Nonlinear time series prediction algorithm based on AD-SSNET for artificial intelligence–powered Internet of Things

Banteng Liu¹,², Wei Chen³, Meng Han⁴, Zhangquan Wang¹, Ping Sun¹, Xiaowen Lv¹, Jiaming Xu⁵ and Zegao Yin²

Abstract
Time series have broad usage in the wireless Internet of Things. This article proposes a nonlinear time series prediction algorithm based on the Small-World Scale-Free Network after the AIC-Optimized Subtractive Clustering Algorithm (AIC-DSCA-SSNET, AD-SSNET) to predict the nonlinear and unstable time series, which improves the prediction accuracy. The AD-SSNET is introduced as a reservoir based on the echo state network to improve the predictive capability of nonlinear time series, and combined with artificial intelligence method to construct the prediction model training samples. First, the optimal clustering scheme of randomly distributed neurons in the network is adaptively obtained by the AIC-DSCA, then the AD-SSNET is constructed according to the intra-cluster priority connection algorithm. Finally, the reservoir synaptic matrix is calculated according to the synaptic information. Experimental results show that the proposed nonlinear time series prediction algorithm extends the feasible range of spectral radii of the reservoir, improves the prediction accuracy of nonlinear time series, and has great significance to time series analysis in the era of wireless Internet of Things.

Keywords
Time series, nonlinear, small world, scale-free, reservoir, prediction accuracy

Date received: 22 February 2021; accepted: 26 February 2021

Handling Editor: Peio Lopez Iturri

Introduction
Numerous data spark in many known or unknown fields with the advent of 5G and the Internet of Things era; these data submerge the world that human beings depend. The analysis of these data has brought great challenges to scholars and has become a hot spot of artificial intelligence research.¹⁻³ Time series data refer to a consistent stream of datasets over the course of a period of time. The time series analysis by machine learning method mainly includes clustering, classification, anomaly detection, and prediction, which will bring significant benefits to all kinds of people in various vertical fields.⁴⁻⁵ Therefore, this article researches...
the time series prediction method to provide more possibilities for analyzing a large number of time series in the wireless Internet of Things.

Time series prediction is widely used in the fields of industry, economy, environment, and so on. However, most of the time series show strong nonlinear characteristics in the real world. Therefore, it is necessary to construct a prediction model by using a nonlinear prediction method to improve the model’s fitting ability to the nonlinear time series data. At present, nonlinear prediction methods are mainly divided into two categories. One is the regression method, which is suitable for time series prediction with slower change, and the nonlinear characteristics of the time series are easily eliminated by its linearization process. The other is to predict by the neural network in machine learning, especially represented by echo state network (ESN). It has a large, sparsely connected reservoir, and its learning method is efficient. The approximation capability of nonlinear time series is mainly ensured through its reservoir. However, the random connection of internal neurons in the reservoir of traditional ESN leads to the randomness of the network structure, making the model training purposeless and poorly adaptable, and unable to meet the requirements of effective prediction to nonlinear time series. Therefore, it is necessary to analyze and improve the network structure of the reservoir.

To improve the performance of the reservoir, some scholars proposed the small-world networks to replace the random network. The small-world network is a kind of network structure that can reflect the real world. It has both a short average characteristic path length (ACPL) and a high average clustering coefficient (ACC) and has the advantage of random network and regular network. The literature proposed a small-world echo state network (SWESN). It used a small-world network to improve the structure of the reservoir and improve prediction accuracy and adaptability. Kawai et al. studied the performance of the reservoir under three different topologies: regular network, small-world network, and random network, which proved the superiority of the small-world network. However, the real-world social network nodes are randomly distributed at birth and gradually form a life circle. Therefore, some scholars applied this idea to improve ESN through clustering methods. The internal neurons of the reservoir are clustered to construct the synaptic matrix according to the clustering information. Furthermore, the predictive model is constructed to improve the nonlinear prediction capability. Deng and Zhang proposed a scale-free highly clustered echo state network (SHESN), whose reservoir is uniformly clustered with both small-world and scale-free characteristics. It was successfully applied to Mackey–Glass (MG) and laser time series prediction and obtained higher prediction accuracy than ESN. The Xue te al. applied SHESN in the financial time series prediction and achieved better prediction performance. It proved that the high clustered scale-free network has strong computing power. Lei et al. proposed a complex ESN based on prior clusters. The results of power spectrum analysis are used as prior knowledge to construct sub-clusters for the prediction problem of traffic flow time series with multi-period characteristics. Najibi and Rostami used the k-means algorithm to optimize the clustering effect of the reservoir in SHESN. However, the number of cluster heads in the reservoir must be pre-set according to prior knowledge.

This article is absorbed in the problem that the above method relies too much on prior knowledge when clustering. In order to improve the clustering performance, the nonlinear time series prediction algorithm based on the Small-World Scale-Free Network after the AIC-Optimized Subtractive Clustering Algorithm (AIC-DSCA-SSNET, AD-SSNET) is proposed. It can adaptively obtain optimal clustering scheme and construct a complex clustering network with small-world scale-free characteristics. The clustering network is used as a reservoir to improve the prediction accuracy of nonlinear time series. Moreover, it has great significance to time series analysis in the era of wireless Internet of Things.

Nonlinear time series prediction algorithm based on AD-SSNET

AD-SSEN architecture

On the basis of ESN architecture, the AD-SSEN model is constructed by using the AD-SSNET as the reservoir to improve the nonlinear approximation capability. The AD-SSEN architecture is shown in Figure 1.

The AD-SSEN architecture has three layers, and the reservoir is AD-SSNET. Its state updated equation and output equation are as follows:

Figure 1. AD-SSEN architecture.
Different parameter combinations $\Gamma = [r_u, \theta]$ of traditional SCA is set to rely heavily on expert experience, and different parameter combinations $\Gamma' = [r_u, \theta]$ will lead to different clustering effects. Therefore, we proposed the DSCA. The DSCA is as follows.

First, the cluster head numbers $k = 1, 2, 3, 4, 5,...$ are set and the SCA to obtain candidate clustering schemes is performed. Let the coordinates of $N$ randomly distributed neurons be $M_1 = (x_1, y_1), M_2 = (x_2, y_2), ..., M_N = (x_N, y_N)$, and the density value of each neuron in the network is calculated. The greater the density value of one neuron, the more the number of neurons included by it, and the higher the probability of becoming a cluster head. When calculating the first cluster head position, the density function $f_1$ of the neuron $i$ is expressed as follows:

$$f_i^1 = \sum_{j=1}^{N} \exp\left(-\frac{\|M_i - M_j\|^2}{(r_u/2)^2}\right)$$

where $r_u$ is a constant, which defines a neighborhood where the density value is significantly reduced and other neurons outside the neighborhood have little effect on the density value of neuron $i$. The maximum density value $f_1^1 = \max\{f_i^1\}$ is selected and the corresponding neuron $M_1^{opt} = (x_1, y_1)$ is the first cluster head.

In order to exclude the influence of the $z$th cluster heads on the density of other neurons, the neuron density function needs to be modified in the selection of the other $(z + 1)$th cluster heads after the first cluster head is determined. The density function $f_i^{z+1}$ of the modified neuron $i$ is expressed as follows:

$$f_i^{z+1} = f_i^z - f_i^{z} \times \sum_{j=1}^{N} \exp\left(-\frac{\|M_i - M_j^{opt}\|^2}{(r_b/2)^2}\right)$$

where $r_b$ is a constant and $r_b = 1.5r_u$. The updated maximum density value $f_i^{z+1} = \max\{f_i^{z+1}\}$ is selected and the corresponding neuron $M_i^{z+1} = (x_j, y_j)$ is the $(z + 1)$th cluster head. When $f_i^{z} = f_i^{z} \times \theta, \theta \in (0, 1)$, the one-time SCA ends and $g$ cluster heads are formed in the network.

If $g$ and $k$ are consistent, the parameter configuration $\Gamma = [r_u, \theta]$ is used as a candidate clustering scheme for $k$ cluster heads and is traversed to repeat the aforementioned SCA until all different candidate clustering schemes of $k$ cluster heads are selected.
Finally, it is necessary to choose the optimal clustering scheme with an evaluation index if there are more than one candidate scheme under the same cluster head numbers. For each candidate scheme, each neuron is assigned to the nearest cluster head by the nearest distance principle. According to the distance from each neuron of the $i$th cluster to the cluster head under $k$ cluster heads, the distance variance $D_j(j = 1, 2, ..., k)$ of the $i$th cluster is calculated as:

$$D_j(j = 1, 2, ..., k) = \sum_{i=1}^{Q} \frac{(d_i - \bar{d})^2}{i}$$  \hspace{1cm} (5)

where $d_i$ denotes the distance from the $i$th member in the $j$th cluster to the cluster head, $\bar{d}$ denotes the average distance from all members of the $j$th cluster to cluster head, and $Q$ denotes the number of all neurons in the $j$th cluster.

The maximum distance variance $\max\{D_j(j = 1, 2, ..., k)\}$ is calculated as the evaluation index. If there are multiple candidate clustering schemes, the optimal clustering scheme is selected under $k$ cluster heads according to the minimum evaluation index.

2. The AIC criterion of DSCA

It is necessary to select the optimal cluster heads after obtaining the optimal clustering scheme under different cluster heads. The AIC is used as an evaluation index proposed by H. Akaike in the study of time series ordering problems. Its distinctive feature is the “principle of parsimony,” and its definition is as follows:

$$AIC = -2 \ln (l) + 2r$$  \hspace{1cm} (6)

where $l$ is the maximum likelihood estimation function of the model, and $r$ is the number of independent parameters of the model. In general, the AIC value decreases when $r$ increases, and the log-likelihood function $\ln (l)$ increases faster. In addition, the AIC value increases, and the model to be over-fitting when $r$ is too large, and the growth rate of $\ln (l)$ is slow. Therefore, the model is best when the AIC value is the smallest.

The AIC criterion of DSCA is as follows. Setting the number of neurons to $N$ and the number of cluster heads to $k$, then the distribution of cluster heads is $M = [M^1_{opt}, M^2_{opt}, ..., M^k_{opt}], \ number\ of\ the\ number\ of\ neurons\ in \ each\ cluster\ is\ Q_i(i = 1, 2, ..., k)$, the maximum distance variance is $v_{max} = \max\{D_i|i = 1, 2, ...k\}$, and the minimum distance variance is $v_{min} = \min\{D_i|i = 1, 2, ...k\}$ in all clusters. Then, the distribution density function of intra-cluster distance variance is as follows:

$$f(v_i) = \frac{Q_i}{N v_{max} - v_{min}} = \frac{k Q_i}{N v_{max} - v_{min}}, i = 1, 2, ..., k$$  \hspace{1cm} (7)

Therefore, according to the log maximum likelihood estimation function, the intra-cluster distance variance likelihood estimation function $l$ is as follows:

$$l = \sum_{i=1}^{k} \frac{Q_i}{N v_{max} - v_{min}}$$  \hspace{1cm} (8)

$$AIC = - \ln (\sum_{i=1}^{k} \frac{Q_i}{N v_{max} - v_{min}}) + 2k$$  \hspace{1cm} (9)

According to equation (9), the cluster head numbers with the smallest $AIC$ are the optimal cluster heads, and the optimal clustering scheme under the optimal cluster number is used as the final optimal clustering scheme. The flow diagram of the AIC-DSCA is shown in Figure 3 and the specific steps are as follows:

- **Step 1:** Set the parameter configuration $\Gamma = [r_a, \theta]$ and $k = 1$.
- **Step 2:** Perform DSCA on $N$ neurons to calculate the cluster head numbers.
- **Step 3:** If cluster head numbers is $k$, proceed to Step 4. Otherwise, modify the parameter configuration $\Gamma = [r_a, \theta]$ and proceed to Step 2.
- **Step 4:** After obtaining all candidate clustering schemes under $k$ cluster heads, the optimal clustering scheme could be selected according to the intra-cluster maximum distance variance.
Construction of SSNET. According to the clustering result and the connections are established with the existing according to the distance from the backbone neurons neighbors. The number of connections of a neuron is called degree. Here, $s_i$ is the degree of current neuron $i$, and $C$ is the candidate neighbor neurons of the new local neuron. Neurons prefer to connect to neurons that already have more connections according to the scale-free criterion. Therefore, the probability that a new local neuron is connected to an existing neuron is proportional to the degree of the existing neurons.

The number of connections of a neuron is called degree. Here, $s_i$ is the degree of current neuron $i$, and $C$ is the candidate neighbor neurons of the new local neuron. Neurons prefer to connect to neurons that already have more connections according to the scale-free criterion. Therefore, the probability that a new local neuron is connected to an existing neuron is proportional to the degree of the existing neurons.

The flow diagram of the SSNET construction is shown in Figure 4 and its specific steps are as follows:

- **Step 1:** Choose a cluster. All neurons are divided into backbone neurons and local neurons.
- **Step 2:** Choose a new local neuron according to the distance from the backbone neurons, and calculate the number of candidate neighbor neurons.
- **Step 3:** Set the maximum number of connections, and let the new local neuron connect to its candidate neighbor neurons according to the intra-cluster priority connection algorithm. If all the new local neurons in the same cluster have been added, proceed to Step 4, otherwise proceed to Step 2.
- **Step 4:** If all the clusters have completed the intra-cluster connections, proceed to Step 5, otherwise, proceed to Step 1.
- **Step 5:** Let all cluster heads make a full connection, and build the small-world scale-free network.

Therefore, the AD-SSESN prediction model construction and training process of the prediction algorithm in this article are as follows:

- **Step 1:** Initialize the model parameters.
- **Step 2:** Obtain the AD-SSESN prediction model by constructing the AD-SSNET as a reservoir.
- **Step 3:** Obtain the internal state matrix $X$ and the corresponding expected output matrix $Y$ by calculating and collecting the internal state vectors $x(t)$ and output vectors $y(t)$ of the reservoir by using the training datasets.

Algorithm 1: AIC-DSCA

**Input:** the parameter configuration $\Gamma = [\alpha, \theta]$ and $k = 1$

**Output:** the optimal clustering scheme and the optimal cluster head number $k$

1: for $k = 1; k \leq N; k ++$ do
2: for $i = 1; i \leq 100; i ++$ do
3: for $j = 1; j \leq 100; j ++$ do
   - Get the cluster head number $g$ by SCA;
   - $g \leftarrow \Gamma = [\alpha(0), \theta(0)]$
4: if $g = k$ then determine the optimal scheme under the optimal cluster head number $k$
5: Calculate $AIC$ value
6: end if
7: end for
8: if $AIC = \text{min}(AIC)$ then determine the optimal scheme and the optimal cluster head number $k$
9: Goto Wait for FINISH
10: end for

- **Step 5:** Calculate the $AIC$ value of the optimal clustering scheme under $k$ cluster heads.
- **Step 6:** If the $AIC$ value is the minimum value, the optimal cluster head number is current $k$, and the final optimal scheme is determined. Otherwise $k = k + 1$ and proceed to Step 2.

The number of connections of a neuron is called degree. Here, $s_i$ is the degree of current neuron $i$, and $C$ is the candidate neighbor neurons of the new local neuron. Neurons prefer to connect to neurons that already have more connections according to the scale-free criterion. Therefore, the probability that a new local neuron is connected to an existing neuron is proportional to the degree of the existing neurons.

The number of connections of a neuron is called degree. Here, $s_i$ is the degree of current neuron $i$, and $C$ is the candidate neighbor neurons of the new local neuron. Neurons prefer to connect to neurons that already have more connections according to the scale-free criterion. Therefore, the probability that a new local neuron is connected to an existing neuron is proportional to the degree of the existing neurons.

The number of connections of a neuron is called degree. Here, $s_i$ is the degree of current neuron $i$, and $C$ is the candidate neighbor neurons of the new local neuron. Neurons prefer to connect to neurons that already have more connections according to the scale-free criterion. Therefore, the probability that a new local neuron is connected to an existing neuron is proportional to the degree of the existing neurons.

The number of connections of a neuron is called degree. Here, $s_i$ is the degree of current neuron $i$, and $C$ is the candidate neighbor neurons of the new local neuron. Neurons prefer to connect to neurons that already have more connections according to the scale-free criterion. Therefore, the probability that a new local neuron is connected to an existing neuron is proportional to the degree of the existing neurons.

The number of connections of a neuron is called degree. Here, $s_i$ is the degree of current neuron $i$, and $C$ is the candidate neighbor neurons of the new local neuron. Neurons prefer to connect to neurons that already have more connections according to the scale-free criterion. Therefore, the probability that a new local neuron is connected to an existing neuron is proportional to the degree of the existing neurons.

The number of connections of a neuron is called degree. Here, $s_i$ is the degree of current neuron $i$, and $C$ is the candidate neighbor neurons of the new local neuron. Neurons prefer to connect to neurons that already have more connections according to the scale-free criterion. Therefore, the probability that a new local neuron is connected to an existing neuron is proportional to the degree of the existing neurons.

The number of connections of a neuron is called degree. Here, $s_i$ is the degree of current neuron $i$, and $C$ is the candidate neighbor neurons of the new local neuron. Neurons prefer to connect to neurons that already have more connections according to the scale-free criterion. Therefore, the probability that a new local neuron is connected to an existing neuron is proportional to the degree of the existing neurons.

The number of connections of a neuron is called degree. Here, $s_i$ is the degree of current neuron $i$, and $C$ is the candidate neighbor neurons of the new local neuron. Neurons prefer to connect to neurons that already have more connections according to the scale-free criterion. Therefore, the probability that a new local neuron is connected to an existing neuron is proportional to the degree of the existing neurons.

The number of connections of a neuron is called degree. Here, $s_i$ is the degree of current neuron $i$, and $C$ is the candidate neighbor neurons of the new local neuron. Neurons prefer to connect to neurons that already have more connections according to the scale-free criterion. Therefore, the probability that a new local neuron is connected to an existing neuron is proportional to the degree of the existing neurons.

The number of connections of a neuron is called degree. Here, $s_i$ is the degree of current neuron $i$, and $C$ is the candidate neighbor neurons of the new local neuron. Neurons prefer to connect to neurons that already have more connections according to the scale-free criterion. Therefore, the probability that a new local neuron is connected to an existing neuron is proportional to the degree of the existing neurons.
• **Step 4:** Calculate the output weight $W_{out}$ by the least-square method, and then the trained AD-SSESN prediction model is obtained to predict nonlinear time series.

### Analysis of simulation

#### Analysis of cluster

In total, 1000 neurons were randomly distributed on the plane of $300 \times 300$. The AIC-DSCA is used for clustering to obtain the optimal clustering scheme. First, the DSCA is used to select the optimal scheme under different cluster head numbers, and the results are shown in Table 1.

Table 1 lists the clustering information under the number of cluster heads from 7 to 12. It can be seen that the clustering scheme is diverse under the same cluster head number. Therefore, the optimal clustering scheme under each cluster head number is selected by an indicator (the smallest “intra-cluster maximum distance variance”).

After obtaining the optimal clustering scheme under each cluster head number, the final optimal cluster head number and its optimal clustering scheme could be selected by the AIC of the DSCA. The $AIC$ values of

![Figure 4. Flow diagram of network construction.](image-url)
the above optimal clustering schemes are calculated and compared, and the results are shown in Table 2.

The number of cluster heads with the minimum $AIC$ value is chosen as the optimal one, so the number of the optimal cluster heads selected by the AIC-DSCA is 10; its clustering result is shown in Figure 5.

**Analysis of network characteristics**

After the optimal clustering scheme is obtained, the AD network is constructed by the method explained in section “Construction of SSNET.” The small-world characteristics and the scale-free characteristics of the AD network are analyzed.

**Analysis of small-world characteristics.** The small-world characteristics in the complex network can be characterized by its ACPL and ACC. When the ACPL is small and the ACC is large, the small-world characteristics of network are better. The ACPL and ACC of the parent network and each subnet based on the AD network are shown in Table 3. The ACPL and ACC of clustering schemes under different cluster head numbers are shown in Table 4. The ACPL and ACC of random-network, small-world network, and high-clustering scale-free network under the same scale are shown in Table 5.

It can be seen from Table 3 that the ACPL of the parent network and its subnets are small, and the ACC is large, so this indicates all have small-world characteristics. The number of members and the small-world characteristics is similar in each subnet, so this indicates that the structures of the AD network are hierarchical and uniform-clustering in terms of small-world characteristics. It can be seen from Table 4 that the ACPL reaches a minimum when the cluster head numbers are 10, and the ACC reaches a maximum when the number of the cluster heads is 10. Therefore, when the number of the cluster heads is 10, the small-world characteristic of the AD network is the best. It can be seen from Table 5 that the ACPL of the AD network is smaller than the small-world network and the highly clustered network.
scale-free network, and the ACC is larger than the random network, the small-world network, and the highly clustered scale-free network. Consequently, the small-world characteristic of the AD network is more significant.

Analysis of scale-free characteristics. The scale-free characteristics in complex networks can be characterized by whether the degree of neurons satisfies the power-law distribution. In the AD network, the degree of each neuron is calculated, and the number of neurons with different degrees is accounted for; its distribution is shown in Figure 6. It is processed logarithmically and fitted linearly, and then the correlation coefficient $R$ is calculated. It is considered that the power-law distribution is satisfied if $R > 0.95$. The logarithmic relationship between the number of neurons and the degree of neurons and their fitted lines are shown in Figure 7. By calculation, the correlation coefficient $R$ is 0.986. Therefore, it indicates that the AD network has scale-free characteristics.

In addition, the correlation coefficient $R$ of the parent network and its subnet are calculated separately. It can be seen from Table 6 that the correlation coefficients $R$ all are bigger than 0.95, so it indicates that they all have the scale-free characteristics. The number of each subnet members is similar; it further indicates that the structure of the AD network is hierarchical and uniform-clustering in terms of the scale-free characteristics.

### Analysis of prediction

**Dataset preparation and testing criterion.** The MG time series and Lorenz time series were used as the dataset of nonlinear time series for prediction, which is generated as follows:

1. The chaotic dynamic formula of the MG system is as follows:

   $$\frac{dx}{dt} = \frac{0.2x(t - \tau) + 0.1y(t)}{1 + x^d(t - \tau)}$$

   where $\tau$ denotes the time delay; the greater the time delay $\tau$, the stronger the nonlinearity of the system, which has chaotic characteristics when $\tau \approx 17$. The time delay is increased from 17 to 31 and the fourth-order Runge–Kutta algorithm is used to solve the MG system, then the 15 nonlinear time series datasets are constructed. The first 2300 points of each dataset are the training set and the last 200 points are the test set.

2. The chaotic dynamic formula of the Lorenz system is as follows:

   $$\begin{align*}
   \frac{dx}{dt} &= 10(-x + y) \\
   \frac{dy}{dt} &= 28x - y - xz \\
   \frac{dz}{dt} &= xy - 8z
   \end{align*}$$

   The Lorenz system is solved by the fourth-order Runge–Kutta algorithm, and the time series of 2500 points are calculated, then the first 2300 points of

---

Table 6. Analysis of scale-free characteristics for parent network and its subnet.

| Type    | Size | $R$     |
|---------|------|---------|
| Parent  | 1000 | 0.9860  |
| Subnet1 | 121  | 0.9862  |
| Subnet2 | 114  | 0.9874  |
| Subnet3 | 112  | 0.9903  |
| Subnet4 | 108  | 0.9921  |
| Subnet5 | 97   | 0.9879  |
| Subnet6 | 83   | 0.9861  |
| Subnet7 | 101  | 0.9936  |
| Subnet8 | 85   | 0.9875  |
| Subnet9 | 87   | 0.9893  |
| Subnet10| 92   | 0.9848  |
The normalized root mean square error (NRMSE) is the performance indicator for all simulation predictions:

\[
NRMSE = \sqrt{\frac{1}{l} \sum_{l=1}^{l} \frac{\sum_{m=n_t}^{n_t+n_e} (y'_d(m) - y'_l(m))^2}{\sum_{m=n_t}^{n_t+n_e} (y'_d(m))^2}}
\]  

where \( l \) denotes the number of the independent repeat tests, and this experiment used 100 independent repeat tests; \( n_t \) and \( n_e \) are the length of the training set and the test set, respectively; \( y'_d(m) \) denotes the true value of the \( m \)th iteration in the \( l \)th independent experiment; and \( y'_l(m) \) denotes the predicted value of the \( m \)th iteration prediction in the \( l \)th independent experiment.

**Analysis of echo state property.** Generally, the prediction model can be trained and predicted normally and stably only when the reservoir has “the echo state.” After normalizing the synaptic matrix, the spectral radius of synaptic matrix \( W_{\text{res}} \) is the largest eigenvalue \( \lambda_{\text{max}} \), and it is used to measure the intensity of the “the echo state” of the reservoir. For such a randomly connected ESN, its spectral radius must satisfy \( |\lambda_{\text{max}}| < 1 \), so that the reservoir could have “the echo state.” In order to analyze “the echo state” of AD-SSESN, the prediction effects under different spectral radius \( \lambda_{\text{max}} \) are obtained by using the MG dataset (\( \tau = 17 \)) and the Lorenz dataset to test the AD-SSESN. Compared with ESN, the results are shown in Figures 8 and 9.

It can be seen from Figures 8 and 9 that the spectral radius of the AD-SSESN is significantly larger than the ESN when NRMSE error increases significantly in the MG dataset and Lorenz data. Therefore, the “echo state” of the AD-SSESN is significantly enhanced, the stability of the prediction time series of the AD-SSESN is maintained over a wider range of spectral radius, and the predictive power is enhanced.

**Approximating nonlinear capability.** The MG dataset and Lorenz dataset are preprocessed through normalization and phase space reconstruction, and the prediction results are de-normalized. The prediction results of
AD-SSESN for MG dataset and Lorenz dataset are shown in Figures 10 and 11.

It can be seen from Figures 10 and 11 that the predicted curves of the AD-SSESN for MG datasets and Lorenz datasets are consistent with the actual curve trend, which indicates that the AD-SSESN has a high fitting ability.

In order to further analyze the error, the small-world scale-free prediction models (X-SSESN) are constructed according to different clustering schemes in Table 4, and 15 MG datasets and Lorenz datasets are predicted respectively. The results are shown in Figure 12 and Table 7.

It can be seen from Figure 12 and Table 7 that the NRMSE errors of AD-SSESN for 15 MG datasets and Lorenz datasets are the minimum. Furthermore, in the prediction of MG datasets, the prediction accuracy can still be maintained with the increase in MG delay time.

Finally, four prediction models of AD-SSESN, ESN, SWESN, and SHESN with the same reservoir size, sparse connectivity, and appropriate spectral radius are constructed respectively according to the different reservoirs in Table 5. And 15 MG datasets and Lorenz datasets are predicted by the above models, and the results are shown in Figure 13 and Table 8.

It can be seen from Figure 13 and Table 8 that AD-SSESN has better prediction results for MG datasets with different time delays and Lorenz datasets.
Table 8. Prediction results of different networks in the Lorenz dataset.

| Type    | NRMSE   |
|---------|---------|
| ESN     | 0.01565 |
| SWESN   | 0.01156 |
| SHESN   | 0.00615 |
| AD-SSESN| 0.00355 |

NRMSE: normalized root mean square error; ESN: echo state network; SWESN: small-world echo state network; SHESN: scale-free highly clustered echo state network.

compared with the other models. In the MG datasets, the AD-SSESN can maintain good prediction performance when $17 \leq \tau \leq 24$. When the time delay $\tau > 24$, the nonlinear enhancement of the MG dataset, the NRMSE error of ESN, SWESN, and SHESN increases rapidly, while the NRMSE error of AD-SSESN increases relatively slowly. In the Lorenz datasets, the AD-SSESN has the minimum NRMSE error and the best prediction performance. To sum up, the clustering performance of the reservoir is optimized, and the ACC of network is improved in the AD-SSESN. In addition, the “echo state” of the AD-SSESN is also significantly enhanced because its spectral radius is extended, so that the highly complex nonlinear dynamic systems can be fitted by the AD-SSESN.

Conclusion
This article proposes a nonlinear time series prediction algorithm based on the AD-SSNET, which improves the prediction accuracy of the prediction model to nonlinear time series data and brings more possibilities for the analysis of a large number of time series in the wireless Internet of Things. The number of the optimal cluster heads is obtained adaptively and its clustering schemes are optimized by the AIC-DSCA, then the AD-SSNET with small-world scale-free characteristics is constructed by the intra-cluster priority connection algorithm. This network is used as a reservoir to construct the AD-SSESN prediction model. Finally, the AD-SSESN prediction model is used to predict MG datasets and Lorenz datasets, respectively. Experimental results show that the NRMSE error of AD-SSESN is the minimum compared with other small-world scale-free network prediction models with different clustering schemes; the NRMSE error of AD-SSESN is also the minimum compared with the other three prediction models with different reservoir networks. The above results show that the highly complex nonlinear dynamic system is more accurately approximated. The prediction accuracy is steadily improved due to the optimized clustering performance of the reservoir, the ACC of the network is improved, and the “echo state” is enhanced significantly in the AD-SSESN.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Public Welfare Technology Research Projects of Zhejiang Province of China (Grant Nos LGG20F010009 and GF21F010018), the Zhejiang Shuren University Basic Scientific Research Special Funds (No. 2020XZ009), and the Project Intelligentization and Digitization for Airline Revolution (No. 2018R02008).

ORCID iDs
Banteng Liu https://orcid.org/0000-0003-3341-3867
Meng Han https://orcid.org/0000-0001-7472-0842

References
1. Witze A. 5G data networks threaten forecasts. Nature 2019; 569(7754): 17–18.
2. Cai Z and He Z. Trading private range counting over big IoT data. In: 2019 IEEE 39th international conference on distributed computing systems (ICDCS), Dallas, TX, 7–10 July 2019. New York: IEEE.
3. Zheng X, Cai Z and Li Y. Data linkage in smart internet of things systems: a consideration from a privacy perspective. IEEE Commun Mag 2018; 56(9): 55–61.
4. Luo Y and Wang Y. A statistical time-frequency model for non-stationary time series analysis. IEEE T Signal Pr 2020; 68: 4757–4772.
5. Cai Z, Zheng X and Yu J. A differential-private framework for urban traffic flows estimation via taxi companies. IEEE Trans Ind Inform 2019; 15: 6492–6499.
6. Su H, Zio E, Zhang J, et al. A hybrid hourly natural gas demand forecasting method based on the integration of wavelet transform and enhanced deep-RNN model. Energy 2019; 178: 585–597.
7. Li T, Luo C and Shao R. The structure and dynamics of granular complex networks deriving from financial time series. Int J Modern Phys C 2020; 31: 2050087.
8. Tian Y, Liu H, Liang T, et al. Fine particulate air pollution and adult hospital admissions in 200 Chinese cities: a time-series analysis. Int J Epidemiol 2019; 48; 1142–1151.
9. Wang S, Gao R and Du R. Bayesian network integrated regression model with super parent time series. J Comput Sci 2017(12): 116–129.
10. Chaouch M. Volatility estimation in a nonlinear heteroscedastic functional regression model with martingale difference errors. J Multivariate Anal 2019; 170: 129–148.
11. Vilela LF, Leme RC, Pinheiro CA, et al. Forecasting financial series using clustering methods and support vector regression. Artif Intel Rev 2019; 52(2): 743–773.
12. He Z, Han M and Han B. Pattern-adaptive time series prediction via online learning and paralleled processing
using CUDA. In: 2019 IEEE 16th international conference on mobile ad hoc and sensor systems workshops (MASSW), Monterey, CA, 4–7 November 2019, pp.31–36. New York: IEEE.
13. Chen Q, Cai Z, Cheng L, et al. Low-latency data aggregation scheduling for cognitive radio networks with non-deterministic structure. IEEE Trans Mobile Comput 2020; 2020: 9710.
14. Liu CL, Hsiao WH and Tu YC. Time series classification with multivariate convolutional neural network. IEEE T Ind Electr 2018; 66(6): 4788–4797.
15. Yang HF and Chen YPP. Hybrid deep learning and empirical mode decomposition model for time series applications. Expert Syst Appl 2019; 120: 128–138.
16. Wang L, Yang C and Qiao J. Research on pm(2.5) prediction based on echo state network. Control Eng 2019; 26(01): 3–7.
17. Peng Y, Wang J and Peng X. Research on time series prediction method based on echo state network. J Electr 2010; 38(01).
18. Chouikhi N, Ammar B, Rokbani N, et al. Pso-based analysis of echo state network parameters for time series forecasting. Appl Soft Comput 2017; 55: 211–225.
19. Miao D, Li J and Cai Z. Maximum reachability preserved graph cut. Theoret Comput Sci 2020; 840: 187–198.
20. Luo G, Yan K, Zheng X, et al. Preserving adjustable path privacy for task acquisition in mobile crowdsensing systems. Inform Sci 2020; 527: 602–619.
21. Jaeger H. Echo state network. Scholarpedia 2007; 2(9): 2330.
22. Cai Z, He Z, Guan X, et al. Collective data-sanitization for preventing sensitive information inference attacks in social networks. IEEE T Depend Secure Comput 2018; 15(4): 577–590.
23. Yu D, Zhang L, Luo Q, et al. Fast skyline community search in multi-valued networks. Big Data Mining Anal 2020; 3(3): 171–180.
24. Cai Z and Zheng X. A private and efficient mechanism for data uploading in smart cyber-physical systems. IEEE T Netw Sci Eng 2020; 7(2): 766–775.
25. Albinali H, Han M, Wang J, et al. The roles of social network mavens. In: 2016 12th international conference on mobile Ad-Hoc and sensor networks (MSN), Hefei, China, 16–18 December 2016, pp.1–8. New York: IEEE.
26. Han M, Han Q, Li L, et al. Maximising influence in sensed heterogeneous social network with privacy preservation. Int J Sensor Netw 2018; 28(2): 69–79.
27. Li H. Nonlinear time series prediction based on echo state network. PhD Thesis, Dalian University of technology, Dalian, China, 2013.
28. Wang J. Topological structure of echo state network. PhD Thesis, Chongqing University, Chongqing, China, 2013.
29. Chen F. Research on network security situation prediction technology based on small world echo state network. PhD Thesis, Lanzhou University, Lanzhou, China, 2014.
30. Lun S, Lin J and Yao X. Time series prediction based on small world echo state network. J Automat 2015(09): 1669–1679.
31. Han M, Yan M, Cai Z, et al. Influence maximization by probing partial communities in dynamic online social networks. Trans Emerg Telecommun Technol 2017; 28(4): e3054.
32. Yu S, Zhang L, Li P, et al. Efficient interest-aware data dissemination in mobile opportunistic networks. J Softw Pract Exp. Epub ahead of print 1 October 2019. DOI: 10.1002/spe.2757.
33. He JS, Han M, Ji S, et al. Spreading social influence with both positive and negative opinions in online networks. Big Data Mining Anal 2019; 2(2): 100–117.
34. Kawai Y, Park J and Asada M. A small-world topology enhances the echo state property and signal propagation in reservoir computing. Neural Netw 2019; 112: 15–23.
35. Liu L, Han M, Zhou Y, et al. LSTM recurrent neural networks for influenza trends prediction. In: International symposium on bioinformatics research and applications, Beijing, China, 8–11 June 2018, pp.259–264. Cham: Springer.
36. Deng Z and Zhang Y. Collective behavior of a small-world recurrent neural system with scale-free distribution. IEEE T Neural Netw 2007; 18(5): 1364–1375.
37. Xue F, Li Q, Li X, et al. The application of SHESN on financial time series prediction. In: 2016 3rd international conference on information science and control engineering (ICISCE), Beijing, China, 8–10 July 2016, pp.692–696. New York: IEEE.
38. Lei M, Peng Y, Guo J, et al. Traffic prediction based on a priori cluster complex echo state network. J Instrum 2011; 10: 32–39.
39. Najibi E and Rostami H. SCESN, SPESN, SWESN: three recurrent neural echo state networks with clustered reservoirs for prediction of nonlinear and chaotic time series. Appl Intel 2015; 43(2): 460–472.
40. Chen M, Fang Y and Chen J. Circular curve fitting by least square method and iterative method. Survey Mapping Sci 2016; 41(001): 194–197202.
41. Xueling Suo Du, et al. Anatomic insights into disrupted small-world networks in pediatric posttraumatic stress disorder. Radiology 2017; 282: 826–834.
42. Chauhan A, Friedrich T and Rothenberger R. Greed is good for deterministic scale-free networks. Algorithmica 2020; 82: 3338–3389.