Hindawi

Research Article

A Forecast Model of City Natural Gas Daily Load Based on Data Mining

Liang Chen and Jijun Zhang (ID)

School of Economics and Management, Southwest Petroleum University, Sichuan, Chengdu 610500, China

Correspondence should be addressed to Jijun Zhang; 201911000159@stu.swpu.edu.cn

Data mining technology is more and more widely used in the daily load forecasting of natural gas systems. It is still difficult to carry out high-precision, timely intraday load forecasting and intraday load dynamic characteristics clustering for natural gas systems. Based on data mining technology, this paper proposes a stable intraday load forecasting method for the natural gas flow state-space model. The load sensitivity under the current operating conditions of the system is obtained by calculation; the sample space of the state space is established through data processing; the partitions under different clustering radii are calculated; and the best intraday load flow is obtained through the state space effectiveness evaluation method. The experimental results show that the model load forecasting accuracy and relative error reached 98.5% and 0.026, respectively, which solved the problem of processing the long-term accumulated historical data of gas intra-day load. At the same time, the amount of data calculation was reduced by 33.6%, which effectively promoted the quantification of intraday load influencing factors and qualitative analysis.

1. Introduction

In recent years, the regional natural gas industry has developed rapidly, and there are factors within the natural gas flow that cause the collapse of the natural gas flow, such as a weak natural gas flow grid and an increase in parallel natural gas containers [1]. At present, since most of the on-load voltage regulator taps have not been put into automatic operation and the natural gas department has adopted the last measure of intraday load prematurely, the problem of natural gas flow stability does not seem to be so prominent. Urban pipeline gas consumption is the daily load of urban pipeline gas consumption. Relying on empirical factors will inevitably reduce the scientific nature of the forecast. The method of applying technical means to extract interesting knowledge and information from a large database can process a large amount of historical data accumulated for a long time in the gas daily load and dig out the important influencing factors of the city gas daily load hidden behind the data. The factors affecting the daily load are analyzed quantitatively and qualitatively, so as to grasp the essential characteristics of the daily load of city gas [2–4]. With the marketization of natural gas, people have higher requirements for natural gas energy quality, and the use of the measure of intra-day load will be restricted. Therefore, the research on the stability of natural gas flow has practical value. Accidents that caused the collapse of the system due to the instability of natural gas flow have occurred many times in some large natural gas flows abroad, causing long-term and large-scale natural gas shutdowns and huge economic losses. The collapse of the natural gas flow of the system is often caused by the instability of the natural gas flow of a certain busbar or a certain area and then spread to the entire system, leading to the collapse of the system. Therefore, how to accurately and quickly determine the weak nodes or weak areas where the natural gas flow of the system is stable has become a concern of the majority of researchers [5–7].

Commonly used natural gas flow stability indicators can be divided into state indicators and margin indicators. Both types of indicators can give a measure of the distance between the current operating point of the system and the point of collapse of natural gas flow. The state index only

uses the current operating state information; the calculation is relatively simple, but there is nonlinearity, including various sensitivity indexes, eigenvalues, singular value indexes, and so on. Starting from a given operating state of the system, the margin index gradually approaches the collapse point of natural gas flow through the increase of daily load or transmission power according to a certain mode and uses the distance from the current operating point of the system to the collapse point of the natural gas flow as the judgment An indicator of the degree of stability of natural gas flow. There are three key factors that determine the margin index: the determination of the breakdown point, the selection of the path from the current operating point to the breakdown point, and the selection of the model [8]. The calculation involves the simulation of the transition process and the determination of critical points and contains a large amount of information. The distance from the operating point of the system to the point of collapse of natural gas flow has a linear relationship with the size of the margin index: the influence of various factors such as constraint conditions, natural gas generator active power distribution, and daily load growth methods during the transition process can be taken into account more conveniently. In order to ensure stable gas supply for residential users and industrial parks when the upstream gas supply indicators are insufficient, how to plan the development direction of the city gas market in the future and formulate an economic and reliable gas supply plan has become a difficult problem for city gas companies. Gas companies need to analyze the influencing factors of gas consumption in their cities to grasp the characteristics of urban gas users' gas consumption and the uneven characteristics of gas consumption so as to optimize and adjust the gas consumption structure. On the other hand, to reasonably do well in emergency gas source scheduling and procurement and to improve the construction of gas storage facilities requires research on short-term daily load forecasting and peak shaving plans for city gas [9–11].

This paper compares the weak nodes and weak regions calculated by the data node system and an actual natural gas flow through the network loss sensitivity, dV/dQ, natural gas flow sensitivity, modal analysis, and state-space methods. At the same time, it has the characteristics of fast speed and good real-time performance compared with the modal analysis. Finally, this paper applies the results obtained by the state-space natural gas flow stability weak node partitioning method to the data node system and a certain actual natural gas flow. Under the condition of natural gas flow collapse, the weak area judged by the operating point before the collapse is implemented urgently to cut the intraday load. We contrast this with cutting the daily load of the entire network step by step without judging weak areas. The results show that the identification of weak areas of natural gas flow can guide operators to make accurate judgments in emergency situations and restore the stable operation of the system at the least possible cost. This verifies the validity and accuracy of the state-space method proposed in this paper and provides a good practical tool for the analysis of weak nodes and weak regions of natural gas flow stability in the natural gas system. The influencing factors of urban gas

consumption and the characteristics of short-term daily load forecasting are complex and changeable, with randomness and uncertainty, and we need to start with their essential laws and their own characteristics. This paper comprehensively considers the limitation, incompleteness, and complexity of the influencing factors of urban pipeline gas historical data and uses data mining methods to process and analyze the data to take a city gas enterprise in the northern region as an example to collect and organize the data. In recent years, the historical data of gas consumption in the city gas pipeline network is used to determine the influencing factors of city gas consumption by using relevant analysis. Through the processing and processing of relevant historical data, the potential connections among the factors influencing the gas consumption of various users of city gas can be unearthed, which will help clarify the cause mechanism of the intraday gas load and provide a basis for the forecast of the intraday gas load of the city.

2. Related Work

The stability of the natural gas system is a research topic that natural gas workers attach great importance to and is especially favored by theoretical researchers. In the past, the most prominent issue of stability was the power angle stability, that is, the stability of the power angle changes that reflect the relative movement of the natural gas generator rotor. Therefore, a large number of natural gas workers mainly focus on power angle stability, and system stability has become synonymous with power angle stability. Due to the development of modern control theory and the introduction of various mathematical methods, especially the widespread application of computers and calculation techniques, the power angle stability has entered a high level of understanding, and various analysis and control methods have become more mature and have been practically applied. So far, the research on the stability of the power angle has achieved recognized and impressive results [12–14].

Dong and Grumbach [15] proposed a spectral clustering method based on matrix perturbation to partition the natural gas flow. Each partition uses the natural gas generator node reactive power change and the daily load node natural gas flow change as indicators to identify the natural gas of the system online. The flow rate is a weak node; then an iterative algorithm based on deviation correction is proposed for the weak node of natural gas flow; the equivalent parameters are tracked in real time; and the indicators of static natural gas flow analysis are obtained. After initial dimensionality reduction of feature attributes using scoring criteria, Huang's et al. [16] improvements were made to the existing principal component analysis methods; the attributes after dimensionality reduction were classified; principal components were extracted; and the various principal components were combined as it. The training input of the support vector machine finally obtains a classifier with stable static natural gas flow. At the same time, an online natural gas flow safety assessment algorithm based on PMU is proposed: the existing data of the natural gas flow database is used to establish a natural gas flow safety

assessment decision tree offline. According to the real-time sampling data of the PMU, the decision tree is dynamically updated to form a dynamic decision tree, which monitors the safety of natural gas flow online and obtains two indicators for static natural gas flow evaluation. Based on the trajectory of the PMU measurement information, Beyca et al. [17] proposed binary data represents a method for transient natural gas flow safety issues. The natural gas flow threshold and the critical natural gas flow excursion duration are used to determine the natural gas flow safety; the primary and secondary fittings are used to convert the critical natural gas flow excursion duration into information expression in the form of natural gas flow. In addition, a method for judging the transient natural gas flow stability with induction natural gas motive is proposed. The natural gas flow drop and admittance change indicators are used as comprehensive criteria for identifying the transient natural gas flow stability and analysis indicators.

Li et al. [18] proposed the use of phase-locked loop PLL technology to track frequency; proposed a synchronization phasor estimation algorithm based on Taylor expansion model; and, using the first-order Taylor model to modify the DFT algorithm, proposed a complete period. Solyali [19] proposed an adaptive sampling method to eliminate the influence of frequency leakage caused by the DFT algorithm at the nonideal fundamental frequency. The methods in the above documents can achieve relatively ideal results and can track the frequency of natural gas flow well. However, after correctly tracking the natural gas flow frequency, no specific plan is given for how to correct the error caused by the DFT transformation, and a related correction plan is proposed. This article derives the amplitude difference formula, uses the uniform approximation method in the functional to correct the expression reasonably, also uses the same method to correct the formula in the amplitude and phase space, and finally gives the quantified modified expression [20–22]. The correction formula is simulated and analyzed, and the approximate correction and the precise correction error are basically within 0.3%, which has a certain engineering application value. After initial dimensionality reduction of feature attributes using scoring criteria, some improvements are made to the existing principal component analysis methods; the attributes after dimensionality reduction are classified; principal components are extracted separately; and the various principal components are combined as a support vector machine training input. This paper uses node natural gas flow and branch loss as attributes, finally obtains a static natural gas flow stable classifier, and conducts simulation analysis on IEEE14 node and IEEE300 node systems. The experimental results show that all three scoring criteria can effectively eliminate the classification. In addition, although the principal component analysis of classification attributes has more principal components than the comprehensive attributes, the relative mass attributes have been greatly reduced, and it can improve accuracy, save memory, and provide new ideas for studying the stability of online natural gas flow [23–25].

3. Daily Load Analysis Based on Data Mining

3.1. Data Mining Hierarchical Topology.
The function of data mining is mainly to use existing data or preprocessed data to build models to identify the types of patterns presented by the attributes of the data set. Some of the patterns are explanatory and are used to explain the correlation between attributes. Other modes are predictive, inferring based on current data to predict the possible values of certain attributes. Generally speaking, data mining is used to identify four main types of patterns: association patterns, prediction patterns, clustering, and sequence patterns.

$$g\{x, t \mid (x, t) \in (1, 1)\} - \sum W^t \times x + b \times \sum W^t \times x - a = 0. \tag{1}$$

Prediction refers to the prediction of future behavior based on data and models. According to different predicted attributes, prediction can be divided into classification prediction (what is predicted is the type of thing) and regression analysis prediction (what is predicted is a specific value). Classification is to analyze the historical data stored in the database through models that describe or identify concepts or categories and predict the categories of other unclassified records and the categories of future events. The classification model is based on the learning and training of the training data set records to identify predictions. Regression analysis prediction is a mathematical statistical analysis of causal data. Prior to this, the causal data should be analyzed to identify attributes that are useless for prediction. These useless attributes should be excluded in advance.

$$\begin{bmatrix} W^t & W^t - 1 & W^t - 2 \end{bmatrix} \times \begin{bmatrix} x \\ x - 1 \\ x - 2 \end{bmatrix} - F(X) \times F(W) = 0. \tag{2}$$

Modal mining analysis technology is to use the system static model to calculate the minimum eigenvalues and eigenvectors specified by the simplified Jacobian matrix. Each eigenvalue is related to the natural gas flow/reactive power change pattern, and its magnitude provides the natural gas flow instability. It analyzes the stability of natural gas flow from the relationship between natural gas flow and reactive power. The data of cluster analysis does not need to consider the known categories. The categories in the cluster are not clear. It divides the events in the data set into multiple parts with some similar characteristics between them and establishes them based on the common characteristics of the things in the data set. The results of clustering will vary according to different state-space algorithms.

3.2. Data Cluster Analysis.
The data cluster analysis results show that the network loss sensitivity is closely related to the trend Jacobian matrix. According to the static stability analysis theory of the natural gas system, when the operating point is close to the collapse point of natural gas flow, the value of the Jacobian matrix determinant tends to zero, and the network loss sensitivity value tends to infinity. This verifies the rationality of the network loss sensitivity index

from another aspect. In actual monitoring, when the network loss sensitivity value exceeds a certain closed value, it can be considered that the system is close to a collapsed state. The numerical calculation of the network loss sensitivity of each node in Table 1 only needs to add a small amount of calculation on the basis of each power flow calculation. Therefore, the network loss sensitivity index has a clear physical concept, fast calculation speed, and good accuracy.

When doing data association analysis, the dimensions of the series should be the same. When the dimensions are different, it is difficult to make effective comparisons. Therefore, the series must be dimensionless. Dimensionless methods commonly used are the initial value method and the average value method. The initial value method refers to dividing all the data by the first data, and then a new dimensionless number sequence is obtained; the average value method is to divide all the data with the average value of the sequence to obtain a number sequence that accounts for the percentage of the average value. Due to a large number of correlation coefficients, the information is too scattered, and it is not convenient to compare. Therefore, it is necessary to concentrate the correlation coefficients at each time into one value. The average value is the method for this kind of information centralized processing, that is, the correlation degree is proposed.

$$\begin{cases} W^t \times x + b - 1 \geq 0, & \text{if,} \quad x < y = 1, \\ W^t \times x + b - 1 \leq 0, & \text{if,} \quad x < y = -1. \end{cases} \tag{3}$$

The sequence pattern is a collection of different sequences. In each sequence, different data are arranged in order, and a specified support threshold is given. The sequence pattern is to find all frequently occurring subsequences. And the frequency of the subsequence in the sequence set cannot be less than the selected support threshold. Each neural unit can receive the input of a large number of other neural units, generate output through the parallel network, and at the same time affect other state-space nodes. The mutual restriction and influence between the networks realize the nonlinear mapping from the input state to the output state space. The overall performance of the network cannot be regarded as the superposition of the performance of all neural units. A network system composed of neural units with thresholds has better performance and improves storage capacity and fault tolerance.

3.3. Intraday Load Analysis Indicators.
A characteristic of the natural gas system's daily load flow collapse is the sudden increase in system network loss. When considering the influence of the changes in active and reactive power at the load node during the day on the system network loss, the network loss sensitivity $dP\text{loss}/dP$ and $dP\text{loss}/dQ$ can be obtained. Regardless of the fact that all the intra-day load nodes are constant impedance intra-day loads, when the system reaches a critical point as the intraday load increases, the sensitivity of the network loss will tend to infinity. It can be seen that the same original data applies the state-space theory, but the prediction results obtained by using different numbers of samples are very different. In particular, the difference between the predicted value obtained by the seven- and four-year modeling by 2019 is nearly twice. It can be seen that although the state-space theory is suitable for predictions with few original data, it still has its own shortcomings and needs to be improved to enhance the accuracy of the prediction in Figure 1.

If the power demand of the load node in the day is increased, the power generated by the natural gas generator cannot meet the load demand in the day no matter how it increases. The power generated by the system will be completely consumed by the transmission line, and the natural gas flow at the load node in the day will drop rapidly. The amount of active and reactive power loss of the line, which includes the information of the transmission capacity of the line and the length of the natural gas distance of the line, is useful information that can measure the stability of the natural gas flow. The greater the transmission power of the line, the greater the distance between natural gas and the greater the power loss. Therefore, the power loss of the line can reflect the participation of the line in the transmission system to a certain extent.

$$\text{for} [i = 1, 2, \ldots, w - 1, w],$$
$$|\min \phi(w)| \sum_{2}^{1} w^t \times \sum w = 0. \tag{4}$$

After the data preprocessing is completed, a suitable modeling method is selected and used on the basis of the data to solve the actual needs. For the task of data mining, there is no recognized best method or algorithm for constructing a model. Various feasible method tools should be considered comprehensively, and the best method or algorithm should be judged through clearly defined experiments and data verification. A good method requires multiple parameter adjustments in the process of building a model, and whether the data preparation phase is sufficient is very important for model building.

3.4. Intraday Load Modal Composition.
For mid-to-long-term intraday load, as time goes by, some disturbance factors in the future will continue to affect the system. The farther the future time is, the larger the gray interval of the predicted value will be, which is a truly meaningful and accurate forecast. The value is the most recent data. Equal-dimensional innovation modeling can first use a series of known data to establish a state-space model to predict a value, then add this predicted value to the known series, and at the same time remove the oldest data so that the formed development sequence is consistent with the original sequence of equal dimensions. After such a gradual prediction, it will be replenished one by one until the predicted age.

$$\sum \frac{\partial(1/2)w^t}{\partial w \times \partial t} - \sum \frac{\partial w}{\partial x \times \partial t} = 0. \tag{5}$$

It is mainly to test the size of indicators related to the prediction error variance and, based on the residuals, to examine the probability of the occurrence of points with

TABLE 1: Data cluster association analysis.

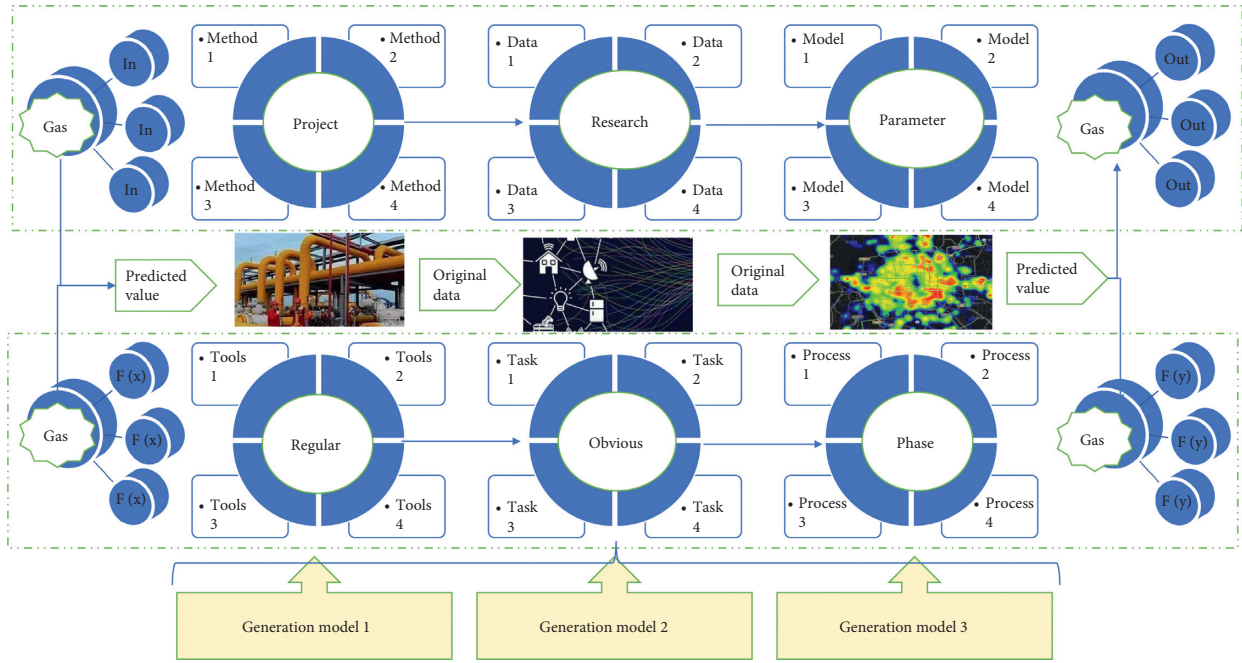| Date record	Gas power flow	Predicted load	Calculation rate (%)	Accuracy error (%)
20191201	529.63	3,906.95	41.60	11.17
20191202	530.25	3,824.26	40.47	10.46
20191203	530.86	3,751.16	38.89	9.66
20191204	531.48	3,665.44	37.45	9.01
20191205	532.10	3,569.00	36.63	8.88
20191206	532.72	3,490.38	36.73	9.61
20191207	533.33	3,474.47	37.76	11.33
20191208	533.95	3,562.61	39.45	13.85

FIGURE 1: Daily load distribution based on data mining.

smaller residuals. A large $s1$ indicates that the original data have a large variance and a large degree of dispersion of the original data. A small $s2$ indicates that the residual variance is small and the residual dispersion is small. A small value of C indicates that although the variance of historical data is large, the difference between the predicted value and the actual value obtained by the model is not too discrete; the larger the probability of small error P, the better, and the larger the value of P; it means that the difference between the residual and the average value of the residual is less than $s1$.

$$\text{if } [f(a,b) > f(a) - f(b)] \longrightarrow \frac{f(W^t \times W + b) - f(W^t \times W - b)}{f(W^t \times W)} \geq 1. \tag{6}$$

It can be seen that the size of each characteristic value determines the vulnerability of the corresponding modal natural gas flow and provides a relative measure close to the instability of natural gas flow. The smaller the eigenvalue, the weaker the corresponding modal natural gas flow. If input = 0, the i-th modal natural gas flow rate will collapse because any change in the modal reactive power will cause an infinite change of the modal natural gas flow rate. If all the eigenvalues of the Jacobian matrix $R\,J$ are positive, the system can be considered as stable natural gas flow. If there is a negative characteristic value, it can be considered that the system is unstable in natural gas flow. The zero eigenvalue of $R\,J$ means that the system is on the boundary of instability, and the smaller eigenvalue in Table 2 determines the degree of instability near the natural gas flow of the system.

The system will select as many influencing factors as possible related to it as indicators for research. The influencing factors involved are generally called indicators. In multivariate statistical analysis, these indicators are also called variables. Each indicator or variable reflects some information of the research target to varying degrees in the actual research process, and there is a certain correlation between the variables, so the information reflected by the variables will overlap to a certain extent. When studying multivariate problems, the selection of more variables not only complicates the problem but also increases the amount of calculation.

Table 2: Description of the degree of stability of the natural gas flow.

	Features text	Stability	Flow ratio	Discharge rate (%)
	Natural gas 1	0.61	38.24	3.98
	Natural gas 2	0.56	37.51	2.97
	Natural gas 3	0.78	36.65	1.96
System accuracy	Natural gas 4	0.85	35.69	0.95
	Natural gas 5	0.95	34.90	−0.06
	Natural gas 6	0.76	34.74	−1.07
	Natural gas 7	0.83	35.63	−2.08

4. Construction of Urban Natural Gas Daily Load Forecasting Model Based on Data Mining

4.1. Data Mining Layout Algorithm. Data mining and state-space output nodes mainly receive data and give the final processing results. Before designing the network, the data for training the network must be sorted out. These data are generally difficult to obtain directly and often require signal processing and feature extraction techniques. The output actually refers to the expected output provided by the network training. The selection is relatively easy and has little effect on the accuracy of the network and the training time. The design of input and state-space output nodes is to reasonably select the number of state-space nodes of input and state-space output nodes, which should be selected and handled according to actual problems.

$$\left\langle \begin{array}{l} \dfrac{\partial L}{\partial W} - \sum \alpha yx(x-1) = 0, \\[2mm] \dfrac{\partial L}{\partial W} + \sum \alpha yx(x+1) = 0. \end{array} \right. \tag{7}$$

This article believes that it is necessary to find out its global mode for offline analysis because the system does not necessarily lose stability in the mode corresponding to the smallest eigenvalue in this state, and as the system deteriorates further, it will lose its global mode. When analyzing online, you do not need to calculate all the characteristic values of RJ but just calculate the minimum characteristic value. Modal analysis technology is to use the eigenvalues of the power flow Jacobian matrix to find a specified number of minimum eigenvalues and their eigenvectors, and each eigenvalue is related to the natural gas flow/reactive power change mode. In order to measure the instability of the natural gas flow, the smallest eigenvalue of the Jacobian matrix can be used as a margin index for the stability of the natural gas flow.

The scientific selection of training sample data and the rationality of data representation in Figure 2 have an extremely important impact on network design. The preparation of sample data is the basis of network design and

training. The design of the input data and the amount of training data have an obvious relationship with the network training time. Generally speaking, the training data of the network should consider the following issues: the training data set must include all patterns; the input data should be as uncorrelated or have little correlation as possible; and the input quantity must be selected to have a large impact on the output quantity.

The probability of simultaneous occurrence of event A and event B is recorded as $P(A \cap B)$, the level of support is proportional to the usefulness of the rule. In each type of data, appropriate consideration should be given to the influence of random noise, the standardization of network input and output data, the consistency of parameter variation ranges, and the normality of sample distribution.

The eigenvectors in Table 3 are used to describe the modes, which provide information about the degree of participation of network elements and natural gas generators in each mode and the mechanism of natural gas flow instability. The largest element of the right eigenvector related to the minimum eigenvalue corresponds to the most sensitive key node of the natural gas flow in the system, and the largest element of the left eigenvector related to the smallest eigenvalue corresponds to the key natural gas that is most sensitive to changes in reactive power in the system machine. Under the premise that A appears, the level of confidence is proportional to the certainty of the rule.

$$\left\langle \begin{array}{c} \alpha yx(x-1) \\ \alpha yx(x+z) \end{array} \middle| \begin{array}{c} \cos x \\ \sin x \end{array} \right\rangle - \left\langle \begin{array}{c} x \\ x-1 \end{array} \middle| \begin{array}{c} 1 \\ -1 \end{array} \right\rangle = 0. \tag{8}$$

The network structure is relatively simple; the signal is directly transformed from input to output; and the information processing is also a simple nonlinear function multiple times. Input nodes do not have computing functions. State-space nodes with computing functions are called computing units. Each computing unit can have any number of inputs but only one output. The input node layer can be called the zeroth layer, and each node layer of the computing unit is called layer 1 to layer N accordingly, which can form an N-layer forward network. The first node layer and the output node layer are called the visible layer. The other intermediate layers are hidden layers, and the nodes of the hidden layer are called hidden nodes.

4.2. Sensitivity of Natural Gas's Daily Load Network Loss. When the natural gas system reaches the critical point with the increase of the daily load, the sensitivity of the network loss will tend to infinity. If the power demand of the load node in the day is increased, the power generated by the natural gas generator will not meet the load demand no matter how it increases. The power generated by the system will be completely consumed by the transmission line, and the natural gas flow at the load node in the day will drop rapidly. The sensitivity of the network loss can be used as an indicator of the collapse of the natural gas flow in the natural
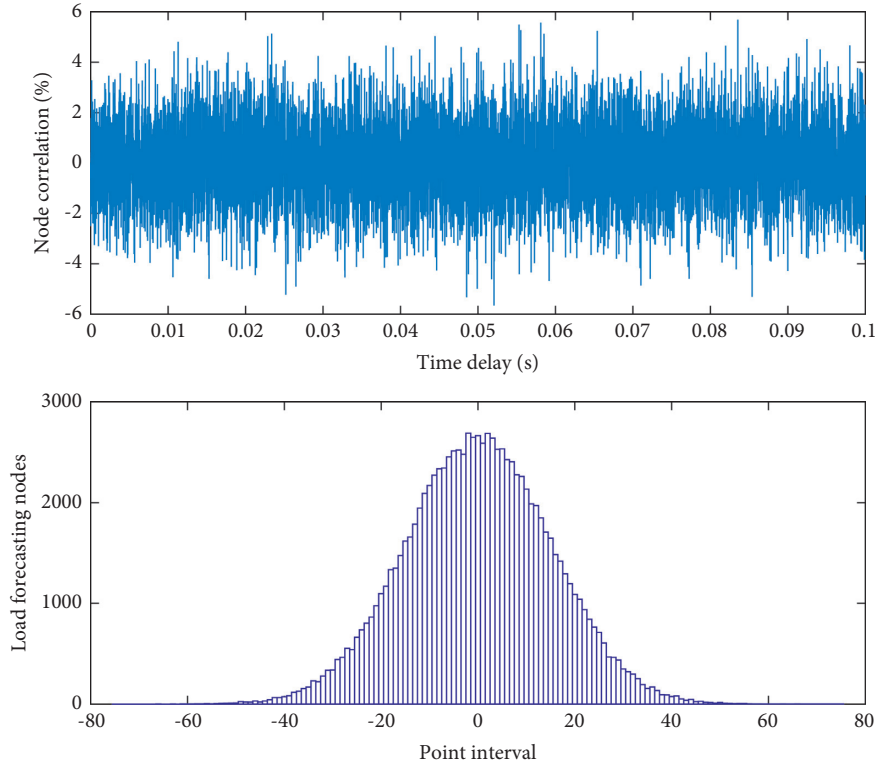
FIGURE 2: Data distribution of nodes in the state space.

TABLE 3: Standardization algorithm of input and output data.

Steps	Algorithm content	Data codes
1	Appropriate consideration in test	Public class test02 {
2	Training sample $w \times x$ data	Public static main(String[]) {
3	A large impact on the $\lim_{x \to \infty} (\sin x)$	Int count = 0;
4	In each type of $[(w \times x) + b]$ data	For (int $i = 101$; $i < 200$; $i+ = 2$) {
5	An obvious relationship $\alpha\beta$	Boolean flag = true;
6	The normality of sample $\sqrt{x^2 + y^2}$	For (int $j = 2$; $j \leq$ Math.sqrt(i); j++)
7	The scientific $x > y$ selection	If ($i\%$j = = 0) {
8	The training data (α, β) of the network	Flag = false;
9	The rationality of data $C > 0$	Public void main(String[] args)
10	The right eigenvector $\sqrt{x^2 + x'^2}$	Nt a,b,c;
11	The mechanism of natural $\text{sgn}(\cos x)$	For(int $i = 101$; $i < 1000$; i++) {
12	The smallest eigenvalue $\sqrt{w^2}$	If(flag = = true) {
13	The critical point with the $3/\sqrt{w^2}$	Count++;
14	The largest element of the $-\cos M$	System.out.println(i);
15	Participation $\forall (a, b)$ of network	If ($a*a*a + b*b*b + c*c*c = = i$)
16	Parameter variation ranges dx	System.out.println (i);
17	Which provide $\sin M$ information	System.out.println (count);
18	The most sensitive key node $(x + 1)^2$	Scanner input = Scanner (System.in)

gas system. Whether the pattern corresponding to the rule frequently uses the following two metrics has to be determined.

The intraday load density method is to determine the daily load per unit building area or per unit area of planned land by analogy with other areas according to the nature of the land in the planned land, and then add the intraday load of the planned area to the prediction method in Table 4; this

forecasting method can be better integrated with urban planning and development.

The intraday load density forecasting method is to predict the occurrence of natural gas from the average natural gas consumption of the regional land area (or building area). Generally, first, we predict the land area (or building area) and the density of natural gas per unit area in a certain period in the future and then multiply the area by

TABLE 4: Daily load forecast data of the planned area.

Node index	Urban planning 1	Urban planning 2	Urban planning 3
101	2.44	2.58	2.72
102	1.87	1.97	2.07
103	1.3	1.36	1.42
104	0.73	0.75	0.77
105	0.16	0.14	0.12
106	−0.41	−0.47	−0.53

the area to obtain the predicted value of natural gas consumption. The natural gas planning goal divides the area to be predicted into multiple functional areas, then uses the daily load density method to predict each functional area, and finally adds up to obtain the total natural gas consumption forecast value.

$$\iint G(w,b,a)\mathrm{d}a\mathrm{d}b - \frac{1}{2}\iint_t W^t \times W \mathrm{d}t\mathrm{d}w = 0. \qquad (9)$$

The method is mainly to study the internal relationship between multiple variables, and then on the basis of exploring the basic data structure in the data to be tested, a few hypothetical variables are proposed to represent the basic data composition, and some guidelines are used to make these. A few hypothetical variables can reflect the main information of the original numerous observed variables so that the interrelationship between these variables to be measured can be explained well. The points of the fitted value and the true value are distributed around the straight line with an intercept of zero and a slope of 45 degrees. From the situation shown in Figure 3, the final fitting effect of the stepwise regression method is average (the residual sum of squares is 13.348).

$$\begin{vmatrix} W^t W & f(w,t) \\ f(w-1,t-1) & W^t W \end{vmatrix} \longrightarrow \begin{vmatrix} W & t \\ t-1 & W^T \end{vmatrix}, for\{(w,t) \longrightarrow (0,0)\}. \qquad (10)$$

The result obtained when the number of hidden state-space nodes is 10, the daily load value in 2020 is 363.13 million cubic meters, which is less than 364.34 million cubic meters in 2019. This is not in line with the current development trend of natural gas. The reason may be analyzed. Because of the "overfitting" problem in the state space, that is, the error for the training set samples can be small, but for the new sample data outside the training set, the error will be very large. Therefore, the predicted result after 2,473 iterations of the state space with 12 nodes in the trailingdm hidden layer state space is finally used in the article. Using the state space method, the predicted value in 2020 was 522.41 million cubic meters; the residual error between the two was 5.41 million cubic meters; and the relative error was 4%. It can be seen that the state-space combination forecasting method in Figure 4 has achieved good results when applied to mid- and long-term intra-day load forecasting of natural gas and has important practical significance.

The basic idea is that the learning process consists of two processes: the forward propagation of the signal and the backward propagation of the error. The input sample data is passed in from the ingress node of the state space, processed by the hidden layer, and the result is transmitted to the state space out node. If the result of the state space out of the node does not match the expected output, the error value of the output result and the expected result is backpropagated. The backpropagation of the error is to pass the error in some form to the state-space ingress node through the hidden layer. Each error signal is used as the basis for correcting the weight of each unit. The continuous adjustment of the weight is also the process of network learning and training. This process continues until the network output error is reduced to an acceptable level.

4.3. Intraday Load Identification in State Space. A three-layer state space can realize any complex nonlinear mapping relationship from input to output, so only one hidden layer is selected. The number of state-space nodes of the hidden layer is obtained according to the formula, where m is the number of output state-space nodes, n is the number of input units, and mouth is a constant between [1, 10]. For the different scales of mid- and long-term intra-day load forecasting data and different training functions, there is no theory to guide which training function is more effective, so this article comprehensively selects six training functions: the basic gradient descent method trangd, the steepest descent algorithm with variable learning rate trangda, gradient descent method trangdm with momentum term, adaptive learning algorithm trangdx with momentum term, and normalized conjugate gradient method transcg.

The RA (corrective measures) module of VSAT can analyze how to most effectively prevent, control, and correct the system's natural gas flow stability problems. If the current operating point is unsafe, the user would like to know what measures can be taken to restore the stable natural gas flow at the operating point. The correction module can determine the most effective measure among all possible control measures specified by the user, such as natural gas generator and SVC natural gas flow control, natural gas container, and natural gas reactor switching, on-load tapping transformer tap adjustment.

The VSAT corrective measures module first tries the best preventive control measures. These control actions will be adopted before the accident. If the designated preventive control measures are not enough to make the operating point safe for all accidents, corrective measures will determine corrective control measures for each critical accident. These measures (turning on and off the natural gas container and reducing the daily load) will be taken after the accident to ensure system security.
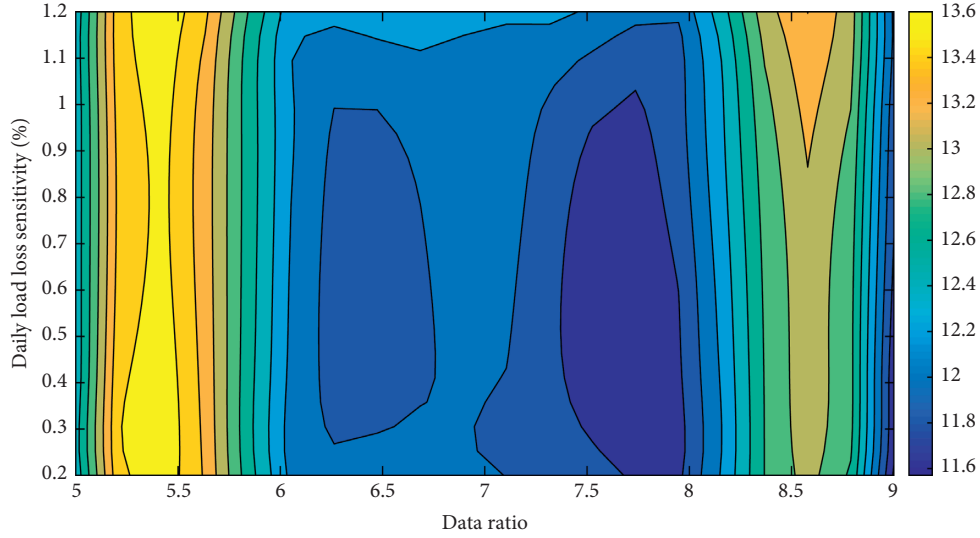
FIGURE 3: Two-dimensional distribution of load loss sensitivity within a day.
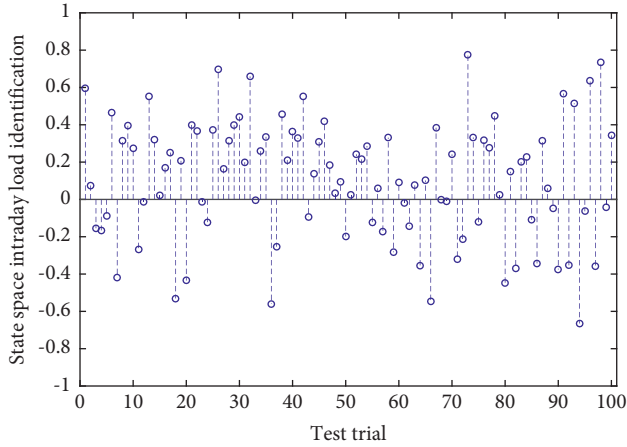

FIGURE 4: Distribution of intra-day load identification in state space.

$$\frac{\sum\left[\partial f\left(W^t \times x + b\right) - 1\right]}{\sum W^t \times x} + \frac{\sum\left[\partial f\left(W^t \times x\right)\right]}{\sum W^t \times x} = 1. \tag{11}$$

The index conversion algorithm for per capita natural gas volume refers to selecting a region at home and abroad that is similar to the region in terms of humanistic and geographical conditions, economic development status, and natural gas structure as the comparison object. By analyzing and comparing the past and present per capita natural gas volume indicators of the two regions, the predicted value of natural gas per capita in this area, combined with population analysis, obtains the predicted value of total natural gas used.

It should be pointed out that because the regional economy is in a period of rapid development and related policies are also in the process of exploration and improvement, there are large changes, many uncertain factors, and weak regularity, so the time series trend model and related analysis models are fitted. The results of historical data prediction are not satisfactory.

5. Application and Analysis of Urban Natural Gas Daily Load Forecasting Model Based on Data Mining

5.1. Data Mining Prediction Processing. To ensure that there are certain differences in the date, weather, temperature, and humidity in the data, the relevant statistical data of the gas companies in City H from October 6 to October 15, 2019 and December 1 to December 10, 2019 are selected. We set the inbound pressure to X1, the outbound pressure to X2, the highest temperature X3, the lowest temperature X4, PM2.5 to X5, date type X6, weather conditions X7, the previous day's gas intra-day load is Y1, and the previous two days' historical intraday load Y2, and the statistical data will be adjusted to the table. Before the intraday load forecasting, the historical data of the city's intraday load and natural gas volume were collected, and its development history, especially in recent years, was studied. An in-depth analysis of the structure of natural gas used was carried out in order to grasp the natural gas. The overall trend and future focus of the development of the quantity and intraday load to guide the work of intraday load forecasting.

$$\lim_{w \longrightarrow \infty} \frac{f(1,1), f(2,2), \ldots, f(w-1, t-1), f(w, t)}{\sum f(w, t)} - \lim_{x \longrightarrow \infty} f(w, w') = 0. \tag{12}$$

The relative error of the calculation is in the range between 6.55% and −2.82% is smaller than the seven-year data modeling. The absolute average relative error between the five-year modeling forecast and the actual value is 3.69%, and the absolute average residual is 9.896 million cubic meters. The posterior difference ratio C = 0.1350, and small error probability P = 1, the two indicators are both good. It can be seen that after the improvement of the equal-dimensional and new-information method, the predicted values of the five- and six-year modeling are relatively close, while the predicted values of the seven-year modeling are quite different from the previous two.

Association analysis includes mining frequent patterns and forming association rules. The reason for this situation is that the equal-dimensional and new-information method is largely affected by new information. The seven-year modeling has a relatively slow update rate due to the large number of modeling data, which results in a large difference between the predicted values obtained from the five- and six-year modeling.

Through software analysis and calculation of the values in Figure 5, it can be seen that KMO = 0.753 and sig = 0. These two indicators meet the constraints of the principal component analysis method, indicating that there is a strong correlation between the original variables, and the state space can be carried out. The absolute value of the value in the component matrix represents the closeness of the relationship between the common factor and the original variable. The larger the absolute value, the closer the relationship between the two.

The component matrix can be rotated to get the rotating component matrix. From the factor loading data in the table, it can be known that the common factor 1 is higher in the
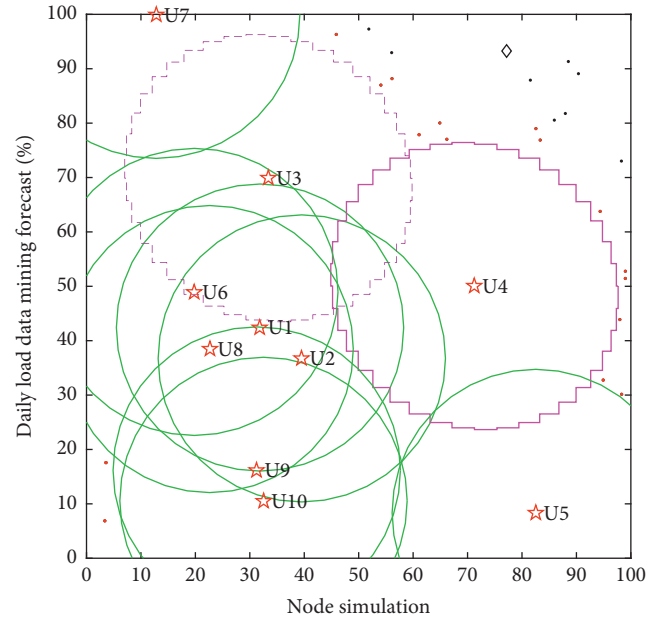

FIGURE 5: Optimization of data mining prediction processing results.

daily load in the previous day, the load in the previous two days, the load in the previous three days, the highest temperature, and the lowest temperature. The common factor 2 has a higher load on the two influencing factors of date type and PM2.5, and the common factor 1 is positively correlated with the historical intra-day load, and negatively correlated with the maximum temperature and the minimum temperature, and the common factor 2 is related to it.

$$\max\{\partial x\,(x-1),\partial\,(x+1)\,(x-0),\partial\,(x+1)\,(x-1)\} \overset{f\,(x),f\,(x')}{\rightleftarrows} \max\{\partial x,\partial\,(x-1)',\partial\,(x+1)'\}. \tag{13}$$

The relative error of the calculation is between 6.78% and 2.54%; the absolute average relative error between the four-year modeling forecast value and the actual value is 2.70%; and the absolute average residual error is 8.04 million cubic meters. The posterior error test result is that the posterior difference ratio C = 0.1328, the probability of small error P = I. Both the two indicators are good. The overall analysis shows that the historical intraday load has a relatively large impact on the gas intraday load forecast. Other variables cannot control the influence of the historical intraday load on the future intraday load. The correlation coefficient between the maximum and minimum temperature and the gas intra-day load is negative. Weather conditions and date types, as well as PM2.5 and gas daily load, have a certain correlation, but the impact intensity is not very large.

5.2. Realization of Simulation of Urban Natural Gas Daily Load Forecast. The original power flow data of the urban natural gas system is as in the text. A simple analysis of the original power flow data shows that the 220 node is at the end of the natural gas flow, and it is a single-line direct-feed natural gas supply. There are only 219 natural gas plants nearby, but its natural gas generator capacity is small. Therefore, the increase in the daily load of the 220 nodes will be transmitted by other natural gas plants over a long distance, so the 220 and 219 areas in the region should be weak areas. The 214, 215, 218, and 216 areas have fewer natural gas sources, and some nodes have heavy daily loads, which will seriously affect the stability of the system's natural gas flow if they are in line with the rapid growth; the 204, 208, 201, and 209 nodes have relatively excessive daily loads. In addition, its powerful reactive natural gas source output is close to its limit.

$$\prod_{x,y} \beta\{y[(w \times x) + b]\} \times \prod_{x,y} \alpha\{y[(w \times x) + b]\} = \begin{cases} \alpha\beta(w \times x), & x > y \\ 1, & x \le y \end{cases}. \tag{14}$$

The maximum daily load utilization hours of the city's natural gas flow are always at a low level of about 4,000 hours, which is mainly determined by the economic structure. The city's tertiary industry occupies a large proportion of the entire national economy. The daily load characteristics of the tertiary industry and natural gas used by residents determine the maximum daily load utilization hours. According to the city's future economic development positioning, it will continue to strengthen the leading role of the industrial economy and the implementation of measures such as strengthening demand-side management. It is expected that the city's maximum daily load utilization hours will remain at the level of about 4,000 hours for a long period of time in the future.

According to the city's national economic and social development, the city's planning characteristics, and the collection of basic data in Figure 6, the overall idea of intraday load forecasting is determined. In the long-term target year, the intraday load density index method is used to predict the load distribution, and on this basis, the total daily load and classified intraday load are counted, and the total natural gas volume and the classified natural gas volume are calculated.

In the mid-term 2019, the prospective natural gas volume and intraday load are used as saturation values, and the method of extrapolating historical data is used to obtain various intraday load forecast results. Combined with the recent classified intraday load forecast results, the intraday load indicators of the prospective blocks can be obtained. Later, the planning system is used to automatically calculate

the intraday load distribution, and the results of the short-term intraday load distribution forecast are obtained.

$$\forall (a, b) \longrightarrow (\alpha, \beta),$$

$$\exists \frac{\partial y[(w \times x) + b]}{\partial y[(w \times x) + a]} - C > 0. \tag{15}$$

Under the same subpopulation migration scale, the greater the number of subpopulations, the more individuals will participate in the migration, and the migrated excellent individuals are likely to be retained and spread in the new subpopulations they move into. Therefore, the more subpopulations, the more taboos. The algorithm evolves slowly, and the evolution process is relatively stable, but it still does not converge when it is iterated to 5,000 generations. At the beginning of the algorithm, the optimal solution was quickly approached, but it quickly fell into the local optimal solution.

5.3. Case Application and Analysis. By using the group-lasso method to estimate the data, the fitted value of the original data can be obtained. From the comparison of the fitted value and the true value below, it can be seen that the scattered points of the fitted value and the true value are distributed in that interception. The distance is zero, and the slope is 45 degrees around the straight line, the closer to that line, the closer the fitted value is to the true value, and the better the fitting effect (residual sum of squares is 0.625), from the display. In terms of the situation, the final fitting effect using the method in this article is very good.

$$\begin{cases} \lim_{x \to \infty} \{\text{sgn}(\sin x) + \text{sgn}(\cos x)\} - \lim_{x \to \infty} \left\{ \sqrt{a^2 + b^2} - \sqrt{x^2 + y^2} \right\} < 0, \\ \lim_{x \to \infty} (\sin x) - \lim_{x \to \infty} \sqrt{x^2 (x - 1)^2} = 0. \end{cases} \tag{16}$$

From the final selection of variables and the corresponding coefficient results below, it can be seen that the three remaining regional variables through data preprocessing are selected, indicating that the regional variables have a relatively large impact on the cross-sectional limit transmission power (TCC). Among the 16 finally selected variables, natural gas flow variables and daily load

variables account for 6 and 5, respectively, and there are only 2 other types of variables. This shows that natural gas flow variables and daily load variables have a relative impact on the cross-sectional limit transmission power (TCC) that has a greater impact; the last is that the signs of the coefficients in front of each variable are both positive and negative.

$$\text{s.t.} \frac{\iint \sqrt{x^2(x-1)^2} \sqrt{x^2(x+1)^2} \, dx dx'}{\iint \sqrt{x^2 + x'^2}} dx dx' - 1 = 0, \, if \begin{cases} x > 0, \\ x' > 0. \end{cases} \tag{17}$$
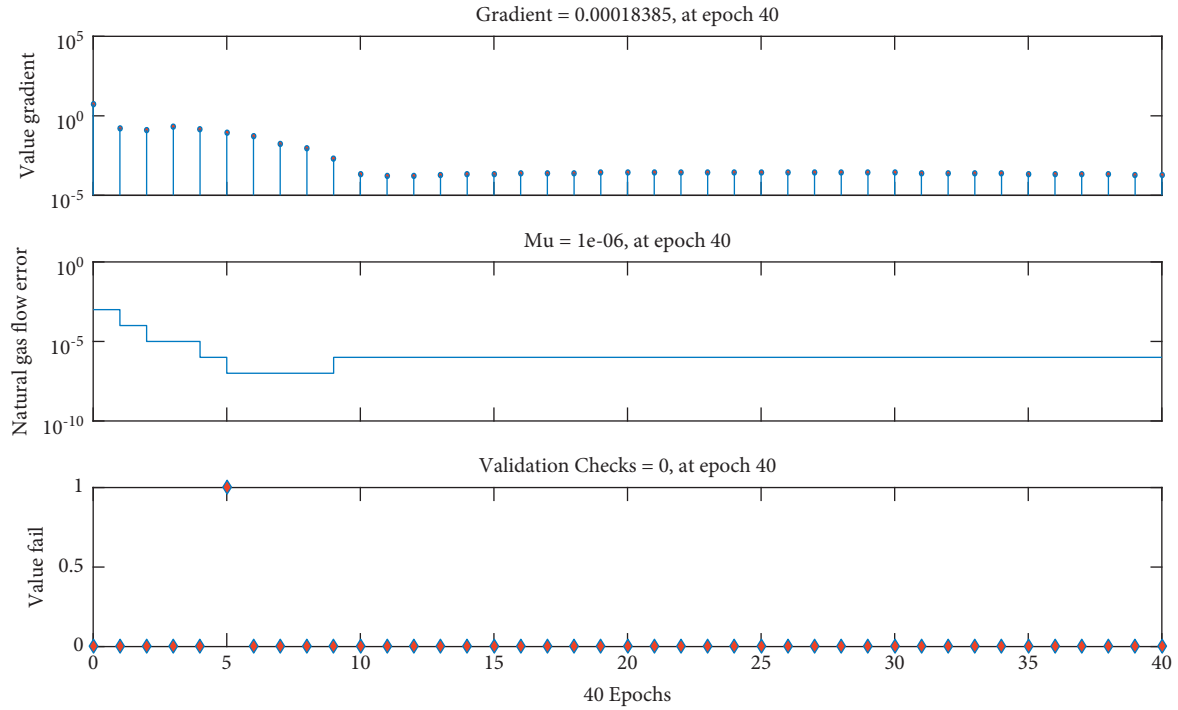
FIGURE 6: Fitted distribution of natural gas daily load forecast.

By comparing the traditional state-space prediction method with the combined prediction method, it is shown that the state-space prediction method has significantly improved the prediction accuracy and reduced the average absolute error and average relative error between the actual intraday load value and the predicted value. The combined method of natural gas avoids the limitations of a single model and reduces the risk of intraday load forecasting.

First, the state-space model based on different historical data is used to predict the long-term intraday load of regional natural gas, and the state-space model is modified with eight parameters, raw data moving average processing, and equal-dimensional and new-information processing. We calculate the predicted values of various improved methods, then use the state-space correlation weighted prediction method to obtain the predicted values of the above-mentioned various state-space models, and finally apply the artificial state space to add the influence of national economic development on the daily load of gas to the prediction model. We make the prediction model in Figure 7 more realistic and reasonable.

We select the optimal node for each independent variable through a few-fold cross-validation (CV) method; the sample size n is 99, divided into 9 groups; then here is the 9-fold cross-validation method; and the degree of freedom (df) is set. The range is 3, 4, 5, 6, 7, 8, and 9; the number of corresponding nodes is 0, 1, 2, 3, 4, 5, and 6; and the purpose is to have 4 to 10 nodes for each covariate to choose the best one among them, use 8 groups for training and 1 group for the test each time, and apply the coefficients obtained from the training samples to the test samples for estimation to obtain the residual sum of squares (MSE).

With the population n = 16 and 32-core hyper-threading on, the cascade calculation takes more time than when the 32-core hyperthreading is turned off. When the number of subpopulations is greater than the number of cores, the calculation time is significantly reduced, but when the number of subpopulations is less than the number of cores, the calculation time consumption is reduced by a small amount or even increased.

$$
\begin{vmatrix}
\sin M - \dfrac{1}{\sqrt{w^2}} & 1 & 1 \\
1 & \sin M - \dfrac{2}{\sqrt{w^2}} & 1 \\
1 & 1 & \sin M - \dfrac{3}{\sqrt{w^2}}
\end{vmatrix}
-
\begin{vmatrix}
\sin M & \cos M & 0 \\
-\cos M & \sin M & \cos M \\
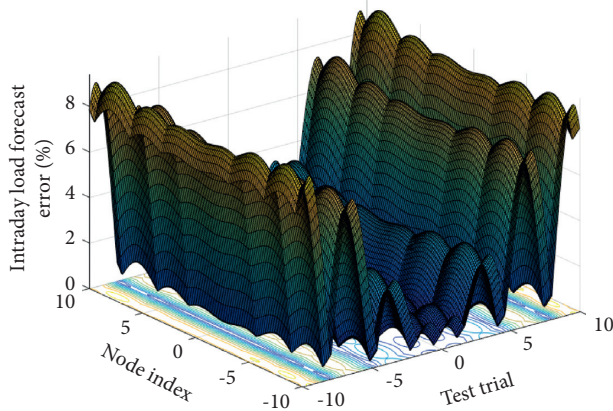0 & -\cos M & \sin M
\end{vmatrix} = 0.
\tag{18}
$$

Figure 7: Distribution of average absolute error of load forecast value within a day.

Because in the process of actual natural gas system operation, the cross-sectional limit transmission power (TCC) will not change due to a certain trend change in a certain variable, but there will generally be a critical point, the so-called inflection point. This situation can also be seen from the relationship between the above independent variables and the cross-sectional limit transmission power (TCC). In addition, when hyperthreading is turned on, the actual speed-up ratio may exceed the theoretical speed-up ratio, indicating that it can efficiently use CPU resources and hyperthreading technology to achieve efficient parallelism of algorithms. This is mainly because the more cores, the greater the impact of system management and CPU communication, the lower the proportion of parallel computing time, and the lower the corresponding efficiency.

6. Conclusion

This article concludes with data mining technology to determine the relevant influencing factors of gas consumption, classify these influencing factors, eliminate factors that have very little impact on short-term daily load, and then use the state-space method to reduce the dimensionality of the preprocessed main influencing factor data. The reconstructed data retains the main information of the original data. On this basis, data mining technology is used to analyze the main influencing factors to find out the law of the city's daily gas load. Aiming at the uncertainty and nonlinear characteristics of the gas intraday load system, this paper selects the state-space method to establish a short-term intraday load forecasting model, uses the extracted principal components as the input variables of the state space, and uses the gas intraday load data of City H for short-term daily load simulation forecast and verification. It is also the basic basis for gas companies' peak shaving and dispatching. In addition, it also plays an important guiding role in city gas planning and design. The simulation experiment takes air quality into consideration when analyzing the influencing factors of urban gas consumption. The factors have repetitiveness and high correlation with the factors that affect the daily load of gas, so it is necessary to incorporate them into the state space and choose the method of state-space

combination. The prediction of gas consumption reduces the amount of calculation and improves the accuracy of the prediction.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] R. Hafezi, A. N. Akhavan, M. Zamani, S. Pakseresht, and S. Shamshirband, "Developing a data mining based model to extract predictor factors in energy systems: application of global natural gas demand," *Energies*, vol. 12, no. 21, p. 4124, 2019.

[2] P. Potočnik, J. Šilc, and G. Papa, "A comparison of models for forecasting the residential natural gas demand of an urban area," *Energy*, vol. 167, pp. 511–522, 2019.

[3] T. Ahmad, H. Chen, J. Shair, and C. Xu, "Deployment of data-mining short and medium-term horizon cooling load forecasting models for building energy optimization and management," *International Journal of Refrigeration*, vol. 98, pp. 399–409, 2019.

[4] G. Sun, C. Jiang, X. Wang, and X. Yang, "Short-term building load forecast based on a data-mining feature selection and LSTM-RNN method," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 15, no. 7, pp. 1002–1010, 2020.

[5] S. Fathi, R. Srinivasan, A. Fenner, and S. Fathi, "Machine learning applications in urban building energy performance forecasting: a systematic review," *Renewable and Sustainable Energy Reviews*, vol. 133, Article ID 110287, 2020.

[6] X. Yang, L. Zhang, and W. Xie, "Forecasting Model for Urban Traffic Flow with BP Neural Network Based on Genetic algorithm," in *Procedings of the Control and Decision Conference (CCDC)*, pp. 4395–4399, IEEE, Kunming China, Auguest 2020.

[7] W. Ma, S. Fang, G. Liu, and R. Zhou, "Modeling of district load forecasting for distributed energy system," *Applied Energy*, vol. 204, pp. 181–205, 2017.

[8] A. Briga-Sá, D. Leitão, J. Boaventura-Cunha, and F. M. Francisco, "Trombe wall thermal performance: data mining techniques for indoor temperatures and heat flux forecasting[J]," *Energy and Buildings*, vol. 252, Article ID 111407, 2021.

[9] M. Su, Z. Zhang, Y. Zhu, and D. Zha, "Data-Driven natural gas spot price forecasting with least squares regression boosting algorithm," *Energies*, vol. 12, no. 6, p. 1094, 2019.

[10] J. Á González Ordiano, S. Waczowicz, V. Hagenmeyer, and M. Ralf, "Energy forecasting tools and services Wiley Interdisciplinary Reviews," *Data Mining and Knowledge Discovery*, vol. 8, no. 2, Article ID e1235, 2018.

[11] T. Ahmad, H. Chen, Y. Guo, and J. Wang, "A comprehensive overview on the data driven and large scale based approaches

for forecasting of building energy demand: a review," *Energy and Buildings*, vol. 165, pp. 301–320, 2018.

[12] J. Peng, A. Kimmig, J. Wang, X. Liu, Z. Niu, and J. Ovtcharova, "Dual-stage attention-based long-short-term memory neural networks for energy demand prediction," *Energy and Buildings*, vol. 249, Article ID 111211, 2021.

[13] J. Liu, S. Wang, N. Wei, X. Chen, H. Xie, and J. Wang, "Natural gas consumption forecasting: a discussion on forecasting history and future challenges," *Journal of Natural Gas Science and Engineering*, vol. 90, Article ID 103930, 2021.

[14] M. Gilanifar, H. Wang, and E. E. Ozguven, R. Arghandeh and Y. Zhou, "Bayesian spatiotemporal Gaussian process for short-term load forecasting using combined transportation and electricity data," *ACM Transactions on Cyber-Physical Systems*, vol. 4, no. 1, pp. 16–25, 2019.

[15] M. Dong and L. Grumbach, "A hybrid distribution feeder long-term load forecasting method based on sequence prediction," *IEEE Transactions on Smart Grid*, vol. 11, no. 1, pp. 470–482, 2019.

[16] Y. Huang, N. Hasan, C. Deng, and Y. Bao, "Multivariate empirical mode decomposition based hybrid model for day-ahead peak load forecasting," *Energy*, vol. 239, Article ID 122245, 2022.

[17] O. F. Beyca, B. C. Ervural, E. Tatoglu, P. G. Ozuyar, and S. Zaim, "Using machine learning tools for forecasting natural gas consumption in the province of Istanbul," *Energy Economics*, vol. 80, pp. 937–949, 2019.

[18] S. Li, R. Bhattarai, R. A. Cooke et al., "Relative performance of different data mining techniques for nitrate concentration and load estimation in different type of watersheds," *Environmental Pollution*, vol. 263, Article ID 114618, 2020.

[19] D. Solyali, "A comparative analysis of machine learning approaches for short-/long-term electricity load forecasting in Cyprus," *Sustainability*, vol. 12, no. 9, p. 3612, 2020.

[20] F. He, J. Zhou, Z.-k. Feng, G. Liu, and Y. Yang, "A hybrid short-term load forecasting model based on variational mode decomposition and long short-term memory networks considering relevant factors with Bayesian optimization algorithm," *Applied energy*, vol. 237, pp. 103–116, 2019.

[21] Y. N. Malek, M. Najib, M. Bakhouya, and M. Essaaidi, "Multivariate deep learning approach for electric vehicle speed forecasting," *Big Data Mining and Analytics*, vol. 4, no. 1, pp. 56–64, 2021.

[22] D. T. Bui, K. Khosravi, M. Karimi et al., "Enhancing nitrate and strontium concentration prediction in groundwater by using new data mining algorithm," *The Science of the Total Environment*, vol. 715, p. 136836, 2020.

[23] T. Ahmad and H. Chen, "Nonlinear autoregressive and random forest approaches to forecasting electricity load for utility energy management systems," *Sustainable Cities and Society*, vol. 45, pp. 460–473, 2019.

[24] A. Tabassum, S. Chinthavali, L. Chen, and A. prakash, "Data mining critical infrastructure systems models and tools[J]," *IEEE Intelligent Informatics Bulletin*, vol. 19, no. 2, pp. 18–22, 2018.

[25] G. Suryanarayana, J. Lago, D. Geysen, P. Aleksiejuk, and C. Johansson, "Thermal load forecasting in district heating networks using deep learning and advanced feature selection methods," *Energy*, vol. 157, pp. 141–149, 2018.