TRAFFIC FLOW TIME-SERIES PREDICTION MODEL BASED ON WOA-VMD-SSA-LSTM ALGORITHM

Fuyou MAO[1,*], Jingke XU[2], Xiyang LIU[3], Jiaojiao LUO[4]

Traffic flow prediction, an integral part of intelligent traffic systems and traffic planning, remains a significant challenge due to the nonlinear nature of time-series traffic flow data. Addressing this challenge, we present a distinctive approach: the WOA-VMD-SSA-LSTM combined prediction model. Harnessing the capacity of the Variational mode decomposition (VMD), an algorithm proven to significantly reduce the complexity of raw data to enhance subsequent model precision, and Sparrow optimization algorithm (SSA), a state-of-the-art intelligent optimization algorithm that has shown excellent performance in achieving superior hyperparameters for the prediction model, we demonstrate a novel methodology for improving prediction precision. We implement the WOA algorithm to determine the optimal number of decomposition modes and penalty parameters for VMD, targeting envelope entropy. The resulting intrinsic mode functions (IMFs) provide inputs for defined LSTM prediction network structures, which we progressively optimized using SSA to minimize mean square error (MSE). Comparative experiments with classic traffic flow time series models affirm the superior precision and performance of our proposed WOA-VMD-SSA-LSTM model in predicting time-series traffic flow.

Keywords: Traffic flow predict; VMD; WOA; LSTM; SSA

1. Introduction

With the emergence of the field of traffic big data, intelligent transportation systems are only important in the construction of smart cities [1]. In recent years, traffic flow has garnered significant attention from scholars and researchers as a crucial measure of route congestion. Deep learning methods have been employed to train historical data for the prediction of future traffic flow, aiming to accurately assess and predict traffic conditions. This approach not only facilitates real-time monitoring of road conditions for travel planning but also contributes to the optimization and design of smart city traffic and aids in predicting emergencies such as traffic congestion [2,3]. Accurate and timely traffic flow prediction is crucial for managing road conditions. Traffic data is influenced by environment, human factors, travel patterns, peak hours, weather, and holidays. Multiple factors interact nonlinearly to impact actual traffic flow. [4-6].

[1] College of Computer Science and Technology, Shenyang Jianzhu University, Shenyang 110168, China, e-mail: mfy15541207057@gmail.com
[2] College of Computer Science and Technology, Shenyang Jianzhu University, Shenyang 110168, China.
[3] College of Computer Science and Technology, Shenyang Jianzhu University, Shenyang 110168, China.
[4] College of Computer Science and Technology, Shenyang Jianzhu University, Shenyang 110168, China.

For the Traffic Flow Data (TFD) problem, some prestigious researchers have tried a variety of methods in the past several years to enhance the prediction precision of traffic flow. Their research methods can be mainly split up into three categories: statistical-based prediction models [7,8], monolithic intelligence-based prediction models [9] and hybrid prediction models [10]. The time series-oriented ARIMA model proposed by Honghui Dong in 2009 is used for TFD prediction, but this model only deals with linear relationships while maintaining an intense degree of stability of the data. Traditional statistical mathematical models face challenges in dealing with complex time-series traffic flow. Therefore, research on time-series traffic flow prediction should be based on neural network optimization algorithms [11-13].Compared with traditional statistical methods, artificial intelligence algorithms have achieved unprecedented development in recent years owing to the availability of big data and the exponential development of computing power [14].Qianqian Zhang analyzed the time-series characteristics of traffic flow in Beijing and used BP neural network algorithm to predict the morning and evening peak traffic flow on *Xueyuan* Road in Beijing, thus, the results achieved high prediction accuracy. Yi-Ming Xing et al. proposed the KELM algorithm, which overcomes the variation problem caused by the random allocation of weights of BP neural networks algorithm and selected the actual traffic flow data in Nanning area for prediction. The results demonstrated that KELM can produce more accurate prediction results [15]. Yan Tian et al. used LSTM to learn irregular sampled traffic flow data through multi-scale time smoothing, and compared it with many traditional TFD prediction models, the results proved that Yan's method had higher prediction accuracy [16]. Then for these traditional machine learning algorithms, they generally used layer descent algorithms when iteration and returning hyperparameters. Traffic flow datasets, often comprising tens of thousands of measurements, can result in lengthy model training and issues such as layer explosion and disappearance in neural networks. As research on swarm intelligence optimization algorithms progresses, these algorithms continue to showcase their biological intelligence characteristics in the global optimization of deep learning hyperparameters. Hybrid forecasting models, combining the strengths of different models, can achieve enhanced forecasting accuracy for complex traffic data.

2. Methodology

2.1 WOA-VMD-SSA-LSTM

The main work of the model WOA-VMD-SSA-LSTM presented in this paper is seen in Fig. 1, and its main innovations are summarized as follows:

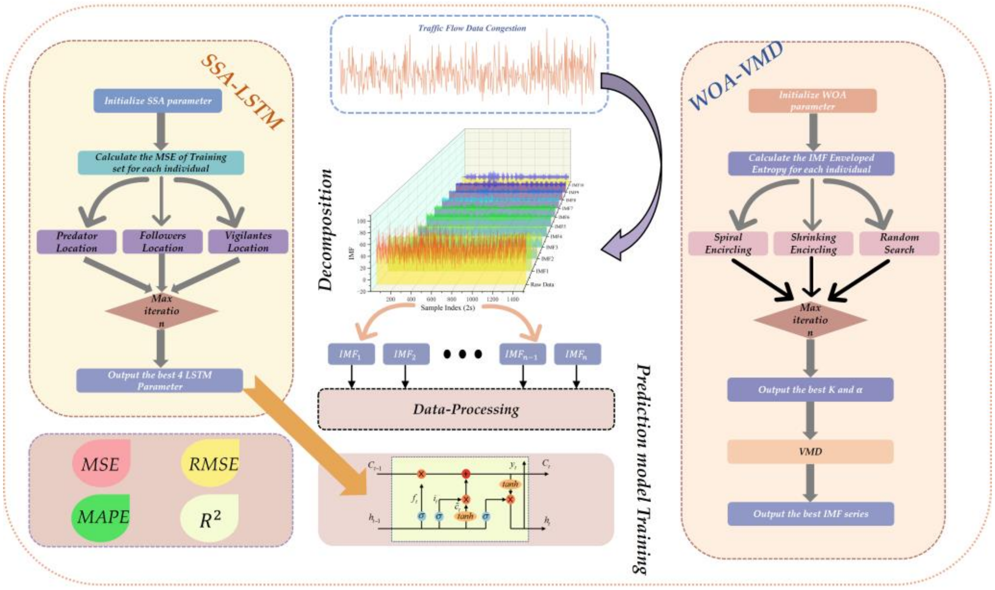

Fig. 1: The whole flowchart of this Method

1. WOA-VMD is rationally adopted to decompose the primitive TFD data in view of the temporal complexity and non-linear attributes of traffic flow and the WOA algorithm is utilized to globally discover the number of decomposition that yields optimal results K and penalty parameters in VMD, followed by VMD decomposition, which significantly refines the accuracy of the later prediction.
2. This paper adopts SSA algorithm to perform global optimization with MSE as the objective function, which achieves faster convergence speed and better network parameters.

The following sections will briefly introduce the four methods mentioned above.

2.2 WOA-VMD

The real traffic flow data is challenging for ideal prediction using a single model, particularly with increased prediction step size. This paper proposes a time-series data decomposition model based on the principles of the whale optimization algorithm to optimize parameters.

VMD is an extremely adaptive signal decomposition rule, invented by Konstantin Dragomiretskiy in 2014, which decomposes the original input according to a predetermined number of components, and then solves and extracts the modal components u_k and central frequencies w_k by the alternating direction multiplier method. VMD takes great advantages in signal feature extraction [17]. The reduction of non-stationary information in the time-series prediction problem is closely related to the prediction precision [18]. The computation model is as follows:

$$\hat{u}_k^{(n+1)}(w) = \frac{\hat{f}(w) - \sum_{i \neq k} \hat{u}_k^{(n)}(w) + \frac{\hat{\lambda}^{(n)}(w)}{2}}{1 + 2\alpha(w - w_k^{(n)})^2} \tag{1.}$$

$$\hat{w}_k^{(n+1)} = \frac{\int_0^\infty w \left| \hat{u}_k^{(n+1)}(w) \right|^2 dw}{\int_0^\infty \left| u_k^{(n+1)}(w) \right|^2 dw} \tag{2.}$$

Wherein, u_k is the first k mode component; w_k is the domain frequency according to the k mode; f is the original signal; α is penalty factor.

Decomposed IMF sub volume will store the modalities and noise of the original data, and then reconstruct the modalities of the data signal containing the main features, thus achieving the effect of denoising the original data and making it smoother. In VMD, the number of model components k has a significant influence on the precision of the later prediction model. A small value of k will contribute to under-decomposition, such that the decomposed sub-series will not fully characterize the original traffic time-series data. A large value of k will lead to over-decomposition of the original data, resulting in model confusion. A large value of the penalty parameter α will result in a loss of band information, conversely, in redundancy of information. Therefore, the VMD algorithm needs to determine a reasonable k and α when decomposing the data. The central frequency observation method has been mostly used to select the k value, but this approach is somewhat contingent and only determines the modes number k but not the penalty parameter α. Based on the problems of parameter selection in the VMD algorithm, this paper introduces the whale optimization algorithm for optimization. WOA is a novel intelligence algorithm invented by Australian scholar Mirjalili in 2016 through the research of the population behaviors of humpback whales. The WOA algorithm has fewer tuning parameters and is robust, with better solution accuracy and stability than particle swarm algorithm (PSO), Gravitational Search Algorithm (GSA), Artificial Butterfly Optimization (ABO), Grey Wolf Optimizer (GWO) and other algorithms in neural network parameter finding and scheduling control optimization [19].

In this paper, the WOA is utilized to depict a global optimization search for the crucial parameters of VMD. Envelope entropy E_p is a function used to represent the sparse nature of the initial data. When there is more disturbed information and less feature information in the IMF, the envelope entropy value is larger, and vice versa, the envelope entropy value is smaller [20]. The mathematical expressions are as follow:

$$\begin{cases} E_p = -\sum_{i=1}^{N} \varepsilon(i) \lg \varepsilon(i) \\ \varepsilon(i) = \dfrac{a(i)}{\sum_{i=1}^{N} a(i)} \end{cases} \tag{3.}$$

$a(i)$ is the envelope flow of the Hilbert demodulation of the k model components decomposed by the VMD. The entropy of the probability distribution sequence $\varepsilon(i)$ denotes the envelope entropy E_p .

The whole framework for the WOA optimized VMD is as followed by Fig. 2.

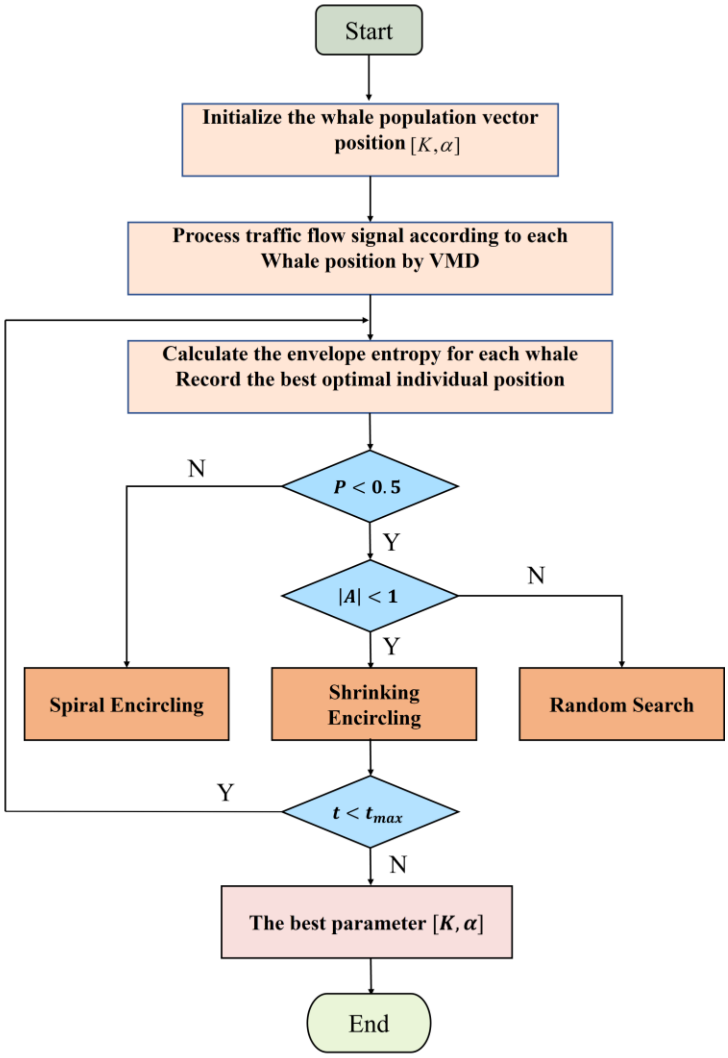

Fig. 2: WOA-VMD flow chart

2.3 SSA-LSTM

LSTM networks are included in the class of neural network models invented by Hochreiter and Schmidhuber in response to the drawbacks of RNN for time-series prediction training as pointed out by Bengio in 1994 [21].

Because the LSTM adds a memory cell component to the hidden layer. It can provide a possible method to tackle the challenges of diminishing and exploding gradients in RNN. The outcome of the input gate is to selectively restore new parameters into the cell state, the forgetting gate is to selectively forget the less important information in the cell, and the output gate is to fetch the pre-stored data into the next neuron [22, 23]. After undergoing VMD, the raw traffic flow data is optimized to reduce the impact of noise. However, the non-linear effects of noise cannot be completely eliminated. When historical data contains non-linear relationships and noise effects, it can lead to overfitting of the LSTM net framework topology during the training process, significantly reducing prediction precision. The original LSTM network is optimized by BPTT for key parameters such as the number of hidden units, maximum training period, initial learning rate, and L2 regularization factor. BPTT utilizes back-propagation to calculate the gradient of the objective function for each layer and modify the parameters. However, BPTT has a large time complexity and can often lead to local convergence problems. The four principal parameters are optimized using SSA technique in this research paper of LSTM temporal prediction with MSE as the objective function: number of hidden units; maximum training period; initial learning rate; L2 regularization factor relying on the excellent performance of SSA in global search, which not only reduces iterations to find the optimal parameters, but also improves the quality of LSTM parameters [24-25]. This paper utilizes SSA to search the best parameter for LSTM.

SSA is an emerging swarm intelligence algorithm presented by Jiankai Xue and Bo Shen in 2020, inspired by sparrow foraging and anti-predation behavior [26]. Compared with some traditional evolutionary class algorithms, such as genetic algorithm (GA) [27], ant colony algorithm (ACO) [28], PSO) [29], etc., SSA has better convergence speed and convergence accuracy due to its unique explorer-follower merit-seeking mechanism, which drives its better aggregation of individuals and reduces the likelihood of the algorithm trapping into local optimum.

3. Experiment

3.1 Data Description

In this paper, this paper utilizes the CHICAGO DATA PORTAL database, w ith the data last accessed on [2023-08-22]. For more information, please visit: http s://data.cityofchicago.org/Transportation/Chicago-Traffic-Tracker-Historical-Con gestion-Esti/sxs8-h27x.

3.2 WOA optimized VMD parameters

To obtain the best VMD parameters, we extracted a subset of traffic flow data to train the WOA-VMD algorithm parameter. Our experiments involved 50 whale populations and the count of 200 maximum iteration. The best VMD parameters k and α are 10 and 1950.

For the original VMD algorithm the dominant frequency method is used to select the most appropriate K value, the main core idea of which is to observe the smallest difference between the central frequency values of adjacent modal components at different K values and select the former as the best K value, but this approach often results in frequency confounding. This paper reports on a large number of decomposition experiments performed on the original data. Therefore, the WOA algorithm tendered in this paper can solve the two important parameters and relatively intelligently and efficiently. the trend diagram of the VMD decomposition of IMF after optimizing the parameter processing is shown in Fig. 3.

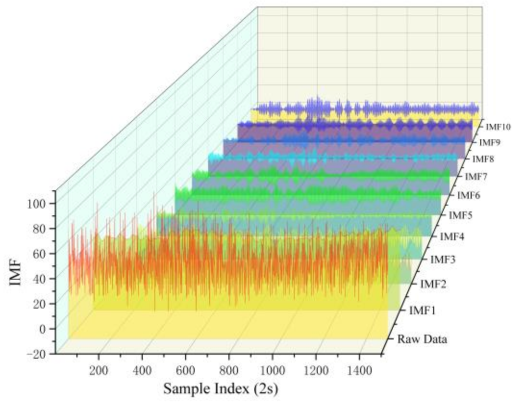

Fig.3: Decomposition trend results of WOA-VMD

3.3 Evaluation indicators

To better identify similarities and differences between the various time-series prediction models, Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), Coefficient of Determination (R^2). The formulas are as follow:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \hat{x}_i\right)^2 \tag{4.}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \hat{x}_i\right)^2} \tag{5.}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{N}\frac{\left|x_i - \hat{x}_i\right|}{\hat{x}_i} \tag{6.}$$

$$R^2 = 1 - \frac{\sum_i^n \frac{\left(\hat{x}_i - x_i\right)^2}{n}}{\sum_i^n \frac{\left(x_i - \overline{x}_i\right)^2}{n}} \tag{7.}$$

n denotes the number of samples taken; x_i denotes the actual value of the traffic flow; $\hat{x}_i$ denotes the predicted value of the model; $\overline{x}_i$ denotes the average of the actual values of the traffic flow.

3.4 Experimental setting

All experiments in this paper were conducted under Window 11 64-bit operating system. The processor of this computer is AMD Ryzen 75800H with Radeson Graphics 4.05 GHz. 24.0 GB of running memory, containing 8 cores and 16 logical processors. The experimental environment is Anaconda 4.13.0 and MATLAB 2022.

3.5 Data pre-processing

To enhance the exactitude of the later model predictions, the raw traffic data is normalized by to constrain the data to be in the range [0,1].

3.6 Experiment Result

Table 1

Evaluation results of models

Prediction models		Processing Times(s)		Evaluation indicators				
		Train	Test	MSE	RMSE	MAE	MAPE	R^2
BP	M1	5406.12	0.0026	305.29978	17.47283	16.9992	0.27367	0.73633
CSO-BP	M2	7792.15	0.0021	199.72246	14.13232	14.8264	0.259562	0.752438
ELM	M3	**4632.51**	0.0024	146.30596	12.0957	14.1948	0.255421	0.759579
SVR	M4	6227.26	0.0063	140.57659	11.8565	13.6932	0.249961	0.766039
LSTM	M5	6510.69	0.0054	126.25893	11.2365	13.1628	0.241976	0.771024
KELM	M6	9142.17	0.0041	120.4616	10.9755	16.9258	0.234688	0.785312
IGWO-SVR	M7	6012.82	0.0095	98.50364	9.9249	14.9016	0.19053	0.82947
POA-ELM	M8	4914.30	0.0015	63.60381	7.9752	10.1508	0.161507	0.864493
GA-KELM	M9	8541.95	0.0027	50.77275	7.1255	9.3028	0.156189	0.873811
EMD-LSTM	M10	7062.83	0.0046	45.86676	6.7725	7.8422	0.127498	0.892502
VMD-LSTM	M11	7486.30	0.0015	18.18425	4.2643	5.2014	0.098896	0.921104
CNN	M12	9421.22	0.0165	11.19906	3.34651	4.5989	0.04003	0.95997

WOA-VMD-SSA-LSTM	M13	8103.77	**0.0009**	**8.57506**	**2.92832**	**1.73981**	**0.023967**	**0.98603**

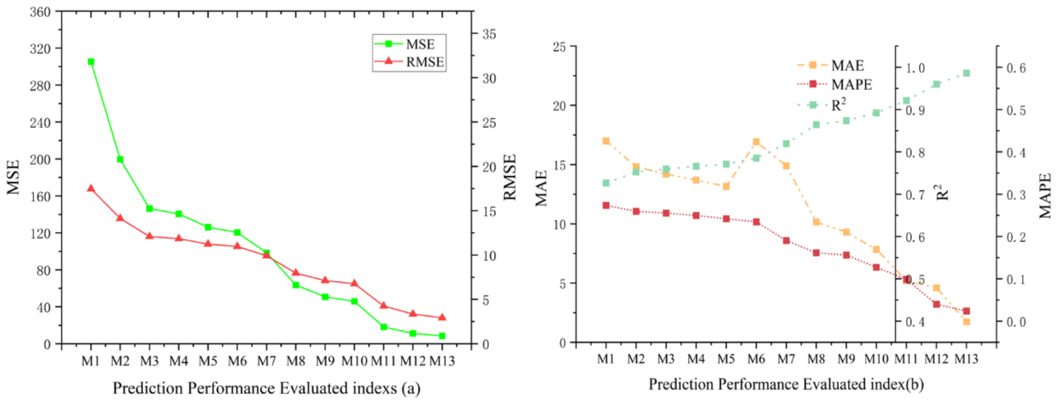

Fig. 4: Prediction Performance of 5 indexes

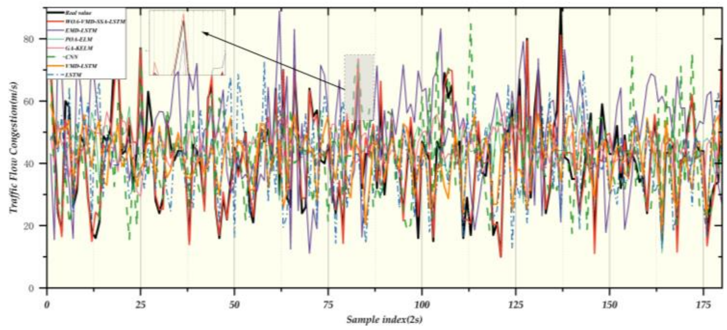

Fig. 5: Prediction Performance compared with M10, M8, M9, M12, M11, M5

Table 1 demonstrates that the M13 model introduced in this paper outperforms the comparison model with regard to prediction accuracy again. The MSE of M13 is 8.57506, which is 23.43%, 52.84%, 81.30%, 83.11%, 86.51% and 93.20% lower compared to M12, M11, M9, M8, M10 and M5 respectively. The RMSE of this model is 2.92832, which is 12.49%, 31.32%, 56.76%, 58.90%, 63.28% and 73.94% lower than that of M12, M11, M9, M8, M10, M5, respectively. The MAE of this model is 1.73981, which is 62.17%, 66.55%, 77.82%, 81.30%, 82.86% and 86.78% lower than that of M12, M11, M9, M8, EMD-LSTM, LSTM respectively. The MAPE of this model is 2.40%, which is 40.13%, 75.77%, 81.20%, 84.65%, 85.16% and 90.10% lower than that of M12, M11, M9, M8, M10 and M5 respectively. For training and testing times for each prediction model. Although the proposed M13 takes slightly longer to train than the M5, it provides better predictions. It uses WOA for preprocessing traffic flow data, predicts with LSTM on decomposed data, and adjusts LSTM parameters using SSA. Despite longer training time, it offers higher prediction accuracy.

The mean prediction time of each model is only 0.0036 seconds, facilitating real-world prediction applications. Although the proposed model's training time is not reduced, it achieves superior prediction parameters, consuming only 0.0009 seconds for practical prediction. With ongoing advancements in computer hardware and parallel computing technology, the actual training time of the model is expected to significantly improve in the future. From the error comparison graphs in Fig.4(a)(b) and Fig. 5 examining the trend of this paper's model against other comparison models, it is a well-known fact that the inaccuracy between the prediction results and the actual traffic flow values is relatively large when the raw traffic data is not processed by data decomposition methods. Among the three algorithms (M10, M11, M13) that processed the data by EMD and VMD, the prediction outcomes of the algorithm presented in this paper, proposed WOA-VMD-SSA-LSTM prediction model, differed the least from the actual values. Comparing the data of M10 and M11, the outcomes indicate that the prediction results using the data separated by VMD are better than those of the EMD decomposed data. The use of VMD decomposition can adequately reduce the volatility of the data, thus improving the accuracy of the prediction method. Meanwhile, this framework innovatively uses the WOA to optimize the parameters of VMD, so that it can achieve the best data decomposition effect when the data is model decomposed. Utilizing the SSA bio-intelligence algorithm, demonstrates significantly improved prediction compared to the model without the swarm intelligence algorithm. Therefore, applying the SSA optimization process for traffic data after VMD processing can further reduce prediction errors. The presented prediction model effectively balances the exploration and local optimization seeking, avoiding the challenge of the target getting trapped in a local optimal solution. VMD efficiently reduces data volatility and extracts more effective features for traffic flow time-series data, while SSA optimized LSTM reduces relative error, demonstrating the precision of the model method. The WOA-VMD-SSA-LSTM model is recommended, as it exhibits the best prediction evaluation index, proving higher prediction precision in the face of non-linearity and data mutation. For M10 and M11, M11 performed better in actual prediction compared to the former. The improvement was 4.75%. This indicates that EMD has certain drawbacks when dealing with noisy traffic data. In response to the human factor interference in the parameter selection of VMD, this paper innovatively utilizes the WOA to globally optimize the selection of decomposition layers and penalty factors of VMD, which greatly reduces the subjective influence of human factors on the model decomposition.

4. Conclusions

Efficient and accurate traffic predictions can significantly mitigate unexpected events and enhance the transport sector's performance. In this paper, our proposed WOA-VMD-SSA-LSTM algorithm shows potent capabilities of handling non-stationary time-series traffic prediction. Our trials spanned 12 sets of comparison experiments, enlisting five evaluation metrics to exhibit its performance. The integration of the cutting-edge VMD algorithm, adept at data decomposing, and the SSA optimization algorithm, known for quickly determining optimal hyperparameters without being trapped in local extremes, forms the innovative backbone of our prediction model. Our discussion covers the significance of temporal prediction models, with LSTM's profound role in long-term dependency recording, with the added advantage of effective parameter tuning through SSA. This work signifies a definite stride in exploiting optimization algorithms to reduce prediction errors in non-stationary time-series data. The hybrid model carving a path of improved precision and efficiency in traffic flow prediction models holds promising potential for applications in the real-world traffic environment.

REFERENCES

[1] Lv, Zhihan & Shang, Wenlong. (2022). Impacts of intelligent transportation systems on energy conservation and emission reduction of transport systems: A comprehensive review. Green Technologies and Sustainability. 100002. 10.1016/j.grets.2022.100002.

[2] Lan T, Zhang X, Qu D, et al. Short-Term Traffic Flow Prediction Based on the Optimization Study of Initial Weights of the Attention Mechanism. Sustainability, 2023, 15 2): 1374.

[3] Shaygan, Maryam & Meese, Collin & Li, Wanxin & Zhao, Xiaoliang & Nejad, Mark. (2022). Traffic prediction using artificial intelligence: Review of recent advances and emerging opportunities. Transportation Research Part C: Emerging Technologies. 145. 103921. 10.1016/j.trc.2022.103921.

[4] Aljuaydi, Fahad & Wiwatanapataphee, Benchawan & Wu, Yong. (2022). Multivariate machine learning-based prediction models of freeway traffic flow under non-recurrent events. Alexandria Engineering Journal. 10.1016/j.aej.2022.10.015.

[5] Khan, Sulaiman & Nazir, Shah & García-Magariño, Iván & Hussain, Anwar. (2021). Deep learning-based urban big data fusion in smart cities: Towards traffic monitoring and flow-preserving fusion. Computers & Electrical Engineering. 89. 106906. 10.1016/j.compeleceng.2020.106906.

[6] Khan, Sulaiman & Nazir, Shah & García-Magariño, Iván & Hussain, Anwar. (2021). Deep learning-based urban big data fusion in smart cities: Towards traffic monitoring and flow-preserving fusion. Computers & Electrical Engineering. 89. 106906. 10.1016/j.compeleceng.2020.106906.

[7] Li Z, Wang X, Yang K. An Effective Self-Attention-Based Hybrid Model for Short-Term Traffic Flow Prediction. Advances in Civil Engineering, 2023, 2023.

[8] YAN J, YU Z Q, ZHU S M. Traffic flow prediction model based on time series decomposition. Journal of Computer Applications, 2023: 0.

[9] H. Lu and F. Yang, "A Network Traffic Prediction Model Based on Wavelet Transformation and LSTM Network," 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS), 2018, pp. 1-4, doi: 10.1109/ICSESS.2018.8663884.

[10] Yang, Hong & Cheng, Yuanxun & Li, Guohui. (2022). A new traffic flow prediction model based on cosine similarity variational mode decomposition, extreme learning machine and iterative error

compensation strategy. Engineering Applications of Artificial Intelligence. 115. 105234. 10.1016/j.engappai.2022.105234.
[11] H. Dong, L. Jia, X. Sun, C. Li and Y. Qin, "Road Traffic Flow Prediction with a Time-Oriented ARIMA Model," 2009 Fifth International Joint Conference on INC, IMS and IDC, 2009, pp. 1649-1652, doi: 10.1109/NCM.2009.224.
[12] Zhang, Yunlong & Zhirui, Ye. (2008). Short-Term Traffic Flow Forecasting Using Fuzzy Logic System Methods. Journal of Intelligent Transportation Systems. 12. 102-112. 10.1080/15472450802262281.
[13] Balsys, K & Valinevicius, Algimantas & Eidukas, D. (2010). Traffic Flow Detection and Forecasting. Elektronika ir Elektrotechnika.
[14] Chen, Yiran & Xie, Yuan & Song, Linghao & Chen, Fan & Tang, Tianqi. (2020). A Survey of Accelerator Architectures for Deep Neural Networks. Engineering. 6. 10.1016/j.eng.2020.01.007.
[15] Xing, Yi-ming & Ban, Xiao-juan & Liu, Ruoyi. (2018). A Short-Term Traffic Flow Prediction Method Based on Kernel Extreme Learning Machine. 533-536. 10.1109/BigComp.2018.00089.
[16] Tian, Yan & Zhang, Kaili & Li, Jianyuan & Lin, Xianxuan & Yang, Bailin. (2018). LSTM-based Traffic Flow Prediction with Missing Data. Neurocomputing. 318. 10.1016/j.neucom.2018.08.067.
[17] Lian, Jijian & Liu, Zhuo & Wang, Haijun & Dong, Xiaofeng. (2018). Adaptive variational mode decomposition method for signal processing based on mode characteristic. Mechanical Systems and Signal Processing. 107. 53-77. 10.1016/j.ymssp.2018.01.019.
[18] Yang, Hong & Cheng, Yuanxun & Li, Guohui. (2021). A denoising method for ship radiated noise based on Spearman variational mode decomposition, spatial-dependence recurrence sample entropy, improved wavelet threshold denoising, and Savitzky-Golay filter. Alexandria Engineering Journal. 60. 3379-3400. 10.1016/j.aej.2021.01.055.
[19] Mirjalili, Seyedali, Lewis, & Andrew. (2016). The whale optimization algorithm. Advances in engineering software.
[20] JIANG Yonghua, TANG Baoping, LIU Wenyi, et al. Feature extraction method based on parameter optimized Morlet wavelet transform. Chinese Journal of Scientific Instrument, 2010, 31(1): 56-60.
[21] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," in Neural Computation, vol. 9, no. 8, pp. 1735-1780, 15 Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.
[22] Alex Sherstinsky. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. Physica D: Nonlinear Phenomena, 2020,404(C): 132306-132306.
[23] Schmidhuber J. Deep learning in neural networks: an overview. Neural Netw. 2015 Jan; 1: 5-117. doi: 10.1016/j.neunet.2014.09.003. Epub 2014 Oct 13. PMID: 25462637.
[24] LI Chun, GAO Fei, WANG Huiqing. Improved fruit fly optimization algorithm for optimizing time series prediction model of CIAO-LSTM network. Computer Engineering and Applications, 2020, 56 (1): 9-134.
[25] Zhang X, Zhong C, Zhang J, et al. Robust Recurrent Neural Networks for Time Series Forecasting. Neurocomputing, 2023.
[26] Xue, Jian-Kai & Shen, Bo. (2020). A novel swarm intelligence optimization approach: sparrow search algorithm. Systems Science & Control Engineering. 8. 22-34. 10.1080/21642583.2019.1708830.
[27] A. Lambora, K. Gupta and K. Chopra, "Genetic Algorithm- A Literature Review," 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), 2019, pp. 380-384, doi: 10.1109/COMITCon.2019.8862255.
[28] Ansari S, Bansal R C, Elsahar Y. Optimization of renewable energy-based transmission system: an application of the genetic, ant colony optimization, Q-learning, and Dijkstra’s algorithms. International Journal of Modelling and Simulation, 2023: 1-20.
[29] J. Kennedy and R. Eberhart, "Particle swarm optimization," Proceedings of ICNN'95 - International Conference on Neural Networks, 1995, pp. 1942-1948 vol.4, doi: 10.1109/ICNN.1995.488968.