

Research Article

Forecast and Analysis of Coal Traffic in Daqin Railway Based on the SARIMA-Markov Model

Cheng Zhang ¹ and Shouchen Liu ^{1,2}

¹School of Transportation and Logistics, East China Jiao Tong University, Nanchang 330013, China

²School of Business Administration, Fujian Business University, Fuzhou 350012, China

Correspondence should be addressed to Shouchen Liu; 20426164@qq.com

Received 9 February 2020; Revised 2 December 2020; Accepted 11 December 2020; Published 23 December 2020

Academic Editor: Hongju Cheng

Copyright © 2020 Cheng Zhang and Shouchen Liu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the continuous advancement of China's supply-side structural reform, the country's energy consumption structure has undergone considerable changes, including an overall reduction in fossil energy use and a rapid increase in clean energy application. In the context of China's coal overcapacity, port and rail capacities are difficult to change in the short term. This study forecasts the monthly coal traffic of Daqin Railway on the basis of the seasonal autoregressive integrated moving-average Markov model and then uses the monthly coal transport data of this railway from September 2009 to November 2019 as samples for model training and verification. Coal traffic from December 2019 to September 2020 is accurately predicted. This study also analyzes the effects of China's industrial structure adjustment, clean energy utilization, and low-carbon usage on the coal transport volume of Daqin Railway. In addition, the characteristics of seasonal fluctuation and the development trend of Daqin Railway's coal traffic are explored. This study provides a reference for adjusting the train operation chart of Daqin Railway's coal transport and developing a special coal train operation plan. It can determine the time of coal transport peak warning, improve the efficiency of coal transport management, and eventually realize a reasonable allocation of resources for Daqin Railway.

1. Introduction

Accounting for 1% of the length of China's railways, Daqin Railway comprises 20% of the country's national railway and 13% of its national coal transport volume. Daqin Railway holds and even constantly breaks the record for the fastest train running speed, the highest running density, the largest single railway volume, and the best transportation efficiency. It is a strategic artery of China's "west-to-east coal transport," which continuously carries coal to the Bohai Sea at a rate of 6.3 tons per second. However, under the background of China's coal production capacity, port and railway transport capacities are excessive and difficult to change within a short period. Moreover, the competition for China's coal supply transport channel is becoming increasingly fierce, and consequently, Daqin Railway is expected to experience pressure from new industry competition. Historical data can be used to predict the changing trend of coal traffic

in Daqin Railway. The railway transportation department can then adjust the train operation chart and formulate a coal train operation plan in accordance with the predicted coal traffic.

At present, local and foreign scholars have focused on railway freight forecasts. Many research achievements have been reported in the prediction of railway freight volume. Commonly used methods include the adaptive flocking algorithm [1], support vector machines [2], the gray model [3], time series models [4], neural networks [5], combination of models [6], multiple models [7], and regression analysis models [8]. Liu and Yu [9] used the seasonal autoregressive integrated moving-average (SARIMA) model to predict and analyze railway freight volume. Zhang [10] applied a time series model and a neural network to predict the annual freight volume of the Guangzhou-Shenzhen railway line. Zhao [11] solved an autoregressive integrated moving-average (ARIMA) model by using EViews software and predicted

China's railway freight volume. Huang et al. [12] and Yuan et al. [13] analyzed the considerable error produced by the gray Verhulst model in predicting railway freight volume and used a Markov chain model to modify the prediction results of the Verhulst model, improving prediction accuracy. Zhang and Zhou [14] used the gray forecast-Markov chain-qualitative analysis method to predict railway freight volume. Tang [4] constructed an improved gray MARKOV prediction model and predicted the future freight demand of China's railways. In Milenković et al. [15], the time series of the monthly passenger flow in Serbian railways from January 2004 to June 2014 was fitted and predicted using the SARIMA method. In Tang and Deng [16], an ARIMA model was established and R programming language was used to solve this model to make reasonable predictions of the future development trend of civil aviation passenger transport.

On the basis of existing studies, local and foreign scholars have conducted considerable research on railway freight transport prediction by using different methods and from various perspectives. However, relatively few studies on railway coal transport prediction have been performed in the context of China's coal production capacity, excessive railway transport capacity, and fierce competition among coal transport channels. In the current work, the actual coal transport volume situation of Daqin Railway is considered, and the SARIMA-Markov model is adopted to predict the monthly coal transport volume of Daqin Railway. This research analyzes the economic and market reasons for coal transportation in Daqin Railway and explores the seasonal fluctuation characteristics of coal transportation in this railway. The results of this study will provide an important reference for the managers of Daqin Railway to adjust their train operation chart and formulate a special operation plan for coal trains. The peak time of coal transport can be determined, and the efficiency of coal transport management can be improved.

The remainder of this paper is organized as follows. In Section 2, we provide the seasonal inspection data of monthly coal volume in Daqin Railway. In Section 3, we describe the SARIMA-Markov prediction model in detail. In Section 4, we verify the applicability of the SARIMA-Markov model by selecting the monthly coal volume of Daqin Railway from January 2009 to September 2019 as sample data for testing the model and forecasting future trends. Finally, we summarize the study in Section 5.

2. Seasonal Inspection of the Monthly Coal Traffic Volume of Daqin Railway

Daqin Railway provides a steady momentum for China's sustained economic development, and it has become a landmark among China's railways since the country's reform and opening up 40 years ago. In the monthly forecast of the coal traffic volume of Daqin Railway, the historical data of the railway's monthly coal traffic volume should be analyzed, and then an appropriate forecast model should be developed. In the current study, the monthly coal traffic volume of Daqin Railway from January 2009 to September 2019 is selected as the

observation data, and the time chart of the coal traffic volume of Daqin Railway is drawn using EVIEWS 10.0, as shown in Figure 1.

As shown in Figure 1, the monthly coal traffic of Daqin Railway exhibited a linear growth trend from 2009 to 2014. As a pilot industry of the supply-side structural reform, the State Council and the National Development and Reform Commission issued corresponding policies to cut coal capacity from 2015 to 2016, resulting in a linear decrease in the coal transport volume of Daqin Railway. After 2017, remarkable achievements were made in adjusting the industrial structure, and the overall benefit of the coal industry recovered steadily. The coal volume of Daqin Railway presented an increasing trend. In the short term, the monthly coal traffic of Daqin Railway exhibits evident seasonal fluctuation characteristics, with a cycle of 12 months. When constructing a prediction model, seasonal, trend growth, and random interference factors should be considered. In addition, the parameters of a matching prediction model should be selected to reduce prediction errors.

3. Modeling the SARIMA-Markov Prediction Model

3.1. SARIMA Prediction Model. The SARIMA model is derived from the ARIMA model. Its basic form is SARIMA $(p, d, q)(P, D, Q)^S$. The monthly coal transport volume time series of Daqin Railway is $\{x_t, t = 0, \pm 1, \pm 2, \dots\}$. The autoregressive (AR) model is as follows:

$$x_t = c + \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \dots + \varphi_p x_{t-p} + \varepsilon_t. \quad (1)$$

The moving-average (MA) model is as follows:

$$x_t = c + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}, \quad (2)$$

where c is a constant, ε_t is the residual sequence, and $\varepsilon_t \sim N(0, \sigma^2)$. When p is the lag order of the time series, the model is referred to as the AR(p) model. When q is the lag order of the residual sequence, the model is referred to as the MA(q) model. Equations (1) and (2) can be simplified as follows:

$$\begin{aligned} \varphi(B)x_t &= c + \varepsilon_t, \\ x_t &= c + \theta(B)\varepsilon_t, \end{aligned} \quad (3)$$

where B is the backward shift operator, and

$$\begin{aligned} B^j x_i &= x_{i-j}, \\ \varphi(B) &= 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p, \\ \theta(B) &= 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q. \end{aligned} \quad (4)$$

Models AR(p) and MA(q) are combined to produce ARMA(p, q),

$$\begin{aligned} x_t &= c + \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \dots + \varphi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} \\ &\quad - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}. \end{aligned} \quad (5)$$

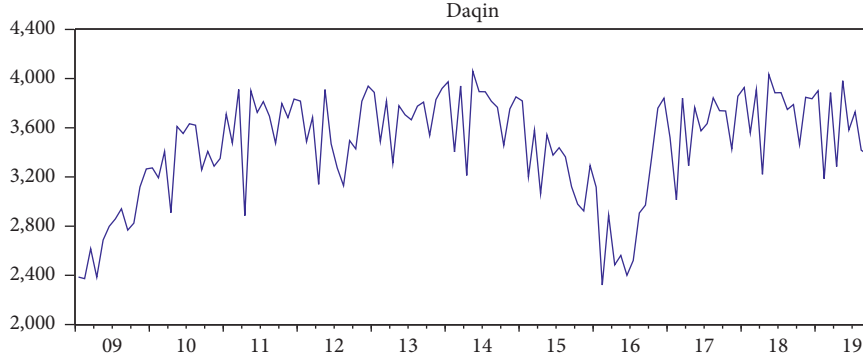


FIGURE 1: Time chart of the monthly coal traffic of Daqin Railway.

Equation (5) is simplified to the following:

$$\varphi(B)x_t = c + \theta(B)\varepsilon_t. \quad (6)$$

The time series should be stationary during analysis; otherwise, false regression will be produced, leading to unreliable predicted results. Obtaining a strictly stationary time series is difficult, and thus, a time series is required to be weakly stationary. The difference method is typically used to change a weakly stationary time series into a stationary sequence after one difference; however, other time series may require multiple differences. A stationary sequence with d -order difference is called a d -order difference sequence. A d -order difference sequence is applied to the ARMA(p, q) model to form the ARIMA(p, d, q) model, which is represented by the following:

$$\varphi(B)(1-B)^d x_t = c + \theta(B)\varepsilon_t. \quad (7)$$

Considering the periodicity of a time series, the SARIMA model can be obtained from the seasonal difference and the seasonal parameters of the ARIMA model. The general form of the SARIMA model is as follows:

$$\varphi(B)\Phi(B^s)(1-B)^d(1-B^s)^D x_t = c + \theta(B)\Theta(B^s)\varepsilon_t, \quad (8)$$

where s is the period of a time series, D is the order of seasonal difference, and B^s is the seasonal shift operator. Then, $\Phi(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_p B^{ps}$, $\Theta(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_q B^{qs}$.

The SARIMA model is denoted as SARIMA(p, d, q)(P, D, Q)^s, where d is the difference order of each period, D is the order of seasonal difference, p is the autoregressive order, q is the moving-average order, P is the seasonal autoregressive order, and Q is the seasonal moving-average order.

The steps of the SARIMA prediction model are as follows:

- (1) The sample data sequence is stabilized. Historical data for the observation period are selected as the sample sequence x_t . The first-order difference of x_t is determined to obtain $dx_t = D(x_t, d, S)$. The trend component in x_t is extracted, where d is the

difference order and S is the number of periodic difference steps.

- (2) The SARIMA(p, d, q)(P, D, Q)^s model is identified. The parameters (p, d, q)(P, D, Q) of the model are determined. The correlation of the difference sequence dx_t is analyzed, and the possible values of p and q are preliminarily identified in accordance with the truncated and trailing autocorrelation coefficient and partial correlation coefficient of dx_t [17]. Then, the significance of the model parameters is tested. The index values of R^2 , the Akaike information criterion (AIC), and the Schwarz criterion (SC) are compared. The optimal model parameters are identified.
- (3) Model adaptability is tested, and parameters are estimated. The correlation of the fitting residual sequence ε_t is analyzed to check if it is a white noise sequence. Whether the SARIMA(p, d, q)(P, D, Q)^s model fully extracts the useful information contained in sequence x_t is verified. If the correlation test result of ε_t is significant, then ε_t is not a white noise sequence, and the model cannot be adopted even if its evaluation index value is higher. By contrast, when the model passes the adaptability test, the least squares method is used to estimate model parameters.
- (4) The model's predictive power is evaluated. SARIMA(p, d, q)(P, D, Q)^s is evaluated. The evaluation index of the predictive power of the model in the sample period is $\text{MAPE} = (1/n)|(\hat{x}_t - x_t)/x_t| \times 100$. If the prediction effect of $\text{MAPE} \leq 5\%$ is excellent, then the prediction effect of $5\% \leq \text{MAPE} \leq 10\%$ is good, that of $10\% \leq \text{MAPE} \leq 20\%$ is qualified, and that of $10\% \leq \text{MAPE} \leq 20\%$ is unqualified [18].

3.2. Markov Prediction Model. The Markov model exerts a nonposterior effect. For a random time series $\{X(t), t \in T\}$, where T is a discrete time set, the nonposterior effect is expressed as follows [19]:

$$P\{X_{n+1} = i_{n+1} | X_1 = i_1, X_2 = i_2, \dots, X_n = i_n\} = P\{X_{n+1} = i_{n+1} | X_n = i_n\}. \quad (9)$$

The steps of the Markov prediction model are as follows:

- (1) The residual sequence e_i is solved as follows:

$$e_i = x_i - f_i, \quad i = 0, 1, 2, \dots, \quad (10)$$

where x_i denotes the original data, and f_i is the predicted value of the SARIMA model [20].

- (2) The states are divided. e_i is divided into r states, with the same spacing width from large to small. The upper limit of the j state in step i is as follows:

$$U_{ij} = \min e_i + \frac{j-1}{r} (\max e_i - \min e_i), \quad (11)$$

and lower limit is as follows:

$$L_{ij} = \min e_i + \frac{j}{r} (\max e_i - \min e_i). \quad (12)$$

- (3) The probability of state transition in k steps is solved. Suppose the state space is $\Omega = \{i_1, i_2, \dots, i_r\}$, and the probability of state transition in step k is $\{i_i = i_1, i_i = i_2, \dots, i_i = i_r\}$ [21].
- (4) The state transition matrix is built. A one-step transfer matrix that can reflect the probability transfer of various states between systems is denoted as follows:

$$P_{(0)} = \begin{pmatrix} p_{11} & \cdots & p_{1r} \\ \vdots & & \vdots \\ p_{r1} & \cdots & p_{rr} \end{pmatrix}. \quad (13)$$

- (5) The predicted value of the Markov model is expressed as follows:

$$\hat{x}_{t+1} = f_{t+1} - \sum_{i=1}^r a_i(t) z_i, \quad (14)$$

where \hat{x}_{t+1} is the predicted value of the SARIMA-Markov model at time $t+1$ [22].

$a_i(t)$ ($i = 1, 2, \dots, r$) is the probability for a one-step state transition of the row vector of the state transition matrix [22–24]. z_i ($i = 1, 2, \dots, r$) is the possible predicted value of each state interval. $z_i = (U_{ij} + L_{ij})/2$ is selected in this study [25].

4. Case Analysis

To verify the applicability of the SARIMA-Markov model, this study selected the monthly coal traffic volume of Daqin Railway from January 2009 to September 2019 as sample data to establish the prediction model. Evaluation indicators were adopted to evaluate the model. Among these, sample

data from October to November 2019 were used for the model test, and trend extrapolation was performed for the period of December 2019 to September 2020.

4.1. SARIMA Model

- (1) The data sequence x_t of the sample period is stabilized. EViews 10.0 software is used to find the first difference of x_t . The mean value of $dx_t = D(x_t, 1, 0)$ is approximately zero, and the trend characteristic disappears, as shown in Figure 2. The linear trend of sequence x_t is fully extracted by the first-order difference [26].

The autocorrelation diagram of the first-order difference sequence dx_t is provided in Figure 3. The autocorrelation coefficients near the 12th and 24th orders are significantly not 0, and thus, a seasonal feature with a period of 12 exists in the first-order difference sequence dx_t . This finding is consistent with the intuitive analysis results of the sequence diagram [27].

After the difference operation with a period of 12 on dx_t , a new sequence $d'x_t = D(x_t, 1, 12)$ is obtained [28]. As shown in Figure 4, the 12th-order autocorrelation coefficient of the second-order difference sequence $d'x_t$ is close to 0, indicating that the periodic factor dx_t is fully extracted [29–31].

An augmented Dickey-Fuller (ADF) test was performed on the second-order difference sequence $d'x_t$, and the results are provided in Table 1.

The unit root statistic is $ADF = -7.635505$, which is less than the critical value with a significance level of 1%. Thus, the null hypothesis that states that the second-order difference sequence $d'x_t$ has a unit root was rejected, indicating that the sequence $d'x_t$ is a stationary sequence. Accordingly, the parameters of the SARIMA $(p, d, q)(P, D, Q)^S$ model are $d = 1, D = 1, S = 12$.

- (2) The SARIMA $(p, d, q)(P, D, Q)^S$ model is determined [32]. As shown in Figure 4, x_t exhibits no significant correlation between different points in the same period. Thus, a simple seasonal model is established for the data sequence SARIMA $(p, 1, q)(1, 1, 1)^{12}$ of the sample period. The 1st-order autocorrelation coefficient of the differential stationary $d'x_t$ is significant, and its 2nd-order partial autocorrelation coefficient is also significant. The autocorrelation and partial autocorrelation diagrams are tailed; thus, five models can be built as follows:

$$\begin{aligned} & \text{SARIMA}(2, 1, 1)(1, 1, 1)^{12}, \\ & \text{SARIMA}(4, 1, 0)(1, 1, 1)^{12}, \\ & \text{SARIMA}(1, 1, 3)(1, 1, 1)^{12}, \\ & \text{SARIMA}(1, 1, 1)(1, 1, 1)^{12}, \\ & \text{SARIMA}(1, 1, 2)(1, 1, 1)^{12}. \end{aligned} \quad (15)$$

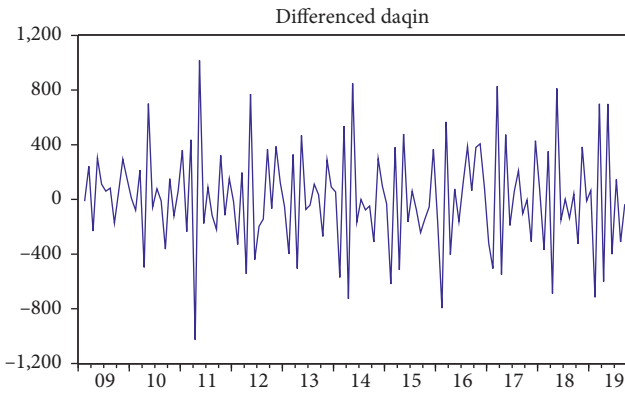


FIGURE 2: First-order difference sequence of x_t .

Sample: 2009M01 2019M09
Included observations: 128

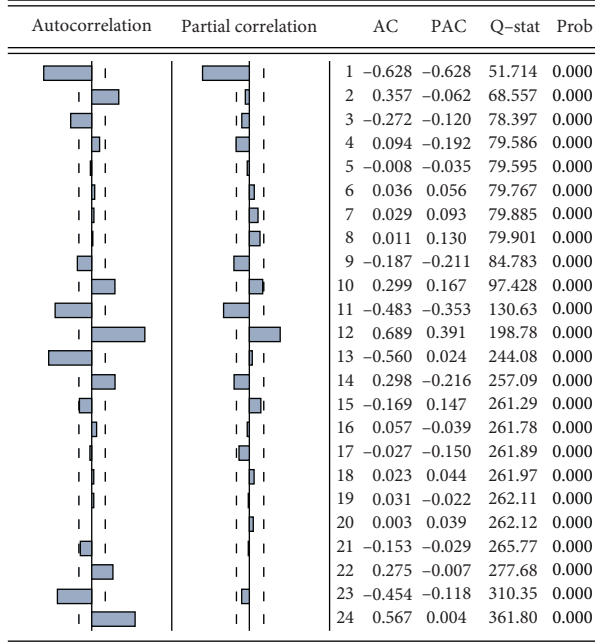


FIGURE 3: Autocorrelation of sequence dx_t .

The five models are tested for parameter significance, and the results are presented in Table 2.

R^2 is the goodness of fit of the entire model; the higher the R^2 value, the better the fit degree, where $R^2 \in [0, 1]$. AIC and SC are information criteria; the lower the value, the better the fitting degree of the model [33]. The test results of five compared models are provided in Table 2. The sequence model of data x_t in the sample period is SARIMA (4, 1, 2) (1, 1, 1)¹².

- (3) Model adaptability is tested, and parameters are estimated. In EViews 10.0, the SARIMA (4, 1, 2) (1, 1, 1)¹² model is used to obtain the fitting residual sequence ε_t , as shown in Figure 5.

Sample: 2009M01 2019M09
Included observations: 114

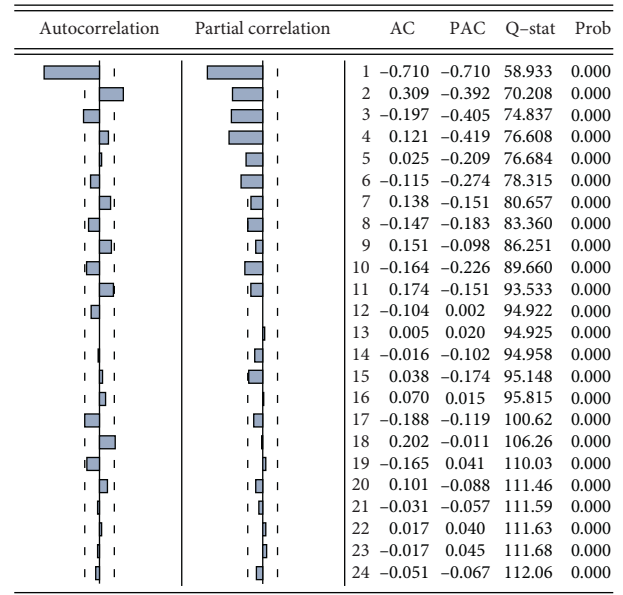


FIGURE 4: Autocorrelation of sequence $d'x_t$.

TABLE 1: ADF test results of sequence $d'x_t$.

	t-statistic	Prob. *
Augmented dickey-fuller test statistic	-7.635505	0.0000
Test critical values	1% level -3.487550	
	5% level -2.886509	
	10% level -2.580163	

TABLE 2: Significance test of model parameters.

Model	R^2	AIC	SC
SARIMA (2, 1, 1) (1, 1, 1) ¹²	0.3977	14.2007	14.2676
SARIMA (1, 1, 2) (1, 1, 1) ¹²	0.4073	14.1851	14.2519
SARIMA (1, 1, 3) (1, 1, 1) ¹²	0.4064	14.1866	14.2534
SARIMA (4, 1, 0) (1, 1, 1) ¹²	0.0088	14.6795	14.7241
SARIMA (4, 1, 2) (1, 1, 1) ¹²	0.6762	13.7281	13.8395

A white noise test is performed on the model residual ε_t of SARIMA (4, 1, 2) (1, 1, 1)¹²; that is, a random test of fitting residual sequence ε_t . The test results are presented in Figure 6.

As shown in Figure 6, the statistical value of Q is 8.963, and the associated probability is 0.34 > 0.05 from line $K = 12$. The null hypothesis states that no correlation exists and ε_t is accepted, indicating that ε_t is a white noise sequence. Therefore, the SARIMA (4, 1, 2) (1, 1, 1)¹² model passes the adaptability test, and its expression is as follows:

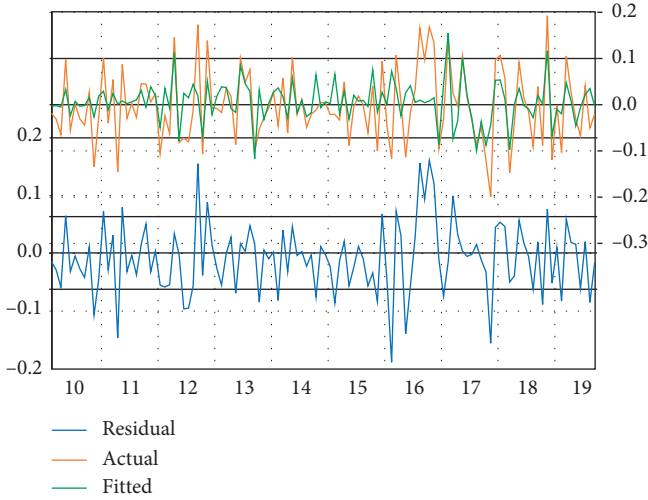


FIGURE 5: Fitting of the SARIMA(4, 1, 2)(1, 1, 1)¹² model.

Date: 11/20/19 Time: 13:22
 Sample: 2009M01 2019M09
 Included observations: 116
 Q-statistic probabilities adjusted for 4 ARMA terms

Autocorrelation	Partial correlation	AC	PAC	Q-stat	Prob
█	█	1	-0.1...	-0.1...	4.148...
█	█	2	-0.0...	-0.0...	4.158...
█	█	3	-0.1...	-0.1...	6.182...
█	█	4	-0.0...	-0.0...	6.182...
█	█	5	0.02...	-0.0...	6.236... 0.01...
█	█	6	-0.0...	-0.1...	7.317... 0.02...
█	█	7	0.07...	0.02...	7.989... 0.04...
█	█	8	-0.0...	-0.0...	8.049... 0.08...
█	█	9	0.02...	-0.0...	8.153... 0.14...
█	█	10	-0.0...	-0.0...	8.413... 0.20...
█	█	11	0.06...	0.05...	8.958... 0.25...
█	█	12	-0.0...	0.00...	8.963... 0.34...

FIGURE 6: Correlation test of the residual sequence ε_t in the SARIMA(4, 1, 2)(1, 1, 1)¹² model.

$$\begin{aligned}
 & \cdot (1 + 0.6569B + 0.9473B^2 + 0.2925B^3 - 0.0381B^4) \\
 & \cdot (1 - 0.148B^{12})\Delta_{12}\lg x_t \\
 & = (1 - 0.4657B - 0.8686B^2)(1 + 0.87B^{12})\varepsilon_t.
 \end{aligned} \tag{16}$$

(4) Model predictive power is assessed. In EViews 10.0, the SARIMA(4, 1, 2)(1, 1, 1)¹² model is used to predict the sample data of the coal traffic volume of Daqin Railway from January 2010 to September 2019. The predicted results are presented in Figure 7. An analysis model prediction level evaluation index, namely, mean absolute percentage error (MAPE) =

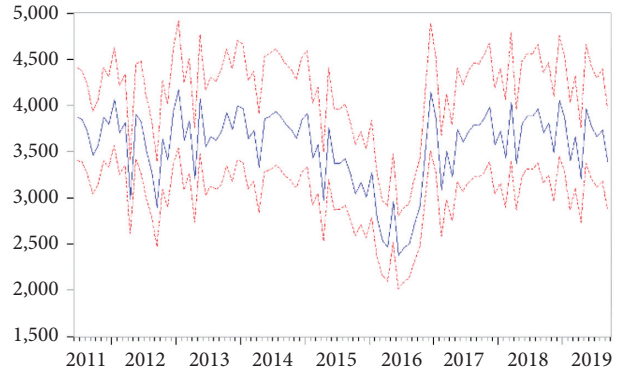


FIGURE 7: Prediction result of the SARIMA(4, 1, 2)(1, 1, 1)¹² model in the sample period.

5.1439%, is used to determine whether the model's predictive capability is good.

The SARIMA(4, 1, 2)(1, 1, 1)¹² model is used to predict the monthly coal transport volume of Daqin Railway from January 2009 to September 2019, and the results are provided in Table 3. The analysis of Table 3 indicates that the error of the single prediction model is still relatively large. To avoid this situation, the Markov model is adopted to correct the residual. This model can reduce the relative residual value and find the internal regularity.

4.2. SARIMA-Markov Model. In accordance with the steps of the Markov model, the range of the residual sequence $e_i \in [-6.884, 5.834]$ is first determined by dividing the residual sequence into E_1, E_2, E_3, E_4 state intervals. The boundary value of state interval E_1 is as follows:

$$U_{ij} = -6.884 + \frac{1-1}{4} (5.834 - (-6.884)) = -6.884, \tag{17}$$

$$L_{ij} = -6.884 + \frac{1}{4} (5.834 - (-6.884)) = -3.705.$$

Similarly,

$$\begin{aligned}
 E_2 & \in [-3.705, -0.525], \\
 E_3 & \in [-0.525, 2.655], \\
 E_4 & \in [2.655, 5.834].
 \end{aligned} \tag{18}$$

In accordance with the state distribution of the residual series, the frequency statistics of the one-step transfer from E_i to E_j are provided in Table 4.

The state transition probability matrix is as follows:

$$P_{(0)} = \begin{pmatrix} 0 & 0 & 0.6000 & 0.4000 \\ 0.0263 & 0.2895 & 0.4737 & 0.2105 \\ 0.0545 & 0.4000 & 0.4909 & 0.0545 \\ 0.0769 & 0.3846 & 0.5385 & 0 \end{pmatrix}. \tag{19}$$

We obtain the following:

TABLE 3: Statistical table of the SARIMA model's predicted value and residual sequence (in million tons).

Time	ACTUAL	SARIMA F	Error
2010/6	35.5380	36.3685	-0.8305
2010/7	36.3230	36.6030	-0.2800
2010/8	36.2110	36.1051	0.1059
2010/9	32.5760	36.0014	-3.4254
2010/10	34.0840	32.7257	1.3583
2010/11	32.8800	35.9553	-3.0753
2010/12	33.4990	34.0188	-0.5198
2011/1	37.1090	32.5815	4.5275
2011/2	34.7510	36.8717	-2.1207
2011/3	39.1160	36.4210	2.6950
2011/4	28.8430	35.7270	-6.8840
2011/5	39.0050	34.0413	4.9637
2011/6	37.2460	38.0273	-0.7813
2011/7	38.1220	38.9142	-0.7922
2011/8	36.9390	36.7624	0.1766
2011/9	34.7450	35.8810	-1.1360
2011/10	37.9690	35.2646	2.7044
2011/11	36.8160	38.5455	-1.7295
2011/12	38.3310	37.8582	0.4728
2012/1	38.1740	39.5214	-1.3474
2012/2	34.8760	37.6725	-2.7965
2012/3	36.8430	37.0296	-0.1866
2012/4	31.4010	31.2652	0.1358
2012/5	39.1010	37.1483	1.9527
2012/6	34.6950	38.1825	-3.4875
2012/7	32.7310	35.4198	-2.6888
2012/8	31.2900	32.1657	-0.8757
2012/9	34.9530	29.1189	5.8341
2012/10	34.2850	35.6686	-1.3836
2012/11	38.1570	34.3911	3.7659
2012/12	39.3840	38.9422	0.4418
2013/1	38.8740	41.9307	-3.0567
2013/2	34.8820	36.7719	-1.8899
2013/3	38.1690	37.8120	0.3570
2013/4	33.1170	32.6547	0.4623
2013/5	37.7950	39.3172	-1.5222
2013/6	37.0570	35.7993	1.2577
2013/7	36.6493	37.0051	-0.3558
2013/8	37.7550	36.3293	1.4257
2013/9	38.0900	36.7567	1.3333
2013/10	35.3830	39.4495	-4.0665
2013/11	38.2820	36.5611	1.7209
2013/12	39.1930	39.4069	-0.2139
2014/1	39.7410	40.0323	-0.2913
2014/2	34.0330	36.6904	-2.6574
2014/3	39.3990	37.0087	2.3903
2014/4	32.1220	33.8711	-1.7491
2014/5	40.6070	38.3192	2.2878
2014/6	38.9360	38.3357	0.6003
2014/7	38.9250	39.6389	-0.7139
2014/8	38.1540	38.4263	-0.2723
2014/9	37.6650	38.3145	-0.6495
2014/10	34.5620	37.6351	-3.0731
2014/11	37.5350	36.0587	1.4763
2014/12	38.5080	38.1776	0.3304
2015/1	38.1860	39.0928	-0.9068

TABLE 4: Statistics of the one-step state transfer frequency.

	$y_i^1 \rightarrow j$				$\sum y_i^1 \rightarrow j$
	$j=1$	$j=2$	$j=3$	$j=4$	
1	0	0	3	2	5
2	1	11	18	8	38
3	3	22	27	3	55
4	1	5	7	0	13

$$\begin{aligned}
 z_1 &= -5.2945, \\
 z_2 &= -2.115, \\
 z_3 &= 1.065, \\
 z_4 &= 4.245,
 \end{aligned}
 \tag{20}$$

from the formula $z_i = (U_{ij} + L_{ij})/2$. The prediction value of the SARIMA-Markov model is calculated in accordance with equations (13) and (14), and the prediction capability of the model is determined to be excellent by the analysis model's prediction level evaluation indicator, i.e., MAPE = 3.8009%. The fitting curves of the actual (ACTUAL), the predicted value of the SARIMA model (SARIMA F), and the predicted value of the SARIMA-Markov model (SARIMA-Markov F) are presented in Figure 8. As shown in the figure, the fitting effect of the SARIMA-Markov model is better than that of the SARIMA model (in million tons).

4.3. *Model Validation.* To verify the prediction accuracy of the established model, the SARIMA-Markov model was used to predict the monthly coal traffic volume of Daqin Railway in October and November 2019. The predicted values for October and November 2019 were 32.5864 and 38.7377 million tons, respectively, when the SARIMA(4, 1, 2)(1, 1, 1)¹² model was used. From Table 3, the state probability transfer vector in September 2019 is E_3 , and the initial state transfer vector is $x_0 = [0.0545 \ 0.4 \ 0.4909 \ 0.0545]$.

In accordance with the initial state probability transfer vector 1 and the state probability transfer matrix 2, the state probability transfer vectors of October and November 2019 can be calculated as follows:

$$\begin{aligned}
 x_1 &= x_0 * p_{(0)} = [0.0415 \ 0.3319 \ 0.4925 \ 0.1328], \\
 x_2 &= x_1 * p_{(0)} = [0.0458 \ 0.3432 \ 0.4954 \ 0.1133].
 \end{aligned}
 \tag{21}$$

Then, from equations (13) and (14), the predicted coal transport volume in October 2019 is 34.1619 million tons and the predicted coal transport volume in November 2019 is 32.8694 million tons. The results are provided in Table 5.

In the current study, mean absolute error (MAE) and MAPE were used to evaluate the fitting effect of the model, as indicated in Table 6. Both indexes were smaller for the optimized model than those before optimization. The fitting

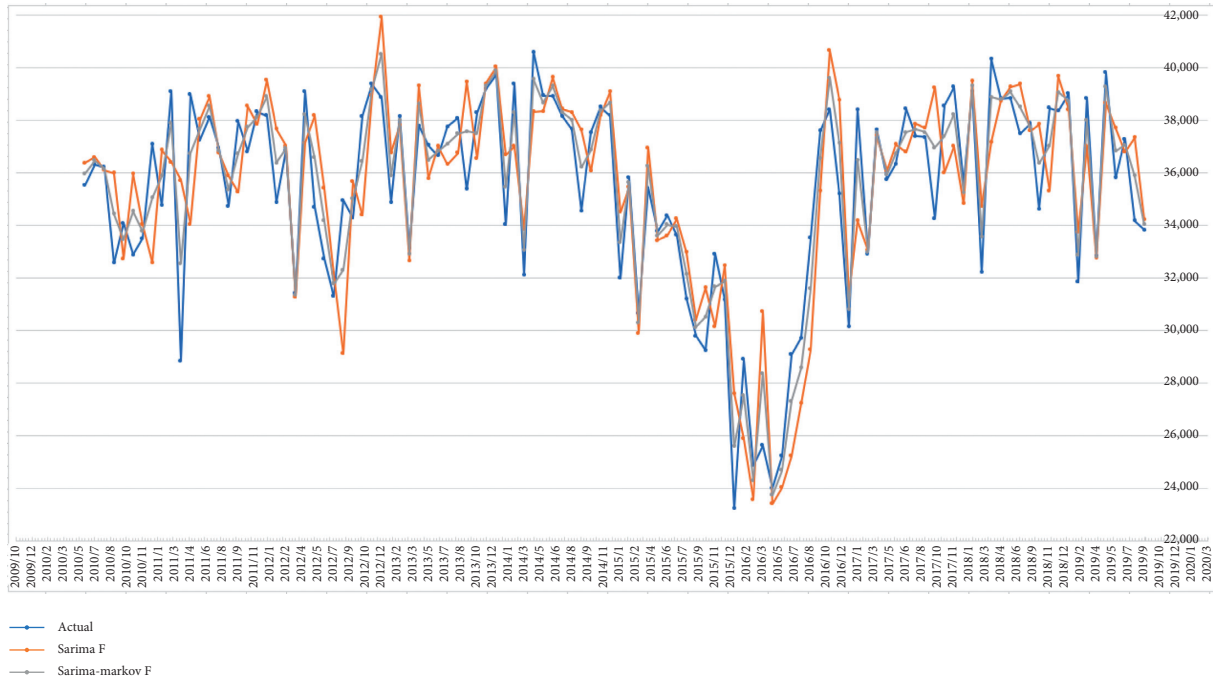


FIGURE 8: SARIMA-Markov model fitting diagram.

TABLE 5: Statistics of the prediction results (in million tons).

Time	Actual	SARIMA F	SARIMA-markov F
2019/10	36.4439	32.8313	34.1619
2019/11	35.4127	38.6018	32.8694

TABLE 6: Model test list (in million tons).

Forecast model	Model fitting		Model forecasting	
	MAE	MAPE	MAE	MAPE (%)
SARIMA	5.7611	5.1439	3.4003	9.4576
SARIMA-markov	4.2570	3.8009	2.4094	6.7126

TABLE 7: Statistics of the forecast results (in million tons).

Time	Company F	SARIMA F	SARIMA MARKOV F
2019/12 (E)	39.4429	40.5264	38.7993
2020/01 (E)	41.5839	39.4894	40.5889
2020/02 (E)	34.7971	36.6173	35.5534
2020/03 (E)	41.0039	40.6742	40.7300
2020/04 (E)	34.2827	34.2780	34.3346
2020/05 (E)	43.4109	41.4161	41.4787
2020/06 (E)	38.8871	40.6402	39.5333
2020/07 (E)	40.3224	39.7717	40.7116
2020/08 (E)	34.8113	39.5051	38.7080
2020/09 (E)	34.5971	34.6678	34.4061

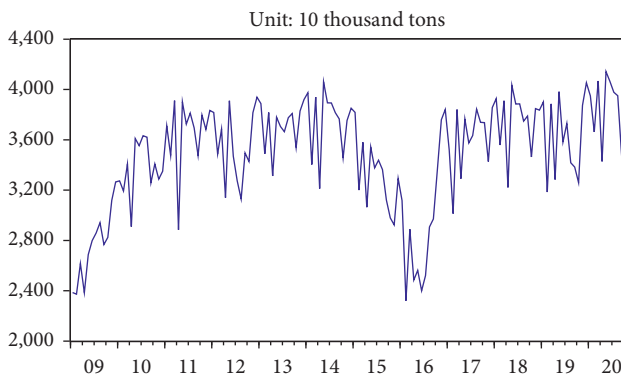


FIGURE 9: Prediction result of the SARIMA (4, 1, 2)(1, 1, 1)¹² model outside the sample period.

accuracy of the SARIMA-Markov model reached 96.1991%, which was higher than the prediction accuracy of the SARIMA model (94.8561%). The prediction accuracy of the

SARIMA-Markov model reached 93.2874%, which was higher than the prediction accuracy of the SARIMA model (90.5424%). Thus, the prediction accuracy of the SARIMA-Markov model is high and meets the requirements.

4.4. Model Prediction. The model was built on the basis of the monthly data of the coal traffic volume of Daqin Railway from January 2009 to September 2019. The predicted sequence x_f and the actual sequence x were drawn in the same diagram for comparison, as shown in Figure 9. In accordance with the state probability transition matrix P and the initial state transition vector x_0 , the state transition vector from December 2019 to September 2020 was determined. Simultaneously, the predicted values of SARIMA (4, 1, 2)(1, 1, 1)¹² from December 2019 to September 2020 were obtained.

The coal traffic volume of Daqin Railway from December 2019 to September 2020 was predicted using the SARIMA-Markov model. The results are provided in Table 7.

5. Conclusion

The SARIMA-Markov model was applied to the monthly coal traffic forecast of Daqin Railway. The SARIMA model comprehensively considered the influence of the seasonal correlation of coal traffic volume on Daqin Railway. Meanwhile, the Markov model used a residual sequence, state partition, and state transition matrix to modify the influence of the sample data on the predicted value. Compared with the simple use of the SARIMA model alone, the combination of the two models can ensure higher prediction accuracy, verifying the scientificity and feasibility of the model and providing a new method for the coal volume prediction of Daqin Railway. Moreover, the coal traffic volume of Daqin Railway from December 2019 to September 2020 was predicted via trend extrapolation, and the forecast results were analyzed as follows.

- (1) China's economy is less dependent on coal. As China's economy moves toward high-quality development, the growth rate of the country's gross domestic product will remain at approximately 6% in the next few years. Meanwhile, the Daqin Railway, as a strategic artery of China's "coal transport from the west to the east," will experience a growth rate of approximately 5%. This conclusion indicates that China's extensive economic development model based on coal energy consumption is beginning to weaken, and the economic form is developing toward the low-carbon, efficient, and green direction. China's coal dependence is decreasing due to the country's effort to adjust its industrial structure, encourage innovations, and increase the intensity of scientific and technological research and development.
- (2) Initial results have been achieved in optimizing the energy structure. The growth rate of the coal transport volume of Daqin Railway will gradually slow down under the influences of energy structure adjustment, industrial structure adjustment, non-fossil energy development, and other factors. This scenario shows a steady decline in the proportion of coal in energy consumption. The rapid development of low-cost nuclear, photovoltaic, and wind power will further reduce the demand for coal, which will account for a smaller share of the country's total energy consumption, while clean, renewable energy will obtain a larger share.
- (3) The process of providing heat from clean energy is proceeding in an orderly manner. China is a developing industrial country that uses coal boilers to provide heating during winter, particularly in the north of Qinling and Huai River. Under the background of the energy revolution, China has effectively promoted clean heating in the north and has replaced coal with gas and electricity for providing heat in an orderly manner, reducing the consumption of coal. In accordance with the local conditions, we expand the variety of clean heating methods to

ensure the balanced development of clean heating. In our future research, the SARIMA model will be combined with the random forest model to improve the accuracy of the model prediction.

Data Availability

Daqin Railway belongs to China Daqin Railway Co., Ltd., which is a listed company. The data required in this paper can be inquired in the monthly or annual financial statements of China Daqin Railway Co., Ltd., which belongs to the public information of the company.

Conflicts of Interest

None of the authors have any conflicts of interest.

References

- [1] Y. Yang, N. Xiong, N. Y. Chong, and X. Défago, "A decentralized and adaptive flocking algorithm for autonomous mobile robots," in *Proceedings of the 3rd International Conference on Grid and Pervasive Computing*, Kunming, China, May 2008.
- [2] X. F. Yu, *Research on the application of machine learning theory in railway freight volume prediction*, Ph.D. thesis, Beijing Jiaotong University, Beijing, China, 2016.
- [3] R. Li, M. R. Dai, and F. Z. Li, "Study on selection methods of key influence factors of railway freight transport volume based on grey correlation," *Railway Freight Transport*, vol. 33, no. 11, pp. 11–14, 2015.
- [4] Z. P. Tang, J. W. Zhu, and J. P. Sun, "Railway freight demand forecasting based on the improved grey MARKOV mod," *Railway Purchasing and Logistics*, vol. 10, no. 3, pp. 57–59, 2015.
- [5] D. Wang and G. J. Mi, "Prediction study of rail-way freight volume based on grey relational analysis and BP neural network," *Journal of Jiangnan University (Natural Science Edition)*, vol. 14, no. 1, pp. 80–84, 2015.
- [6] W. T. Zhu, "Forecasting railway freight volume based on improved BP neural network model," *Journal of Shijiazhuang Tiedao University (Natural Science Edition)*, vol. 27, no. 2, pp. 79–82, 2014.
- [7] Q. Zhang, C. Zhou, N. Xiong, Y. Qin, X. Li, and S. Huang, "Multimodel based incident prediction and risk assessment in dynamic cybersecurity protection for industrial control systems," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 46, no. 10, pp. 1429–1444, 2015.
- [8] Y. Yan and Z. K. Wu, "A study on a freight volume forecast model at harbin railway hub based on grey-linear regression model," *Railway Freight Transport*, vol. 36, no. 11, pp. 1–5, 2018.
- [9] M. T. Liu and J. L. Yu, "Forecast analysis of China railway freight volume based on SARIMA model," *Journal of Guizhou Normal College*, vol. 31, no. 12, pp. 43–47, 2015.
- [10] X. Zhang, *Research of railway freight volume forecasting based on data mining*, Ph.D. thesis, Southwest Jiaotong University, Chengdu, China, 2016.
- [11] N. Q. Zhao, "Forecast of railway freight volume based on ARIMA model," *Think Tank Era*, vol. 22, p. 187, 2019.
- [12] K. Huang, Q. Zhang, C. Zhou, N. Xiong, and Y. Qin, "An efficient intrusion detection approach for visual sensor networks based on traffic pattern learning," *IEEE Transactions on*

- Systems, Man, and Cybernetics: Systems*, vol. 47, no. 10, pp. 2704–2713, 2017.
- [13] S. Q. Yuan, B. Xue-ying, and W. Qi-cai, “Research on railway freight volume prediction based on Markov verhulst model,” *Railway Standard Design*, vol. 60, no. 10, pp. 27–30, 2016.
- [14] C. Zhang and X.-F. Zhou, “Prediction of railway freight volumes based on gray forecast-Markov chain-qualitative analysis,” *Journal of the China Railway Society*, vol. 29, no. 5, pp. 15–21, 2007.
- [15] M. Milenković, L. Švadlenka, and V. Melichar, “SARIMA modelling approach for railway passenger flow forecasting,” *Transport*, vol. 33, no. 5, pp. 1113–1120, 2018.
- [16] X. Tang and G. Deng, “Prediction of civil aviation passenger transportation based on ARIMA model,” *Open Journal of Statistics*, vol. 6, no. 5, pp. 824–834, 2016.
- [17] A. Shahzad, M. Lee, Y.-K. Lee et al., “Real time MODBUS transmissions and cryptography security designs and enhancements of protocol sensitive information,” *Symmetry*, vol. 7, no. 3, pp. 1176–1210, 2015.
- [18] W. Wu, N. Xiong, and C. Wu, “Improved clustering algorithm based on energy consumption in wireless sensor networks,” *IET Networks*, vol. 6, no. 3, pp. 47–53, 2017.
- [19] W. Zhang, D. Chen, H. Si, and N. N. Xiong, “RTDCM: a coding preemption collection system for key data prioritization with hierarchical probability exchange mechanism in mobile computing,” *IEEE Access*, vol. 8, pp. 4629–4639, 2020.
- [20] C. Chen, N. N. Xiong, X. Guo, and J. Ren, “The system identification and prediction of the social earthquakes burst in human society,” *IEEE Access*, vol. 8, pp. 103848–103859, 2020.
- [21] M. Wu, L. Zhong, L. Tan, and N. Xiong, “The sequential fusion estimation algorithms based on gauss-Newton method over multi-agent networked systems,” *IEEE Access*, vol. 8, pp. 114315–114329, 2020.
- [22] B. Yin, X. Wei, J. Wang, N. Xiong, and K. Gu, “An industrial dynamic skyline based similarity joins for multidimensional big data applications,” *IEEE Transactions on Industrial Informatics*, vol. 16, no. 4, pp. 2520–2532, 2020.
- [23] W. Guo, Y. Shi, S. Wang, and N. N. Xiong, “An unsupervised embedding learning feature representation scheme for network big data analysis,” *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 1, pp. 115–126, 2020.
- [24] Y. Zhang, J. Chen, B. Liu et al., “COVID-19 public opinion and emotion monitoring system based on time series thermal new word mining,” *Computers, Materials and Continua*, vol. 64, no. 3, p. 11458, 2020.
- [25] A. Fu, X. Zhang, N. Xiong, Y. Gao, and H. Wang, “VFL: a verifiable federated learning with privacy-preserving for big data in industrial IoT,” *IEEE Transactions on Industrial Informatics*, p. 13585. In press.
- [26] G. Liu, C. Guo, L. Xie, W. Liu, N. Xiong, and G. Chen, “An intelligent CNN-VAE text representation technology based on text semantics for comprehensive big data,” *CoRR Abs/2008*, In press.
- [27] Y. Ren, W. Liu, T. Wang, X. Li, N. N. Xiong, and A. Liu, “A collaboration platform for effective task and data reporter selection in crowdsourcing network,” *IEEE Access*, vol. 7, pp. 19238–19257, 2019.
- [28] H. Cheng, Z. Xie, Y. Shi, and N. Xiong, “Multi-step data prediction in wireless sensor networks based on one-dimensional CNN and bidirectional LSTM,” *IEEE Access*, vol. 7, pp. 117883–117896, 2019.
- [29] N. Xiong, Y. Yang, H. Jing, and Y. He, “On designing QoS for congestion control service using neural network predictive techniques,” in *Proceedings of the 2006 IEEE International Conference on Granular Computing*, pp. 299–304, Atlanta, GA, USA, May 2006.
- [30] Y. Liu, N. Xiong, Y. Li, K. Xu, J. H. Park, and C. Lin, “A secure model for controlling the hubs in P2P wireless network based on trust value,” *Computer Communications*, vol. 33, no. 8, pp. 997–1004, 2010.
- [31] N. N. Xiong, H. Cheng, S. Hussain, and Y. Qu, “Fault-tolerant and ubiquitous computing in sensor networks,” *International Journal of Distributed Sensor Networks*, vol. 9, no. 10, 2013.
- [32] M. Wu, L. Tan, and N. Xiong, “A structure fidelity approach for big data collection in wireless sensor networks,” *Sensors*, vol. 15, no. 1, pp. 248–273, 2014.
- [33] P. yang, N. Xiong, and R. Jingli, “Data security and privacy protection for cloud storage: a survey,” *IEEE Access*, vol. 8, pp. 131723–131740, 2020.