannot be infinite. Generally, it is required that the number of point pairs in each distance should not be less than 20 within the variation range, so as to ensure that the experimental variogram value within the variation range can accurately reflect the spatial variability of regional variables. The sampling time and update frequency of mobile observation equipment are inconsistent, resulting in too few samples in a single time slice, and the fitting spatial variation function is not representative, which cannot describe the spatial variation characteristics of the current region. In this paper, multi period superposition method is used to fit variogram. The formula for calculating the variogram value of single sub periodic space experiment is as follows:

$$\gamma_s^*(h_s) = \frac{1}{2mn}\sum_{j=0}^{m}\sum_{i=1}^{n}\left[Z(s_i,t) - Z(s_i + h_s, t \pm j)\right]^2 \tag{7}$$

The original spatial experiment variogram value of a single time slice is extended to both sides along the time dimension moderately (the expansion interval m is determined according to the time variogram, usually no more than 3). The sample points in the small interval are used to calculate the spatial test variogram value, and then the results of the test variogram value in each sub cycle are superposed to get the fitting result $\gamma_s(h_s)$. In a short time, the experimental variogram value between the sampling points largely eliminates the interference of time change. When the fitting results of multiple sub periods overlap, the number of fitting spatial variogram is ensured, and the influence of too large time variation between sampling points on the fitting accuracy of variogram is avoided.

(7) Construction and interpolation of spatiotemporal variation function

The spatiotemporal variogram is constructed by the product combination model (a common non separation model). The construction method is as follows:

$$\begin{aligned} \gamma(h_s,h_t) = {} & (k_1C_t(0)+k_2)\gamma_s(h_s) + (k_1C_s(0)+k_3)\gamma_t(h_t) \\ & - k_1\gamma_s(h_s)\gamma_t(h_t) \end{aligned} \tag{8}$$

The calculation results of spatiotemporal variation function are substituted into Kriging equation to solve the spatiotemporal weight λ_i of spatiotemporal sampling points. Finally, the interpolation results are obtained by spatiotemporal Kriging interpolation.

3. Experiments

3.1. Purpose of the Experiment

Stock index can predict the trend of stock market, and it is a reference index to reflect the change of stock market. In the process of financial globalization, since 2019, we have selected 36 major stock markets in the world as the research object, which ensure the availability of data. Therefore, according to the relevant data released by tonghuashun database and Yahoo Finance website, this paper collects the weekly stock index inventory of 36 stock markets from January 2019 to December 2019 as the research variable.

3.2. Experimental Data Set

The stock market selected in this paper is scattered in all regions of the world, so as to ensure that the stock market network constructed is more scientific and meets the global concept. Among them, the number of stock markets in East Asia accounts for 14% of the total; the number of stock markets in North America, West Asia and Southeast Asia accounts for 11% of the total; while the number of stock markets in Africa, West Asia, South Asia and Central Asia is only 3% of the total, as shown in Table 1. This is because this paper attempts to conduct a linkage study on the global stock market, which has been playing a prominent role in the process of financial globalization since 2019.

Table 1. List of stock market names and index codes

Number	Stock index code	Number	Stock index code	Number	Stock index code
1	AEX	13	N225	25	BVSP
2	NYSE	14	HIS	26	MXX
3	LXIC	15	CSS	27	IPSA
4	ATX	16	STI	28	MERV
5	BFX	17	AXJO	29	MICEX
6	OSEAX	18	TWII	30	HERMES
7	OMXSPI	19	KS11	31	TA100
8	SSMI	20	NZ50	32	MADX
9	GDAT	21	KLSE	33	ISE
10	FTSE	22	JKSE	34	SET
11	GDAXI	23	BSESN	35	OMXPHI
12	FCHI	24	GSPTSE	36	DFMGI

What's important is that for the global stock markets, at the same panel time, there is a phenomenon that one stock market has just opened while another stock market has closed.

4. Discussion

4.1. Data Processing of Stock Market

First of all, through the statistical analysis method, this paper makes a descriptive statistical analysis on the weekly return of 36 stock markets from January 2012 to December 2015, as shown in Figure 2. Through descriptive statistics to understand the basic timing of sample data, improve the scientificity of follow-up empirical analysis.

Figure 2. Descriptive statistics of stock index yield

In the long run, the average returns of these stock markets are close to 0, indicating that the stock market has a long-term balance. Among them, the stock market index with the largest average yield is the market index average yield of a stock market, reaching 0.47%. Moreover, these stock market variances are very small, which means that the volatility of stock index return is small, the stock market has strong self-regulation ability and long-term stability. It is worth noting that the variance of return rate of B stock market index is the largest, reaching 0.0295, that is to say, compared with other stock markets, B stock market is more volatile, reflecting the greater risk of its stock market. In addition, the kurtosis value of most stock markets in this stock market is far greater than the value of 3. The phenomenon of sending seed reflects that the fluctuation distribution of most stock market indexes has the feature of “thick tail”.

Based on the descriptive statistics of 36 stock market indexes, this paper finds the basic characteristics of the volatility of index return in these four years. However, in order to further understand the relationship between the fluctuation trend of 36 stock market indexes and the fluctuation of these stock markets in the past four years, this paper constructs the fluctuation trend chart of stock market indexes from January 2019 to December 2019 as shown in Figure 3. Through observation, we can see that the volatility range and direction of these stock market index returns are consistent, except for some abnormal value fluctuations of individual stocks.

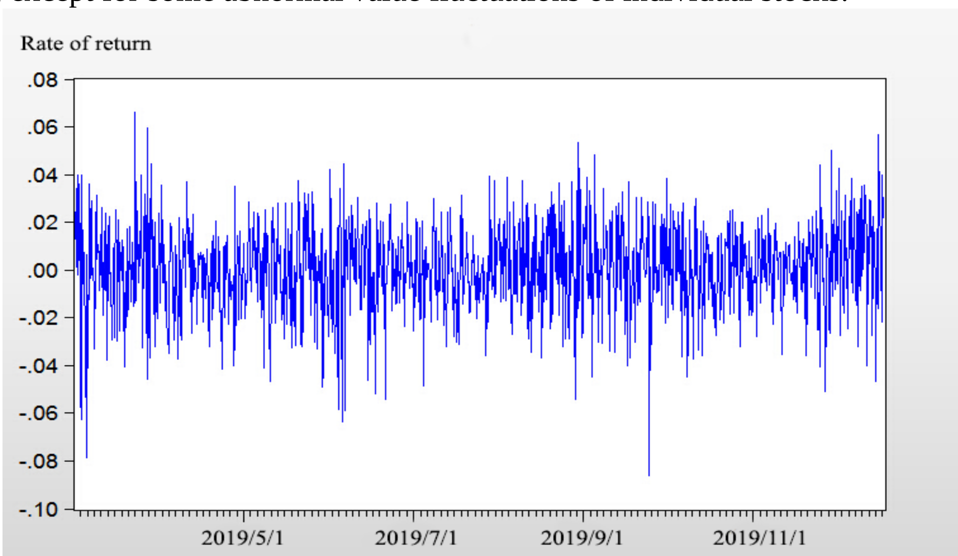

Figure 3. Volatility trend of stock index return from January 2019 to December 2019

4.2. Construction of Correlation Coefficient Model of Stock Market Index

Kriging method is used to calculate the spatiotemporal variation function, C_{st} (0,0) is estimated to be 0.062, and the spatiotemporal variation function is constructed. Because the spatiotemporal variation function is determined by time-varying function and time-varying function, the spatiotemporal distance between sampling points is zero respectively. According to the change of stock market reference value in different periods, we can get the change of different stock market reference value, the same experience in the same period, and then get the corresponding change function value. Through fitting, the function of pure time and pure space change is established. Because the construction accuracy of variogram is affected by the value of variogram and the fitting accuracy, in order to establish the best spatiotemporal variogram, the larger value of variogram should be eliminated. Through comparison, fitting time variation function and space to select the appropriate model. The spatial variability function and exponential model of the spherical model are used to fit the spatial variability function, and the distribution of the variation function is shown in Figure 4-5.

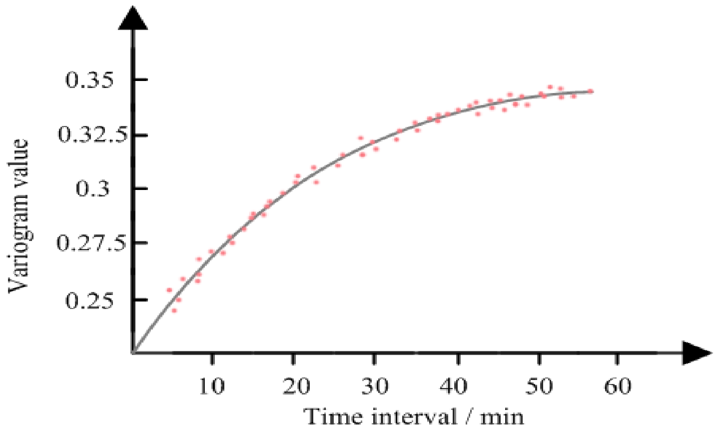

Figure 4. Time variation function distribution

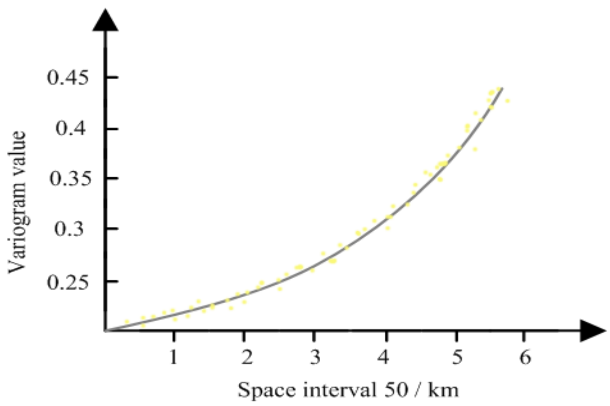

Figure 5. Spatial variation function distribution

4.3. Average Path and Clustering Coefficient of Stock Market Network

The average path and clustering coefficient of stock market network and random network are shown in Figure 6.

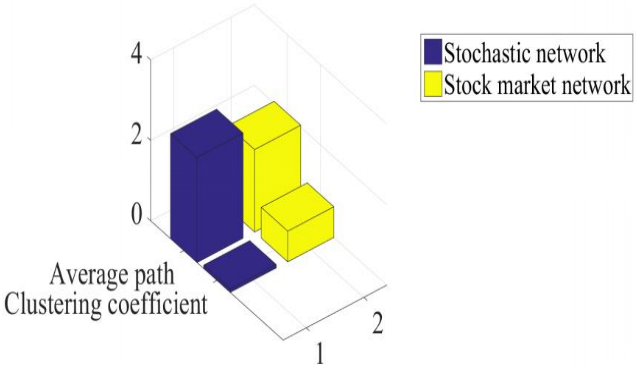

Figure 6. Average path and clustering coefficient of stock market network and random network

From the average path and clustering coefficient of the stock market network and a random network of the same size, we can see that the average path of the stock market network is 2.0476, which is close to the average path of the random network of 2.6095. At the same time, the clustering coefficient of stock market network is 7619, which is much larger than that of random network. For any real network, if it satisfies the following two conditions at the same time, it can be called a small world effect. That is: first, the average path of the network is close to the average path of a random network of the same size; second, the clustering coefficient of the network is far less than the clustering coefficient of the random network. That is to say, the stock market network satisfies the basic characteristics of the small world effect of the complex network, and it can be concluded that the stock market network has the small world effect.

5. Conclusion

Complex network theory is a new research hotspot in recent years, and its application research also has a wide range of fields. Stock market is an important research field of complex network. The main work of this paper is to use the complex network method to study the dynamic changes of China's stock market. Based on the correlation of the daily closing price fluctuation of each stock in the stock market, a complex network model of China's stock market is constructed by using the spatiotemporal Kriging algorithm, and the actual operation of the stock market is analyzed by studying the topological structure change of the stock market network.

In this paper, the international stock market is taken as the research object, the stock market of each country is taken as the network node, and the network model between the stock markets of each country is established by using the spatial Kriging machine learning algorithm through the correlation of the representative index fluctuations of each stock market. Then MATLAB is used to simulate the stock market. There is a certain relationship between the stock markets of different countries. Finally, this paper studies the data processing of the stock market to build the stock market index correlation coefficient model.

In this paper, the basic properties of the network and the classical network model are briefly summarized, including the graph representation of the network, the network centrality of nodes, the average path length and clustering coefficient, regular network, random network model, small world network model and scale-free network model. At the same time, this paper studies Kriging method and spatiotemporal Kriging algorithm, constructs the stock market network based on spatiotemporal Kriging algorithm, and analyzes the stock market using the network method.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1]Ibrahim Fayad, Nicolas Baghdadi, Jean-Stéphane Bailly. Regional Scale Rain-Forest Height Mapping Using Regression-Kriging of Spaceborne and Airborne LiDAR Data: Application on French Guiana. Remote Sensing, 2016, 8(3):240. https://doi.org/10.3390/rs8030240

[2]Nussaïbah B. Raja, Olgu Aydin, Necla Türkoğlu. Space-time kriging of precipitation variability in Turkey for the period 1976–2010. Theoretical & Applied Climatology, 2017, 129(1-2):293-204. https://doi.org/10.1007/s00704-016-1788-8

[3]Junyong Jang, Jin Min Lee, Su-gil Cho. Space-Time Kriging Surrogate Model to Consider Uncertainty of Time Interval of Torque Curve for Electric Power Steering Motor. IEEE Transactions on Magnetics, 2017, PP(99):1-4. https://doi.org/10.1109/TMAG.2017.2755459

[4]Mohammed Falah Allawi, Othman Jaafar, Firdaus Mohamad Hamzah. Operating a reservoir system based on the shark machine learning algorithm. Environmental Earth Sciences, 2018, 77(10):366. https://doi.org/10.1007/s12665-018-7546-8

[5]Sobar, R. Machmud, A. Wijaya. Behavior determinant based cervical cancer early detection with machine learning algorithm. Advanced Science Letters, 2016, 22(10):3120-3123. https://doi.org/10.1166/asl.2016.7980

[6]Feras Hatib, Zhongping Jian, Sai Buddi. Machine-learning Algorithm to Predict Hypotension Based on High-fidelity Arterial Pressure Waveform Analysis. Anesthesiology, 2018, 129(4):1.

[7]Gang Luo. Toward a Progress Indicator for Machine Learning Model Building and Data Mining Algorithm Execution: A Position Paper. Acm Sigkdd Explorations Newsletter, 2017, 19(2):13-24. https://doi.org/10.1145/3166054.3166057

[8]Christian Tesche, Carlo N. De Cecco, Stefan Baumann. Coronary CT Angiography–derived Fractional Flow Reserve: Machine Learning Algorithm versus Computational Fluid Dynamics Modeling. Radiology, 2018, 288(1):171291. https://doi.org/10.1148/radiol.2018171291

[9]Yun Zou, Qing-Zi Xing, Bai-Chuan Wang. Application of the asynchronous advantage actor–critic machine learning algorithm to real-time accelerator tuning. Nuclear Science and Techniques, 2019, 30(10):158. https://doi.org/10.1007/s41365-019-0668-1

[10]James E. Dobson. Can An Algorithm Be Disturbed?: Machine Learning, Intrinsic Criticism, and the Digital Humanities. College Literature, 2015, 42(4):543-564. https://doi.org/10.1353/lit.2015.0037

[11]Flavio Luis de Mello, Jose Antonio Moreira Xexeo. Cryptographic Algorithm Identification Using Machine Learning and Massive Processing. IEEE Latin America Transactions, 2016, 14(11):4585-4590. https://doi.org/10.1109/TLA.2016.7795833

[12]Neha Patel, Divakar Singh. An Algorithm to Construct Decision Tree for Machine Learning based on Similarity Factor. International Journal of Computer Applications, 2015,

111(10):22-26. https://doi.org/10.5120/19575-1376

[13]A Nicolae, L Lu, G Morton. SU-G-201-09: Evaluation of a Novel Machine-Learning Algorithm for Permanent Prostate Brachytherapy Treatment Planning. Medical Physics, 2016, 43(6):3624-3625. https://doi.org/10.1118/1.4956882

[14]Ghanshyam Pilania, P. V. Balachandran, J. E. Gubernatis. Classification of ABO3 perovskite solids: A machine learning study. Acta Crystallographica, 2015, 71(5):507-513.

[15]Luke Macyszyn, Hamed Akbari, Jared M. Pisapia. Imaging patterns predict patient survival and molecular subtype in glioblastoma via machine learning techniques. Neuro-Oncology, 2016, 18(3):417-425. https://doi.org/10.1093/neuonc/nov127

[16]Xueqiang Zeng, Gang Luo. Progressive sampling-based Bayesian optimization for efficient and automatic machine learning model selection. Health Information Science & Systems, 2017, 5(1):2. https://doi.org/10.1007/s13755-017-0023-z

[17]P. Alcaide-Leon, P. Dufort, A.F. Geraldo. Differentiation of Enhancing Glioma and Primary Central Nervous System Lymphoma by Texture-Based Machine Learning. American Journal of Neuroradiology, 2017, 38(6):1145. https://doi.org/10.3174/ajnr.A5173

[18]Huisung Kim, Olumide Awofeso, SoMi Choi. Colorimetric analysis of saliva–alcohol test strips by smartphone-based instruments using machine-learning algorithms. Applied Optics, 2017, 56(1):84. https://doi.org/10.1364/AO.56.000084

[19]Wei Xiao. A Probabilistic Machine Learning Approach to Detect Industrial Plant Faults. International Journal of Prognostics & Health Management, 2016, 7(1):11. https://doi.org/10.36001/ijphm.2016.v7i1.2360

[20]Kreuchauff, Florian, Korzinov, Vladimir. A patent search strategy based on machine learning for the emerging field of service robotics. Scientometrics, 2015, 111(2):1-30. https://doi.org/10.1007/s11192-017-2268-3

[21]Ashfaqur Rahman, Greg Timms, Md. Sumon Shahriar. Association Between Imaging and XRF Sensing: A Machine Learning Approach to Discover Mineralogy in Abandoned Mine Voids. IEEE Sensors Journal, 2016, 16(11):1-1. https://doi.org/10.1109/JSEN.2016.2546241

[22]Bingjie Wang. Quantum algorithms for machine learning. Xrds Crossroads the Acm Magazine for Students, 2016, 23(1):20-24. https://doi.org/10.1145/2983535

[23]Ronay Ak, Olga Fink, Enrico Zio. Two Machine Learning Approaches for Short-Term Wind Speed Time-Series Prediction. IEEE Transactions on Neural Networks & Learning Systems, 2015, 27(8):1734-1747. https://doi.org/10.1109/TNNLS.2015.2418739

[24]Rc Kessler, Hm Van Loo, Kj Wardenaar. Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. Mol Psychiatry, 2016, 21(10):1366-1371. https://doi.org/10.1038/mp.2015.198