Optimization and Decomposition Methods in Network Traffic Prediction Model: A Review and Discussion

JINMEI SHI[1,2], YU-BENG LEAU[1], (Member, IEEE), KUN LI[3], YONG-JIN PARK[1], (Life Senior Member, IEEE), AND ZHIWEI YAN[4]

[1]Faculty of Computing and Informatics, Universiti Malaysia Sabah, Kota Kinabalu 88400, Malaysia
[2]College of Information Engineering, Hainan Vocational University of Science and Technology, Haikou 571126, China
[3]College of Engineering, Bohai University, Jinzhou 121013, China
[4]National Engineering Laboratory for Next Generation Internet Interconnection Devices, CNNIC, Beijing Jiaotong University, Beijing 100044, China

Corresponding author: Yu-Beng Leau (lybeng@ums.edu.my)

ABSTRACT The 21st century is a high-tech information era in which our lives are closely linked by computer networks. Hence, how to effectively supervise networks and reduce the frequency of network security incidents has now become a research hotspot in cyberspace. Specifically, researchers have shown an increased interest in predicting the network traffic before any untoward incident happens. Optimization and decomposition technologies are the core components of network traffic prediction model which plays an important role in network management. This article discusses past network traffic prediction research and critically examines the optimization and decomposition technologies used in the model, lists the model parameter structure based on the research methodology, the data set used, the evaluation criteria and so on. By comparison, digging out the Particle Swarm Optimization (PSO) algorithm and the Variational Mode Decomposition (VMD) decomposition technique will effectively solve the network traffic model predictive difficulties that have proven to be crucial to improving predictive accuracy and convergence speed strategy.The comprehensive review reveals that PSO and VMD are the most suitable optimization algorithm and decomposition technology for network traffic prediction modeling.

INDEX TERMS Decomposition technology, network traffic prediction, optimization algorithm, particle swarm optimization, variational mode decomposition.

I. INTRODUCTION

The huge tide of the Internet is driving the rapid development of society, whereby computer network has become an important technical means of the information society. To ensure network service quality, the increasing network traffic has triggered an alarm for network monitoring. These rapid changes have serious effects so much so that the detection of traffic abnormalities has become harder. A single detection can no longer provide a reliable network for an organization. Instead, prevention/intervention should be foreseen before abnormal events occur. In network security monitoring, network traffic is an important parameter to evaluate the running state of the network [1]. Normal network traffic indicates the ordered and safety of the network operation. On the contrary, malicious and abnormal network traffic

represents the network failure and abnormal behavior which call for preventive measures. Given the fact that preventing an incident requires accurate analysis and prediction, network security communities are constantly vigilant and monitors the incoming traffic in their networks even before taking any precautions. Therefore, a reliable network traffic prediction mechanism is desired to increase the level of security in a network communication [2].

Network traffic prediction is a process which analyzes the characteristics of traffics in the past and present. It uses optimization algorithm and decomposition technique to obtain the rules of internal structure, conducts modelling and produces the result in order to forecast feature and future trends of network traffic [3]. Recently, many researchers have focused on designing networktraffic prediction modelling and simulation which can characterize and simulate the traffic rules of internal structure. However, due to too many parameters and too large computational complexity,etc., their

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

model is difficult to converge or falls into local optimal solutions, their findings show that their models are inadequate to reveal subtle internal features of the traffic, unable to predict and take preventive measures accurately and quickly. Traffic modelling and prediction with self-similarity, chaos and mutability have serious effects on the network. Therefore, how to precisely eliminate interference and predict multi-scale characteristics of network traffic in a prediction model has become a worthy research topic.

This article first briefly states the importance of network traffic prediction. Then, it proceeds with reviewing the optimization algorithm in the network traffic prediction mechanism. The third section of the paper discusses the decomposition technology in the traffic time series model. The paper concludes by reiterating the significant role of optimization and decomposition methods in network traffic prediction model.

II. OPTIMIZATION ALGORITHM

In an attempt to better describe the traffic characteristics and improve the accuracy of the prediction model, effective network protection measures, prediction model and its key should be taken in advance. The core of solving this problem lies in the selection and improvement of optimization algorithm, researchers have tended to adopt various optimization algorithms to optimize the model. This article reviews some of these algorithms namely Genetic Algorithm (GA), Quantum Genetic Algorithm (QGA), Fruit Fly Optimization Algorithm (FOA) and Particle Swarm Optimization (PSO), which is shown in figure 1.
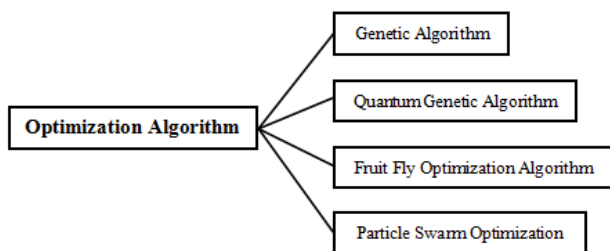

FIGURE 1. Optimization algorithm in network traffic prediction model.

A. GENETIC ALGORITHM

Genetic Algorithm (GA) is a search heuristic algorithm combining genetic and evolutionary computing. It was developed by an American professor, J. Holland, [4] in 1975. In 1989, Goldberg's work made a comprehensive and systematic summary and discussion of the genetic algorithm which laid the foundation for the genetic algorithm which based on the nature of the pattern of change [43]. The purpose of GA is to compare many different individual solutions taken from a population and select the one that fits the data best in terms of a characteristic value. Selection simulates the natural 'law of survival of the fittest.

For the event and the data set contained in the event, Genetic algorithm first collects the characteristic quantity of the data model, and then finds the best individual. The process of genetic algorithm is as follows:

(1) The data of the initial population is coded to determine its code length, and the initial population is determined by the random number group output.

(2) Calculate the fitness functions of the initial population to determine whether the current requirements are satisfied. If they are, proceed to step (4). Otherwise, go to step (3).

(3) The population in the dataset is selected for replication, crossover and variation, and then output the next generation population. At this time, the number of iterations increases by one, and the genetic algebra increases by one. Then the fitness function of the new population is calculated to determine whether the current requirements are met. If they are, go to step. (4)Otherwise, proceed to step (3).

(4) Return the current optimal individual and complete the entire process.

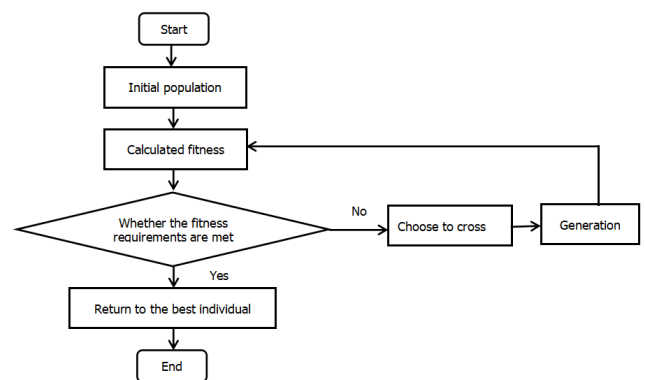The process as discussed above is encapsulated in figure 2.

FIGURE 2. Flow chart of Genetic algorithm.

Genetic algorithm can optimize the local optimal solution situation through the optimization process to get better algorithm performance. It is usually used to improve the ability of data processing. However, the algorithm's ability to explore new space is limited and it is easy to converge to the local optimal solution. Moreover, the computation time is a problem when the problem is complex, involving a large number of individuals. Despite that, researchers in recent years have successfully developed a quantum genetic algorithm based on genetic algorithm which does not only address the shortcomings of genetic algorithm but also produces good performance.

B. QUANTUM-INSPIRED GENETIC ALGORITHM

Quantum Genetic Algorithm (QGA) is a combination of quantum computing and genetic algorithm. This algorithm was proposed by M. Moore and Narayannan from Exeter University in 1996 [5]. QGA combines quantum bits of data and processes the data through quantum chromosomes which is the classical theory of genetic algorithm. New quantum

chromosomes such as crossover and mutation were adopted for relevant processing and analysis so as to ensure that the rules in the data could be well observed so as to finally obtain efficient system model and data with high prediction accuracy. Utilizing the quantum algorithm to solve the gradient explosion problem of BP, Kun Zhang proposed the QPSO-BP prediction model [6]. On the other hand, Ying Han adopted Quantum mechanics theory to optimize five important parameters of ESN, and proposed QFOA-ESN model mechanism to provide model accuracy [7].

The steps of quantum genetic algorithm are mainly divided into three steps:

Step 1: Initialize. The feature quantity corresponding to the initial data is loaded, and then the feature quantity is extracted.

Step 2: Measure. Measuring each chromosome in population $Q(t)$ and analyze the probability of different genes to obtain the corresponding definite solution which called $P(t)$.

Step 3: Evolution. Select, cross and mutate to determine whether the entire algorithm process meets the termination condition, and return the current best individual if it meets the condition; otherwise, carry out continuous iterative calculation until the corresponding condition is met.

Pseudo-code of quantum genetic algorithm is shown in Table 1.

TABLE 1. Pseudo-Code of quantum genetic algorithm.

Step1	Initialize $Q(t)$; $t = 0$
	Make P(t) by observing Q(t) states;
	Evaluate P(t)；store the best solution among P(t);
Step2	While (not termination-condition) do
	1）$t = t + 1$;
	2）Make P(t)by observing Q(t − 1)states;
	3）Evaluate P(t);
Step3	4）Update $Q(t)$ using quantum revolving gates $U(t)$
	5）Return the best solution among P(t)；
	End

The fundamental parts of quantum genetic algorithms are quantum coding and quantum gates. Quantum coding itself is combined with quantum bits, which can construct quantum coding for the corresponding dataset or multiple individuals. It also pays attention to the whole search problem so as to solve the local optimal solution. Quantum gate makes the algorithm more efficient, has better data results and global characteristic optimization, and also increases the possibility of data variation so that the data utilization scale is further amplified. It can be seen that the quantum genetic algorithm has the ability to search all spatial solutions and has fast convergence speed and high performance. Moreover, this algorithm can prevent getting into the situation of local

optimal solution. One weakness, however, it's very heavy computation [8]. Hui Tian used the Quantum Genetic Algorithm (QGA) with efficient global search capability to model and predict the network traffic, then proposes QGA-BP hybrid model to predict the development trend of network traffic. However, the model ignores the problem of the QGA slide rule, which increases the complexity of the hybrid model, resulting in lower generalization performance and lower convergence rate [36].

C. FRUIT FLY OPTIMATION ALGORITHM

Fruit Fly Optimization Algorithm (FOA) is an intelligent algorithm proposed by Dr. Pan Wenchao in 2011 [9]. Fruit Fly Optimization Algorithm, with its sensitive sensory system, can fly to the target source quickly and efficiently. Its sense of smell can not only search different smells, but also can search dozens of kilometres away. It can smell the approximate location of food. After a blind flight for a period of time, it can closely observe the surrounding individuals, determine the exact location of food according to the information source and vision, and then fly to target location.

Generally speaking, when looking for food from (0,0), individual fruit flies first use their own olfactory organs to smell the smell of food and send smell information to the surrounding fruit flies, or receive smell information from the surrounding fruit flies (X,Y).Then the flies used their visual organs to compare the location of the flies that collected the best odor information in the current group (X2,Y2). The other flies in the group all flew to this location and continued to search.The process of Fruit Fly searching for target food is shown in figure 3.
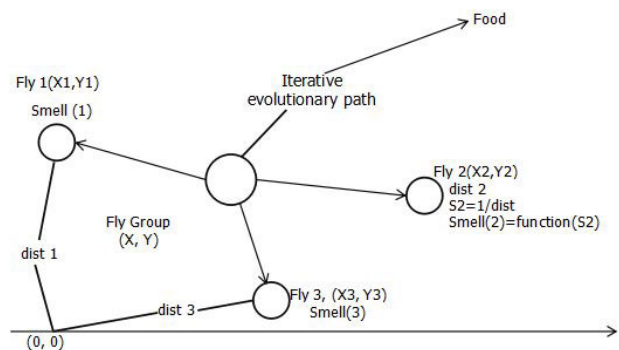

FIGURE 3. The process of Fruit Fly searching for target food.

Fruit fly searches for food in two broad stages:

(1) Blind flight stage: at a distance from the food source, the fly uses its unique olfactory advantage to quickly locate the general location of the fly, and then fly to the food source.

(2) Precise flight stage: after the previous stage of flight, the fruit fly has been in the vicinity of the food source. At this stage, in addition to observing the surrounding population to obtain information, the most important thing for it to do is to directly fly by means of sharp visual positioning.

The three main features of fly flight are as follows:

(1) Feature 1: fruit fly goes near the group position.

(2) Feature 2: fruit fly according to odor concentration and the whole population, flies to individuals with higher odor concentration.

(3) Feature 3: fruit fly population as a whole is close to the food source.

According to the food search stage and flight feature of fruit fly, the final fruit fly algorithm has four parameters, including initial value, iteration step length, sizepop and maxgen. The following is a detailed introduction of the influence of each parameter on the algorithm.

(1)The initial value: it is generated by a random function, and the random() function generates the population diversity that determines the entire fruit fly's ability to search and whether it can find the target food.

(2) Iteration step length: the iteration step length of fruit fly runs through the whole process of the whole algorithm search. When the iteration step length of fruit fly is relatively long, each flight distance is relatively long. Accordingly the whole fruit fly search process will be accelerated. However, it is easy to miss the global optimization in the process of flight and fall into the local optimization. When the step size is smaller, fruit fly flies slowly, but it is not easy to fall into local optimization. At this time, we should choose a proper iterative step size so that it can find the global optimal value with high flying efficiency of drosophila.

(3)The sizepop: the number of individual fruit fly searching for food is related to the size of the fruit fly searching ability. If the number of individual fruit fly is larger, each individual is making its own contribution to the group, then the target food should be located more quickly. However, if the number of individual is large, the memory and running time will be affected. The selection of individuals should also be appropriate when using the fruit fly algorithm.

(4) The maxgen: obviously, the more iterations, the longer the running time is. However, in the case of fewer iterations, it is possible that fruit fly has not yet reached the target source. Therefore, the selection of the number of iterations will also affect the performance of the fruit fly algorithm.

Based on the above analysis, the FOA algorithm has a faster convergence speed and does not rely on manual experience. Using the FOA method, Ying Han managed to solve and optimize the four parameters of ESN. He eventually proposed a network traffic combined model which improves the prediction accuracy [10]. Unfortunately, in the iterative process, it is easy to fall into the ''premature'' linearity and local optimal solution, which can in turn affect the optimization performance of the algorithm.

D. PARTICLE SWARM OPTIMIZATION

Particle Swarm Optimization (PSO) was proposed by Kennedy and Eberhart in 1995 [11]. PSO is a heuristic search algorithm for natural clustering such as clustering the behaviors of birds and insects. This algorithm is similar to genetic algorithm. It is based on the mutual transmission, cooperation and competition among individuals of information, and seeks the optimal value in the global space. The PSO algorithm starts from the initial random solution, that is, the particle is iterated from the beginning, and the optimal solution is obtained by changing the direction, speed and position of the search. In basic particle swarm optimization, particle updating its state is usually based on three rules: maintaining inertia; changing the position and state of the optimized particles through experience; and referring to the experience of other particles, the static change of the global optimal position. Tao He put forward the PSO-RBF combined model based on optimized PSO algorithm to realize traffic prediction [12]. Fei Han used the adaptive PSO algorithm to optimize ELM network which led to the APSO-ELM hybrid model [13]. Yi Yang utilized PSO to optimize the two parameters of Least Squares Support Vector Machines(LSSVM) to predict the network traffic [14].

1) ALGORITHM PRINCIPLE OF PSO

Suppose in a search space with dimension D, the population $X_i = (X_1, X_2, \ldots, X_n)$ is made up of n particles. Suppose the ith particle can be represented as a D-dimensional vector $X_i = (X_{i1}, X_{i2}, \ldots, X_{iD})^T$, which represents the position of the ith particle in the D-dimensional search space, is a potential solution to the problem. Once the objective function is determined, we can calculate the fitness value corresponding to the position X_i of each particle according to its formula. The velocity of the ith particle is $V_i = (V_{i1}, V_{i2}, \ldots, V_{iD})^T$, whose individual extreme value is $P_i = (P_{i1}, P_{i2}, \ldots, P_{iD})^T$, the population extreme value of the population is $P_g = (P_{g1}, P_{g2}, \ldots, P_{gD})^T$.

In each iteration of the algorithm, the particle will adjust its position and speed according to the size of individual extreme value and population extreme value, as shown in formula (1) and (2).

$$V_{id}^{k+1} = \omega V_{id}^k + c_1 r_1 \left(P_{id}^k - X_{id}^k \right) + c_2 r_2 \left(P_{gd}^k - X_{id}^k \right) \quad (1)$$

$$X_{id}^{k+1} = X_{id}^k + V_{id}^{k+1} \quad (2)$$

where, ω is the inertia weight; $d = 1, 2, \ldots, D$; $i = 1, 2, \ldots, n$; k is the number of iterations; V_{id} is the velocity of the particle swarm; c_1 and c_2 are constants that greater than zero ($c_1 \geq 0, c_2 \geq 0$), defined as the acceleration factor; r_1 and r_2 are random numbers distributed in the interval of [0,1]. In order to prevent the particle from searching blindly, it is generally recommended to limit its position and speed to a certain range $[-X_{max}, CX_{max}]$ and $[-V_{max}, CV_{max}]$.

2) BASIC FLOW OF PSO ALGORITHM

Particle swarm optimization algorithm (PSO) is often used to find the optimal solution between the maximum and minimum optimization problems. The general processing steps are as follows:

Step 1: Initialize the particle swarm (set the population size as M);

Step 2: Calculate the fitness value of the particle;

Step 3: Compare the fitness value of each particle with the best position P_{best} that it passes through. If the fitness value of the particle is good, the best position P_{best} of the particle will be replaced by the particle with the current fitness value.

Step 4: For each particle, compare its fitness value with the best position G_{best} in the whole population. If the fitness value is good, the best position G_{best} in the current population will be replaced by the particle with the current fitness value.

Step 5: Adjust the velocity and position of particles according to formulae (1) and (2);

Step 6: If the end condition is not met, go to step 5.

To summarize these steps, a flow chart of particle swarm optimization algorithm is depicted in figure 4.
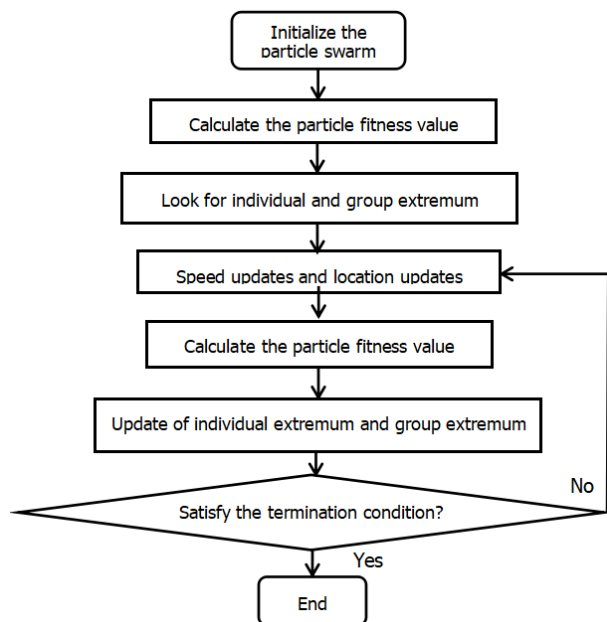

FIGURE 4. Flow chart of particle swarm optimization algorithm.

The analysis based on PSO algorithm shows that the learning factor, the elasticity coefficient of velocity update and population update and the parameter iteration velocity coefficient are all adjustable coefficients in PSO. Moreover, the principle is simple and easy to realize. Weijie Zhang improved the PSO algorithm to optimize the three parameters of RBF as well as proposed the network traffic combination prediction model. By adjusting the inertia weight and learning factor to improve the global search ability in the global extremum search, he was able to solve and avoid local optimal solution of PSO. Then he optimized the four parameters of the model network so as to obtain the accuracy of the prediction model. However, it is easy to fall into the local optimal solution in the iteration process [15].

Table 2 summarizes the strengths and limitations of the optimization algorithms. The application of optimization algorithm in network traffic prediction model is also analyzed and compared as shown in table 3.

TABLE 2. Strengths and limitations of optimization algorithms.

NO.	Year	Author	Name	Strengths	Limitations
1	1975 1989	Holland[4] Goldberg [43]	Genetic Algorithm (GA)	Good data processing ability; optimize the local optimal solution.	Low efficiency and high overhead.
2	1995	Kennedy, Eberhart [11]	Particle Swarm Optimization (PSO)	High identification, few parameters, simple principle and converges quickly.	Local optimal solution.
3	1996	Moore, Narayannan [5]	Quantum Genetic Algorithm (QGA)	Global optimization, fast speed, and high performance	The calculation is large and complex
4	2011	Wenchao Pan[9]	Fruit Fly Optimization Algorithm (FOA)	Fast convergence; cannot rely on manual experience.	Local optimal solution and "early maturity".

Based on the comparison table above, it can be surmised that with PSO algorithm it is easy to fall into the problem of local optimal solution in the iterative process. Yet it has the advantages of very fast convergence, few parameters, simple principle and easy implementation, PSO also has a very good ability to identify network traffic characteristics [12].

Compared with other optimization algorithms, PSO is found to be feasible and effective for constructing traffic prediction model. This in turn lays a good foundation for follow-up research and algorithm optimization. At the same time, how to solve the local optimal solution problem of PSO will be another key issue as well as an optimization strategy for constructing the prediction model is much needed. In the future research, we will design the algorithm to solve the problem of PSO local optimal solution, and then build the network traffic prediction mechanism.

III. DECOMPOSITION TECHNIQUE
Network traffic has the characteristics of nonlinearity and nonstationarity such as suddenness, self-similarity, multifractal, chaos and so on. Time series analysis has been found to be an effective method to solve the nonlinearity and non-stationarity characteristics of traffic sequence [42] as it provides a modeling idea based on time frequency analysis for traffic analysis. Hence, researchers have introduced time-frequency analysis into traffic law analysis by applying the signal analysis theory to traffic time series analysis.

TABLE 3. Application of optimization algorithm in network traffic prediction model.

TABLE 3. (Continued.) Application of optimization algorithm in network traffic prediction model.

NO.	Year	Author	Original model	Combined Model	Practical	Theoretical	Data set Used	Evaluation Criteria	Strength	Limitation
1	2013	Kun Zhang et al.[6]	Particle Swarm Optimization / Quantum / BP Neural Network	QPSO-BP	√		Monitor the traffic between clients in our campus network and servers	RMSE\MSE\MAPE	Use quantum algorithm to optimize PSO which solves BP gradient explosion.	Focus on model optimization, but the quantum algorithm is difficult to calculate; easy to fall into the local optimal solution.
2	2013	Fei Han et al.[13]	Particle Swarm Optimization / More-Penrose generalized inverse algorithm / Extreme Learning Machine	APSO-ELM		√	—	RMSE	1. Optimize the input weight and deviation of ELM with APSO. 2.Solve the existing problems of ELM and PSO's local optimal solution.	In the modeling, adaptive algorithm and More-Penrose generalized inverse algorithm were adopted, which increased the computational complexity and reduced the response speed of the model. Thus, it will be impossible to find the optimal solution.
3	2014	Yi Yang et al.[14]	Seasonal Transform / Particle Swarm Optimization / Least Squares Support Vector Machine	SA-PSO-LSSVM	√		Lanzhou University data set	MAE\MAPE	1. Sequence elimination by SA reduces the interference of seasons on component nts; 2. Use PSO to optimize two parameters of LSSVM and provided prediction accuracy.	1. With the increase of network nodes, the performance of the model decreases. This makes it suitable only for small-scale networks. 2. The model only considers the seasonal characteristics of traffic.
4	2016	Tao He et al.[12]	Radial Basis Function Neural Network / Particle Swarm Optimization	PSO-RBF	√		Iris data set	MSE	Improve the performance of RBF network structure; Based on the use of PSO, it optimizes the weight, width and center position parameters.	PSO iteratively optimizes the three parameter processes of RBF, which makes it difficult to approach the global optimal.
5	2017	Dengfeng Wei et al.[16]	Gravity search algorithm / Radial Basis Function Neural Network / Particle Swarm Optimization	IGSA-RBF	√		Lorenz time series	RMSE	1.Adopted PSO to improve the speed selection formula of GSA; 2.Three parameters of RBF were optimized by PSO.	GSA lacks theoretical guidance; RBF optimization parameters are too many; the structure is complex, and occur local optimal solution problem.
6	2017	Ying Han et al.[7]	Echo State Network / Fruit Optimization Algorithm / Quantum	QFOA-ESN	√		BC-Oct89Ext4	RMSE\MAE	Use quantum mechanics to optimize FOA, and then use QFOA to optimize 5 important parameters of the ESN model and provide model accuracy.	It can deal with the optimization of FOA algorithm well, but it has too many optimization parameters and too much computation, which affects the convergence rate.
7	2017	Ying Han et al.[10]	Echo State Network / Black Hole / Fruit Fly Optimization Algorithm / Opposition-Based Learning	OBL-FOA-ESN	√		Mackey-Glass chaotic time series	RMSE\MAE	Adopt OBL optimize FOA to become OBLFOA, which was used to optimize 4 important parameters of ESN model to provide model accuracy.	The optimization of FOA algorithm is well handled, but the optimization parameters are too many;the calculation is large, the structure is complex, and the convergence speed is affected.

TABLE 3. *(Continued.)* Application of optimization algorithm in network traffic prediction model.

8	2018	Weijie Zhang et al. [15]	Radial Basis Function Neural Network	IPSO-RBF	√	CIS-CO6509's related nodes based on SNMP protocol	RMSE	Using mutation method to avoid local minimum problem of PSO and then three parameters of RBF were optimized by PSO.	It is difficult to approach the global optimal solution due to too many parameters when using IPSO to optimize RBF.
			Particle Swarm Optimization						
			Adjusting inertia weight and learning factor method						

Time-frequency analysis is to treat the traffic sequence as a signal sequence formed by the accumulation of sinusoidal wave elements with different periods. Firstly, appropriate technologies are adopted for identification and decomposition. Then, by analyzing the subtle features contained in the signal sequence, the internal variation rules of the traffic sequence are mastered. Finally, the trend and characteristics of the future traffic are predicted. Wavelet Analysis and Modal Decomposition are effective applications of time-frequency analysis. In recent years, Wavelet Analysis and Modal Decomposition have been widely used in traffic sequence prediction modeling, and they are effective tools for signal decomposition. This article will first analyze several common signal decomposition technologies such as Wavelet Analysis, Empirical Mode Decomposition(EMD), Ensemble Empirical Mode Decomposition(EEMD) and Variational Mode Decomposition(VMD). It then proceeds to seek a strong scientific and technological basis for sequence signal decomposition, as shown in figure 5.
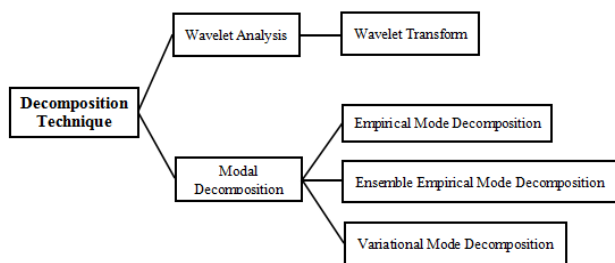

FIGURE 5. Decomposition techniques in network traffi prediction model.

A. WAVELET TRANSFORM

Wavelet Analysis (WA) and its application are the product of the development of Fourier to a certain stage, and are a relatively common Analysis method in signal processing also known as Wavelet Transform (WT). It was developed by French engineer, Goupillaud *et al.* [17] in 1984. In 1986, a famous mathematician, Y.Meyer, cooperated with Mallat [18] to establish a unified method for constructing wavelet bases Wavelet transform mainly changes the signal by changing the wavelet function. The wavelet function has the characteristics of fast fading and continuous oscillation which is similar to the Fourier transform. Through this processing method the characteristics of the signal in the time domain and frequency domain can be determined. Hence the signal can be analyzed through these characteristics. In recent years, wavelet transform has become one of the most powerful signal processing tools in non-stationary signal decomposition and has been widely applied in the field of network traffic prediction [19], [20].

Assuming that in any space $L^2(R)$, F (t) function passes through wavelet basis expansion, we call this expansion as continuous wavelet transformation, and the expression is shown in formula (3).

$$WT_x(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t) \Psi\left(\frac{t - b}{a}\right) dt = \langle x(t), \Psi_{a,b}(t)\rangle \tag{3}$$

In the above formulas, the most important one is that the wavelet basis has two parameters of scale a and translation b. Because of the definition of continuous wavelet transform, wavelet transform is also an integral transform. In this case, we call $WT_x(a, b)$ the wavelet transform coefficient.

If a and b in formula (4) are discretized simultaneously, and $a = a_0^i$, $b = ka_0^i b_0$, the discrete wavelet transform of x (t) can be obtained, as shown in formula 4).

$$WT_x(j, k) = \int_{-\infty}^{+\infty} x(t) \Psi_{j,k}(t) dt, \ Cj, k \in Z \tag{4}$$

This transform not only saves the computation time, but also reduces the information redundancy after the continuous wavelet transform.

With the development of research, Mallat *et al.* proposed a fast algorithm for calculating discrete orthogonal wavelet transform, namely Mallet algorithm [18], based on the analysis of multi-resolution signals. The principle is shown in figure 6.
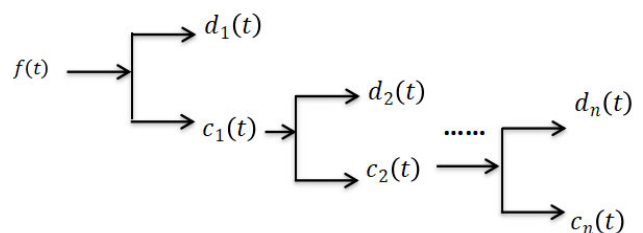

FIGURE 6. Schematic diagram of the signal sequence of Mallet algorithm analysis.

As can be seen from figure 6, the Mallet algorithm decomposes the low-frequency part, and the high-frequency part is obtained from the low-frequency part, and is not considered

after the decomposition. The decomposition formula is indicated in formula (5).

$$x(t) = \sum_{j-1}^{n} d_j(t) + c_n(t) \tag{5}$$

WT can analyze the characteristics of signal in time domain and frequency domain and ensure certain accuracy when obtaining signal characteristics [21]. At the same time, the scale parameters and multiresolution of wavelet transform are the key of signal decomposition.

Recent research has reported that the wavelet transform was confused by decomposition and was not suitable for high-frequency signal analysis [22]. At the same time, due to the limitation of wavelet basis length, after scaling decomposition, the result of wavelet transform will be a signal at a certain scale. Further, its frequency component is only related to the sampling frequency, but has nothing to do with the signal itself, thus making it difficult to decompose the flow signal [23].

B. EMPIRICAL MODE DECOMPOSITION

Empirical mode decomposition (EMD) is the first Modal Decomposition analysis method, which is a new method of signal processing. This method was proposed by Huang, N E *et al.* in 1998 [24].The essence of EMD is to smooth the time series, decompose the original signal into several Intrinsic Mode functions (IMF), and determine the intrinsic characteristics of effective signals in data according to experience. It can more effectively reflect the spatial (or temporal) scale distribution of energy in physical processes. By EMD decomposition, the non-stationary complex signals of each IMF component are stable. Thus, EMD is an effective method to decompose non-stationary signals [25].

EMD decomposes the original sequence into several subsequence components (IMF). IMF components. After decomposition, the IMF not only have the internal characteristics of the original signal, but also show stronger regularity and stability than the original signal. The sub-mode IMF obtained by EMD decomposition needs to meet two conditions: (1) The number of extreme points and zero as the same, almost in 1; and (2) The mean value of the upper and lower envelope is zero.

For a given signal $x(t)$, EMD decomposition algorithm is as follows:

(1) Need to know the extremum of $x(t)$. CSI functions are used to construct envelope $a_u(t)$ and $a_t(t)$ of maxima and minima, respectively.

(2) The mean values $m(t)$ of the upper and lower envelop lines are calculated as shown in formula (6).

$$m(t) = \frac{a_u(t) + a_t(t)}{2} \tag{6}$$

(3) We can calculate $h(t) = x(t) - m(t)$. If $h(t)$ meets the IMF condition, so $IMF_i(t) = h(t)$, continue to step(4); otherwise, $x(t) = h(t)$, return to step (1).

(4) Subtract $IMF_i(t)$ from the original signal $x(t)$ to get the new signal $m_i(t)$ with the high-frequency components

removed, as shown in formula (7).

$$m_i(t) = x(t) - IMF_i(t) \tag{7}$$

(5) If $m_i(t)$ is monotone or constant, then the residual $R(t) = m_i(t)$ is decomposed. $x(t)$ based on EMD can be expressed as shown in formula (8).

$$x(t) = \sum_{i=1}^{n} IMF_i(t) + R(t) \tag{8}$$

Evidently, multiple IMFs are obtained after signal decomposition by EMD to display more internal characteristics of traffic data. Its basis function is derived from the signal itself, which has obvious self-adaptability and becomes a breakthrough in the field of time series signal decomposition. It also eliminates the influence of remote dependence and noise signals, and solves the problem of multiresolution and decomposition scale of wavelet. Thus, EMD is more suitable than WT for the decomposition of nonlinear and nonstationary signals [26]. Regrettably, with the deepening of the research, Deering *et al.* stated that EMD caused modal aliasing and endpoint effect due to the intermittent driving mechanism [27].

C. ENSEMBLE EMPIRICAL MODE DECOMOPSITION

In addressing the problems in EMD, Zhang *et al.* proposed a set of Ensemble Empirical Mode Decomposition (EEMD) algorithm in 2010 [28]. As a new auxiliary data analysis method, EEMD uses statistical characteristics of white noise to decompose the original signal. The EEMD method improves the decomposition accuracy and has been rapidly applied to machine monitoring and fault diagnosis [29].

EEMD technology is to decompose the signal containing intermittent signal into IMF1,IMF2,..., IMFn. IMF1 contains intermittent high frequency signals and partial low frequency sinusoidal signals which will all appear in IMF2. Obviously, the modal mixing problem occurs between different IMFn. Mixing has obvious negative effects on subsequent decomposition. The waveform frequencies of IMF2 and IMF3 are similar, but they are located in different IMF components. Thus, it can be seen that intermittency not only causes severe aliasing in time-frequency distribution, but also makes the physical meaning of a single IMF vague [25].

In discussing the optimal EEMD, Du *et al.* pointed out that the traditional EEMD was mainly due to the decomposition process [25]. If the amplitude of the additional noise is too small relative to the original signal, modal mixing cannot be effectively eliminated. Inversely, if the increased noise amplitude is too large, some extra IMF components will be generated, leading to misinterpretation of the analysis results. To completely offset the effect of adding white noise, numerous integration tests are required, and too many tests increase the computational cost. At the same time, if the amplitude of the white noise is set unreasonable, it will also produce unsatisfactory results.

It can be seen here that the performance of EEMD depends to a large extent on the selection of two important

parameters: the amplitude of white noise and the number of overall tests [25].Therefore, EEMD can decompose the original data signals into several IMF, and eliminate pattern aliasing. But, it over-dependence on the selection and optimization of parameters which affects the effective decomposition of signals and cause lower accurate prediction.

D. VARIATIONAL MODE DECOMPOSITION
1) OVERVIEW OF THE VARIATIONAL MODE DECOMPOSITION

In light of the existing problems of WT, EMD and EEMD, K. Dragomiretskiy et al. proposed the Variational Mode Decomposition (VMD) in 2014. [30]. It is a new decomposition technique for signal feature extraction. The theoretical basis of VMD includes The Hilbert-Huang transformation (HHT) and frequency mixing and other. The calculation process is the solution of the variational problem.

Compared with WT, EMD and EEMD, VMD can overcome the problems of mode aliasing and endpoint effect. It also has a better decomposition effect which has been applied in the field of signal research. Wang et al. analyzed the friction fault signal and identified the validity of the friction-induced signature by comparing the experimental data in VMD, Wavelet Transform (WT), Ensemble Empirical Mode Decomposition (EEMD) and EMD. Their findings subsequently proved the robustness of VMD over the other methods [31].

Similarly, Ma based on the VMD method, decomposed the traffic and carried out fault detection and concluded that VMD indeed showed better performance, lower noise and higher detection accuracy than EMD and EEMD [32].

2) PRINCIPLE AND ALGORITHM PROCESS OF VARIATIONAL MODE DECOMPOSITION

VMD transforms the signal decomposition process into the solution of the variational problem. Then find the optimal solution and update the modal function. Finally, the sub-mode is converted to IMF mode in time domain by inverse Fourier transform.

If the IMF has a limited bandwidth of the central frequency, we can change the variational problem into solving N modal functions $U_n(t)$, C ($n = 1, 2, \ldots, N$), and the IMF has the minimum estimated bandwidth and constraint conditions for solving the problem, and the sum of each IMF is taken as the input signal for the modeling. The specific steps are as follows:

(1) According to the HHT processing mode function $U_n(t)$ analytic signal, we can calculate its unilateral spectrum, as shown in formulae (9) and (10).

$$P(t) = \left(\delta(t) + \frac{j}{\pi t}\right) U_n(t) \tag{9}$$

$$\delta(t) = \left\{ \begin{array}{l} 0\, t \neq 0 \\ \infty\, t = 0 \end{array}, \int_{-\infty}^{+\infty} \delta(t)\, dt = 1 \right\} \tag{10}$$

where, $\delta(t)$ is the impact function; $\{U_n\}$ is the modal function $U_1, U_2, \ldots, U_n$.

(2) The spectrum of each mode is modulated to the corresponding base frequency band based on the mixing-estimated center frequency $e^{-jw_n(t)}$ of the analytical signals of each mode, as shown in formula (11).

$$q(t) = \left[\left(\delta(t) + \frac{j}{\pi t}\right) U_n(t)\, e^{-jw_n(t)}\right] \tag{11}$$

where, $\{\omega_k\}$ is the central frequency $\omega_1, \omega_2, \ldots, \omega_n$.

(3) Solve the square norm L^2 of the above signal gradient. We need to estimate the bandwidth of each mode signal, and the expression of the constrained variational problem is shown in formula (12).

$$\left\{ \begin{array}{l} \min\limits_{\{U_n\}, \{\omega_n\}} \left\{ \sum_{n=1}^{n} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t}\right) U_n(t)\right] e^{-jw_n(t)} \right\|_2^2 \right\} \\ \text{s.t. } \sum_{n=1}^{n} U_n = f(t) \end{array} \right. \tag{12}$$

To find an optimal solution of the above constrained problem, introduce the quadratic penalty factor C and Lagrance multiplication operator θ_t. The improved Lagrance expression is shown in formula (13).

$$L(\{u_n\}, \{w_n\}, \theta)$$
$$= C \sum_{n=1}^{n} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t}\right) U_n(t)\right] e^{-jw_n(t)} \right\|_2^2$$
$$+ \left\| f(t) - \sum_{n=1}^{n} U_t(t) \right\|_2^2 + \langle \theta(t), f(t) - \sum_{n=1}^{n} U_t \rangle \tag{13}$$

Use VMD based on Alternating Direction Method of Multipliers(ADMM) algorithm to solve the variational problem, at the same time to update u_n^{m+1}, ω_n^{m+1} and θ_n^{m+1} (m says the number of iterations), search to improve Lagrance expression of saddle points. The u_n^{m+1} can be transformed into the frequency domain by Fourier isometric transform, as shown in formula (14).

$$\hat{u}_n^{m+1}$$
$$= \arg \min_{\hat{u}_{u}, u_n \in X} \left\{ C \left\| j\omega \left\{ \left[1 + \text{sgn}(\omega + \omega_n)\right] \hat{u}_n(\omega + \omega_n) \right\} \right\|_2^2 \right.$$
$$\left. + \left\| \hat{f}(\omega) - \sum_{n=1}^{n} \hat{u}_n(\omega) + \frac{\hat{\theta}(\omega)}{2} \right\|_2^2 \right\} \tag{14}$$

where, ω is the random frequency, finish $\omega - \omega_n \to \omega$ substitution, we can calculate the integral expression of the non-negative frequency interval, as shown in formula (15).

$$\hat{u}_n^{m+1} = \arg \min_{\hat{u}_{u}, u_n \in X} \left\{ \int_0^{\infty} \left[4C(\omega - \omega_k)^2 \left| \hat{u}_k(\omega) \right|^2 \right. \right.$$
$$\left. \left. + 2 \left| \hat{f}(\omega) - \sum_{n=1}^{n} \hat{u}_n(\omega) + \frac{\hat{\theta}(\omega)}{2} \right|^2 \right] d\omega \right\} \tag{15}$$

At this point, the solution of the quadratic optimization problem is shown in formula (16).

$$\hat{u}_n^{m+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{n=1}^{n} \hat{u}_n(\omega) + \frac{\hat{\theta}(\omega)}{2}}{1 + 2C(\omega - \omega_n)^2} \tag{16}$$

According to the same calculation principle, update the ω_n^{m+1} center frequency and get the formula (17).

$$\omega_n^{m+1} = \frac{\int_0^\infty \omega \left| \hat{u}_k(\omega) \right|^2 d\omega}{\int_0^\infty \left| \hat{u}_k(\omega) \right|^2 d\omega} \qquad (17)$$

where, $\hat{u}_n^{m+1}(\omega)$ is the Wiener filter of the remaining $\hat{f}(\omega) - \sum_{n=1}^{n} \hat{u}_n(\omega) . \omega_n^{m+1}$ is the center of the current IMF power spectrum. $\hat{u}_k(\omega)$ is Fourier inverse, and the real part is $\{u_n(t)\}$.

Based on the above analysis of VMD variational problem solving, it can be concluded that penalty factor C and Lagrance multiplication operator θ_t play a key role in VMD variational problem solving.

The VMD algorithm flow can be summarized in these five steps as follows:

(1) Initialization parameter $\{u_n^1\}$, $\{\omega_n^1\}$, $\{\hat{\theta}^1\}$ and n.

(2) Update u_n and ω_n according to formulae (18) and (19).

$$\hat{u}_n^{m+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{n=1}^{n} \hat{u}_n(\omega) + \frac{\hat{\theta}(\omega)}{2}}{1 + 2C(\omega - \omega_n)^2} \qquad (18)$$

$$\omega_n^{m+1} = \frac{\int_0^\infty \omega \left| \hat{u}_k(\omega) \right|^2 d\omega}{\int_0^\infty \left| \hat{u}_k(\omega) \right|^2 d\omega} \qquad (19)$$

(3) Update θ as shown in formula (20).

$$\hat{\theta}^{n+1} \leftarrow \hat{\theta}^n(\omega) + \tau \left[\hat{f}(\omega) - \sum_{n=1}^{n} \hat{u}_n^{m+1}(\omega) \right] \qquad (20)$$

(4) Repeat steps (2) and (3). If $e > 0$ exists, $\sum_{n=1}^{n} \frac{\| \hat{u}_n^{m+1} - \hat{u}_n^m \|_2^2}{\| -\hat{u}_n^m \|_2^2}$, so the loop will stop.

(5) The decomposed results will be output, and we can obtain the each Intrinsic Mode functions (IMF).

According to the above analysis, the VMD decomposition technique involves two main parameters: the number of modes and the iterative factor. The values of the signals need to be pre-set before they can be decomposed. In order to reduce the influence of artificial subjective experience and prior knowledge on the value, an optimization method must be adopted to realize the adaptive selection of parameters to adapt to the decomposition of different signals, so as to build a prediction model mechanism with high accuracy.

Based on the above statement of time series decomposition technology, this article summarizes and compares the strengths and limitations of WT, EMD, EEMD and VMD (see table 4), in order to ascertain the correct data signal processing method constructed by network traffic mechanism In addition, the application of decomposition technology in network traffic prediction model is analyzed and compared as shown in Table 5.

The literature reviews reveal that VMD is less susceptible to noise interference in decomposing noise signals. VMD is does not only overcome the multiresolution and decomposition scale problem in traditional Wavelet Transform, but it can also solve the problem of mode aliasing and white noise amplitude in EMD and EEMD methods [30]. Due to

TABLE 4. Strengths and limitations of decomposition techniques.

NO.	Year	Author	Name	Strength	Limitation
1	1984	J.Morlet [17] Y.Meyer S.Mallat [18]	Wavelet Transform, WT	The most powerful signal processing tool in nonstationary signal decomposition	Multi-resolution, decomposition scale is difficult.
2	1998	By Norden E Huang et al.[24]	Empirical mode decomposition, EMD	Eliminate remote dependence and noise signals. Solve the problem of multiresolution and decomposition scale in wavelet.	Modal aliasing, endpoint effects, low robustness
3	2010	Zhang, J and Kwon [28]	Ensemble Empirical Mode Decomposition, EEMD	Eliminate modal aliasing and endpoint effects; High decomposition accuracy and calculation efficiency.	Depends on the amplitude of white noise, and sets the number of tests.
4	2014	Konstantin Dragomiretskiy [30]	Variational Mode Decomposition, VMD	Overcome modal aliasing and endpoint effect. Not easily disturbed by noise and high robustness.	It is affected by decomposition layer, penalty factor and Lagrange multiplier.

this good decomposition effect and robustness, VMD is more suitable as the core decomposition technology of network traffic prediction model.

VMD can adaptively determine the relevant bands and estimate the corresponding patterns, then appropriately balancing the errors between them. According to the narrow-band characteristics corresponding to the current IMF definition, VMD finds a set of patterns and reconstructs and gives them in an optimal process to obtain the best input signal.

Distinctly, the decomposition of VMD and the reconstruction of IMF sub-pattern is an extremely important process, The decomposition modes, penalty factors, and Lagrange multiplier update steps are the core parameters to determine the optimal process. Therefore, an algorithm must be designed to obtain and optimize the performance of VMD in future research. To seek the best three parameters of VMD, improve the performance of VMD, and strive to realize the network traffic model prediction mechanism with higher accuracy and faster convergence speed.

To sum up, the VMD decomposition technology and PSO optimization algorithm for constructing the prediction

TABLE 5. Application of decomposition technology in network prediction model.

TABLE 5. *(Continued.)* Application of decomposition technology in network prediction model.

NO.	Year	Author	Original model	Combined Model	Research methods Practical	Research methods Theoretical	Data set used	Evaluation Criteria	Strength	Limitation
1	2017	Laisen Nie et al. [33]	Discrete Wavelet Transform; Gaussion Model; Deep Belief Network	DWT-DBN-GM	√		A real network traffic data set with 2016 times is sampled on a time scale of 5 minutes	SRE\TRE	It is divided into low and high-pass components by DWT; Gaussian model predicts high-pass component, while DBN method predicts low-pass component.	Ignores the difficulty of wavelet decomposition scale and the complexity of the new model structure, leading to low generalization performance
2	2018	Laisen Nie et al. [34]	Discrete Wavelet Transform; Deep Belief Network; Spatiotemporal Compress-ive Sensing	DWT-DBN-SCS	√		A real network traffic data set with 2016 time slots is sampled on a time scale of 5 minutes.	SRE\TRE	It is divided into low and high-pass components by DWT;DBN model predicts low-pass component, whereas SCS method predicts high-pass component	
3	2017	Lina Pan et al.[35]	Echo State Network; Variational Mode Decompo-sition; Bat Algorithm	VMD-BA-ESN		√	A5M(ISP with centres in 11 European cities); B5M(United Kingdom academic network backbone)	MAPE\NRMSE\coefficient of determination	Decomposes the complex sequence into several simple components by VMD; global optimization is used to determine the initial weight of ESN and reduces the influence.	Decomposition layers affect the decomposition effect and ultimately affect the accuracy of prediction.
4	2018	Hui Tian et al.[36]	Quantum genetic algorithm; Wavelet Transform; BP Neural Network	WT-QGA-BP	√		BC-poct16 data from a state-owned enterprise in this section and data from Abilene network.	SMAPE\NRMSE	Determines embedded dimension and associated dimension based on the C-C and G-P algorithm; reconstructs and optimizes BP based on the WT.	WT can easily cause signal noise interference, and decomposition scale is difficult; new model is too complex and lowers the generalization performance.
5	2018	Tian Zhongda [37]	Empirical Mode Decompo-sition; Extreme Learning Machine; Seasonal Loess Trend decomposition process method	EMD-ELM	√		Lorenz time series and Shenyang University of Technology's data	RMSE\MAE\MAPE	Decomposes the traffic into stable data signals with different characteristic scales by EMD; combines the advantages of ELM's fast learning speed to improve the accuracy of network traffic prediction.	The ELM random input weights deviations and mode aliasing of EMD can inversely impact the accuracy of the model.
6	2018	Rishabh Madan et al.[38]	Discrete Wavelet Transform; AR/ARIMA Model; Recurrent neural network	DWT-ARIMA-RNN		√	The time series that have been used, consists of internet traffic (in bits) as recorded by an ISP and it corresponds to transatlantic link.	NRMSE	Decomposes reconstruction traffic into details and approximates components by iDWT;ARIMA model is used to predict the low component and RNN is used to predict the high component.	The signal decomposition and reconstruction issues are sufficiently addressed, but the calculation of RNN is difficult and the WT decomposition scale is complex.
7	2018	Haipeng Lu et al. [39]	Long Short Term Memory; Wavelet Transform	WT-LSTM		√	Historical network traffic in a DNS server of China Mobile	MAE\RMSE\MAPE	Decomposes the traffic by WT and predicts based on the WT-LSTM hybrid model.	The scale difficulty of WT and the local optimal solution of LSTM are not well dealt with.
8	2019	Wenbo Chen et al.[40]	Empirical Mode Decompo-sition; Support Vector Machine; Particle Swarm Optimizat-ion	EMD-PSO-SVM	√		Network Center of Lanzhou University	RMSE\MAE\MRE	Eliminates noise from data based on the EMD and optimizes SVM based on the PSO.	The influence of modal aliasing in EMD on model accuracy is ignored.

TABLE 5. *(Continued.)* **Application of decomposition technology in network prediction model.**

9	2019	Xinghan Xu et al.[41]	BP Neural Network / Variational Mode Decomposition / Ensemble Empirical Mode Decomposition / Firefly algorithm	CEEMD-VMD-FA-BPNN	√	The experimental data is collected from the real world	RMSE\NRMSE\MAPE\SMAPE	Improves EEMD defects by increasing adaptive white noise amplitude and forms a two-stage decomposition with VMD technology; optimizes the threshold and weight of BP neural node based on FA to improve the ability of function approximation to neural network	Ignores the number of VMD decomposition layers which can influence the decomposition effect, thus reducing the prediction accuracy of the model.
10	2019	YING HAN et al.[42]	Variational Mode Decomposition / Phase Space Reconstruction / Levy's Flight Function / Cloud generator / Fruit Fly Optimization Algorithm / Echo State Network	VMD-PSR-IFOA-ESN	√	MAWI Working Group	RMSE\MAE\MAPE	Improves FOA by Cloud model and Levy flight function to obtain VMD modes and iterative factor parameters; the sub-modes after VMD decomposition are reconstructed by PSR;ESN multiple subreservoirs is automatically built according to optimized decomposition results by VMD.	The new combined model needs longer calculation time, which due to synchronous optimization inside and outside of the multiple sub reservoirs. This will impact the training time of the model and affect the convergence speed and the performance of model prediction

mechanism of combined model with faster convergence rate and higher accuracy still need to be improved and promoted.

IV. CONCLUSION

In a nutshell, optimization algorithm and decomposition technique play an important role in combined traffic prediction model so as to ensure prediction accuracy and increase the convergence speed. On the one hand, PSO is an optimization strategy for network model construction, it can better identify network traffic time sequence which has the strengths of simple principle, small calculation scale, fast speed and so

forth. On the other hand, VMD is the most important core strategy in constructing network traffic prediction mechanism. It processes network time sequence which has the strengths of reducing signal transmission error, eliminating mode aliasing, reducing endpoint effects among others. It is worth noting that an adaptive algorithm must be designed to solve the problem for PSO and VMD are limited to the acquisition of optimal parameters, overcome the defects of human experience, improve the local optimal solution phenomenon and the decomposition effect of data signals, so as to enhance the prediction accuracy of the model.

In closing, the decomposition technology and optimization algorithm for constructing a combined network traffic prediction model with faster convergence speed and higher accuracy is a topic which still warrants future research attention.

REFERENCES

[1] G. O. Young, "Synthetic structure of industrial plastics," in *Plastics*, vol. 3, J. Peters, Ed., 2nd ed. New York, NY, USA: McGraw-Hill, 1964, pp. 15–64.

[2] W.-K. Chen, *Linear Networks and Systems*. Belmont, CA, USA: Wadsworth, 1993, pp. 123–135.

[3] J. P. Dickinson, "Some comments on the combination of forecasts," *Oper. Res. Quart.*, vol. 26, no. 1, pp. 205–210, 1975.

[4] J. H. Holland, "Genetic algorithms and the optimal allocation of trials," *SIAM J. Comput.*, vol. 2, no. 2, pp. 88–105, Jun. 1973.

[5] A. Narayanan and M. Moore, "Quantum-inspired genetic algorithms," in *Proc. IEEE Int. Conf. Evol. Comput.* May 1996, pp. 61–66.

[6] K. Zhang, L. Liang, and Y. Huang, "A network traffic prediction model based on quantum inspired PSO and neural network," in *Proc. 6th Int. Symp. Comput. Intell. Design*, vol. 2, Oct. 2013, pp. 219–222, doi: 10.1109/ISCID.2013.168.

[7] Y. Han, Y. Jing, and K. Li, "Multi-step prediction for the network traffic based on echo state network optimized by quantum-behaved fruit fly optimization algorithm," in *Proc. 29th Chin. Control Decis. Conf. (CCDC)*, May 2017, pp. 2270–2274, doi: 10.1109/CCDC.2017.7978893.

[8] S. Nayak, S. Nayak, and P. J. P. Singh, "An introduction to quantum neural computing," *J. Global Res. Comput. Sci.*, vol. 2, no. 8, pp. 50–55, 2011.

[9] P. A. N. Wen-chao, "Using fruit fly optimization algorithm optimized general regression neural network to construct the operating performance of enterprises model," *J. Taiyuan Univ. Technol. (Social Sci. Ed.)*, vol. 4, pp. 1–5, Apr. 2011.

[10] Y. Han and K. Li, "Improved FOA-ESN method using opposition-based learning mechanism for the network traffic prediction with multiple steps," in *Proc. Chin. Autom. Congr. (CAC)*, Oct. 2017, pp. 7183–7187, doi: 10.1109/CAC.2017.8244074.

[11] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *Proc. 6th Int. Symp. Micro Mach. Hum. Sci.*, Oct. 1995, pp. 39–43.

[12] T. He, T. Dan, Y. Wei, H. Li, X. Chen, and G. Qin, "Particle swarm optimization RBF neural network model for Internet traffic prediction," in *Proc. Int. Conf. Intell. Transp., Big Data Smart City (ICITBS)*, Dec. 2016, pp. 521–524, doi: 10.1109/ICITBS.2016.146.

[13] F. Han, H.-F. Yao, and Q.-H. Ling, "An improved evolutionary extreme learning machine based on particle swarm optimization," *Neurocomputing*, vol. 116, pp. 87–93, Sep. 2013, doi: 10.1016/j.neucom.2011.12.062.

[14] Y. Yang, Y. Chen, C. Li, X. Gui, and L. Li, "Network traffic prediction based on LSSVM optimized by PSO," in *Proc. IEEE 11th Int. Conf Ubiquitous Intell. Comput. IEEE 11th Int. Conf. Autonomic Trusted Comput. IEEE 14th Int. Conf Scalable Comput. Commun. Associated Workshops*, Dec. 2014, pp. 829–834, doi: 10.1109/UIC-ATC-ScalCom.2014.100.

[15] W. Zhang and D. Wei, "Prediction for network traffic of radial basis function neural network model based on improved particle swarm optimization algorithm," *Neural Comput. Appl.*, vol. 29, no. 4, pp. 1143–1152, Feb. 2018, doi: 10.1007/s00521-016-2483-5.

[16] D. Wei, "Network traffic prediction based on RBF neural network optimized by improved gravitation search algorithm," *Neural Comput. Appl.*, vol. 28, no. 8, pp. 2303–2312, Aug. 2017.

[17] P. Goupillaud, A. Grossmann, and J. Morlet, "Cycle-octave and related transforms in seismic signal analysis," *Geoexploration*, vol. 23, no. 1, pp. 85–102, Oct. 1984.

[18] S. Mallat, *A Wavelet Tour of Signal Processing, the Sparse Way*. New York, NY, USA: Academic, 2009.

[19] R. H. Riedi, M. S. Crouse, V. J. Ribeiro, and R. G. Baraniuk, "A multifractal wavelet model with application to network traffic," *IEEE Trans. Inf. Theory*, vol. 45, no. 3, pp. 992–1018, Apr. 1999.

[20] H. Xiao, H. Sun, B. Ran, and Y. Oh, "Fuzzy-neural network traffic prediction framework with wavelet decomposition," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1836, no. 1, pp. 16–20, Jan. 2003, doi: 10.3141/1836-03.

[21] V. Kuková and P. C. Kainen, "Functionally equivalent feedforward neural networks," *Neural Comput.*, vol. 6, no. 3, pp. 543–558, May 1994.

[22] Y. Yu and C. Junsheng, "A roller bearing fault diagnosis method based on EMD energy entropy and ANN," *J. Sound Vibrat.*, vol. 294, nos. 1–2, pp. 269–277, Jun. 2006, doi: 10.1016/j.jsv.2005.11.002.

[23] Z. K. Peng, P. W. Tse, and F. L. Chu, "An improved Hilbert–Huang transform and its application in vibration signal analysis," *J. Sound Vibrat.*, vol. 286, nos. 1–2, pp. 187–205, Aug. 2005, doi: 10.1016/j.jsv.2004.10.005.

[24] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zhang, N. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London. A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, 1998.

[25] S.-C. Du, T. Liu, D.-L. Huang, and G.-L. Li, "An optimal ensemble empirical mode decomposition method for vibration signal decomposition," *J. Vibrat. Acoust.*, vol. 139, no. 3, Jun. 2017, Art. no. 031003, doi: 10.1115/1.4035480.

[26] Y. Sun, C. Zhuang, and Z. Xiong, "Transform operator pair assisted Hilbert–Huang transform for signals with instantaneous frequency intersections," *J. Vibrat. Acoust.*, vol. 137, no. 6, Dec. 2015, Art. no. 061016, doi: 10.1115/1.4031407.

[27] R. Deering and J. F. Kaiser, "The use of a masking signal to improve empirical mode decomposition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, vol. 4, Mar. 2005, pp. iv/485–iv/488. [Online]. Available: https://doi.org/10.1109/ICASSP.2005.1416051.

[28] J. Zhang, R. Yan, R. X. Gao, and Z. Feng, "Performance enhancement of ensemble empirical mode decomposition," *Mech. Syst. Signal Process.*, vol. 24, no. 7, pp. 2104–2123, Oct. 2010, doi: 10.1016/j.ymssp.2010.03.003.

[29] W. Guo and P. W. Tse, "Enhancing the ability of ensemble empirical mode decomposition in machine fault diagnosis," in *Proc. Prognostics Syst. Health Manage. Conf.*, Jan. 2010, pp. 1–7, doi: 10.1109/PHM.2010.5413421.

[30] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014, doi: 10.1109/TSP.2013.2288675.

[31] Y. Wang, R. Markert, J. Xiang, and W. Zheng, "Research on variational mode decomposition and its application in detecting rub-impact fault of the rotor system," *Mech. Syst. Signal Process.*, vols. 60–61, pp. 243–251, Aug. 2015, doi: 10.1016/j.ymssp.2015.02.020.

[32] P. Ma, H. Zhang, W. Fan, and C. Wang, "Early fault detection of bearings based on adaptive variational mode decomposition and local tangent space alignment," *Eng. Comput.*, vol. 36, no. 2, pp. 509–532, Mar. 2019, doi: 10.1108/EC-05-2018-0206.

[33] L. Nie, D. Jiang, S. Yu, and H. Song, "Network traffic prediction based on deep belief network in wireless mesh backbone networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, San Francisco, CA, USA, 2017, pp. 1–5, doi: 10.1109/WCNC.2017.7925498.

[34] L. Nie, X. Wang, L. Wan, S. Yu, H. Song, and D. Jiang, "Network traffic prediction based on deep belief network and spatiotemporal compressive sensing in wireless mesh backbone networks," *Wireless Commun. Mobile Comput.*, vol. 2018, pp. 1–10, Jan. 2018.

[35] L. Pan, J. Cheng, H. Li, Y. Zhang, and X. Chen, "An improved echo state network based on variational mode decomposition and bat optimization for Internet traffic forecasting," in *Proc. IEEE 17th Int. Conf. Commun. Technol. (ICCT)*, Oct. 2017, pp. 417–421.

[36] H. Tian, X. Zhou, and J. Liu, "A hybrid network traffic prediction model based on optimized neural network," in *Proc. 18th Int. Conf. Parallel Distrib. Comput., Appl. Technol. (PDCAT)*, Dec. 2017, pp. 284–287, doi: 10.1109/PDCAT.2017.00053.

[37] T. Zhongda, M. Chengcheng, W. Gang, and R. Yi, "Approach for time series prediction based on empirical mode decomposition and extreme learning machine," in *Proc. Chin. Control Decis. Conf. (CCDC)*, Jun. 2018, pp. 3119–3123.

[38] R. Madan and P. S. Mangipudi, "Predicting computer network traffic: A time series forecasting approach using DWT, ARIMA and RNN," in *Proc. 11th Int. Conf. Contemp. Comput. (IC3)*, vol. 1, Aug. 2018, pp. 2–4.

[39] H. Lu and F. Yang, "A network traffic prediction model based on wavelet transformation and LSTM network," in *Proc. IEEE 9th Int. Conf. Softw. Eng. Service Sci. (ICSESS)*, Nov. 2018, vol. 1, no. 3, pp. 1–4.

[40] W. Chen, Z. Shang, and Y. Chen, "A novel hybrid network traffic prediction approach based on support vector machines," *J. Comput. Netw. Commun.*, vol. 2019, pp. 1–10, Feb. 2019.

[41] X. Xu and W. Ren, "A hybrid model based on a two-layer decomposition approach and an optimized neural network for chaotic time series prediction," *Symmetry*, vol. 11, no. 5, p. 610, May 2019.

[42] 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*, vol. 7, pp. 138364–138377, 2019.

[43] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, MA, USA: Addison-Wesley, vol. 13, no. 7, 1989, pp. 2104–2116.

[44] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 841–851, Jan. 2019, doi: 10.1109/TSG.2017.2753802.

JINMEI SHI is currently pursuing the Ph.D. degree in computer science with the Faculty of Computing and Informatics, University Malaysia Sabah. She has been engaged in research and has always been in the forefront of scientific research, which she has been teaching for 12 years with the Hainan Vocational University of Science and Technology. She plays an active role in the construction of computer science and personnel training, and has good practical experience and academic foundation. Her main research interests include network traffic prediction, algorithm analysis, and software. In recent years, she has published ten articles, three sponsored projects, three books, and ten awards above the provincial level.

YU-BENG LEAU (Member, IEEE) received the B.Sc. degree in multimedia technology from Universiti Malaysia Sabah, Malaysia, the M.Sc. degree in information security from Universiti Teknologi Malaysia, and the Ph.D. degree in Internet infrastructures security from Universiti Sains Malaysia. He is currently a Senior Lecturer of computer science with the Faculty of Computing and Informatics, Universiti Malaysia Sabah. His current research interests include intrusion alert detection and prediction, network security situation awareness, IPv6 security, the Internet of Things (IoT), and information centric networks (ICNs).

KUN LI received the B.Sc. degree from the Shandong University of Science and Technology, in 2005, the M.Sc. degree from Liaoning Technical University, in 2008, and the Ph.D. degree from Northeastern University, Shenyang, China, in 2013. He is currently an Associate Professor with the College of Engineering, Bohai University. His current research interests include complex industrial process modeling, intelligent optimal control, machine learning, and their application.

YONG-JIN PARK (Life Senior Member, IEEE) received the B.E., M.E., and Ph.D. degrees in electronic engineering from Waseda University, Tokyo. He has worked with Hanyang University, Seoul, for 33 years. He was a Professor with Waseda University, for six years. In 2016, he joined University Malaysia Sabah, where he is currently a Professor with the Faculty of Computing and Informatics. He was a Visiting Associate Professor with the Department of Computer Science, University of Illinois at Urbana–Champaign, and a Research Fellow with the Computing Laboratory, University of Kent, U.K. He is also a Professor Emeritus with Hanyang University. He is a Fellow of IEICE. He was the President of the Open Standards and ICT Association (OSIA) and the Chairman of the IEEE Seoul Section. He was the President of the Korea Institute of Information Scientists and Engineers (KIISE) and the Director of the IEEE Region 10.

ZHIWEI YAN received the Ph.D. degree from the National Engineering Laboratory for Next Generation Internet Interconnection Devices, Beijing Jiaotong University. He joined the China Internet Network Information Center (CNNIC), in 2011, where he is currently a Professor. Since April 2013, he has been an Invited Researcher with Waseda University. His research interests include mobility management, network security, and next-generation Internet. He is active in IETF, APNIC, and ICANN. He is the ICANN RSSAC Caucus Member and published RFC8191 in IETF.

• • •