Environmental Monitoring and Predicting Using Graph-Based Series Decomposition Model

Mughair Aslam Bhatti[1], Uzair Aslam Bhatti[1*], Muhammad Aamir[5], Sarkar Sabarata Kumar[2], Maqbool Khan[3], Yonis Gulzar[4]

[1]School of information and Communication Engineering, Hainan University, China
[2]Department of Mechatronics Engineering Engineering, Rajshahi University of Engineering & Technology, Bangladesh
[3]Pak-Austria Fachhochschule-Institute of Applied Sciences and Technology, Mang, Haripur, 22621, Pakistan
[4]Department of Management Information Systems, King Faisal University, P.O. Box 380, Al-Ahsa, 31982, Saudi Arabia
[5]College of Computer Science, Huanggang Normal University, Huanggang 438000, China
uzair@hainanu.edu.cn

Abstract. Better methods of forecasting the concentrations of air pollution are important in developmental of sound policies and decisions at a time when environmental concerns and the need for mitigating climate change has become important. This paper describes how climate pollution can be predicted from climate data streams by the integrated design of GCNs and EEMD with PSO. The ability of the PSO-EEMD-GCN model to predict the said six major air pollutants: $PM_{2.5}$, PM_{10}, CO, NO_2, O_3, and SO_2. Closely, the effectiveness of the current model is assessed with aid of vital parameters including RMSE, MAE, MSE, MAPE, and R-Squared. Drawing from the evaluation results, success reveals that PSO-EEMD-GCN is an efficient predictive model for climate pollution. For the Chinese consumers, the "Willingness to Pay Model" budget accuracy is 23, stated in terms of RMSE. 92 and an MAE of 16. The high correlation means explains why the correlation coefficients for SO_2 and PM_{10} are much higher for the rural areas, and the low RMSE means that is why the RMSE for the total values is 86 for PM_{10}, but an RMSE of 17. As for the MAE, we analyzed data derived from 85 state examination results and attained an average of 65 and an MAE of 11. 71 for $PM_{2.5}$. The values of the model also show the practical significance with RMSE of 0. 28 and the MAE is 0. 16, which proves that it is appropriate at predicting the CO levels. Even for ozone (O_3), which is considered as the most challenging chemical, the predictor gives a pretty good RMSE of 16.09 and an MAE of 12.17 and nitrogen dioxide with an Root Mean Square Error (RMSE) of 11.24 and an MAE of 7.50. It is seen that for sulfur dioxide (SO_2) forecasts the error metrics are higher but the system remains predictive. Therefore solving, the PSO-EEMD-GCN model has shown great promise in being a helpful tool for forecasting climate pollution.

W. Guo et al. (eds.), *Proceedings of the 3rd International Conference on Green Building, Civil Engineering and Smart City (GBCESC 2024)*, Advances in Engineering Research 264,
https://doi.org/10.2991/978-94-6463-728-1_69

Keywords: Climate pollution prediction, PSO, EEMD, GCN, Air pollutant concentrations

1 Introduction

During the present generation where technology is rapidly developing, coupled with the increased rate at which data is being produced, the application usage of time series forecasting is becoming increasingly significant in almost all fields. Examples of time series include stock prices, temperature, or the amount of product sold at a particular time; a time series is defined as a series of data points that are recorded at a given time [1]. Thus, the specific aim of time series forecasting is to identify future trends and patterns from historical data in order to support decision making. Although, due to the general properties of time series data as being stochastic and non homogeneous in most cases, additional challenges arise and to overcome these complex time series data requires the use of more statistical and machine learning algorithm to arrive at valid prediction models [2][3]. The data here has a lot of complexities which include; seasonal variations, cyclic trends and normally considered as noisy data and this has to be managed to be able to predict future trends.

More specifically, deep learning has appeared in the last few years as revolutionary methodology able to build multilevel representations for the initial input data through combining the simple nonlinear modules [4]. Thus, each module of deep neural networks improves representation by taking it to a higher level of abstraction. Such repeated transformations help in effectively capturing as many features as possible from the input data, this is according to deep learning [5]. Deep learning is used in multiple fields of natural language processing [6], speech recognition [3], and even computer vision [4], [5]. This results in traditional forecasting methods formulating ineffective strategies and plans seeing that they are unable to adequately analyze the specific characteristics of time series [7][8]. To this end, the present paper presents a new predictive model work based on deep learning technique in combination with signal decomposition and parameter optimization to improve the quality of prediction in time series context. It also enables comprehensive learning of sequence data features over the interval length of the sequence data itself. Our key contributions include:

1. Presenting a new forecasting model based on deep learning and parameter optimization with a new signal decomposition approach.
2. Employing EEMD method to analyse the obtained data which in turn scales down the complexity of the data. Subsequently, utilizing GCN for the determination of data features and optimization of the model parameters with the help of PSO algorithm to enhance the prediction capability of the model.
3. More specifically, the model is tested and the prediction accuracy and stability are examined by performing experiments on a meteorological dataset of Hubei Province.

Finally, the proposed model has great potential to further enrich the field of time series forecasting as the data set can be substantially non-linear and non-stationary.

2 Related work

When delving into the realm of time series prediction, three primary avenues of exploration emerge: mathematical methods, artificial algorithms, and combinations of the two. The widening of the search goes from the standard and the incorporation of the machine learning augmenting a vast deep learning array. Signal decomposition as a preliminary step followed by time series prediction, gives a holistic approach to the prediction of time series data. However, in the earlier periods during the time series forecasting research, statistical approaches formed the basis. Other scholars such as Fong L C et al. [9] and Khan MM et al., [10] applied the ARMA model in the determination of short-term rainfall and tourist demand. Further development of the field led to the migration to the ARIMA model used by Chen B & Wu J. [11] for air transportation demand and Tehrani S et al. [12] for electricity price prediction. Nevertheless, statistical models are not always effective to handle time series problems even the AM model has achieved many significant success factors [13].

In the cases of non-linear data extraction, Malvoni et al. [14] established that SVM can be used to predict photovoltaic power generation, as well as the fluctuations in stock prices. Therefore, using the power of Fitting of Complex data with Artificial Neural Networks (ANN) S V Lakshminarayana et al. [15] predicted well for rainfall. Machine learning, particularly deep learning is today's most widely embraced technique because of its ability to handle complex trends in data. Other types of neural networks include Recurrent Neural Networks (RNN) and the modified optimized versions of the same, including Long Short-Term Memory (LSTM) Networks, which have been proved to be very successful in many time series tasks, as reenacted by Y Tian & L Pan [16] for traffic flow prediction. Even more, relatively new directions of signal decomposition, such as EMD, have been complemented by their modifications, as well as integrated with machine learning approaches, and have improved the predictive ability of models even further. For example, Chen N et al present a WEF system by applying EMD and SVM [17], while Zhao Z et al [18] utilizes VMD with CNN and GRU for feature extraction and modelling.

Over the course of the last years, the interest in the usage of ensemble approaches for time series forecasting has risen steadily. These models usually incorporate signal analysis and predictability where tools like EEMD and VMD are used. For instance, Qin Q et al. employed EEMD in conjunction with LSTM to forecast wind speed [19][20] and Y Qiu et al. used EEMD with Graph Convolutional Networks, for precipitation estimation [21]. These kinds of mixed models have shown better performances over single measure models, thereby implying that by integrating signal decomposition with a more progressive machine learning machinery, the considerably challenging characterization of time series' might just be possible.

3 Method

3.1 PSO-EEMD-GCN

Therefore, this study put forward the model of EEMD signal decomposition approach and the GCN deep learning technology and the Particle Swarm Optimization Algorithm to obtain high precision in predicting the time series data. Time series data is complex inherently because of high level of noise, nonlinearity and non-stationarity. With help of this approach, EEMD separates these series into several Intrinsic Mode Functions (IMFs), that enhances decomposition's capability to reveal periodicity and local properties. With this decomposition, a more refined knowledge of the multitude of frequency components in the data becomes feasible. GCNs are quite effective in capturing complex relation between nodes (data points) of a time series when the data is graph structured. Therefore, from the viewpoint of the data structure, GCNs provide a comprehensive view of the adjacency relationships while acquiring information.

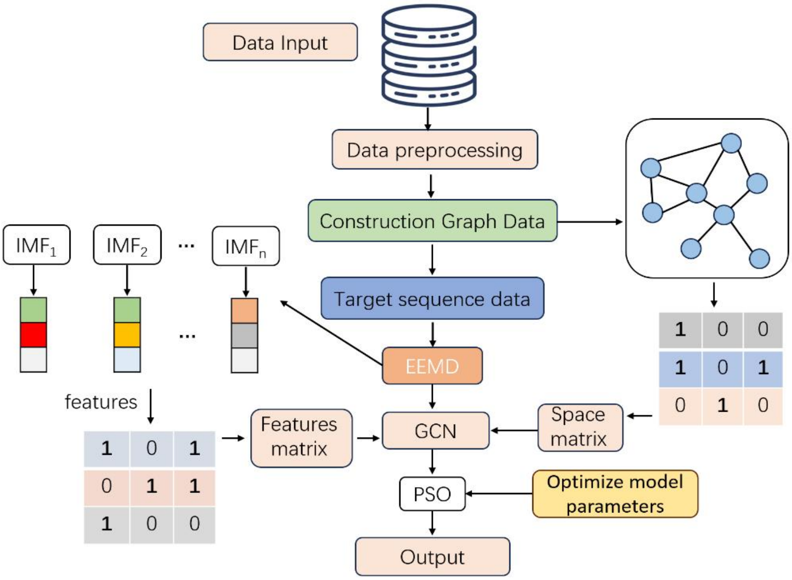

Fig. 1. The flow chart of the proposed model

The PSO is also helpful in model parameter fine-tuning since this algorithm cyclically searches for values that optimize the expected performance by fine-tuning the model to the data being used.

The modeling process consists of the following specific steps: The modeling process consists of the following specific steps:

1. To apply EEMD, decompose raw time series data into IMF components.
2. By employing these components of the IMF as input features to the introduced Graph Neural Network.

3. Evaluating and assembling a model, where information is obtained, intrinsic connections are determined, and prospective results are forecasted with the help of a Graph Neural Network.

4. As a measure to improve the accuracy of the forecast, the Particle Swarm Optimization (PSO) of model parameters will have to be done.

The structure of the entire proposed model is illustrated in figure 1.

4 Results

The performance measure of the different models that were built to predict the concentration of the air pollutants with the help of PSO, EEMD, GCN is depicted in figure 2. Air pollutants which have been considered in this study entails $PM_{2.5}$ some which include 5, PM_{10}, CO, NO_2, O_3 as well as SO_2. Also, it highlights the Root Mean Square Error (RMSE), which shows the quantity of the prediction error of the model in question, where the lower amount here is preferable.

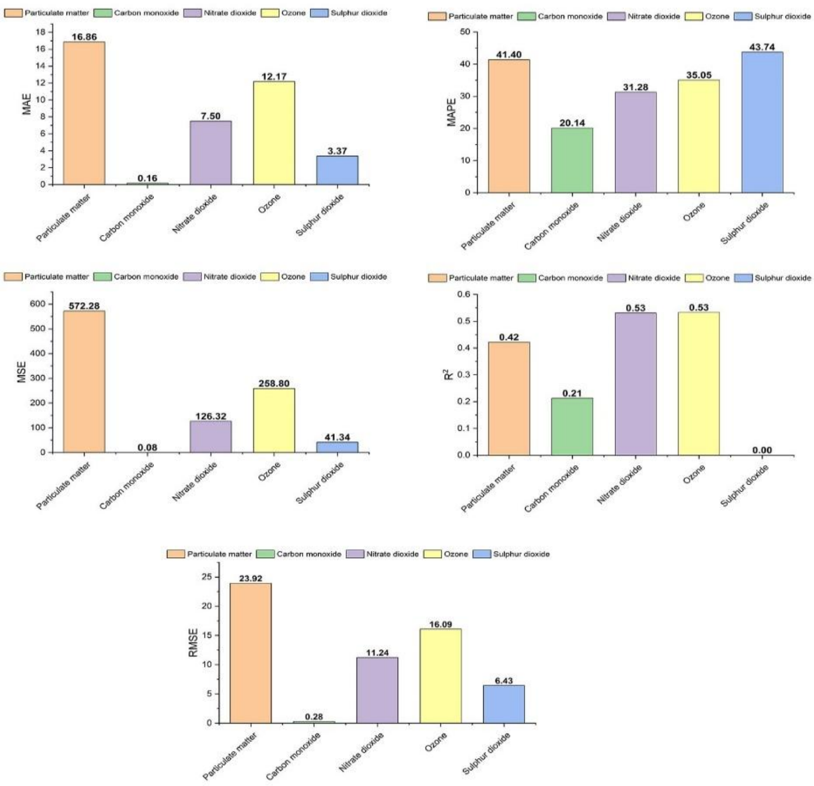

Fig. 2. The flow chart of the proposed model

For instance, the difference between the predicted and actual numbers manifested in RMSE is relatively low for $PM_{2.5}$ in the PSO-EEMD-GCN model is 17.650768, of which the figure indicates moderate prediction errors. The Mean Absolute Error (MAE), another measure of accuracy, averages the absolute differences between predicted and actual values, with lower values being preferable. The CO model's MAE of 0.162368 indicates relatively accurate predictions. Mean Squared Error (MSE) penalizes larger errors more heavily and, like RMSE, prefers lower values. The O_3 model's MSE of 258.797242 reflects significant prediction errors. Mean Absolute Percentage Error (MAPE) expresses the error as a percentage, with lower values being better. The CO model's MAPE of 20.138403 suggests predictions are off by about 20% on average. R-squared (R^2) measures how well the model explains the variance in the data, ranging from 0 to 1. The NO_2 model's R^2 of 0.530644 indicates it explains 53% of the variance in NO_2 concentrations.

The Mean Absolute Error (MAE), an accuracy measure, is the mean of the absolute differences between the predictions and the actual values and is preferred at a lower value. Autonomy's CO model's MAE was a little higher at 0.162368 is regarded as moderate and close to the real situation of predictions. Mean Squared Error (MSE) also discourages larger errors and like RMSE – the lower the better. At that time the O_3 model still has the lowest MSE of 258.797242 In the present analysis, Mean Absolute error is another commonly used measure of accuracy to denote the error in terms of the percentage and it is calculated by dividing the mean by hundred. On the average, the CO model has an average absolute percentage error (MAPE) of 20% regarding the number of retail outlets 20.138403 implies that predictions are quite off by approximately 20% on the average. Adjusted R-squared tests for the goodness of the model concerning the variance of the data and obtains its maximum value of one. The NO^2 model's R^2, arrived at through 250 iterations of steepest descent, is 0. Analyzing the correlation for 0.530644' it shows that it accounts to 0.53 meaning that it can explain 53% variation in NO_2 levels.

5 Conclusion

This research work has therefore brought a positive change in the advancement of environmental science especially in the climate pollutant prediction. To this end, PSO, EEMD, and GCN are combined innovatively for the development of a robust model for predicting air pollutant concentrations. The proposed PSO-EEMD-GCN model performs well in the prediction of levels of main air pollutants including $PM_{2.5}$, PM_{10}, CO, NO_2, O_3, and SO_2 of the study area data and evaluate the model's performance using four statistical measures, RMSE, MAE, MSE, and MAPE, as well as a statistical significance test and R-squared coefficient. By using the findings of this model's reliability, one can conclude about its progressive applicability as the tool for environment monitoring and decision-making in air quality regulation, strengthening the predictive qualities that are necessary for effective problem-solving.

References

1. Allioui, H.; Mourdi, Y. Exploring the Full Potentials of IoT for Better Financial Growth and Stability: A Comprehensive Survey. Sensors. 23 (19), 8015, (2023).
2. Kashpruk N, Piskor-Ignatowicz C, Baranowski J. Time Series Prediction in Industry 4.0: A Comprehensive Review and Prospects for Future Advancements. Applied Sciences. 13(22), 12374, (2023).
3. Hahn Y, Langer T, Meyes R, Meisen T. Time Series Dataset Survey for Forecasting with Deep Learning. Forecasting. 2023, 5(1), 315-335. (2023)
4. Ahmed, S.F., Alam, M.S.B., Hassan, M. et al. Deep learning modelling techniques: current progress, applications, advantages, and challenges. Artificial Intelligence Review. 56, 13521–13617, (2023).
5. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, Santamaría J, Fadhel MA, Al-Amidie M, Farhan L. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Journal of Big Data. 8(1), 53, (2021).
6. Yang Y, Wu Z, Yang Y, Lian S, Guo F, Wang Z. A Survey of Information Extraction Based on Deep Learning. Applied Sciences. 12(19), 9691, (2022).
7. Linardatos P, Papastefanopoulos V, Kotsiantis S. Explainable AI: A Review of Machine Learning Interpretability Methods. Entropy. 23(1), 18, (2021).
8. Bhatti, Uzair & Mengxing, Huang & Li, Jingbing & Bazai, Sibghat & Aamir, Muhammad. (2023). Deep Learning for Multimedia Processing Applications: Volume One: Image Security and Intelligent Systems for Multimedia Processing. 10.1201/9781003427674.
9. Fong-Lin Chu, Forecasting tourism demand with ARMA-based methods, Tourism Management, 30(5), 740-751, (2009).
10. Khan, M. M. H., Mustafa, M. R. U., Hossain, M. S., Shams, S., & Julius, A. D. Short-Term and Long-Term Rainfall Forecasting Using ARIMA Model. International Journal of Environmental Science and Development, 14(5), 292-298, (2023).
11. Chen B, Wu J. Predicting Model for Air Transport Demand under Uncertainties Based on Particle Filter. Sustainability. 14(24), 16694, (2022).
12. Tehrani S, Juan J, Caro E. Electricity Spot Price Modeling and Forecasting in European Markets. Energies.15(16), 5980, (2022).
13. Bhatti, Uzair & Tang, Hao & Khan, Asad & Ghadi, Yazeed & Bhatti, Mughair & Khan, Khalid. (2024). Investigating the nexus between energy, socio-economic factors and environmental pollution: A geo-spatial multi regression approach. Gondwana Research. 130. 308-325. 10.1016/j.gr.2024.02.007.
14. Liu Y-W, Feng H, Li H-Y, Li L-L. An Improved Whale Algorithm for Support Vector Machine Prediction of Photovoltaic Power Generation. Symmetry. 13(2), 212, (2021).
15. S V, Lakshminarayana & Singh, P.K. & Kothari, Prof. Mahesh & Sharma, Deepak & Yadav, K.K. Rainfall Forecasting using Artificial Neural Networks (ANNs): A Comprehensive Literature Review. 8(4), 589-599, (2020).
16. Y. Tian and L. Pan, "Predicting Short-Term Traffic Flow by Long Short-Term Memory Recurrent Neural Network," 2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity), Chengdu, China, 2015, pp. 153-158.
17. Chen N, Sun H, Zhang Q, Li S. A Short-Term Wind Speed Forecasting Model Based on EMD/CEEMD and ARIMA-SVM Algorithms. Applied Sciences. 12(12), 6085, (2022).
18. Zhao, Zeni & Yun, Sining & Jia, Lingyun & Guo, Jiaxin & Meng, Yao & He, Ning & Li, Xuejuan & Shi, Jiarong & Yang, Liu. Hybrid VMD-CNN-GRU-based model for short-

term forecasting of wind power considering spatio-temporal features. Engineering Applications of Artificial Intelligence. 121(4610), 105982, (2023).

19. Bhatti, Uzair & Bhatti, Mughair & Tang, Hao & M.S., Syam & Awwad, Emad & Sharaf, Mohame & Ghadi, Yazeed. (2023). Global production patterns: Understanding the relationship between greenhouse gas emissions, agriculture greening and climate variability. Environmental Research. 245. 118049. 10.1016/j.envres.2023.118049
20. Qin Q, Lai X, Zou J. Direct Multistep Wind Speed Forecasting Using LSTM Neural Network Combining EEMD and Fuzzy Entropy. Applied Sciences. 9(1), 126, (2019).
21. Qiu, Y., Lu, Z., & Tang, H. A short-term regional precipitation prediction model based on wind-improved spatiotemporal convolutional network. Earth and Space Science, 9(9), e2022EA002411, (2022).