IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received June 1, 2021, accepted June 15, 2021, date of publication June 21, 2021, date of current version July 2, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3091162

Forecast Methods for Time Series Data: A Survey

ZHENYU LIU, ZHENGTONG ZHU, JING GAO, AND CHENG XU
College of Computer and Information Engineering, Inner Mongolia Agricultural University, Hohhot 010018, China
Inner Mongolia Autonomous Region Key Laboratory of Big Data Research and Application for Agriculture and Animal Husbandry, Hohhot 010018, China

Corresponding author: Jing Gao (gaojing@imau.edu.cn)

This work was supported in part by the Natural Science Foundation of Inner Mongolia Autonomous Region under Grant 2019MS03014, in part by the Major Science and Technology Project in Inner Mongolia Autonomous Region under Grant 2019ZD016 and Grant 2020ZD0007, and in part by the Science and Technology Plan Project in Inner Mongolia Autonomous Region under Grant 2019GG372.

ABSTRACT Research on forecasting methods of time series data has become one of the hot spots. More and more time series data are produced in various fields. It provides data for the research of time series analysis method, and promotes the development of time series research. Due to the generation of highly complex and large-scale time series data, the construction of forecasting models for time series data brings greater challenges. The main challenges of time series modeling are high complexity of time series data, low accuracy and poor generalization ability of prediction model. This paper attempts to cover the existing modeling methods for time series data and classify them. In addition, we make comparisons between different methods and list some potential directions for time series forecasting.

INDEX TERMS Time series analysis, forecasting, modeling.

I. INTRODUCTION

Time series data refers to the results of observing a certain process at a given sampling rate in an equally spaced time period. The core of time series analysis is to discover laws from data and predict future values based on historical observations, which can provide reference and basis for decision-making. In daily life, time series observation data has spread across various fields, such as agriculture, industry, finance, meteorology, military, etc. And time series data is being generated at a rapid rate.

The research of time series data forecasting has become one of the hot spots, and has been well applied in the fields of meteorological and weather forecasting, industrial production forecasting, and stock trend forecasting. It can help decision makers avoid risks and make more favorable decisions.

Traditional time series forecasting methods based on probability and statistics have achieved great results in the fields of meteorology, finance, industry, etc. However, with the development of the era of big data, massive, non-linear time series data that obey multiple distribution patterns are constantly being produced, which also poses higher challenges to time series forecasting methods. Therefore, people apply machine learning and deep learning to highly complex time series data prediction methods, and have achieved great results.

In this article, we intend to classify various existing time series forecasting methods.

The rest of this paper is organized as follows: Section II proposes the current problems in time series forecasting; Section III introduces the time series forecasting methods proposed in recent years. Then, Section IV discusses some unresolved problems and future time series methods Research direction. Finally, section V summarizes this paper.

II. PREDICTION ISSUE OF TIME SERIES

With the development of the data age, although great results have been achieved in the research of time series forecasting methods, there are still many problems. We will introduce some questions from the data aspect and the model aspect.

A. IN TERM OF DATA

It is well known, the accuracy of data analysis and modeling were seriously affected by data quality. Therefore, the quality of data plays a decisive role in data analysis. However, in reality, the original data is not perfect. On the one hand, time series data is generated from sensors, smart terminal equipment, acquisition systems, and other experimental instruments, etc. Due to various factors such as objective conditions and the stability of data collection equipment, the data often

The associate editor coordinating the review of this manuscript and approving it for publication was Giambattista Gruosso.

TABLE 1. Comparative analysis of model characteristics.

Model	Characteristic	PACF correlogram	ACF correlogram	Data haracteristic
AR(p)	1. y_t depends on its own past values 2. P is computed using PACF function	Spikes till p^{th} lag then cuts off to zero	Spikes then decays to zero	Data should be stationary in nature
MA(q)	1. y_t depends on error term which follows a white noise process 2. q is computed using ACF function	Spikes then decays to zero	Spikes till p^{th} lag then cuts off to zero	Data should be stationary in nature
ARMA (p, q)	1. ARMA = AR+MA 2. Value of p and q are determined using AIC criteria	Spikes then decays to zero	Spikes then decays to zero	Data should be stationary in nature
ARIMA (p, d, q)	1. Data is made stationary by differencing it 2. Box-Jenkins approach is used to determine model	Spikes then decays to zero	Spikes then decays to zero	Data should be non-stationary in nature

contains abnormalities such as noise and missing values. How to deal with outliers and missing values based on the distribution pattern of sample data, which poses higher challenges for large-scale time series data preprocessing methods.

B. IN TERM OF MODEL

At present, researchers usually construct predictive models based on historical data to for predictive analysis of future data. However, in practical applications, it is found that the accuracy and performance of the prediction model gradually decrease over time. The main reason is the gap between historical data and real-time data. And new data features are not incrementally learned by the model and applied to predictions.

In summary, the research on the revised model of time series forecasting based on online incremental data has become an urgent problem to be solved.

C. IN TERM OF REAL_TIME CALCULATION

Due to the rapid growth of time series data in various fields, online time series real-time analysis has become a development requirement.

Real-time calculation. Due to the rapid growth of time series data in various fields and the demand for timeliness of data, the real-time analysis of online for time series data has become a development requirement. Now time series analysis models usually adopt stand-alone mode. And it improves the operating efficiency by used high-performance GPU servers. On the one hand, GPU server is expensive, which leads to high research cost. On the other hand, due to the influence of computing resources and data scale, real-time computing cannot be realized.

III. TAXONOMY OF THE PREDICTION APPROACHES FOR TIME SERIES

According to the development process of time series forecasting methods, we classify the existing popular prediction methods of time series into three categories: classical forecasting method of time series, prediction methods of machine learning and deep learning, and hybrid forecasting methods of time series. In the subsequent parts of this section, we will introduce the existing time series data forecasting methods in detail.

A. EQUATIONS CLASSICAL FORECASTING METHOD OF TIME SERIES

The classic forecasting method of time series is based on mathematical and statistical modeling. As shown in TABLE 1, the classic linear models mainly include: autoregressive (AR) model, moving average (MA) model, autoregressive moving average (ARMA) model and autoregressive integrated moving average (ARIMA) model. And the classic nonlinear linear models mainly include Threshold Autoregressive (TAR) model [1], [2], Constant Conditional Correlation (CCC) model, conditional heteroscedasticity model [3]. In addition, there are some important classical prediction methods based on exponential smoothing, such as: Simple Exponential Smoothing (SES), Holt's linear trend method, Holt-Winters' multiplicative method, Holt-Winters' additive method and Holt-Winters' damped method.

1) THE CLASSIC LINEAR MODELS

a: PREDICTION MODEL FOR STATIONARY DATA

For the prediction of stationary time series, Yule [4] first introduced the concept of randomness into the time series, regarded each time series as the realization of a random process, and proposed an autoregressive(AR) model, regarded each time series as the realization of a random process, and proposed an the AR model. Then, the Moving Averaging (MA) model is proposed by researcher, and Wold proposed the famous Wold decomposition theorem [5]. It laid the foundation for the research of time series forecasting. Until 1970, Box and JenKins proposed the ARMA model [6], which includes three basic models: AR model, MA model and ARMA model. It is widely used in the modeling of stationary time series.

The ARMA model is the most common model for predicting of stationary time series data. It regards time series data as a random sequence, and the dependence of these random variables reflects the continuity of the original data in time.

For example, the variables include $x_1, x_2, \ldots, x_p$, which can be obtained by regression analysis:

$$Y_t = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + Z \quad (1)$$

where Y is the observed value of the prediction object, and Z is the error. As the target of prediction, Y_t is affected by its own changes, and its law can be embodied by the following formula:

$$Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + Z_t \quad (2)$$

The error has a dependent relationship in different periods and is expressed by the following formula:

$$Z_t = \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \ldots + \alpha_q \varepsilon_{t-q} \quad (3)$$

Therefore, the ARMA model expression is as follows:

$$\begin{aligned} Y_t = \ & \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} \\ & + \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \ldots + \alpha_q \varepsilon_{t-q} \end{aligned} \quad (4)$$

If the time series y_t satisfies the above formula, the time series y_t obeys ARMA(p,q).

At present, the ARMA model is widely used in time series forecasting in various fields, and achieved great results. For example, Rounaghi and Zadeh [7] used the ARMA model to make monthly and annual forecasts on the time series stock returns of the Standard & Poor's 500 Index and the London Stock Exchange. The statistical analysis of S&P 500 shows that the ARMA model for S&P 500 outperforms the London stock exchange and it is capable for predicting medium or long horizons using real known values. Chen *et al.* [8] established an adaptive ARMA model for short-term power load forecasting of power generation systems. Experiments prove that the adaptive ARMA model is more accurate than the traditional ARMA model in forecasting 24 hours and one week in advance. Zheng [9] based on the data of tuberculosis incidence and air pollution variables (PM2.5, PM10, SO2, CO, NO2, O-3) in Urumqi, the ARMA (1, (1, 3)) + model was established by time series ARMA model method, cross-correlation analysis, and principal component regression method, and its predictive performance was superior to that of the ARMA (1, (1, 3)) model based on tuberculosis historical data. During the analysis, it was found that the higher the concentration of O-3, the higher the incidence of tuberculosis. Aiming at the construction of prediction interval of ARMA (p, q) model with unknown order, Lu and Wang [10] proposed a prediction interval bootstrap algorithm based on bootstrap distribution (p, q), which significantly improves the coverage accuracy of the prediction interval compared to the methods using pre-estimated values of orders. Inoue [11] derive a closed-form expression for the finite predictor coefficients of multivariate ARMA processes, and apply the expression to determine the asymptotic behavior of a sum that appears in the autoregressive model fitting and the autoregressive sieve bootstrap. The results are new even for univariate ARMA processes. For the modeling of categorical time series, both nominal or ordinal time series, an extension of the basic discrete autoregressive moving-average (ARMA) models is proposed by WEISS C H. It uses an observation-driven regime-switching mechanism, leading to the family of RS-DARMA models. This RS-DAR(1) model constitutes a parsimoniously parameterized type of Markov chain, which has an easy-to-interpret data-generating mechanism and may also handle negative forms of serial dependence.

With the Granger causality [12] being proposed, researchers extend the univariate time series model to multivariate time series analysis. A multivariate promotion model named Vector Autoregressive Moving Average (VARMA) was proposed based on the ARMA model, which can flexibly represent Vector Autoregressive (VAR) and Vector Moving Average (VMA) model. When constructing the VARMA model, the time series data is required to be a stationary series. If the data is a non-stationary time series, you need to do first-order difference processing to obtain stationary data. However, the trend information existing in the time series will be ignored in the difference processing. In order to solve the above problems, the Vector Error Correction Model (VECM) is proposed by Engle and Granger [13], which can well consider the co-integration relationship between time series.

In summary, the ARMA time series forecasting model has made great achievements in the prediction of stationary time series. However, there is almost no purely stationary data in real time series data. Therefore, the application of this model is limited by the characteristics of the data, and its versatility is poor.

b: PREDICTION MODEL FOR NONSTATIONARY DATA

A sequence that contains characteristics such as trend, seasonality, or periodicity is called a non-stationary sequence. It may contain only one component or several components. A non-stationary time series means that after its local level or trend is eliminated, it shows homogeneity. At this time, some parts of the series are very similar to other parts. This non-stationary time series can be converted into a stationary time series after difference processing. ARIMA(p, d, q) is a well-known non-stationary time series model, which can reflect the changes of different data patterns, and the model requires fewer parameters to estimate. Therefore, the ARIMA model has been widely used.

The ARIMA model is described as follows, denote ∇ as a difference operator, we can get,

$$\nabla^2 y_t = \nabla(y_t - y_{t-1}) = y_t - 2y_{t-1} + y_{t-2} \quad (5)$$

delay operator B,

$$y_{t-p} = \mathrm{B}^p y_t, \quad (\forall p \geq 1) \quad (6)$$

$$\nabla^k = (1 - \mathrm{B})^k \quad (7)$$

If y_t is a second non-stationary time series of order d, then $\nabla^d y_t$ is a stationary time series, which can be set as an ARMA(p, q) model, as follows:

$$\lambda(\mathrm{B})(\nabla^d y_t) = \theta(\mathrm{B})\varepsilon_t \quad (8)$$

where $\lambda(B) = 1-\lambda_1 B-\lambda_2 B^2-\ldots\lambda_p B^p$ is the autoregressive coefficient polynomial, and $\theta(B) = 1-\theta_1 B-\theta_2 B^2-\ldots-\theta_p B^p$ is the moving average coefficient polynomial. ε_t is a white noise sequence of zero-mean. If the above conditions are met, the model can be called an autoregressive integrated moving average (ARIMA(p, d, q)) model. When the value of d is 0, the ARIMA model is equivalent to the ARMA model. Therefore, we can also judge whether the sequence is a stationary sequence by whether the d value is 0.

It is necessary to determine the order before constructing the ARIMA model. Autocorrelation Function(ACF) and Partial Autocorrelation Function (PACF) are often used to determine the order of the ARIMA model. In addition, Akaike Information Criterion (AIC) [14] and Bayesian Information Criterion (BIC) [15] are also used for reference. The information criterion not only considers the prediction residuals, but also adds a penalty for overfitting. After determining the order of the model, there are many methods to estimate the parameters of the ARIMA model [16], such as Least Squares Estimation (LSE), Maximum Likelihood Estimation (MLE), Bayesian estimation, etc.

At present, the ARIMA model has made good achievements for the time series forecasting of relatively simple data. For example, Valipour [17] used the ARIMA model to predict the precipitation in key areas of Iran based on 50 years of rainfall data, and proved that the ARIMA model is an effective annual precipitation prediction model. Zhang *et al.* [18] conducted an effective prediction and trend analysis of PM2.5 concentration in Fuzhou, China based on the ARIMA model. Viccione *et al.* [19] used the ARIMA model to predict and analyze the time series of the water level of the Secin tank in Avellino, Italy. And it describes in detail the change trend of the reservoir level over time. Malki *et al.* [20] predicted the possible end time and possible second outbreak time of COVID-19 in 2020 based on the ARIMA model. Zheng *et al.* [21] used the ARIMA model to predict the total health expenditure (THE), government health expenditure (GHE), social health expenditure (SHE) and out-of-pocket health expenditure (OOP) in China from 2018 to 2022. It provides a valuable theoretical basis for the adjustment of the Chinese government's health policy.

At the same time, researchers have achieved better results by improving the ARIMA model or integrating other models. For example, Li *et al.* [22] established an ARIMA-EGARCH mixed prediction model for visceral leishmaniasis. The root mean square error (RMSE) is used to evaluate the model, which proves that the model has a better predictive effect in the short-term prediction of visceral leishmaniasis, and its performance is higher than others. Lopes *et al.* [23] constructed a hybrid model based on ARIMA and neural network to predict the electric field of the vertical path near the digital television transmitter in the Amazon area. It is verified that the model is better than the classic least square method. Huang *et al.* [24] proposed a hybrid prediction model combining ARIMA and adaptive filtering, and verified the time series data set of emergency department (ED) visits, and proved that the model is more accurate than the traditional ARIMA model and is suitable for short- and medium-term use. Huang *et al.* [24] proposed a hybrid prediction model that combines ARIMA and adaptive filtering, and verified it with the emergency department (ED) time series data set. It is proved that the model has higher accuracy than the traditional ARIMA model and better applicability in the short and medium term.

In summary, the model of ARIMA has a simple structure. When the ARIMA model is applied, the time series data is required to be stationary, or stable after being differentiated. Therefore, ARIMA can only capture linear relationships, but not non-linear relationships. As shown in TABLE 1, a comparative analysis of the classic linear time series model is carried out. The characteristics, application scenarios, *et al.* of each model are summarized separately.

2) THE CLASSIC NONLINEAR MODELS

The linear prediction model is easy to understand and simple to implement. However, the linear model needs to be constructed under the linear assumption, which is less effective for nonlinear time series data. Therefore, the nonlinear time series forecasting model was proposed. The famous classic nonlinear models include Threshold Autoregressive (TAR) model, Autoregressive Conditional Heteroscedasticity (ARCH) model and Constant Conditional Correlation (CCC) model, etc.

Threshold Autoregressive (TAR) model is one of the classic nonlinear models, It uses linear autoregressive models on different intervals to describe the characteristics of nonlinear changes in the entire interval. The core idea of threshold autoregression is to introduce $L-1$ threshold values $r_j(j = 1, 2, 3, \ldots, L-1)$in the value range of the observation sequence $\{x_i\}$, and divide it into L intervals. According to the number of delay steps d, $\{x_i\}$ is assigned to different threshold intervals according to the value of $\{x_{i-d}\}$, and then $x_i$in different intervals is described by different autoregressive models. The sum of these autoregressive models is a nonlinear dynamic representation of the entire $\{x_i\}$. The formula is as follows:

$$x_i = \sum_{i=1}^{n_j} \varphi_i^{(j)} \alpha_i^{(j)}, \quad r_{j-1} < x_{i-d} \le r_j,\ j = 1, 2, \ldots L \tag{9}$$

$r_j(j = 1, 2, 3, \ldots, L-1)$ is the threshold value; L is the number of thresholds; d is the number of delay steps; $\{a_i^{(j)}\}$ is the white noise series with variance σ_j^2, and each $\{\alpha_i^{(j)}\}(j = 1, 2, 3, \ldots, L-1)$ is independent of each other, φ_i^j is the autoregressive coefficient of the j threshold interval, n_j is the order of the j threshold model.

TAR was not well promoted in the early days. The main reason is that the steps to construct the TAR model are complicated, and the number of thresholds and each threshold need to be determined first. Until Tsay proposed a convenient method, and put forward the estimation method of threshold number and the value of threshold [25]. The TAR model is widely used.

The parameters of the threshold autoregressive model change with the threshold value. When the threshold variable is selected as the delay of the time series itself, the model becomes a special kind of threshold model, called a self-exciting threshold autoregressive (SETAR) model. Then the Multivariate Threshold Autoregressive (MTAR) model was proposed by Tsay [26] based on TAR, which is a multivariate promotion model. In order to solve the problem of discontinuous dynamic transitions between different stages of the TAR model, the Smooth Transition Autoregressive (STAR) model was proposed by Terasvirta [27].

In the financial field, time series data has the characteristics of volatility clustering, such as large fluctuations followed by large fluctuations, and small fluctuations followed by small fluctuations. Aiming at the phenomenon of volatility clustering, Engle [28] proposed an Autoregressive Conditional Heteroscedasticity (ARCH) model, in which the conditional variance is a function of the square of the past return.

In the linear regression model $y_t = x_t'\beta + \varepsilon_t$, the conditional variance of the disturbance termε_t is $\sigma_t^2 \equiv Var(\varepsilon_t|\varepsilon_{t-1}, \ldots)$, which can change with time. Inspired by volatility clusters, i $\sigma_t^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2$ is the disturbance term of ARCH(1), so we can get, the $\sigma_t^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \ldots + \alpha_p\varepsilon_{t-p}^2$ is the disturbance term of ARCH(p).

Bollerslev [3] proposed a generalized autoregressive conditional heteroscedasticity (GARCH) model based on the ARCH model. Based on the ARCH model, GARCH adds the dependence of conditional variance on past conditional variances, and has the representation of the ARMA model. After that, a series of promotion models were proposed for GARCH model, including EGARCH (Exponential GARCH) model [29], APGARCH model [30], GJR-GARCH model [31] and FIGARCH model [32], etc.

In addition, Bollerslev [33] also proposed a Constant Conditional Correlation (CCC) model for nonlinear time series prediction. The CCC model divides the conditional covariance matrix into two parts: conditional variance and conditional correlation matrix. And the maximum likelihood estimation method is used to estimate the parameters. The CCC model has fewer parameters and has certain advantages in ensuring the positive definiteness of the conditional variance. After that, Engle [34] and Tsui [35] respectively proposed the promotion model Dynamic Conditional Correlation (DCC) model based on the CCC model. Compared with CCC, DCC models the conditional correlation matrix and makes the conditional correlation matrix change over time.

In summary, the classical nonlinear forecasting methods were used for nonlinear time series forecasting in the early days. Researchers have proposed a lot of generalization models, which are used to deal with more complex time series data

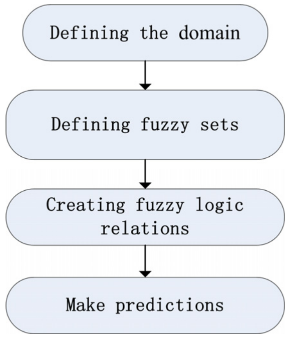


FIGURE 1. The construction of fuzzy time series prediction model.

in a specific field. And it has made very good achievements in specific fields.

B. MACHINE LEARNING FORECASTING MODEL OF TIME SERIES

The classical forecasting models of time series can well capture the linear relationships in time series, and achieve good results when the data set is small. However, it is poor in effect when applied to large-scale complex nonlinear time series. Therefore, researchers pay more attention to the time series prediction methods of machine learning or deep learning. Artificial neural networks such as multi-layer perceptrons (MLPs) networks [36] and radial basis function (RBF) networks [37] have adaptive and self-organizing learning mechanisms. They are the earliest neural network models used in nonlinear time series forecasting in different fields, and it has achieved good effect. In addition, fuzzy theory, Gaussian process regression, decision tree, support vector machine, LSTM, etc. are also used in time series forecasting, and also have good forecasting ability on nonlinear time series.

1) FUZZY TIME SERIES FORECASTING METHOD

Fuzzy time series forecasting can solve nonlinear problems and is one of the research hotspots in the field of forecast analysis. It is often used for time series forecasting of small data sets or data sets with missing values. Song and Chissom [38] proposed a fuzzy time series forecasting method based on the fuzzy set theory proposed by Zadeh [39]. Then, in order to solve the problem of complicated calculation and long calculation time, Chen [40] proposed a simple and effective method for constructing fuzzy time series model. As shown in FIGURE 1, it includes the following steps.

At first, determine the domain of the time series and divide the domain. For example, the Domain is defined as $U = \{u_1, u_2, \ldots, u_n\}$, m_i is the midpoint of n. And its fuzzy set is $Ai = \{i = 1, 2, \ldots, n\}$.

Secondly, define the fuzzy set according to the result of the division of the domain. The fuzzy set A_i can be expressed as $A_i = (a_{i1}, a_{i2}, \ldots, a_{in,})$, Where $a_{ij} \in [0, 1]$, u_j is the membership degree in A_i. If $a_{ij} = \{a_{i1}, a_{i2,} \ldots, a_{in,}\}$, then the data at time t should be classified into the j-th category.

Here, define the fuzzy set as shown in formula:

$$\begin{cases} A_1 = \frac{1}{u_1} + \frac{0.5}{u_2} + \frac{0}{u_3} + \ldots + \frac{0}{u_i} + \ldots + \frac{0}{u_{n-1}} + \frac{0}{u_n} \\ A_2 = \frac{0.5}{u_1} + \frac{1}{u_2} + \frac{0.5}{u_3} + \frac{0}{u_4} + \ldots + \frac{0}{u_i} + \ldots \\ \quad + \frac{0}{u_{n-1}} + \frac{0}{u_n} \\ A_3 = \frac{0}{u_1} + \ldots + \frac{0}{u_{i-2}} + \frac{0.5}{u_{i-1}} + \frac{1}{u_i} + \frac{0.5}{u_{i+1}} + \frac{0}{u_{i+2}} \\ \quad + \ldots + \frac{0}{u_n} \\ A_4 = \frac{0}{u_1} + \frac{0}{u_2} + \frac{0}{u_3} + \ldots + \frac{0}{u_i} + \ldots + \frac{0}{u_{n-2}} + \frac{0.5}{u_{n-1}} \\ \quad + \frac{1}{u_n} \end{cases} \tag{10}$$

Formula (10) defines the membership degree of $A_i = (i = 1, 2, \ldots, n)$ where the time is i and the value is x_i.

$$\mu_{A_i}(t) \begin{cases} 1, & if\ (i = 1)\&(x_i \leq m_1) \\ 1, & if\ (i = n)\&(x_i \leq m_n) \\ \max\{0,\ 1 - \frac{|x_i - m_i|}{2 \times l_{in}}\}, & others \end{cases} \tag{11}$$

where x_i is the observed value at time i and l_{in} is the length of the interval.

Thirdly, the original time series is fuzzified according to the fuzzy set, and the fuzzified time series is obtained. For example, the fuzzy logic relationship GTS(M, N) of the model can be divided into M $\times$ N order relationship matrix. And it is expressed as $R^{(k,l)}(k = 1, 2, \ldots, M, l = 1, 2, \ldots, N)$.

Finally, make predictions based on the fuzzy logic relationship, and defuzzify the obtained prediction results. Let $F(t-k) = (\mu_1(t-k), \mu_2(t-k), \ldots, \mu_n(t-k))$, and let the l-th maximum degree of membership be expressed as $\mu_{tl}(t-k)$. Here are the cross-fuzzy logic relationships: $\wedge_N(A_{t_1}^{(k,1)}, \ldots, A_{t_l}^{(k,l)}, A_{t_N}^{(k,N)})$.

Suppose $F^k(t) = \wedge_N(A_{t_1}^{(k,1)}, \ldots, A_{t_l}^{(k,l)}, \ldots, A_{t_N}^{(k,N)}) \circ (m_1, m_2, \ldots, m_n)^T$, where "$\circ$" is a composite operator used for prediction, and it has the following principles: if the sum of $\wedge_N(A_{t_1}^{(k,1)}, \ldots, A_{t_l}^{(k,l)}, A_{t_N}^{(k,N)})$ is equal to 0, the predicted value corresponds to the midpoint m_{t_1} of A_{t_1}; otherwise, the predicted value is the weighted sum of$m_{t_1}, \ldots, m_{t_N}$. Then, according to the given M, the predicted value of time t can be obtained by the following formula(12):

$$\tilde{F}(t) = \sum_{k=1}^{M} F^k(t) * \omega_k \tag{12}$$

where the k-th predicted adjustment parameter $\omega_k(k = 1, 2, 3, \ldots, M)$ can be obtained by minimizing the root mean square error of the training data set or other evaluation criteria.

In the study of fuzzy time series forecasting, the research focuses on the division method of the universe, the dimension of time series (from one element to multiple), fuzzy rules (from first order to higher order), and the length of prediction (from Single-step to multi-step prediction) etc. For example, in terms of domain division, Huamgt [41] improves model prediction accuracy by adjusting the fuzzy interval; Yu [42] proposed a method of dividing fuzzy intervals according to the characteristics of data distribution; Wang *et al.* [43], [44], and Lu [45] *et al.* divide the universe of discourse through the idea of information granules to better reflect the characteristics of the data; Wang and Liu [46], [47] used automatic clustering algorithm to divide the universe for the first time, and proposed a new fuzzy time series model based on Axiomatic Fuzzy Sets (AFS) classification algorithm. In the study of model order, Chen [48], [49] proposed a high-order fuzzy time series forecasting model. For long-term forecasting, Wang *et al.* [50] made improvements based on the forecasting method proposed by Dong and Pedrycz [51]. Since the information granule divides the time series unequal. Firstly, the dynamic time warping is combined with the fuzzy C-means clustering algorithm. Then, the obtained time series fragments with unequal length are clustered. Next, the fuzzy time series prediction idea is adopted. Finally, based on the idea of fuzzy time series forecasting, a long-term forecast of the time series is made. Yuan *et al.* [52] proposed a time series prediction model Kernel-HFCM based on kernel mapping and high-order fuzzy cognitive map (HFCM). He mapped the original one-dimensional time series to a multi-dimensional feature time series, and then extracted the key features of time series. Finally, the inverse kernel mapping is used to map the characteristic time series back to the predicted one-dimensional time series. And the effectiveness of this method in time series forecasting is proved at the base of the experiments with 7 benchmark data sets. Hanapi *et al.* [53] proposed a fuzzy linear regression sliding window GARCH (FLR-FSWGARCH) model, which combines fuzzy linear regression with GARCH to estimate parameters. This model improves the consistency of parameter estimation and prediction. Therefore, it improves the prediction accuracy of the model. And the experiments on two data sets prove that the model has a good degree of fit and reliability for time series forecasting.

In summary, due to the characteristics and advantages of fuzzy time series forecasting, it has become a unique research topic and is widely used in different fields. Researchers study on fuzzy time series forecasting methods to improve the accuracy and interpretability of the model, which will have important research significance and application value in the field of time series forecasting.

2) ARTIFICIAL NEURAL NETWORK (ANN)

Artificial Neural Network (ANN) is a data-driven predictive model. It has strong self-organization, self-learning and good nonlinear approximation capabilities. Therefore, it has attracted much attention from researchers in the field of time series forecasting. It has become one of the effective tools for nonlinear modeling and is widely used in various

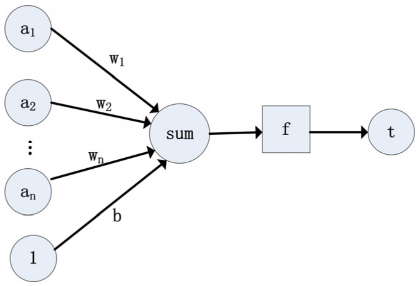


FIGURE 2. The basic structure of the artificial neural network. Where the $a_1 \sim a_n$ are the components of the input vector; $w_1 \sim w_n$ are the weights of the synapses of the neurons; b is the bias; f is the transfer function, and t is the neuron output.

fields [54]–[58]. In the ANN model, it is assumed that the number of training samples tends to be infinite, but the actual training samples are limited. Therefore, the training model often has over-fitting phenomenon, especially in the small sample modeling, the performance deterioration of the model is more obvious.

The basic structure of the artificial neural network is shown in FIGURE 2. The $a_1 \sim a_n$ are the components of the input vector; $w_1 \sim w_n$ are the weights of the synapses of the neurons; b is the bias; f is the transfer function, and t is the neuron output. The formula is as follows:

$$t = f(WA' + b) \tag{13}$$

where W is the weight vector, A is the input vector, and A' is the transpose of the A vector. Therefore, the function of a neuron is to obtain the inner product of the input vector and the weight vector, and then obtain a result through a nonlinear transfer function.

At present, neural network algorithms are widely used in the field of time series forecasting. DeGroot and Würtz [59] first verified the prediction of one-element nonlinear time series based on feedforward neural network (FFNN). Then, Chakraborty *et al.* [60] conducted an empirical study on multivariate time series based on artificial neural networks, and studied one-step and multi-step forecasting. After that, Aras and Kocakoç [61] proposed a selection strategy of neural network prediction model. The first step is to determine the number of input and hidden units; the second step is to consider the performance of each neural network model and select the neural network model from different initial weights. Finally, it is proved that the effect of this method is significantly improved on both the simulated data set and the real data set. And it has good robustness. Fotso *et al.* [62] adopted an improved BP neural network training algorithm and used it to optimize the structure of feedforward and recurrent neural networks. The wind speed is predicted by inputting the configuration of different variables, which proves that the model is better than the existing model in terms of prediction accuracy.

The methods based on computational intelligence are very popular for parameter optimization of artificial neural network models, which can give better model parameters, so as to build a more robust model. Akpinar *et al.* [63] trained the structure of artificial neural network based on artificial bee colony algorithm (ABC) to determine the weight and deviation value. The results on real time series data show that the ANN training based on ABC achieves good results in prediction. Cheng *et al.* [64] proposed a hybrid model for forecasting hydrological monthly runoff. The swarm intelligence method Gray Wolf Optimizer (GWO) and the Moore-Penrose generalized inverse method are used to optimize the input and output implicit weights of the extreme learning method (ELM). It is verified by the Three Gorges hydrological data and simulation data, which proves that the method is superior to the traditional forecasting method in many indicators. A new bootstrapped hybrid artificial neural network is proposed by Eğrioğlu *et al.* [65] for forecasting. The network has three parts: linear, non-linear and a combination with associated weights and biases. These weights are used to test the input significance, linearity and nonlinearity hypotheses with this new method providing empirical distributions for forecasts and weights. At the same time, researchers have also built a mixed model prediction model by combining artificial neural networks with classic linear models, which can improve the performance of the model. A time series analysis model (ARIMA) and an artificial neural network model (the multilayer perceptron neural network, MLPNN) were proposed for predicting wastewater inflow by Zhang *et al.* [66]. A case study of the Barrie Wastewater Treatment Facility in Barrie, Canada, was carried out to demonstrate the performance of the proposed models.

In summary, although the artificial neural network has achieved good results in time series forecasting. However, under the influence of the scale of the data, the over-fitting often occurs. Therefore, artificial neural networks are not suitable for time series forecasting of small data samples. The mixed model is better than the single model in prediction. And the research shows that the mixed model performs better than the single model in prediction.

3) GAUSSIAN PROCESS REGRESSION

Gaussian process (GP) [67] is a machine learning method based on Bayesian neural network. It is a set of random variables. Any number of random variables in the set obey the joint Gaussian distribution and are uniquely determined by the mean function and covariance function. The formula is as follows:

$$\begin{cases} m(x) = E[f(x)] \\ k(x, x') = E[(f(x) - m(x))(f(x') - m(x'))] \end{cases} \tag{14}$$

where $x, x' \in R^d$ are any random variable factor. Therefore GP can be defined as $f(x) \sim GP(m(x), k(x, x'))$.

And the model formula to solve the regression problem is as follows:

$$y = f(x) + \varepsilon \tag{15}$$

where x is the input vector, f is the function, r is the noise and y is the data polluted by noise. Assuming noise $\varepsilon \sim N(0, \sigma_n^2)$, the prior distribution of y can be obtained as the formula:

$$y \sim N(0, K(X,X) + \sigma_n^2 I_n) \tag{16}$$

And the joint distribution probability of the observed value y and the predicted value f_* is expressed as follows:

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(X,X) + \sigma_n^2 I_n & K(X,x_*) \\ K(x_*,X) & k(x_*,x_*) \end{bmatrix}\right) \tag{17}$$

where $K(X,X) = K_n = (k_{ij})$ is the covariance matrix of order $n \times n$ symmetric positive definite, and the matrix element $k_{ij} = k(x_i, x_j)$ is used to measure the correlation between x_i and x_j. $K(X,x_*) = K(x_*,X)^T$ is the $n \times 1$ order covariance matrix between the observation point x_* and the input X of the training set. $k(x_*,x_*)$ is the covariance of observation point x_* itself. I_n is the n-dimensional identity matrix.

From the above, the posterior distribution of the predicted value f can be obtained, and the formula is as follows:

$$f_*|X, y, x_* \sim N(\bar{f}_*, \mathrm{cov}(f_*)) \tag{18}$$

where,

$$\begin{aligned} \bar{f}_* &= K(x_*,X)[K(X,X) + \sigma_n^2 I_n]^{-1} y, \\ \mathrm{cov}(f_*) &= k(x_*,x_*) - K(x_*,X) \\ &\quad \times \left[K(X,X) + \sigma_n^2 I_n\right]^{-1} K(X,x_*) \end{aligned} \tag{19}$$

The $\hat{\mu}_* = \bar{f}_*$ *and* $\hat{\sigma}_{f_*}^2 = \mathrm{cov}(f_*)$ are the mean and variance of the predicted value f_* corresponding to observation point x_*.

Gaussian Process Regression (GPR) [68] is a non-parameteric model that uses Gaussian Process (GP) priors to perform regression analysis on data. The model assumptions of GPR include noise (regression residual) and Gaussian process prior. It is solved according to Bayesian inference [69]. GPR is a universal approximator of any continuous function in a compact space. In addition, GPR can provide a posterior to predict the outcome. When the likelihood is normally distributed, the posterior has an analytical form. Therefore, GPR is a probabilistic model with versatility and resolvability [70].

At present, the application of Gaussian process regression model has spread to many fields, such as tool loss prediction [71], population model non-parametric identification [72], regression and classification [73]–[75], etc. For example, Raghavendra *et al.* [75] based on the Gaussian Process Regression (GPR) model to predict the monthly groundwater level fluctuation probability of the water layer in the Kumaradara River Basin in Suryataruk, India. And the simulation experiment proves that the method has high accuracy. Hu *et al.* [76] predicted the slope displacement based on the Gaussian Process Regression (GPR) model. The displacements of the slope of the Three Gorges permanent ship lock, the slope of Wolong Temple and the high slope of Longtan Hydropower Station were predicted. It is proved that the model has good adaptability to the prediction of the nonlinear time series of slope displacement. Richardson *et al.* [77] used the GPR model to predict the health of the battery, and the experiments on the lithium-ion battery-based cycling data set proved the effectiveness of GPR for short-term and long-term predictions. Sun *et al.* [78] predicted runoff one month in advance based on the GPR model and the MOPEX hydrometeorological time series data set. The prediction results were compared with the linear regression model (LR) and the ANN model, respectively, showing that GPR predicts results in most cases Better than the latter two. It is proved that the prediction results of GPR are better than linear regression model (LR) and ANN model in most cases.

In addition, in order to achieve better time series forecasting results, researchers have proposed some improved or integrated models of GPR. For example, Liu *et al.* [79] proposed a multivariate forecasting model based on GPR and multiple imputation method for time series with missing data. The model is used for wind power forecasting in wind farms, and it is proved that the model has good forecasting performance. A hybrid GMM-IGPR [80] model is proposed by using an improved Gaussian process regression (GPR) based on Gaussian mixture model (GMM) and a variant of the basic particles swarm optimization (PSO). And the effectiveness of the proposed algorithm is verified by means of a numerical example and a real industrial winding process. Statistical tests of experimental results compared with other popular prediction models demonstrate the good performance of the proposed model. In order to reduce computational complexity and improve prediction accuracy. Yan *et al.* [81] proposed a time series prediction method for changing Gaussian processes, and proved the effectiveness of this method by predicting wind power generation in Irish wind farms. Zhang *et al.* [82] proposed a hybrid model based on GPR to predict wind speed and proved the validity of the model through data from wind farms in eastern China.

In summary, Gaussian Process Regression (GPR) is a Bayesian machine learning method. Due to the high computational cost of GPR, it is usually used for low-dimensional and small-sample regression problems. At the same time, compared with methods such as ANN and SVM, GPR requires fewer parameters for modeling and more kernel functions available, so GPR is more flexible.

4) SUPPORT VECTOR MACHINE

Support Vector Machine (SVM) is an important classification algorithm first proposed by the Vapnik team [83], which has unique advantages for small samples and nonlinear problems. It is widely used in research on classification, pattern recognition, and time series forecasting.

Support vector machines were originally applied to the two-dimensional linearly separable case. And the core idea is to find the optimal hyperplane on the basis of classification accuracy.

For example, as show in FIGURE 3, the dots and the stars represent two different types of samples, and H is the classification line (linear hyperplane), H1 and H2 are the lines parallel to the hyperplane H and passing through the samples closest to H in the two types of samples. The distance between

H1 and H2 is called the classification interval. The optimal hyperplane to be searched by the SVM algorithm refers to the maximum classification interval. It means that while accurately separating the two types of data, the distance between H1 and H2 is the largest. And the samples on the H1 and H2 lines are called support vectors. The general function form of the linear discrimination that separates the two types of samples is as follows:

$$f(x) = \omega \cdot x + b \tag{20}$$

where ω *and* b are regression coefficients. And the classification hyperplane (H) is expressed as:

$$\omega \cdot x + b = 0 \tag{21}$$

In the linear separable data sets (x_i, y_i), $i = 1, 2, \ldots, l$, $x \in R^n$; *and* $y_i \in (-1, +1)$ denote sample categories. The two samples can be expressed as follows:

$$\begin{cases} dots : \omega \cdot x + b > 1, & y_i = 1 \\ stars : \omega \cdot x + b < -1, & y_i = -1 \end{cases}$$
$$\begin{cases} hyperplane\ H_1 : x + b = 1 \\ hyperplane\ H_2 : x + b = -1 \end{cases} \tag{22}$$

The classification interval is:

$$|H_1 H_2| = \frac{2}{\parallel \omega \parallel}. \tag{23}$$

According to the basic idea of support vector machine: finding the maximum classification interval is equivalent to finding the minimum $\parallel \omega \parallel$, therefore, it is transformed into a constrained quadratic programming problem:

$$\begin{cases} \min \ \frac{1}{2} \parallel \omega \parallel^2 \\ s.t.\ y_i(\omega \cdot x + b) - 1 \geq 0, i = 1, 2, \ldots, l \end{cases} \tag{24}$$

Lagrange function is used to solve the problem of support vector machine constrained quadratic programming, the formula is as follows:

$$L(\omega, b, a) = \frac{\parallel \omega \parallel^2}{2} - \sum_{i=1}^{l} a_i [y_i(\omega \cdot x_i + b)] \tag{25}$$

Then it is transformed into a dual problem based on the Lagrange optimization method, which is to maximize:

$$W(a) = \sum_{i=1}^{l} a_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} a_i a_j y_i y_j x_i x_j \tag{26}$$

Then, satisfy the constraints:

$$\begin{cases} a \geq 0, i = 1, 2, \ldots, l \\ \sum_{i=1}^{l} a_i y_i = 0 \\ a = \{a_1, a_2, \ldots, a_l\} \end{cases} \tag{27}$$

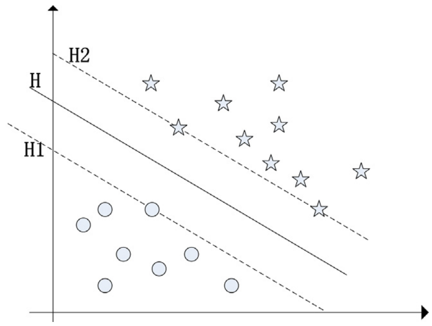


FIGURE 3. Two dimensional vector separable graph. Where the dots and the stars represent two different types of samples, and H is the classification line (linear hyperplane), H1 and H2 are the lines parallel to the hyperplane H and passing through the samples closest to H in the two types of samples.

Here we get the following formula:

$$\begin{cases} \omega^* = \sum_{i=1}^{l} a_i^* y_i x_i \\ b^* = y_i - \sum_{i=1}^{l} y_i a_i^* (x_i \cdot x) \end{cases} \tag{28}$$

When the result is not zero, the sample is the support vector. When it is zero, it indicates that the contribution of the sample vector is 0. Therefore, the optimal classification function is constructed as follows:

$$\begin{aligned} f(x) &= \mathrm{sgn}[(\omega \cdot x) + b] \\ &= \mathrm{sgn}[\sum_{i=1}^{n} a_i^* \cdot y_i (x_i \cdot x) + b] \end{aligned} \tag{29}$$

However, a single support vector machine model is not as good as traditional prediction models in the field of prediction. For example, Ashrafzadeh *et al.* [84] predict water evapotranspiration based on seasonal autoregressive integrated moving average (SARIMA), support vector machine (SVM) and data processing grouping method (GMDH) models. And verified by the time series data of 4 weather stations. SARIMA and GMDH have their own advantages on different data sets, and are better than SVM model. Then the researchers combined other methods with SVM and achieved certain results. For example, Yu *et al.* [85] proposed a residual-based deep least squares support vector machine (RBD-LSSVM). And it is proved that the RBD-LSSVM model has good performance in nonlinear time series modeling and prediction based on the annual sunspot number series and monthly total ozone column series data. Pattanayak *et al.* [86] proposed an interval unequal length determination method based on fuzzy c-means clustering, and adopted support vector machine (SVM) for modeling, while considering the value of membership. The prediction accuracy of this model is better than other methods through 10 different time series data sets. Then the researchers combined other methods with SVM and achieved certain results.

In summary, the main advantage of support vector machines is data classification. In terms of classification, it is suitable for small sample data sets, which simplifies classification and regression problems. Its computational complexity is determined by the number of support vectors, thus avoiding the ''curse of dimensionality''. In addition, its result is determined by a few support vectors. It is not sensitive to outliers, so it can capture key samples well, and has good robustness and generalization ability. However, in terms of prediction, a single support vector machine model is not effective, and a mixed model is usually used for prediction, and certain effects can be achieved.

5) RECURRENT NEURAL NETWORKS

Recurrent Neural Networks (RNN) take sequence data as input, recursively in the evolution direction of the sequence, and all nodes are connected in a chain. It is good at processing sequence and correlation data, and is widely used in the fields of pattern recognition and time series prediction. RNN is more sensitive to time series and has memory in data transmission. For example, the input at time t needs to refer to the result at time $t-1$, and an output value is obtained based on a complex function transformation. The structure of the RNN model is shown in the FIGURE 4. Where X is the input value, A is the value of the hidden layer, Y is the output value, U is the weight from the input layer to the hidden layer, W is the weight of the hidden layer, and V is the weight from the hidden layer to the output layer. From the expanded diagram of RNN, it can be seen that at time t, the input value, hidden value, and output value of the network are: X_t, A_t, Y_t. It is worth noting that the current ht is determined by the current input X_t and the hidden value A_{t-1} at the previous moment. Therefore, the calculation formula of RNN is as follows:

$$\begin{cases} y_t = g(VA_t) \\ A_t = f(Ux_t + WA_{t-1}) \end{cases} \tag{30}$$

where g is the activation function of output layer, and f is the activation function of the hidden layer. Substituting formula A_t into formula Y_t iteratively, we can get the following formula:

$$\begin{aligned} Y_t &= g(VA_t) \\ &= f(UX_t + WA_{t-1}) \\ &= Vf(UX_t + Wf(UA_{t-1} + WA_{t-2})) \\ &= Vf(UX_t + Wf(UX_{t-1} + Wf(UX_{t-2} + WA_{t-3}))) \\ &= Vf(UX_t + Wf(UX_{t-1} + Wf(UX_{t-2} \\ &\quad + Wf(UX_{t-3} + \ldots)))) \end{aligned} \tag{31}$$

From formula (31), we can know that the output value Y_t of RNN network is affected by the current input A_t and the historical input A_{t-1}, A_{t-2}, A_{t-3},.... Therefore, the RNN can rely on any number of historical input values.

However, the longer the input sequence, the more timing references are needed, which leads to the deeper network. When the sequence is long, it is difficult for the gradient to propagate back from the latter sequence to the previous sequence. As a result, the problem of vanishing gradients arises, so it is impossible to deal with the problem of long-term dependence.

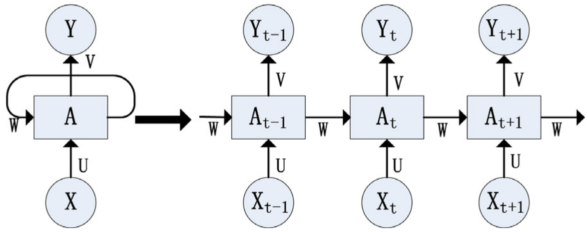


FIGURE 4. The structure of the RNN model. Where X is the input value, A is the value of the hidden layer, Y is the output value, U is the weight from the input layer to the hidden layer, W is the weight of the hidden layer, and V is the weight from the hidden layer to the output layer.

In order to solve the problem of vanishing and exploding gradient of RNN, a long short-term memory (LSTM) model was proposed. It is a special kind of RNN, which effectively avoids the phenomenon of gradient dispersion by creating a ''retention effect'' between input and feedback. As shown in the FIGURE 5, the three logic gate control units of input gate, output gate and forget gate are used by LSTM to control the transmission state. The mathematical formula of the LSTM unit is as follows:

$$\begin{cases} f_t = \sigma(W_{fA}A_{t-1} + W_{fX}X_t + b_f) \\ i_t = \sigma(W_{iA}A_{t-1} + W_{iX}X_t + b_i) \\ \tilde{c}_t = \tanh(W_{cA}A_{t-1} + W_{cX}X_t + b_c) \\ c_t = f_t \circ c_{t-1} + i_t \circ \tilde{c}_t \\ Y_t = \sigma(W_{YA}A_{t-1} + W_{YX}X_t + b_Y) \\ A_t = Y_t \circ \tanh(c_t) \end{cases} \tag{32}$$

where f_t is the output of the forget gate, i_t is the output of the input gate, $\tilde{c}_t$ is the state of the current input unit, W_{fA} is the weight from the forgetting gate to the output of the unit, W_{fX} is the weight of the forget gate to the unit input, W_{iA} is the weight from the input gate to the unit output, W_{iX} is the weight from the input gate to the unit input, W_{cA} is the weight from the current input unit state to the unit output, W_{cX} is the weight from the current input unit state to the unit input, W_{YA} is the weight from the output gate to the unit output, W_{YX} is the weight of the output forget the door to the unit input, b_f is the bias of the forget gate; b_i is the bias of the input gate, b_c is the bias of the current input unit state, b_Y is the bias of the output gate; σ is the sigmoid function, and the symbol ''$\circ$'' represents the point-by-point multiplication of 2 vectors. Each LSTM has a forget gate, an input gate, and an output gate. The forget gate and input gate are used to control the status of the unit, the forgetting gate controls the amount of unit state information saved at time t-1 to time t, the input gate controls the amount of input saved to the current unit state information at time t, and the output of the output gate control unit state output to the LSTM output at time t.

Compared with the superimposed memory f RNN, LSTM can remember the information that needs to be memorized for a long time, and forget unimportant information. Therefore, LSTM has better performance in long series prediction.

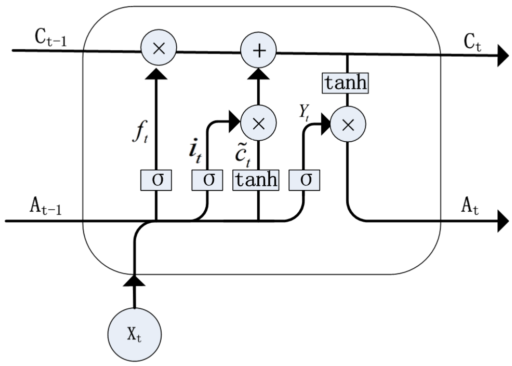


FIGURE 5. The structure of LSTM model. It has a different internal structure from RNN. Where f_t is the output of the forget gate, i_t is the output of the input gate, $\tilde{c}_t$ is the state of the current input unit, W_{fA} is the weight from the forgetting gate to the output of the unit, W_{fX} is the weight of the forget gate to the unit input, W_{iA} is the weight from the input gate to the unit output, W_{iX} is the weight from the input gate to the unit input, W_{cA} is the weight from the current input unit state to the unit output, W_{cX} is the weight from the current input unit state to the unit input, W_{YA} is the weight from the output gate to the unit output, W_{YX} is the weight of the output forget the door to the unit input, b_f is the bias of the forget gate; b_i is the bias of the input gate, b_c is the bias of the current input unit state, b_Y is the bias of the output gate; σ is the sigmoid function, and the symbol "∘" represents the point-by-point multiplication of 2 vectors.

At present, LSTM is widely used in the field of time series forecasting. For example, Sorkun *et al.* [87] proposed a multi-prediction model based on long-short-term memory (LSTM). And it finds the best model for global solar radiation data based on a combination of different meteorological variables (such as temperature, humidity and cloud cover). The results prove that the multivariate prediction method has better performance than the previous univariate model. And it is found that temperature and cloud cover are the most effective parameters for predicting future solar radiation. Liu *et al.* [88] used the LSTM model to predict lime sequences with long-term and short-term correlations. And the experiment proves that the prediction result of LSTM is better than the traditional autoregressive model.

In addition, in order to improve the effect of the model, the researchers improved the LSTM based on the characteristics of the data. For example, Karevan *et al.* [89] proposed the transfer LSTM (T-LSTM) prediction model, which is based on the long-term dependence of LSMT and the large influence of nearby samples on model fitting. And the experiments on two different climate data sets prove that T-LSTM has good performance in prediction tasks. Sagheer *et al.* [90] proposed the structure of deep long short-term memory (DLSTM). And the genetic algorithm is used to optimize the configuration of DLSTM. Based on the production data of two actual oilfields, different metrics have been used to prove that the DLSTM model is superior to other methods. At the same time, researchers also improve the prediction performance of LSTM by integrating models. For example, Sun *et al.* [91] proposed the AdaBoost-LSTM ensemble learning method. And, the AdaBoost-LSTM ensemble method based on financial data experiments is proved to be better than the single prediction method. And it has good advantages for non-linear and irregular time series data forecasting.

In summary, the LSTM model is widely used in the field of time series forecasting. Because it optimizes the gradient problem, it is suitable for long series data analysis. And the research shows that the performance of the integrated LSTM model is usually better than that of a single model.

6) TRANSFORMER

Transformer is a classic NLP model proposed by the Google team in 2017. The popular Bert model is also based on Transformer. The Transformer model uses the Self-Attention mechanism instead of the sequential structure of RNN, so that the model can be trained in parallel and can obtain global information. It is also currently used for time series forecasting.

As shown in FIGURE 6, the structure of transformer is also composed of encoder and decoder. The encoder part takes historical time series as input, and in the decoder part, the future value is predicted based on autoregressiveness. The decoder is connected to the encoder with an attention mechanism. In this way, the decoder can learn to "pay attention" to the most useful part of the historical value of the time series before making a prediction. The decoder uses masked self-attention. In this way, the network will not obtain future values during training, and it can avoid information leakage.

At present, Transformer is used in the field of time series forecasting. For example, Neo Wu et al. [92] developed a novel method that employs Transformer-based machine learning models to forecast time series data. This approach works by leveraging self-attention mechanisms to learn complex patterns and dynamics from time series data. Moreover, it is a generic frame work and can be applied to univariate and multivariate time series data, as well as time series embeddings. Using influenza-like illness (ILI) forecasting as a case study, they shows that the forecasting results produced by the approach of Transformer-based are favorably comparable to the state-of-the-art. Wu *et al.* [93] propose a new time series forecasting model - Adversarial Sparse Transformer (AST), based on Generative Adversarial Networks (GANs). Specifically, AST adopts a Sparse Transformer as the generator to learn a sparse attention map for time series forecasting, and uses a discriminator to improve the prediction performance at a sequence level. Extensive experiments on several real-world datasets show the effectiveness and efficiency of this method. And Wu *et al.* [94] develop a hierarchically structured Spatial-Temporal Transformer network (STtrans) which leverages a main embedding space to capture the inter-dependencies across time and space for alleviating the data imbalance issue. To make the latent spatial-temporal representations be reflective of the relational structure between categories, a cross-category fusion transformer network is applicated. Extensive experiments on the imbalanced spatial-temporal datasets from NYC and Chicago

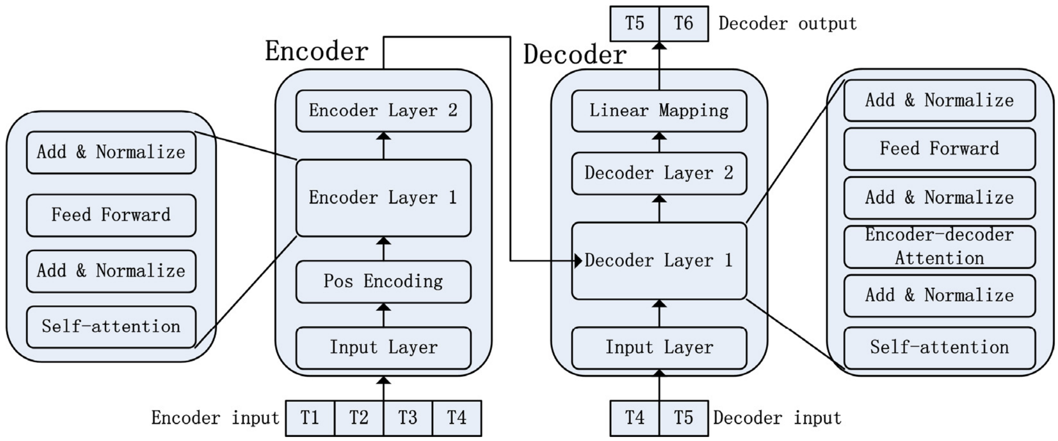


FIGURE 6. Architecture of transformer-based forecasting model.

demonstrate the superiority of this method over various state-of-the-art baselines.

In summary, transformer is used in time series forecasting and has achieved good results. It can well capture the complex dependencies between time series and can be calculated in parallel. However, the method transformer cannot capture long-distance information between sequences and the amount of calculation is large.

As shown in TABLE 2, a comparative analysis of different time series model is carried out. The Data haracteristic, Advantage, and the Disadvantage of each model are summarized separately.

C. HYBIRD FORECASTING MODEL

In the field of time series forecasting, the forecasting methods of classical and machine learning have their respective advantages. However, real time series data have the following characteristics: (1) It is difficult to determine whether the time series data is linear or non-linear, and it is impossible to judge the validity of a specific model to the data; (2) In reality, there are few purely linear or purely nonlinear time series data, and time series data is usually a combination of linear and nonlinear; (3) In the field of time series forecasting, no model is suitable for any situation, and a single model cannot capture different time series patterns at the same time.

Therefore, in order to capture different distribution patterns in data, a hybrid model that combines classic and machine learning has become a development trend. The hybrid model can capture the complex distribution pattern of the time series and improve the accuracy and generalization ability of the model.

1) HYBRID MODEL BASED ON ARMA AND MACHINE LEARNING

The hybrid algorithm combining ARMA and machine learning has been applied in different fields and has achieved good results. For example, Wang *et al.* [95] proposed a clustering hybrid prediction model C-PSO-SVM-ARMA based on ARMA and PSO-SVM models. And the experiments with historical data of wind farms have proved that the model has high computational efficiency and good wind power forecasting performance. In order to improve the prediction accuracy of mine safety production situation, an empirical model decomposition gray autoregressive moving average model EMD-GM-ARMA was proposed by Wu *et al.* [96]. And it proves that the model has high precision and high stability forecasting effect in short and medium term forecasting. Through real-time wind speed prediction experiments, it is proved that VMD-PRBF-ARMA-E has high prediction accuracy and can better and truly reflect the characteristics of wind speed.

2) HYBRID MODEL BASED ON ARIMA AND MACHINE LEARNING

In addition, the hybrid method combining ARIMA and machine learning is widely used in the field of time series forecasting. For example, Kao *et al.* [97] proposed an EEMD-ARIMA-GA-SVR hybrid forecasting model based on integrated empirical mode decomposition (EEMD) to predict primary energy consumption economy. Compared with other prediction mechanisms, this hybrid framework has better prediction accuracy. Unnikrishnan and Jothiprakash [98] proposed a hybrid model (SSA-ARIM-ANN). Singular Spectrum Analysis (SSA) is used as a preprocessing tool to separate stationary and non-stationary components from rainfall data, and ARIMA is used to model stationary components, and ANN is used to model non-stationary components. The results of daily rainfall forecasts in the Koyna watershed in Maharashtra prove that the SSA-ARIMA-ANN hybrid model is highly accurate. Fraiha Lopes *et al.* [99] proposed a hybrid model based on ARIMA and artificial neural network (ANN) to predict electromagnetic wave propagation in dense urban forest areas. And it is proved that the hybrid model is more accurate than the classical least square method (LS). Ozoegwu *et al.* [100] proposed a hybrid method based on non-linear autoregressive and artificial neural network, and verified that the hybrid method has high accuracy for global solar forecasting based on different statistical

TABLE 2. Comparative analysis of different models.

Method	Data haracteristic	Advantage	Disadvantage
Fuzzy method	The amount of data is small, the data is incomplete and there are missing values	Good at handling incomplete data and uncertain data.	(1)Low accuracy The (2)interpretability is weak.
ANN	Large-scale sample data set	(1)Self-learning ability (2)The ability to quickly find the optimal solution	(1)interpretability is weak. (2) Large amount of calculation (3)large-scale data sets is required
GPR	Low-dimensional and small sample data	Different kernel functions can be specified, and the model is flexible	(1)complete sample data is requested (2)High-dimensional space failure
SVM	High-dimensional small-scale data set	(1)Insensitive to outliers, (2)Excellent robustness and generalization ability	(1) It is difficult to implement large-scale sample data; (2) Sensitive to parameter and kernel function selection.
RNN	Short dependency, correlation sequence data	Good at handling short dependent sequence data	There is a problem of gradient explosion and gradient disappearance
LSTM	long dependency, correlation sequence data	(1)Good at handling long dependent sequence data (2)Can construct deeper neural networks	(1)Large amount of calculation (2) Calculation time-consuming
Transformer	Different length sequence data of complex dependence	Can capture complex dependencies, Can be calculated in parallel	(1)Large amount of calculation (2)Can not capture long distance information

criteria. Wu *et al.* [101] proposed a hybrid method combining ARIMA time series and support vector machine to forecast corn futures prices. And it is verified that the method has outstanding advantages in the forecasting of corn futures prices. A hybrid model based on the combination of autoregressive integrated moving average (ARIMA) and long short-term memory (LSTM) was proposed by Deng *et al.* [102]. It predicts the number of outpatient visits in the hospital, which proves that the hybrid model has higher prediction accuracy and more stable model performance than the single LSTM and ARIMA model. Wang *et al.* [103] proposed a hydrological time series prediction model based on wavelet denoising and ARIMA-LSTM. The wavelet is used to remove interference factors in hydrological time series, and ARIMA is used for data fitting and prediction. It proves that the model is more suitable for hydrological time series forecasting and has better forecasting effect.

3) HYBRID MODEL BASED ON MACHINE LEARNING

At the same time, the researchers also proposed some integrated prediction models based on machine learning and deep learning, and proved that it works very well in practical applications. For example, Chan *et al.* [104] proposed a hybrid time prediction model that combines convolutional neural networks (CNN) and support vector machines (SVM). And it is proved that the model solves the short-term load forecasting problem. Livieris *et al.* [105] proposed the CNN-LSTM model to predict the price and trend of gold. The convolutional layer is used to extract useful knowledge and learn the internal features of time series data. And the LSTM layer is used to identify short-term and long-term correlations of data. It is proved that the convolutional layer can significantly improve the prediction performance of LSTM. Sahoo *et al.* [106] compared the applicability of long and short memory recurrent neural network (LSTM-RNN), CNN model and naive method in low-traffic time series prediction. And different models are verified through different evaluation indicators, which proves that the LSTM-RNN model is superior to the RNN model and the naive method.

In summary, with the development of big data, massive time series data are rapidly being produced in different fields. And the data is complex nonlinear data and obeys different distribution patterns. When a simple statistical model or a machine learning model is used to process complex time series data, the performance is poor and the generalization ability is weak. However, the hybrid model is better than the single model in terms of accuracy and generalization ability. Therefore, when you don't know which model is appropriate, a hybrid time series forecasting model is one of the best choices.

In addition, there are some other interesting time series analysis algorithms. For example, the time series analysis methods based on complex networks [107]. Complex networks provide a new method for nonlinear time series analysis, and has been widely used. And according to the different definitions of network nodes and connected edges, the network can be divided into three main methods: recursive network, viewable network and transformed network. These methods provide supplementary knowledge of phase space geometry and dynamics for the original nonlinear time series analysis. At present, time series network method is one of the research hotspots. And it still has a lot of room for development, especially the comparative study of network methods and traditional methods. And another interesting method is a fuzzy interval time-series forecasting model on the basis of network-based multiple time frequency-spaces and the

TABLE 3. The common methods for missing data.

Method classification	Method name	Disadvantage
Direct delete method	Direct delete method	Loss of important data information
Nearest interpolation method	Nearest interpolation method	Ignore changes between adjacent data
Filling method of statistics	Mean filling, Median filling, Constant value filling	Ignore timing information between adjacent data
	Linear filling	Ignore data distribution mode information
Filling method of machine learning	KNN（K-Nearest Neighbor） RNN（Recurrent Neural Networks） EM（Expectation-Maximization） MF（Matrix Factorization）	Ignore timing information between adjacent data

induced-ordered weighted averaging aggregation (IOWA) operation, and it was proposed for energy and finance forecasting modeling by Liu *et al.* [108]. It was proved based on a hydrological time series from the Biliuhe Riverin China and two well-known sets of financial time-series data, Taiwan Stock Exchange Capitalization Weighted Stock Index and Hang Seng Index. And the results of this approach can achieve better performance than the existing models.

IV. FUTURE DIRECTIONS AND OPEN ISSUES

The third section summarizes the time series forecasting methods proposed in recent years. In recent years, time series forecasting research has made great achievements. However, due to the rapid growth of data scale and the generation of highly complex time series data, it has brought huge challenges to the existing forecasting methods and affected the calculation efficiency of the forecasting methods. Some potential research directions and trends of time series forecasting research in the future can be summarized as follows.

A. DATA PREPROCESSING

As we all know, the quality of data has an important impact on data modeling. However, there are some missing data in the process of generating massive time series data. For the problem of missing data, the common methods are as follows in TABLE 3. These methods usually ignore the correlation between adjacent time series data. And filling in the data blindly will destroy the distribution pattern of the sample data. In order to capture the pattern distribution information of the sample data and supplement the missing data based on the pattern information of the sample data, the deep learning method will be one of the effective methods. For example, filling the missing data according to the sample data distribution pattern based on the generative adversarial neural network will obtain a complete time series that is as similar as possible to the original data, which will help the subsequent work of time series analysis.

At present, we have not found a high-precision and efficient method for filling missing values in time series. However, with the need for time series data analysis, we believe that filling in missing values based on the original data distribution pattern will become a very promising direction in the field of time series analysis research.

B. MODEL CONSTRUCTION

Combine the characteristics of real-time incremental data with the model to improve the accuracy and robustness of the model. At present, people usually build forecasting models based on historical time series data and use them for future data forecasting. However, with the passage of time and the influence of various objective factors, the model needs to be revised based on the characteristics of the new data and the data distribution pattern to improve its accuracy and robustness. Otherwise, as time goes by, the prediction accuracy and performance of the model will not meet our needs. Therefore, the revised time series forecasting model based on incremental data will become a potential research direction.

So far, we have not found a revised time series forecasting model based on online incremental data. With the accumulation of large-scale time series and the demand for forecast analysis, we believe it will become a very promising direction in the field of time series forecasting.

C. PARALLEL COMPUTING

With the rapid growth of time series, online real-time analysis of time series will become a development requirement. At present, the time series analysis model is constructed based on the stand-alone mode. Researchers usually use high-performance GPU servers to improve computing efficiency. On the one hand, GPU server equipment is expensive, which increases the cost of research. On the other hand, it is limited by the influence of computing resources and data scale, and cannot realize real-time predictive analysis. Therefore, the parallel computing of time series data based on big data technology will become a potential research direction.

V. CONCLUSION

With the development of the era of big data, forecasting research based on time series data has become one of the hot spots. More and more time series data are produced in various fields, which provides a data basis for the research of time series analysis methods. It promotes the further development of the field of time series analysis. Due to the complex pattern distribution of large-scale time series data, more and more researchers are capturing complex time series distribution patterns based on hybrid forecasting models to obtain better forecasting accuracy and performance.

This paper first presents the concept of time series and summarizes the relevant issues in the current research field of time series forecasting. Then, the time series forecasting methods are introduced by classification. On this basis, we summarized several potential research directions and unsolved problems, such as data preprocessing, incremental data model construction, and parallel computing.

By this comprehensive survey, we expect that our classification and analysis of existing time series forecasting methods will provide a reference for research in the field of time series forecasting.

REFERENCES

[1] H. Tong and K. S. Lim, "Threshold autoregression, limit cycles and cyclical data," *J. Roy. Stat. Society: Ser. B (Methodological)*, vol. 42, no. 3, pp. 245–268, Jul. 1980.

[2] H. Tong, "On a threshold model in pattern recognition and signal processing," *Sijthoff Noordhoff*, vol. 5, no. 3, pp. 575–586, 1978.

[3] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *J. Econometrics*, vol. 31, no. 3, pp. 307–327, Apr. 1986.

[4] G. U. Yule, "On a method of investigating periodicities in distributed series, with special reference to Wolfer's sunspot numbers," *Philos. Trans. Roy. Soc. London A*, vol. 226, nos. 636–646, Jan. 1927, doi: 10.1098/rsta.1927.0007.

[5] H. Wold, "On prediction in stationary time series," *Ann. Math. Statist.*, vol. 19, no. 4, pp. 558–567, 1948.

[6] G. Janacek, "Time series analysis forecasting and control," *J. Time Ser. Anal.*, vol. 3, no. 3, pp. 131–133, Dec. 2009.

[7] M. M. Rounaghi and F. N. Zadeh, "Investigation of market efficiency and financial stability between S&P 500 and London stock exchange: Monthly and yearly forecasting of time series stock returns using ARMA model," *Phys. A, Stat. Mech. Appl.*, vol. 456, pp. 10–21, Aug. 2016.

[8] J.-F. Chen, W.-M. Wang, and C.-M. Huang, "Analysis of an adaptive time-series autoregressive moving-average (ARMA) model for short-term load forecasting," *Electr. Power Syst. Res.*, vol. 34, no. 3, pp. 187–196, Sep. 1995.

[9] Y. Zheng, "Predictive study of tuberculosis incidence by ARMA model combined with air pollution variables," *Complexity*, vol. 2020, May 2020, Art. no. 3619063.

[10] X. Y. Lu and L. H. Wang, "Bootstrap prediction interval for ARMA models with unknown orders," *REVSTAT-Stat. J.*, vol. 18, no. 3, pp. 96–375, 2020.

[11] A. Inoue, "Closed-form expression for finite predictor coefficients of multivariate ARMA processes," *J. Multivariate Anal.*, vol. 1761, pp. 44–52, Mar. 2020.

[12] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, Aug. 1969.

[13] R. F. Engle and C. W. J. Granger, "Cointegration and error-correction: Representation, estimation and testing," *Econometrica*, vol. 55, no. 2, pp. 251–276, Aug. 1987.

[14] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. AC-19, no. 6, pp. 716–723, Dec. 1974.

[15] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, no. 2, pp. 5–15, Mar. 1978.

[16] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, "Time series analysis: Forecasting and control, 5th edition," *J. Oper. Res. Soc.*, vol. 22, no. 2, pp. 199–201, 2015.

[17] M. Valipour, "Critical areas of Iran for agriculture water management according to the annual rainfall," *Eur. J. Sci. Res.*, vol. 84, no. 4, pp. 600–608, 2012.

[18] L. Zhang, J. Lin, R. Qiu, X. Hu, H. Zhang, Q. Chen, H. Tan, D. Lin, and J. Wang, "Trend analysis and forecast of $PM_{2.5}$ in Fuzhou, China using the ARIMA model," *Ecol. Indicators*, vol. 95, pp. 702–710, Dec. 2018.

[19] G. Viccione, C. Guarnaccia, S. Mancini, and J. Quartieri, "On the use of ARIMA models for short-term water tank levels forecasting," *Water Supply*, vol. 20, no. 3, pp. 787–799, May 2020.

[20] Z. Malki, E.-S. Atlam, A. Ewis, G. Dagnew, A. R. Alzighaibi, G. ELmarhomy, M. A. Elhosseini, A. E. Hassanien, and I. Gad, "ARIMA models for predicting the end of COVID-19 pandemic and the risk of second rebound," *Neural Comput. Appl.*, vol. 33, no. 3, pp. 2929–2948, 2021.

[21] A. Zheng *et al.*, "An application of ARIMA model for predicting total health expenditure in China from 1978–2022," *J. Global Health*, vol. 10, no. 1, pp. 1–8, 2020.

[22] H.-L. Li, R.-J. Zheng, Q. Zheng, W. Jiang, X.-L. Zhang, W.-M. Wang, X. Feng, K. Wang, and X.-B. Lu, "Predicting the number of visceral leishmaniasis cases in Kashgar, Xinjiang, China using the ARIMA-EGARCH model," *Asian Pacific J. Tropical Med.*, vol. 13, no. 2, pp. 35–44, 2020.

[23] R. L. F. Lopes, S. G. C. Fraiha, V. D. Lima, H. S. Gomes, and G. P. S. Cavalcante, "Hybrid ARIMA and neural network modelling applied to telecommunications in urban environments in the Amazon region," *Int. J. Antennas Propag.*, vol. 2020, pp. 1–14, Feb. 2020.

[24] Y. Huang *et al.*, "Medical service demand forecasting using a hybrid model based on ARIMA and self-adaptive filtering method," *BMC Med. Inform. Decis. Making*, vol. 20, no. 1, pp. 220–237, 2020.

[25] R. S. Tsay, "Testing and modeling multivariate threshold models," *J. Amer. Stat. Assoc.*, vol. 93, no. 443, pp. 1188–1202, Sep. 1998.

[26] R. S. Tsay, "Identifying multivariate time series models," *J. Time Ser. Anal.*, vol. 10, no. 4, pp. 357–372, Jul. 1989.

[27] T. Teräsvirta, "Specification, estimation, and evaluation of smooth transition autoregressive models," *J. Amer. Stat. Assoc.*, vol. 89, no. 425, pp. 208–218, Mar. 1994.

[28] R. F. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of U," *Econometrica*, vol. 89, no. 425, pp. 987–1008, 1981.

[29] D. B. Nelson, "Conditional heteroskedasticity in asset returns: A new approach," *Econ., J. Econ. Soc.*, vol. 59, no. 2, pp. 37–64, 1991.

[30] Z. Ding, C. W. J. Granger, and R. F. Engle, "A long memory property of stock market returns and a new model," *J. Empirical Finance*, vol. 1, no. 1, pp. 93–106, 1993.

[31] R. Glosten, R. Jagannathan, and E. Runkle, "On the relation between the expected value and the volatility of the nominal excess return on stocks," *J. Econometrics*, vol. 48, no. 5, pp. 1779–1801, 1993.

[32] R. T. Baillie, T. Bollerslev, and H. O. Mikkelsen, "Fractionally integrated generalized autoregressive conditional heteroskedasticity," *J. Econometrics*, vol. 74, no. 1, pp. 3–30, 1996.

[33] T. Bollerslev, "Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model," *Rev. Econ. Statist.*, vol. 72, no. 3, pp. 498–505, 1990.

[34] R. Engle, "Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models," *J. Bus. Econ. Statist.*, vol. 20, no. 3, pp. 339–350, Jul. 2002.

[35] T. Tsui, "A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations," *J. Bus. Econ. Statist.*, vol. 20, no. 3, pp. 351–362, 2002.

[36] A. S. Lapedes and R. M. Farber, "How neural nets work," in *Proc. Neural Inf. Process. Syst.* DBLP: Denver, CO, USA, 1987, pp. 331–346.

[37] J. Park and I. W. Sandberg, "Approximation and radial-basis-function networks," *Neural Comput.*, vol. 5, no. 2, pp. 305–316, Mar. 1993.

[38] S. Qiang and B. S. Chissom, "Forecasting enrollments with fuzzy time series—Part I," *Fuzzy Sets Syst.*, vol. 62, no. 1, pp. 1–8, 1991.

[39] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, 1965.

[40] S.-M. Chen, "Forecasting enrollments based on fuzzy time series," *Fuzzy Sets Syst.*, vol. 81, no. 3, pp. 311–319, Aug. 1996.

[41] K.-H. Huarng, T. H.-K. Yu, and Y. W. Hsu, "A multivariate heuristic model for fuzzy time-series forecasting," *IEEE Trans. Syst., Man Cybern., B, Cybern.*, vol. 37, no. 4, pp. 836–846, Aug. 2007.

[42] H.-K. Yu, "A refined fuzzy time-series model for forecasting," *Phys. A, Stat. Mech. Appl.*, vol. 346, nos. 3–4, pp. 657–681, Feb. 2005.

[43] L. Wang, X. Liu, and W. Pedrycz, "Effective intervals determined by information granules to improve forecasting in fuzzy time series," *Expert Syst. Appl.*, vol. 40, no. 14, pp. 5673–5679, Oct. 2013.

[44] L. Wang, X. Liu, W. Pedrycz, and Y. Shao, "Determination of temporal information granules to improve forecasting in fuzzy time series," *Expert Syst. Appl.*, vol. 41, no. 6, pp. 3134–3142, May 2014.

[45] W. Lu, X. Chen, W. Pedrycz, X. Liu, and J. Yang, "Using interval information granules to improve forecasting in fuzzy time series," *Int. J. Approx. Reasoning*, vol. 57, pp. 1–18, Feb. 2015.

[46] W. Wang and X. Liu, "Fuzzy forecasting based on automatic clustering and axiomatic fuzzy set classification," *Inf. Sci.*, vol. 294, pp. 78–94, Feb. 2015.

[47] X. Tian, L. I. Jing, and X. Liu, "Axiomatic fuzzy set theory and its applications," in *Proc. Inf. Sci., 2nd Int. Conf. Inf. Syst. Sci.*, Dalian, China, 2008, pp. 934–945.

[48] S.-M. Chen, "Forecasting enrollments based on high-order fuzzy time series," *Cybern. Syst.*, vol. 33, no. 1, pp. 1–16, Jan. 2002.

[49] S.-M. Chen and C.-D. Chen, "Handling forecasting problems based on high-order fuzzy logical relationships," *Expert Syst. Appl.*, vol. 38, no. 4, pp. 3857–3864, Apr. 2011.

[50] W. Wang, W. Pedrycz, and X. Liu, "Time series long-term forecasting model based on information granules and fuzzy clustering," *Eng. Appl. Artif. Intell.*, vol. 41, pp. 17–24, May 2015.

[51] R. Dong and W. Pedrycz, "A granular time series approach to long-term forecasting and trend forecasting," *Phys. A, Stat. Mech. Appl.*, vol. 387, no. 13, pp. 3253–3270, May 2008.

[52] K. Yuan, J. Liu, S. Yang, K. Wu, and F. Shen, "Time series forecasting based on kernel mapping and high-order fuzzy cognitive maps," *Knowl.-Based Syst.*, vol. 206, Oct. 2020, Art. no. 106359.

[53] A. L. M. Hanapi, M. Othman, R. Sokkalingam, N. Ramli, A. Husin, and P. Vasant, "A novel fuzzy linear regression sliding window GARCH model for time-series forecasting," *Appl. Sci.*, vol. 10, no. 6, p. 1949, Mar. 2020.

[54] L. Wang, Z. Wang, H. Qu, and S. Liu, "Optimal forecast combination based on neural networks for time series forecasting," *Appl. Soft Comput.*, vol. 66, pp. 1–17, May 2018.

[55] X. Chen, X. Chen, J. She, and M. Wu, "A hybrid time series prediction model based on recurrent neural network and double joint linear–nonlinear extreme learning network for prediction of carbon efficiency in iron ore sintering process," *Neurocomputing*, vol. 249, pp. 128–139, Aug. 2017.

[56] C. Zhang, H. Wei, J. Zhao, T. Liu, T. Zhu, and K. Zhang, "Short-term wind speed forecasting using empirical mode decomposition and feature selection," *Renew. Energy*, vol. 96, pp. 727–737, Oct. 2016.

[57] H. D. N. Camelo, P. S. Lucio, J. B. V. Leal, Jr., P. C. M. D. Carvalho, and D. V. G. D. Santos, "Innovative hybrid models for forecasting time series applied in wind generation based on the combination of time series models with artificial neural networks," *Energy*, vol. 151, pp. 347–357, May 2018.

[58] X. Dong, Y. Lian, and Y. Liu, "Small and multi-peak nonlinear time series forecasting using a hybrid back propagation neural network," *Inf. Sci.*, vol. 424, pp. 39–54, Jan. 2018.

[59] C. de Groot and D. Würtz, "Analysis of univariate time series with connectionist nets: A case study of two classical examples," *Neurocomputing*, vol. 3, no. 4, pp. 177–192, Nov. 1991.

[60] K. Chakraborty, K. Mehrotra, C. K. Mohan, and S. Ranka, "Forecasting the behaviour of multivariate time series using neural networks," *Neural Netw.*, vol. 5, no. 6, pp. 961–970, 1992.

[61] S. Aras and İ. D. Kocakoç, "A new model selection strategy in time series forecasting with artificial neural networks: IHTS," *Neurocomputing*, vol. 174, pp. 974–987, Jan. 2016.

[62] F. H. R. Fogno, K. C. V. Aloyem, and G. D. Kenmoe, "Optimal input variables disposition of artificial neural networks models for enhancing time series forecasting accuracy," *Appl. Artif. Intell.*, vol. 34, no. 11, pp. 792–815, 2020.

[63] M. Akpinar, M. F. Adak, and N. Yumusak, "Time series forecasting using artificial bee colony based neural networks," in *Proc. Int. Conf. Comput. Sci. Eng. (UBMK)*, Oct. 2017, pp. 554–558.

[64] X. Cheng, Z.-K. Feng, and W.-J. Niu, "Forecasting monthly runoff time series by single-layer feedforward artificial neural network and grey wolf optimizer," *IEEE Access*, vol. 8, pp. 157346–157355, 2020.

[65] E. Eğrioğlu and R. Fildes, "A new bootstrapped hybrid artificial neural network approach for time series forecasting," *Comput. Econ.*, vol. 1, pp. 1–29, Nov. 2020.

[66] Q. Zhang, Z. Li, S. Snowling, A. Siam, and W. El-Dakhakhni, "Predictive models for wastewater flow forecasting based on time series analysis and artificial neural network," *Water Sci. Technol.*, vol. 80, no. 2, pp. 243–253, Jul. 2019.

[67] M. Seeger, "Gaussian processes for machine learning," *Int. J. Neural Syst.*, vol. 14, no. 2, pp. 69–106, 2004.

[68] X. Chen, J. Zhou, B. Jia, X. Yang, and C. Zhou, "Characterizing the hydraulic connection of cascade reservoirs for short-term generation dispatching via Gaussian process regression," *IEEE Access*, vol. 8, pp. 145489–145502, 2020.

[69] B. A. Warner and R. M. Neal, "Bayesian learning for neural networks (Lecture notes in statistical vol. 118)," *J. Amer. Stat. Assoc.*, vol. 92, no. 438, pp. 791–792, 1997.

[70] D. Duvenaud, "Automatic model construction with Gaussian processes," Ph.D. dissertation, Dept. Eng., Univ. Cambridge, Cambridge, U.K., 2014.

[71] D. Kong, Y. Chen, and N. Li, "Gaussian process regression for tool wear prediction," *Mech. Syst. Signal Process.*, vol. 104, pp. 556–574, May 2018.

[72] M. Neve, G. De Nicolao, and L. Marchesi, "Nonparametric identification of pharmacokinetic population models via Gaussian processes," *IFAC Proc. Volumes*, vol. 38, no. 1, pp. 295–300, 2005.

[73] C. K. I. Williams, *Prediction With Gaussian Processes: From Linear Regression to Linear Prediction and Beyond*. Dordrecht, The Netherlands: Springer, 1998, pp. 599–621.

[74] C. K. I. Williams and D. Barber, "Bayesian classification with Gaussian processes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 12, pp. 1342–1351, Dec. 1998.

[75] N. S. Raghavendra and P. C. Deka, "Multistep ahead groundwater level time-series forecasting using Gaussian process regression and ANFIS," in *Advanced Computing and Systems for Security*. New Delhi, India: Springer, 2016, pp. 289–302.

[76] B. Hu, G. Su, J. Jiang, J. Sheng, and J. Li, "Uncertain prediction for slope displacement time-series using Gaussian process machine learning," *IEEE Access*, vol. 7, pp. 27535–27546, 2019.

[77] R. R. Richardson, M. A. Osborne, and D. A. Howey, "Gaussian process regression for forecasting battery state of health," *J. Power Sources*, vol. 357, pp. 209–219, Jul. 2017.

[78] A. Y. Sun, D. Wang, and X. Xu, "Monthly streamflow forecasting using Gaussian process regression," *J. Hydrol.*, vol. 511, pp. 72–81, Apr. 2014.

[79] T. Liu, H. Wei, and K. Zhang, "Wind power prediction with missing data using Gaussian process regression and multiple imputation," *Appl. Soft Comput.*, vol. 71, pp. 905–916, Oct. 2018.

[80] T. Liu, H. Wei, S. Liu, and K. Zhang, "Industrial time series forecasting based on improved Gaussian process regression," *Soft Comput.*, vol. 24, no. 20, pp. 15853–15869, Oct. 2020.

[81] J. Yan, K. Li, E. Bai, Z. Yang, and A. Foley, "Time series wind power forecasting based on variant Gaussian process and TLBO," *Neurocomputing*, vol. 189, pp. 135–144, May 2016.

[82] C. Zhang, H. Wei, X. Zhao, T. Liu, and K. Zhang, "A Gaussian process regression based hybrid approach for short-term wind speed prediction," *Energy Convers. Manage.*, vol. 126, pp. 1084–1092, Oct. 2016.

[83] V. N. Vapnik and A. Chervonenkis, "A note on one class of perceptrons," *Automat. Remote Control*, vol. 25, pp. 1–10, Jan. 1964.

[84] A. Ashrafzadeh, O. Kisi, P. Aghelpour, S. M. Biazar, and M. A. Masouleh, "Comparative study of time series models, support vector machines, and GMDH in forecasting long-term evapotranspiration rates in northern Iran," *J. Irrigation Drainage Eng.*, vol. 146, no. 6, pp. 1–10, 2020.

[85] Y. Yu and J. Li, "Residuals-based deep least square support vector machine with redundancy test based model selection to predict time series," *Tsinghua Sci. Technol.*, vol. 24, no. 6, pp. 706–715, Dec. 2019.

[86] R. M. Pattanayak, S. Panigrahi, and H. S. Behera, "High-order fuzzy time series forecasting by using membership values along with data and support vector machine," *Arabian J. for Sci. Eng.*, vol. 45, no. 12, pp. 10311–10325, Dec. 2020.

[87] M. C. Sorkun, Ö. D. Incel, and C. Paoli, "Time series forecasting on multivariate solar radiation data using deep learning (LSTM)," *TURKISH J. Electr. Eng. Comput. Sci.*, vol. 28, no. 1, pp. 211–223, Jan. 2020.

[88] Y. Liu, Z. Su, H. Li, and Y. Zhang, "An LSTM based classification method for time series trend forecasting," in *Proc. 14th IEEE Conf. Ind. Electron. Appl. (ICIEA)*, Jun. 2019, pp. 402–406.

[89] Z. Karevan and J. A. K. Suykens, "Transductive LSTM for time-series prediction: An application to weather forecasting," *Neural Netw.*, vol. 125, pp. 1–9, May 2020.

[90] A. Sagheer and M. Kotb, "Time series forecasting of petroleum production using deep LSTM recurrent networks," *Neurocomputing*, vol. 323, pp. 203–213, Jan. 2019.

[91] S. Sun, Y. Wei, and S. Wang, "AdaBoost-LSTM ensemble learning for financial time series forecasting," in *Computer Science—ICCS*, vol. 10862. Cham, Switzerland: Springer, 2018, pp. 590–597.

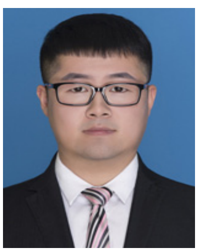
[92] N. Wu, B. Green, X. Ben, and S. O'Banion, "Deep transformer models for time series forecasting: The influenza prevalence case," in *Proc. 37th Int. Conf. Mach. Learn.*, Vienna, Austria, 2020, pp. 1–10.

[93] S. Wu, X. Xiao, Q. Ding, P. Zhao, Y. Wei, and J. Huang, "Adversarial sparse transformer for time series forecasting," in *Proc. 34th Conf. Neural Inf. Process. Syst. (NeurIPS)*, vol. 33. Electr Network, 2020, pp. 1–11.

[94] X. Wu, C. Huang, C. Zhang, and N. V. Chawla, "Hierarchically structured transformer networks for fine-grained spatial event forecasting," in *Proc. Web Conf. (WWW)*, Apr. 2020, pp. 2320–2330.

[95] Y. Wang, D. Wang, and Y. Tang, "Clustered hybrid wind power prediction model based on ARMA, PSO-SVM, and clustering methods," *IEEE Access*, vol. 8, pp. 17071–17079, 2020.

[96] M. Wu, Y. Ye, N. Hu, Q. Wang, H. Jiang, and W. Li, "EMD-GM-ARMA model for mining safety production situation prediction," *Complexity*, vol. 2020, Jun. 2020, Art. no. 1341047.
[97] Y.-S. Kao, K. Nawata, and C.-Y. Huang, "Predicting primary energy consumption using hybrid ARIMA and GA-SVR based on EEMD decomposition," *Mathematics*, vol. 8, no. 10, p. 1722, Oct. 2020.
[98] P. Unnikrishnan and V. Jothiprakash, "Hybrid SSA-ARIMA-ANN model for forecasting daily rainfall," *Water Resour. Manage.*, vol. 34, no. 11, pp. 3609–3623, Sep. 2020.
[99] R. L. F. Lopes, S. G. Fraiha, H. S. Gomes, V. D. Lima, and G. P. Cavalcante, "Application of hybrid ARIMA and artificial neural network modelling for electromagnetic propagation: An alternative to the least squares method and ITU recommendation P.1546-5 for Amazon urbanized cities," *Int. J. Antennas Propag.*, vol. 2020, Mar. 2020, Art. no. 8494185.
[100] C. G. Ozoegwu, "Artificial neural network forecast of monthly mean daily global solar radiation of selected locations based on time series and month number," *J. Cleaner Prod.*, vol. 216, pp. 1–13, Apr. 2019.
[101] W. Shengchao, S. Fengjing, and S. Rencheng, "Corn futures price forecast based on ARIMA time series and support vector machine," in *Proc. 4th Int. Conf. Syst. Comput. Big Data*, 2019, no. 5, pp. 41–49.
[102] Y. Deng, H. Fan, and S. Wu, "A hybrid ARIMA-LSTM model optimized by BP in the forecast of outpatient visits," *J. Ambient Intell. Humanized Comput.*, vol. 8, pp. 1–11, Oct. 2020.
[103] Z. Wang and Y. Lou, "Hydrological time series forecast model based on wavelet de-noising and ARIMA-LSTM," in *Proc. IEEE 3rd Inf. Technol., Netw., Electron. Automat. Control Conf.*, Mar. 2019, no. 6, pp. 1697–1701.
[104] S. Chan, I. Oktavianti, and V. Puspita, "A deep learning CNN and AI-tuned SVM for electricity consumption forecasting: Multivariate time series data," in *Proc. IEEE 10th Annu. Inf. Technol., Electron. Mobile Commun. Conf. (IEMCON)*, Oct. 2019, pp. 488–494.
[105] I. E. Livieris, E. Pintelas, and P. Pintelas, "A CNN–LSTM model for gold price time-series forecasting," *Neural Comput. Appl.*, vol. 32, no. 23, pp. 17351–17360, Dec. 2020.
[106] B. B. Sahoo, R. Jha, A. Singh, and D. Kumar, "Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting," *Acta Geophysica*, vol. 67, no. 5, pp. 1471–1481, Oct. 2019.
[107] S. Mao and F. Xiao, "Time series forecasting based on complex network analysis," *IEEE Access*, vol. 99, no. 1, pp. 1–10, 2019.
[108] G. Liu, F. Xiao, C.-T. Lin, and Z. Cao, "A fuzzy interval time-series energy and financial forecasting model using network-based multiple time-frequency spaces and the induced-ordered weighted averaging aggregation operation," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 11, pp. 2677–2690, Nov. 2020.



ZHENYU LIU is currently pursuing the Ph.D. degree with the School of Computer and Information Engineering, Inner Mongolia Agricultural University, China. His current research interest includes the forecast analysis of time series data.

ZHENGTONG ZHU is currently a Graduate Student at the School of Computer and Information Engineering, Inner Mongolia Agricultural University, China. His research interests includes the forecast analysis of time series data.

JING GAO received the Ph.D. degree from Beihang University, in 2009. She is currently a Professor at the School of Computer and Information Engineering, Inner Mongolia Agricultural University, China. She serves as the Director of the Key Laboratory of Large Data Research and Application of Agriculture and Animal Husbandry in Inner Mongolia. She published over 40 international journal articles and conference papers cited by EI and SCI. Her research interests include time series mining, big data, cloud computing, and artificial intelligence technology in the field of agriculture and animal husbandry application research. She led the team to complete more than 20 software, big data, and cloud computing platform projects, which were applied to the fields of agriculture and animal husbandry, and won the first and second prizes of the Inner Mongolia Autonomous Region for scientific and technological progress. She serves as the Vice Chairman of the "Inner Mongolia Autonomous Region Data Science and Big Data Society" and a member of the Academic Committee of the "Inner Mongolia Autonomous Region Discipline Inspection and Supervision Big Data Laboratory."

CHENG XU is currently a Graduate Student at the School of Computer and Information Engineering, Inner Mongolia Agricultural University, China. His research interest includes the forecast analysis of time series data.