Received: 31 March 2023 | Revised: 15 May 2023 | Accepted: 5 June 2023

DOI: 10.4218/etrij.2023-0136

ORIGINAL ARTICLE

ETRI Journal WILEY

Network traffic prediction model based on linear and nonlinear model combination

Lian Lian

College of Information Engineering, Shenyang University of Chemical Technology, Shenyang, China

Correspondence
Lian Lian, College of Information Engineering, Shenyang University of Chemical Technology, Shenyang, China.
Email: lianlian_syuct@163.com

Funding information
the Doctoral Scientific Research Foundation of Liaoning Province, Grant/Award Number: 20180540050

Abstract

We propose a network traffic prediction model based on linear and nonlinear model combination. Network traffic is modeled by an autoregressive moving average model, and the error between the measured and predicted network traffic values is obtained. Then, an echo state network is used to fit the prediction error with nonlinear components. In addition, an improved slime mold algorithm is proposed for reservoir parameter optimization of the echo state network, further improving the regression performance. The predictions of the linear (autoregressive moving average) and nonlinear (echo state network) models are added to obtain the final prediction. Compared with other prediction models, test results on two network traffic datasets from mobile and fixed networks show that the proposed prediction model has a smaller error and difference measures. In addition, the coefficient of determination and index of agreement is close to 1, indicating a better data fitting performance. Although the proposed prediction model has a slight increase in time complexity for training and prediction compared with some models, it shows practical applicability.

KEYWORDS

combined prediction, improved slime mold algorithm, linear model, network traffic, nonlinear model

1 | INTRODUCTION

The Internet is ever-expanding driven by its popularization and the development of networking technologies, and network services and applications are becoming increasingly diverse. Network traffic can reflect user activities and allow to evaluate the network load and operation status [1]. Through network traffic prediction, researchers can manage the network operation, find bottlenecks, detect potential threats and faults, optimize configuration, and perform intrusion detection and fault management according to complex characteristics and change laws [2]. Therefore, network traffic prediction has become a research hotspot.

Network traffic exhibits complex linear and nonlinear characteristics [3]. Therefore, combining linear and nonlinear components for network traffic prediction seems reasonable to increase its accuracy. By extracting the linear and nonlinear components of network traffic separately and selecting appropriate models for prediction, correct prediction of network traffic can be achieved when adding the predictions of linear and nonlinear models.

We propose a combined prediction model for network traffic considering linear and nonlinear components. The linear component of network traffic is described by an autoregressive moving average (ARMA) prediction model. In addition, an echo state network (ESN) optimized by an improved slime mold algorithm (ISMA) is introduced to predict the residual of the network traffic after removing the linear components predicted by the ARMA model. Finally, the predictions of the ARMA model and ISMA-optimized ESN are added to obtain the prediction results. Two real network traffic datasets are considered as the research object, and various evaluation indicators allow to verify the effectiveness of the proposed prediction model. The results indicate that the proposed prediction model is superior to existing models for all performance indicators.

The main contributions of this study are as follows:

1. A combined network traffic prediction model based on both linear and nonlinear components is proposed.
2. ARMA is selected as the prediction model for the linear component of network traffic.
3. The ISMA-optimized ESN is adopted as the prediction model for the nonlinear component of network traffic.

2 | RESEARCH STATUS

Network traffic prediction is based on objective and scientific mathematical methods to analyze and describe related trends [4]. Currently, linear, nonlinear, and combined prediction models are the mainstream research directions in network traffic prediction. Such models have promoted the development and progress of network traffic prediction.

2.1 | Linear prediction models

Traditional models are based on the periodicity of linear changes. They are simple and fast and allow to suitably interpret and describe linear data. Early network traffic prediction assumed that network traffic follows a strict periodic law, and the prediction was obtained by averaging historical data. Exponential smoothing also assumes that network traffic is stable and regular and that the calculations are not intensive. For prediction, the weight of historical data converges to zero, and the prediction is corrected by calculating an exponential smoothing parameter [5]. Researchers have considered that network traffic follows a Poisson distribution [6] or resembles a Markov process [7]. Other models include the Markov time-varying [8], ARMA [9], and autoregressive integrated moving average (ARIMA) [10] models. However, accurately describing the complex characteristics of network traffic remains difficult using traditional linear prediction models.

2.2 | Nonlinear prediction models

Network traffic has complex characteristics such as nonlinearity, self-similarity, multifractality, and periodicity. With the continuous development of machine learning, numerous nonlinear models have been applied to network traffic prediction. Because of their characteristics, support vector machines (SVMs) and least-squares SVMs are particularly suitable for network traffic prediction with small sample sizes [11, 12]. In many cases, these models outperform ordinary neural networks. However, both SVM and least-squares SVM can suffer from the curse of dimensionality by large amounts of data and are very sensitive to model parameters [13]. Another widely used nonlinear model is the neural network, which has a strong nonlinear mapping ability and can process highly nonlinear data. Typical application models of neural networks include ESNs [14], extreme learning machines [15], fuzzy neural networks [16], and radial basis function neural networks [17]. Several experimental simulation studies have demonstrated that a nonlinear prediction model based on an SVM or neural network provides more accurate predictions than a traditional linear prediction model.

The continuous improvement in computing power has opened new directions for applying deep learning to network traffic prediction [18, 19]. A deep learning model is very suitable for nonlinear network traffic modeling and prediction. Nevertheless, deep learning has some defects, such as uneven structure, difficulty in determining hyper-parameters, and susceptibility to fall into local optima. In addition, it requires algorithm optimization during training and has a large calculation cost, requiring massive computing power from graphics processors.

2.3 | Combined prediction models

Separate linear or nonlinear models fail to directly describe all the characteristics of real network traffic, thus affecting the final prediction. Therefore, the combination of prediction models has been explored. Combined prediction models are divided into three main approaches. The first approach is combining two or more single prediction models to increase the accuracy of prediction by iterating multiple prediction results. Examples include the Kalman filter, ESN [20], deep neural network, and hidden Markov model [21]. The second approach introduces an optimization algorithm into the model to improve prediction. These models include particle swarm optimization and Elman neural networks [22]. The third approach introduces an algorithm to decompose the network traffic and then selects the appropriate prediction model for each obtained component. These models include wavelet transform with multiple model fusion [23], ensemble empirical mode decomposition with multiple models [24], and variational mode decomposition (VMD) with multi-reservoir ESN [25]. Overall, the feasibility of combined prediction models has been confirmed.

In practice, network traffic involves a mixture of linear and nonlinear components. When selecting a prediction model, we should extract the different components of network traffic, input them into the corresponding linear or nonlinear model for independent prediction, and combine the prediction results.

3 | PROPOSED PREDICTION MODEL

In a complex system, external influencing information causes changes in the system. For network traffic, the influencing factors are complex. Therefore, a prediction model can only describe the main trend of network traffic owing to diverse influencing factors. The prediction error has been mostly neglected as a factor influencing prediction. Moreover, for a network traffic sample, the performance of different models varies. Nevertheless, any single model can extract information from a sample, and its prediction error can be considered valuable. Therefore, different models can be applied to the same data. If the change law of network traffic can be determined from historical data, the prediction error can be estimated, and the obtained prediction can be compensated to improve the overall prediction accuracy. Hence, we first use a linear ARMA prediction model to fit the network traffic data. Then, the ISMA-optimized ESN with good nonlinear prediction ability is used to fit the prediction error. The final prediction is derived from the sum of the predictions of the two models.

Let a network traffic sample be denoted as $T(k)$, where k is the current sampling time. Linear and nonlinear components are included in the sample:

$$T(k) = L(k) + N(k), \tag{1}$$

where $L(k)$ and $N(k)$ are the linear and nonlinear components of the network traffic sample, respectively.

If an ARMA model is adopted to forecast network traffic, the prediction of the linear component is $\overline{L}(k)$. The prediction error, including the nonlinear components shown in (2), can then be obtained.

$$N(k) = T(k) - \overline{L}(k). \tag{2}$$

The ISMA-optimized ESN forecasts nonlinear component $N(k)$, where $\overline{N}(k)$ is the prediction. Hence, the final prediction, $\overline{T}(k)$, of the combined prediction model can be obtained as follows:

$$\overline{T}(k) = \overline{L}(k) + \overline{N}(k). \tag{3}$$

A diagram of the proposed network traffic prediction model is shown in Figure 1. The model performs a one-step prediction. In practice, network traffic should be predicted at multiple moments in the future. We obtain multistep prediction by using one-step prediction and cyclic iteration. Specifically, the model predicts network traffic at a sampling time in the future. The prediction is considered as the real value and placed into the head of the network traffic sample queue, whereas the earliest network traffic data point at the end of the queue is discarded. Hence, network traffic values at multiple sampling times can be predicted. As shown in Figure 1, the predictions of the ARMA model and ESN are added with weights of 1.

3.1 | Modeling

We first collected network traffic samples and divided them into disjoint training and test sets. The ARMA model was determined using the training samples. The obtained ARMA model was used to forecast the network traffic training set and obtain predictions. The real values of the network traffic sample minus the predicted values

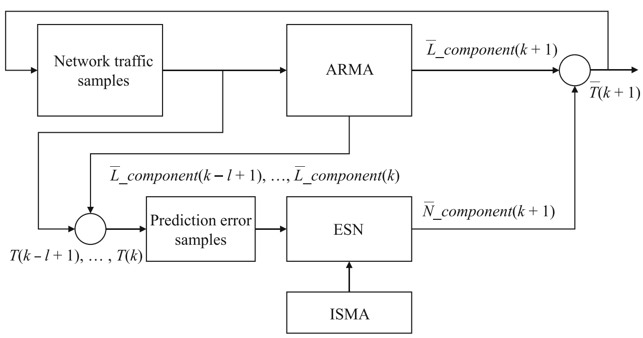


FIGURE 1 Diagram of proposed network traffic prediction model.

of the ARMA model were used to obtain prediction error samples, which were then used to train the ESN, whose reservoir parameters were optimized by the ISMA.

Step 1 For current sampling time k, the inputs of the network traffic test set are $T(k-l+1)$, $T(k-l+2)$, ..., $T(k-1)$, where l is the length of input sample. Then, prediction $\overline{L}(k+1)$ of the ARMA model is obtained.

Step 2 Predictions $\overline{L}(k-l+1)$, ..., $\overline{L}(k-1)$ are subtracted from $T(k-l+1)$, $T(k-l+2)$, ..., $T(k-1)$ to obtain the prediction error samples. These samples are input into the ESN to obtain prediction $\overline{N}(k+1)$.

Step 3 The final prediction, $\overline{T}(k+1)$, at the next sampling time is obtained by adding $\overline{L}(k+1)$ and $\overline{N}(k+1)$.

Step 4 The input network traffic sample sequence is updated. The last value of the sequence is discarded, and $\overline{T}(k+1)$ is added to the head of the sequence.

Step 5 Steps 2–4 are repeated until all the predictions are made.

4 | FORMULATIONS

In this section, the formulations of the ARMA model, ESN, and ISMA for the proposed prediction model are presented.

4.1 | ARMA model

The ARMA model allows to evaluate the correlation between periodic and nonstationary data and accurately predict linear components. Thus, this model is appropriate for predicting linear components in network traffic. The stationary and linear series are modeled using the ARMA model as follows:

$$y_t = \theta_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots \theta_q \varepsilon_{t-q}, \tag{4}$$

where y_t represents the data, ϕ_i and θ_i are polynomial parameters, and ε_t is white noise that follows a normal distribution. The order determination of the ARMA model is based on the Akaike information criterion defined as

$$\mathrm{AIC} = -2 \ln L + 2n, \tag{5}$$

where L and n are model parameters. Hence, p and q in (4) can be obtained as the orders that can minimize (5).

4.2 | ISMA-optimized ESN

4.2.1 | ESN

The sparse ESN consists of an input layer, middle layer (dynamic reservoir), and output layer. Several hidden neurons constitute the dynamic unit of the ESN, which establishes short-term memory. The prediction performance of the ESN is several times better than that of conventional neural networks. Hence, we use the ESN to predict the nonlinear components of network traffic. Its output, $\overline{N}(k)$, can be expressed as

$$\overline{N}(k) = \sum_{i=1}^{L} \mathbf{w}_i^{\mathrm{out}} x_i(k). \tag{6}$$

This formula implies that the weight matrix expected by the ESN should minimize the mean square error of the system:

$$\mathbf{W}_{\mathrm{out}} = \left(\mathbf{M}^{-1} \times \mathbf{T}\right)^{\mathrm{T}}, \tag{7}$$

where $\mathbf{M}$ and $\mathbf{T}$ are the input and output matrices, respectively.

The ESN performance is drastically affected by the reservoir parameters. Four parameters are considered: connection weight spectrum radius SR, reservoir size N, input scale IS, and sparsity SD. We optimize these parameters using the ISMA.

4.2.2 | ISMA

The slime mold algorithm (SMA) performs swarm intelligence optimization. The algorithm imitates the behavioral characteristics of the slime mold during foraging and introduces weights to simulate the correlation between the contraction mode of its venous wall and its shape change [26]. Compared with conventional optimization algorithms, such as particle swarm optimization, differential evolution, and genetic algorithms, the SMA has a higher solving efficiency and faster convergence, being suitable for solving practical engineering optimization problems. Therefore, we use the ISMA to optimize the ESN.

Slime molds approach food by smell that propagates in the air. The position update for a slime mold approaching food can be expressed as

$$I_{new} = \begin{cases} \text{rand}().(UC-LC)+LC, & \text{rand}()<z \\ I_b(t)+mb\cdot(W\cdot w\cdot I_A(t)-I_B(t)), & r<p, \\ mc\cdot I(t), & r\geq p, \end{cases} \tag{8}$$

where

$$p = \tanh|j(t)-DE|, \tag{9}$$

LC and UC are the lower and upper bounds of the search space, respectively, mb is in range $[-k, k]$, mc decreases linearly from 1 to 0, t is the current iteration number, I_b is the currently highest concentration of food, I is the current position of the slime mold, I_A and I_B represent two randomly selected slime molds, W is the weight of the slime mold, $j(t)$ is the fitness value of I, and DE is the optimal fitness value. We calculate k as follows:

$$k = \text{arc}\tanh\left(-\frac{1}{\max T}+1\right), \tag{10}$$

where $\max T$ denotes the maximum number of iterations. The weight of the slime mold, W, is updated as follows:

$$W(\text{SmellIndex}(i)) = \begin{cases} 1+r.\log\left(\frac{bF-j(i)}{bF-wF}+1\right), & A, \\ 1-r.\log\left(\frac{bF-j(i)}{bF-wF}+1\right), & B, \end{cases} \tag{11}$$

$$\text{SmellIndex} = \text{sort}(j), \tag{12}$$

where A represents the top half of the population after ranking, B represents the remaining population, r is a random value between 0 and 1, bF and wF are the currently best and worst fitness values, respectively, and *SmellIndex* is the sequence of fitness values.

We propose a weight updating method, in which the iteration time of the SMA is shortened owing to random characteristics. This improvement is aimed to introduce a weight into the standard SMA individual update and improve the diversity of newly generated individuals through random weights, thereby avoiding an imbalance between exploration and exploitation. The update of the random weight is given by (13). The update of the population position after introducing the random weight is given by (14).

$$w = w_{\min} + (w_{\max} - w_{\min})\cdot\text{rand}() + \sigma\cdot\text{randn}(), \tag{13}$$

$$I_{\text{new}} = \begin{cases} \text{rand}().(UC-LC)+LC, & \text{rand}()<z, \\ w\cdot I_b(t)+mb\cdot(W\cdot w\cdot I_A(t)-I_B(t)), & r<p, \\ mc\cdot w\cdot I(t), & r\geq p, \end{cases} \tag{14}$$

where $w_{\min}$ and $w_{\max}$ are the minimum and maximum of the random weight, respectively, rand() is a random number uniformly distributed between 0 and 1, and σ is the standard deviation used to determine the error between the weight and expected value. The improved updating method described by (13) and (14) allows the SMA to jump out of a local optimum at the beginning of search and search over a wider range. In late stages, the improved strategy completes search within a small range, thereby enhancing the optimization ability of the SMA.

4.2.3 | Optimization

The process to obtain the ISMA-optimized ESN is as follows. The four parameters of the ESN dynamic reservoir are regarded as individuals in the ISMA, and the optimal value is determined iteratively. Optimization proceeds as follows:

Step 1 The parameters are initialized, including the population size, N, maximum number of iterations, $\max T$, and parameter z. The initial population position is also generated.

Step 2 The fitness function per iteration is given by (15), and the fitness values of all individuals are evaluated. *SmellIndex* is generated by ranking the fitness values. Then, bF and wF are obtained. The weight of the mold is calculated using (11).

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-\bar{y}_i))^2}, \tag{15}$$

where N is the number of samples and y_i and $\bar{y}_i$ are the real and predicted values of the sample, respectively.

Step 3 The random weight is recalculated based on (13). The individual position is updated using (14). Subsequently, the parameters that should be optimized, I_b, are calculated. In this iteration, by rejudging the size of the two fitness values, the smaller one is taken as the optimal value. Simultaneously, the current position is set as the optimal position for the population.

Step 4 The individual fitness is calculated, and the global optimal position of the population is updated. Parameter k can be obtained using (10), and parameter b is calculated using (16). The global optimal position of the population is the updated.

$$b = 1 - \frac{T}{\max T}. \tag{16}$$

Step 5 When the condition for jumping out of the loop is satisfied, the last iteration index is considered as the maximum number of iterations. If the termination conditions are satisfied, the four optimized dynamic reservoir parameters (SR, N, IS, and SD) are output. Otherwise, steps 1–5 are repeated to continue the optimization.

5 | EVALUATION AND RESULTS

We verified the effectiveness of the proposed network traffic prediction model through comparisons considering two types of network traffic data.

5.1 | Datasets

Two datasets were collected in this study. Dataset A was obtained from fixed network traffic of a router on campus with a sampling period of 1 h. Dataset B was obtained from mobile network traffic of a mobile base station with a sampling period of 5 min. We collected 300 samples per dataset. The training and test sets were obtained by splitting in a ratio of 5:1. The two network traffic datasets exhibited periodicity and randomness, as shown in Figure 2. Within a short period, the network traffic had no obvious change law. When considering a larger period, the network traffic showed a similar change trend. Network traffic has both nonlinear and linear characteristics, which are the premises for designing the prediction model.

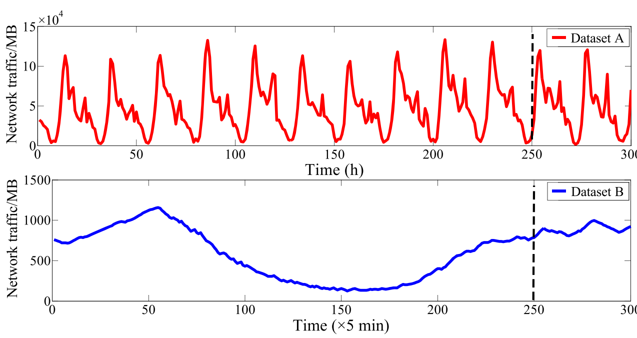


FIGURE 2 Network traffic datasets collected in this study.

5.2 | Performance indicators

The following performance indicators were used for comparison between different models to verify the effectiveness of the proposed model.

Root mean square error (RMSE):

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(T(k) - \overline{T}k))^2}. \tag{17}$$

Mean absolute error (MAE):

$$\mathrm{MAE} = \frac{1}{N}\sum_{k=1}^{N} | T(k) - \overline{T}(k) | . \tag{18}$$

Mean absolute percentile error (MAPE)

$$\mathrm{MAPE} = \frac{1}{N}\sum_{k=1}^{N} | T(k) - \overline{T}(k) | \times 100/T(k). \tag{19}$$

Relative RMSE (RRMSE):

$$\mathrm{RRMSE} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(\frac{T(k) - \overline{T}(k)}{T(k)}\right)^2}. \tag{20}$$

Square sum error (SSE):

$$\mathrm{SSE} = \sum_{k=1}^{N}\left(T(k) - \overline{T}(k)\right)^2. \tag{21}$$

Coefficient of determination (R^2):

$$R^2 = 1 - \frac{\sum_{k=1}^{N}\left(T(k) - \overline{T}(k)\right)^2}{\sum_{k=1}^{N}\left(T(k) - Tm\right)^2}. \tag{22}$$

Theil inequality coefficient (TIC):

$$\mathrm{TIC} = \frac{\sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(T(k) - \overline{T}(k)\right)^2}}{\sqrt{\frac{1}{N}\sum_{k=1}^{N}T(k)^2} + \sqrt{\frac{1}{N}\sum_{k=1}^{N}\overline{T}(k)^2}}. \tag{23}$$

Index of agreement (IA):

$$\mathrm{IA} = 1 - \frac{\sum_{k=1}^{N}\left(T(k) - \overline{T}(k)\right)^2}{\sum_{k=1}^{N}\left(|\overline{T}(k) - Tm| + |T(k) + Tm|\right)^2}. \tag{24}$$

In (17)–(24), N is the number of samples and $T(k)$, $\overline{T}(k)$, and Tm are the real, predicted, and average network traffic values, respectively.

The effectiveness of the prediction model can also be evaluated statistically using methods such as the Wilcoxon signed-rank and rank-sum tests. These nonparametric hypothesis tests can easily determine whether a prediction is consistent with the real value.

5.3 | Comparison models

Three single prediction models and two combined prediction models were selected for comparison. The single models were an ARIMA model [10], SVM [11], and ESN [14]. The two combined models were wavelet transform with multiple model fusion [23] and VMD with multi-reservoir ESN [25]. The parameters of the comparison models were selected according to either a conventional algorithm or the values recommended in the literature.

For the ISMA adopted in this study, we set the population size to 20, maximum number of iterations to 100, and z in (14) to 0.03. The ranges of the reservoir parameters to be optimized were $SR, IS, SD \in [0.01, 1)$ and $N \in [10, 200]$. The sample sequence had a length l of 30. The detailed parameters of the proposed and comparison models are listed in Table 1.

5.4 | Results

The ARMA model was applied to the network traffic training set and then to predict that set again. The difference between the prediction and original training data was used to generate the training set of prediction errors for the ISMA-optimized ESN. The prediction results of the ARMA model and training set of prediction errors for datasets A and B are shown in Figures 3 and 4, respectively. The prediction of the ARMA model showed a linear trend, indicating that it suitably fit the linear component of network traffic.

After obtaining the model parameters of the ARMA model and ESN, the two network traffic test sets were predicted. Figures 5 and 6 show the predictions of the ARMA model for network traffic and ESN for network traffic prediction error, respectively. The ARMA model and ESN achieved good prediction results.

The predictions of the ARMA model and ESN were added to obtain the final prediction. The predictions of the five comparison models were also obtained. The comparison results for datasets A and B are shown in Figures 7 and 8, respectively. The proposed prediction model better described changes and internal

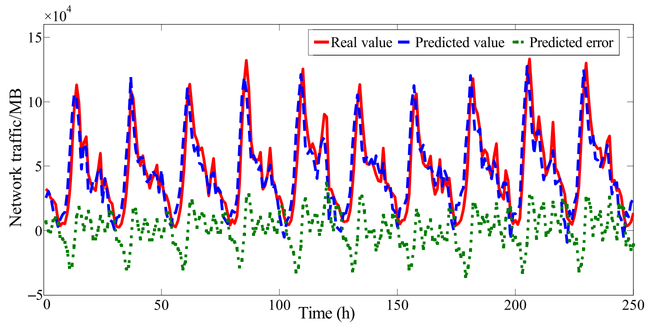


FIGURE 3 Prediction results of autoregressive moving average (ARMA) model on training set and corresponding prediction error for dataset A.

TABLE 1 Prediction model parameters.

Prediction model	Dataset A	Dataset B
Proposed model	ARMA model: $p = 4$, $q = 2$. ESN: $SR = 0.4642$, $N = 65$, $IS = 0.4630$, $SD = 0.5563$	ARMA model: $p = 4$, $q = 3$. ESN: $SR = 0.5128$, $N = 52$, $IS = 0.5152$, $SD = 0.4903$
ARIMA model	$p = 5, d = 2, q = 3$	$p = 5, d = 2, q = 3$
SVM	$C = 3.5562$, $g = 12.4561$	$C = 7.6652$, $g = 10.3305$
ESN	$SR = 0.3748$, $N = 52$, $IS = 0.5241$, $SD = 0.5108$	$SR = 0.3845$, $N = 61$, $IS = 0.4125$, SD $= 0.4251$
Wavelet transform with multiple model fusion	Ca: ARIMA (5, 3, 2). Cd1: $\gamma = 20.7651$, $\delta^2 = 9.0083$, $m = 28$. Cd2: $\gamma = 11.0052$, $\delta^2 = 27.3302$, $m = 26$. Cd3: $\gamma = 7.0870$, $\delta^2 = 19.8892$, $m = 25$	Ca: ARIMA (5, 2, 2). Cd1: $\gamma = 16.0093$, $\delta^2 = 21.9836$, $m = 30$. Cd2: $\gamma = 14.8762$, $\delta^2 = 18.9063$, $m = 28$. Cd3: $\gamma = 13.0072$, $\delta^2 = 13.7084$, $m = 28$
VMD with multi-reservoir ESN	No. decomposition layers = 6, population size = 20, maximum number of iterations = 100, iterative adjustment coefficient = 0.2, exponential factors = 2, 1.5	No. decomposition layers = 5, population size = 20, maximum number of iterations = 100, iterative adjustment coefficient = 0.2, exponential factors = 2, 1.5

Abbreviations: ARIMA, autoregressive integrated moving average; ESN, echo state network; SVM, support vector machine; VMD, variational mode decomposition.

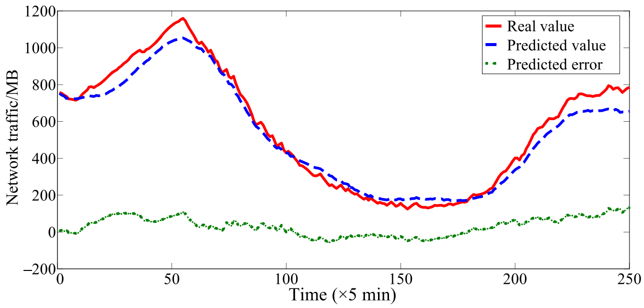


FIGURE 4 Prediction results of autoregressive moving average (ARMA) model on training set and corresponding prediction error for dataset B.

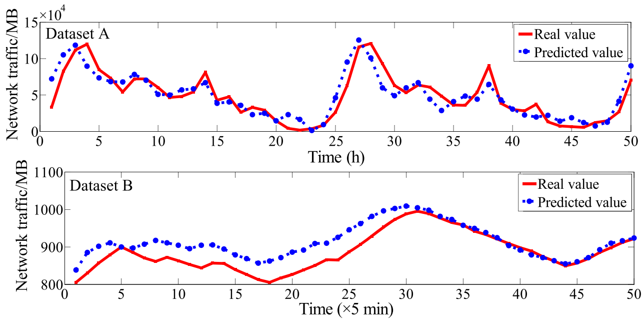


FIGURE 5 Prediction results of autoregressive moving average (ARMA) model on two datasets.

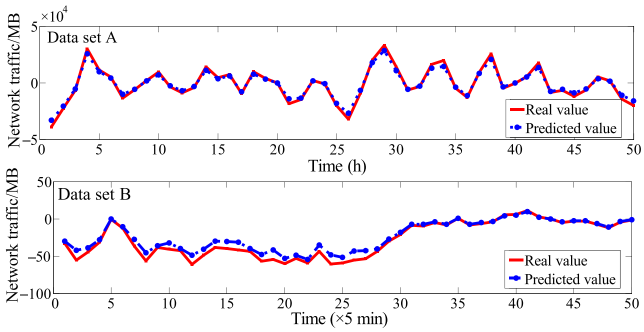


FIGURE 6 Prediction results of echo state network (ESN) on two datasets.

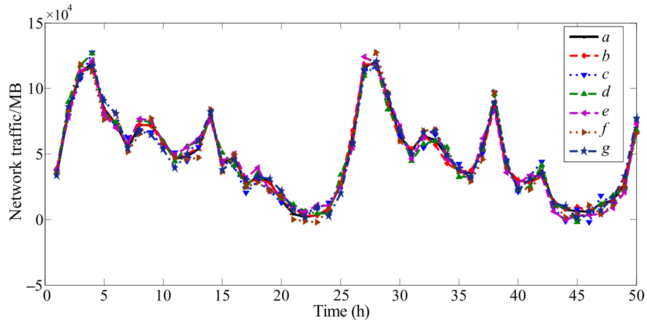


FIGURE 7 Results of prediction models for dataset A (*a*, real value; *b*, proposed model; *c*, autoregressive integrated moving average [ARIMA] model; *d*, support vector machine [SVM]; *e*, echo state network [ESN]; *f*, wavelet transform with multiple model fusion; g, variational mode decomposition [VMD] with multi-reservoir ESN).

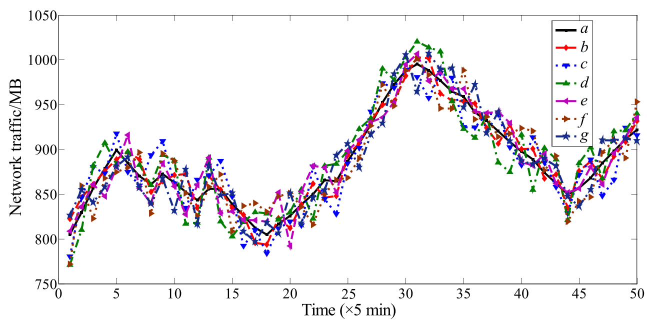


FIGURE 8 Results of prediction models for dataset B (*a*, real value; *b*, proposed model; *c*, autoregressive integrated moving average [ARIMA] model; *d*, support vector machine [SVM]; *e*, echo state network [ESN]; *f*, wavelet transform with multiple model fusion; g, variational mode decomposition [VMD] with multi-reservoir ESN).

characteristics of network traffic. Therefore, our model was superior to the comparison models.

Figures 9 and 10 show the distribution of prediction errors of the proposed and comparison models for datasets A and B, respectively. The prediction error of the proposed model was smaller, indicating its higher fitting ability.

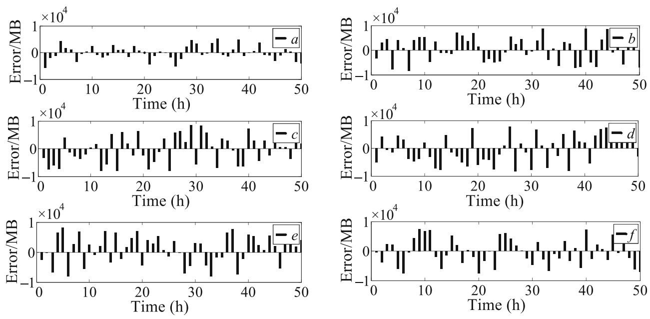


FIGURE 9 Prediction error of proposed and comparison models for dataset A (*a*, proposed model; *b*, autoregressive integrated moving average [ARIMA] model; *c*, support vector machine [SVM]; *d*, echo state network [ESN]; *e*, wavelet transform with multiple model fusion; *f*, variational mode decomposition [VMD] with multi-reservoir ESN).

Tables 2 and 3 list the performance indicators given by (17)–(24) of the evaluated prediction models for datasets A and B, respectively. The RMSE, MAE, MAPE, RRMSE, SSE, and TIC of the proposed model were smaller than those of the comparison models. In addition, the IA and R^2 value of the proposed model were closer to 1 compared with the other models. These results confirm the excellent performance of the proposed prediction model, indicating its high regression performance.

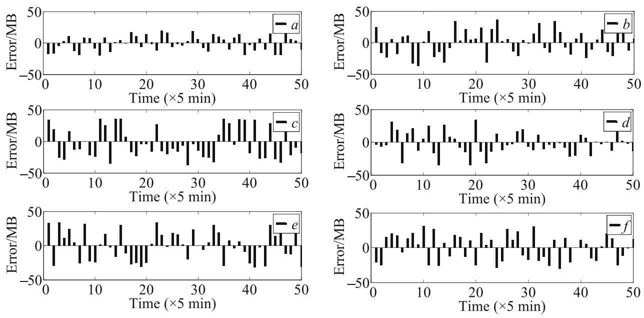


FIGURE 10 Prediction error of proposed and comparison models for dataset B (*a*, proposed model; *b*, autoregressive integrated moving average [ARIMA] model; *c*, support vector machine [SVM]; *d*, echo state network [ESN]; *e*, wavelet transform with multiple model fusion; *f*, variational mode decomposition [VMD] with multi-reservoir ESN).

Table 4 lists the results of the Wilcoxon signed-rank and rank-sum tests of the prediction models. The significance threshold was set at $p < 0.05$. The p-value of the Wilcoxon signed-rank test of the proposed prediction model was higher than that of the comparison models. The median difference between the predicted and real values was smaller for the proposed model than for the comparison models. Furthermore, the p-value of the Wilcoxon rank-sum test of the proposed model was greater than that of the comparison models. Thus, compared with the other models, the predicted value of the proposed model was closer to the average of the real values. The results of the two tests showed that the predicted values of the proposed prediction model were more consistent with the real network traffic.

Table 5 shows the computation times for training and prediction using the evaluated models for dataset A. The results were obtained using a computer equipped with an Intel(R) Core (TM) i5-6200U @ 2.30 GHz processor with 4 GB in memory. The proposed prediction

TABLE 2 Performance indicators of prediction models for dataset A.

Prediction model	RMSE (MB)	MAE (MB)	MAPE (%)	RRMSE	SSE (MB^2)	TIC	R^2	IA
Proposed model	2.5375e + 3	2.0013e + 3	9.2811	0.2086	0.3219e + 9	0.0220	0.9937	0.9994
ARIMA model	5.0117e + 3	4.4454e + 3	25.5006	0.5027	1.2558e + 9	0.0434	0.9753	0.9976
SVM	5.0500e + 3	4.4014e + 3	23.2335	0.5352	1.2751e + 9	0.0435	0.9749	0.9976
ESN	5.0982e + 3	4.2489e + 3	27.0833	0.5591	1.2996e + 9	0.0438	0.9744	0.9975
Wavelet transform with multiple model fusion	4.8205e + 3	4.1974e + 3	21.9443	0.4334	1.1619e + 9	0.0416	0.9771	0.9977
VMD with multi-reservoir ESN	4.3741e + 3	3.7007e + 3	18.1170	0.3618	0.9566e + 9	0.0377	0.9812	0.9982

Abbreviations: ARIMA, autoregressive integrated moving average; ESN, echo state network; IA, index of agreement; MAE, mean absolute error; MAPE, mean absolute percentile error; RMSE, root mean square error; RRMSE, relative RMSE; SSE, square sum error; SVM, support vector machine; TIC, Theil inequality coefficient; VMD, variational mode decomposition.

TABLE 3 Performance indicators of prediction models for dataset B.

Prediction model	RMSE (MB)	MAE (MB)	MAPE (%)	RRMSE	SSE (MB^2)	TIC	R^2	IA
Proposed model	11.7464	10.1857	1.1575	0.0134	0.6898e + 4	0.0066	0.9455	0.9993
ARIMA model	20.0600	17.0789	1.9417	0.0229	2.0120e + 4	0.0113	0.8411	0.9971
SVM	24.0732	13.6563	2.4636	0.0273	2.8976e + 4	0.0135	0.7711	0.9974
ESN	16.8391	13.6563	1.5580	0.0194	1.4178e + 4	0.0095	0.8880	0.9976
Wavelet transform with multiple model fusion	21.3097	18.3393	2.0970	0.0245	2.2705e + 4	0.0120	0.8206	0.9982
VMD with multi-reservoir ESN	18.5844	16.7293	1.8927	0.0210	1.7269e + 4	0.0105	0.8636	0.9985

Abbreviations: ARIMA, autoregressive integrated moving average; ESN, echo state network; IA, index of agreement; MAE, mean absolute error; MAPE, mean absolute percentile error; RMSE, root mean square error; RRMSE, relative RMSE; SSE, square sum error; SVM, support vector machine; TIC, Theil inequality coefficient; VMD, variational mode decomposition.

TABLE 4 Results of Wilcoxon signed-rank and rank-sum tests.

Prediction model	Signed-rank test (dataset A)	Rank-sum test (dataset A)	Signed-rank test (dataset B)	Rank-sum test (dataset B)
Proposed model	0.9064	0.9259	0.9211	0.9194
ARIMA model	0.6592	0.6698	0.7130	0.7767
SVM	0.7397	0.7972	0.7840	0.8266
ESN	0.8216	0.8288	0.8266	0.8485
Wavelet transform with multiple model fusion	0.8480	0.8577	0.8430	0.8753
VMD with multi-reservoir ESN	0.8662	0.8933	0.8769	0.8949

Abbreviations: ARIMA, autoregressive integrated moving average; ESN, echo state network; SVM, support vector machine; VMD, variational mode decomposition.

TABLE 5 Computation time for training and prediction using evaluated models (dataset A as an example).

Prediction model	Training time (s)	Prediction time (s)
Proposed model	165.3350	0.1032
ARIMA model	17.2268	0.0324
SVM	72.0364	0.0431
ESN	21.3629	0.0617
Wavelet transform with multiple model fusion	216.3642	0.2249
VMD with multi-reservoir ESN	244.0240	0.2038

Abbreviations: ARIMA, autoregressive integrated moving average; ESN, echo state network; SVM, support vector machine; VMD, variational mode decomposition.

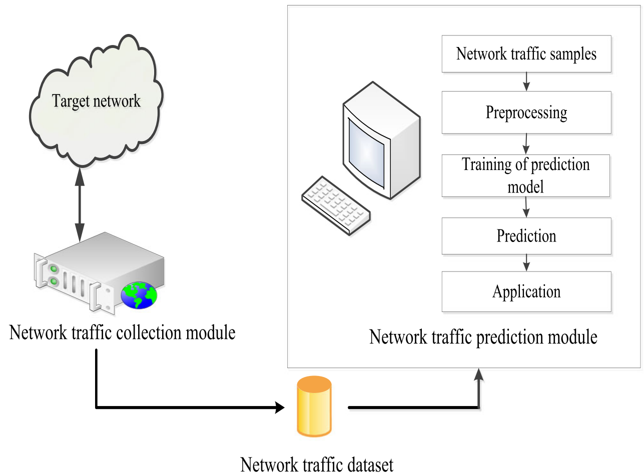


FIGURE 11 Diagram of network traffic prediction system.

model improved the overall prediction performance and accuracy, but the computation time was slightly increased compared with some comparison models. Nevertheless, considering the sampling time of common network traffic, the proposed prediction model can achieve real-time performance.

Comprehensively considering the evaluation results including the comparisons of fitting, prediction error, histogram distribution, and performance and statistical indicators, we can conclude that the proposed model is superior to other prediction models.

5.5 | Discussion

The proposed prediction model has various advantages. We can predict the network traffic by combining linear and nonlinear models. The linear ARMA model can predict linear sequences, whereas the ISMA-optimized ESN with a good fit determines nonlinear components. The combination of these two models improves the accuracy of network traffic prediction. Owing to model characteristics, separate linear and nonlinear models cannot guarantee a good prediction performance for network traffic that contains both linear and nonlinear components. When predicting network traffic, the proposed prediction model adequately extracts linear and nonlinear components of network traffic as well as valuable information from the error to achieve more accurate network traffic prediction. In contrast, existing models have neglected useful information contained in the prediction error. Therefore, combining linear and nonlinear models for network traffic prediction has important theoretical and practical value.

Although the proposed network traffic prediction model achieved good performance, the ESN was selected as the prediction model for the prediction error, and its parameters were optimized using the ISMA. As a model of the nonlinear prediction error, the ESN substantially influences the final prediction performance. The parameters optimized by the ISMA can improve the regression performance of the ESN, but the ESN reservoir is generated randomly. Consequently, a reservoir with the same number of neurons and spectral radius will show notable differences in performance owing to its internal structure and affect the nonlinear description of network traffic by the ESN.

Figure 11 illustrates the implementation of the network traffic prediction system in practice. This system comprises three main parts. The first part is the network traffic acquisition module, which collects network traffic from a target network using hardware or software tools. The second part is the traffic data storage module, which stores the collected network traffic in a dataset and retrieves it when necessary. The third part is the most important network traffic prediction module. First, network traffic samples are collected and preprocessed. Then, network traffic is predicted using the pretrained model. Finally, the prediction results are applied to practical scenarios such as congestion early warning and network optimization.

6 | CONCLUSIONS

To suitably plan and optimize a network, enable early warning of network problems, and guarantee security, stability, and normal operation of a network, network traffic should be predicted with high accuracy. Constructing an appropriate mathematical model and improving the prediction accuracy of network traffic are research hotspots in network management. We introduce a network traffic prediction model considering the combination of linear and nonlinear components. These components are treated differently, and the corresponding prediction models are established. An ARMA model predicts the linear components, and an ISMA-optimized ESN predicts the nonlinear components. The final prediction result is jointly determined by adding the prediction results of the ARMA model and ESN. The proposed prediction model was verified, and the prediction of real network traffic data provided good results. The proposed prediction model has theoretical support and practical applicability for improving network management and performance indicators. In future work, we will focus on improving the structure of the ESN reservoir and strengthening the connectivity of neurons in the reservoir to further improve the fitting ability for nonlinear data.

ACKNOWLEDGMENTS

The author would like to thank the editor and all anonymous reviewers for their helpful feedback.

CONFLICT OF INTEREST STATEMENT

The author declares that there are no conflicts of interest.

ORCID

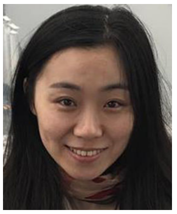
Lian Lian https://orcid.org/0000-0002-5093-2450

REFERENCES

1. S. Izadi, M. Ahmadi, and A. Rajabzadeh, *Network traffic classification using deep learning networks and Bayesian data fusion*, J. Netw. Syst. Manag. **30** (2022), no. 2, 25.
2. Z. D. Tian and F. H. Li, *Network traffic prediction method based on autoregressive integrated moving average and adaptive Volterra filter*, Int. J. Commun. Syst. **34** (2021), no. 12, e4891.
3. Z. D. Tian, *Chaotic characteristic analysis of network traffic time series at different time scales*, Chaos Solitons Fractals **130** (2020), 109412.
4. K. Zhou, W. Wang, L. Huang, and B. Liu, *Comparative study on the time series forecasting of web traffic based on statistical model and generative adversarial model*, Knowl. Based Syst. **213** (2021), 106467.
5. Q. T. Tran, L. Hao, and Q. K. Trinh, *A comprehensive research on exponential smoothing methods in modeling and forecasting cellular traffic*, Concurrency Comput. Pract. Exper. **32** (2020), no. 23, e5602.
6. U. Premaratne and U. S. Premarathne, *A sum of Bernoulli sources approximation for packet switched network traffic in backbone links*, IEEE Commun. Lett. **24** (2020), no. 1, 141–145.
7. A. Domański, J. Domańska, K. Filus, J. Szyguła, and T. Czachórski, *Self-similar Markovian sources*, Appl. Sci. Basel **10** (2020), no. 11, 3727.
8. Y. Xie, J. Hu, Y. Xiang, S. Yu, S. Tang, and Y. Wang, *Modeling oscillation behavior of network traffic by nested hidden Markov model with variable state-duration*, IEEE Trans. Parallel Distrib. Syst. **24** (2013), no. 9, 1807–1817.
9. M. Laner, P. Svoboda, and M. Rupp, *Parsimonious fitting of long-range dependent network traffic using ARMA models*, IEEE Commun. Lett. **17** (2013), no. 12, 2368–2371.
10. Q. Yu, L. Jibin, and L. R. Jiang, *An improved ARIMA-based traffic anomaly detection algorithm for wireless sensor networks*, Int. J. Distrib. Sens. Netw. **12** (2016), no. 1, 9653230.
11. S. Dong, *Multi class SVM algorithm with active learning for network traffic classification*, Expert Syst. Appl. **176** (2021), 114885.
12. J. X. Liu and Z. H. Jia, *Telecommunication traffic prediction based on improved LSSVM*, Int. J. Pattern Recognit. Artif. Intell. **32** (2018), no. 3, 1850007.
13. V. K. Chauhan, K. Dahiya, and A. Sharma, *Problem formulations and solvers in linear SVM: a review*, Artif. Intell. Rev. **52** (2019), no. 2, 803–855.
14. J. Zhou, H. Wang, F. Xiao, X. Yan, and L. Sun, *Network traffic prediction method based on echo state network with adaptive reservoir*, Softw. Pract. Exper. **51** (2021), 2238–2251.
15. X. L. Zheng, W. Lai, H. Chen, S. Fang, and Z. Li, *A study of cellular traffic data prediction by kernel ELM with parameter optimization*, Appl. Sci. Basel **10** (2020), no. 10, 3517.
16. Y. Hou, L. Zhao, and H. W. Lu, *Fuzzy neural network optimization and network traffic forecasting based on improved differential evolution*, Internat. J. Engrg. Sci. **81** (2018), 425–432.
17. D. F. Wei, *Network traffic prediction based on RBF neural network optimized by improved gravitation search algorithm*, Neural Comput. Appl. **28** (2017), no. 8, 2303–2312.
18. K. T. Selvi and R. Thamilselvan, *An intelligent traffic prediction framework for 5G network using SDN and fusion learning*, Peer-to-Peer Netw. Appl. **10** (2022), 7003–7015.

19. M. Emec and M. H. Ozcanhan, *A hybrid deep learning approach for intrusion detection in IoT networks*, Adv. Electr. Comput. Eng. **22** (2022), no. 1, 3–12.
20. Y. Han, Y. W. Jing, and G. M. Dimirovski, *An improved fruit fly algorithm-unscented Kalman filter-echo state network method for time series prediction of the network traffic data with noises*, Trans. Inst. Meas. Control. **42** (2020), no. 7, 1281–1293.
21. X. C. Tan, H. Ma, S. Yu, and J. Hu, *Recognizing the content types of network traffic based on a hybrid DNN-HMM model*, J. Netw. Comput. Appl. **142** (2019), 51–62.
22. M. L. Yuan, *Jitter buffer control algorithm and simulation based on network traffic prediction*, Int. J. Wirel. Inf. Netw. **26** (2019), no. 3, 133–142.
23. Z. D. Tian, *Network traffic prediction method based on wavelet transform and multiple models fusion*, Int. J. Commun. Syst. **33** (2020), no. 11, e4415.
24. L. Lian and Z. D. Tian, *Network traffic prediction model based on ensemble empirical mode decomposition and multiple models*, Int. J. Commun. Syst. **34** (2021), no. 17, e4966.
25. Y. Han, Y. Jing, K. Li, and G. M. Dimirovski, *Network traffic prediction using variational mode decomposition and multi-reservoirs echo state network*, IEEE Access **7** (2019), 138364–138377.
26. L. L. Ren, A. A. Heidari, Z. Cai, Q. Shao, G. Liang, H.-L. Chen, and Z. Pan, *Gaussian kernel probability-driven slime mould algorithm with new movement mechanism for multi-level image segmentation*, Measurement **192** (2022), 110884.

AUTHOR BIOGRAPHY



Lian Lian received a PhD degree in Control Theory and Control Engineering from Northeastern University, China, in 2016. She is currently a Lecturer at the College of Information Engineering, Shenyang University of Chemical Technology, China. Her research interests include time series prediction and swarm intelligence optimization.

How to cite this article: L. Lian, *Network traffic prediction model based on linear and nonlinear model combination*, ETRI Journal **46** (2024), 461–472. DOI 10.4218/etrij.2023-0136