DWT Denoising for Multi-variate Time Series Forecasting

Citation for published version:
Wan, Z, Gong, P, Li, X, Zhang, Z, Yang, F & Pang, W 2022, DWT Denoising for Multi-variate Time Series Forecasting. in X Jiang, W Tao, D Zeng & Y Xie (eds), *Fourteenth International Conference on Digital Image Processing (ICDIP 2022).*, 1234237, Proceedings of SPIE, vol. 12342, SPIE, 14th International Conference on Digital Image Processing 2022, Wuhan, China, 20/05/22. https://doi.org/10.1117/12.2644384

Digital Object Identifier (DOI):
10.1117/12.2644384

Published In:
Fourteenth International Conference on Digital Image Processing (ICDIP 2022)

DWT denoising for multi-variate time series forecasting

Zhifeng Wan, Peng Gong, Xin Li, Zhen Zhang, Fuhao Yang, et al.

Zhifeng Wan, Peng Gong, Xin Li, Zhen Zhang, Fuhao Yang, Wei Pang, "DWT denoising for multi-variate time series forecasting," Proc. SPIE 12342, Fourteenth International Conference on Digital Image Processing (ICDIP 2022), 1234237 (12 October 2022); doi: 10.1117/12.2644384

Event: Fourteenth International Conference on Digital Image Processing (ICDIP 2022), 2022, Wuhan, China

DWT Denoising for Multi-variate Time Series Forecasting

Zhifeng Wan[a], Peng Gong[a], Xin Li[a*], Zhen Zhang[a], Fuhao Yang[a], Wei Pang[b]
[a]Beijing Institute of Technology, No. 5, South Street, Beijing, China;
[b]Heriot-Watt University, Edinburgh, Scotland, UK

ABSTRACT

Multivariate time series data is ubiquitous in the real world, and the study of its modeling and analysis is a popular research topic in meteorology, transportation, finance and other fields. In these studies, classical statistical methods are primarily aimed at single time series analysis, while deep learning demonstrates the power to mine patterns from massive amounts of data. A major application of these studies is to analyze collected historical sequence information to predict what will happen over time in the future. Currently, recurrent neural network-based models and time-convolution-based models realize the predictive power of multivariate time series, but these deep models perform mediocrely at predicting long-sequence tasks. On the one hand, due to the accumulation of errors, on the other hand, the fact that the collected sequence contains a large amount of high-frequency noise. In order to improve the prediction accuracy of the model and mine more valuable features from the series, we propose a novel multivariate time series prediction framework ADWT for time series modeling. By designing an adaptive filtering module in the characteristics of the signal frequency domain, our model removes noise from some of the time series and builds an end-to-end framework by fusing it with the prediction module of deep learning. Experimental results show that our model can effectively improve the prediction accuracy of multivariate time series, and its performance in the three benchmark data sets is competitive with the latest spatial-temporal series prediction model, and has good interpretability.

Keywords: Discrete Wavelet Transform, Multi-variate time series forecasting, Deep Learning, Data Mining

1. INTRODUCTION

Time series forecast analysis is to use the characteristics of an event time in the past time to predict the characteristics of the event in the future period of time, multi-time series is a current research hotspot, in finance, e-commerce, security supervision and many other fields have a large number of applications.

At present, there are two main methods: first, from a statistical point of view combined with deep learning methods; by calculating the statistical characteristics of the series and the statistical characteristics between the series, using deep learning methods to predict, for example, DeepAR[1], by training LSTM based model to predict the probability distribution parameters of each series, from a statistical point of view to give the distribution probability of the future values of each sequence variable; second, from the perspective of the graph to explore the relationship between the sequences, the multivariate time series as a "spatial-temporal graph" / "dynamic graph" from the perspective of graph theory, instantiate the connection between variables; suppose that some variables have a certain relationship at a certain time, construct a graph that changes with time between variables, and then combine neural network methods of the temporal dimension, such as RNN, and aggregation methods of spatial dimensions, such as the graph neural network (GNN) will converge and update the time series information of the variables themselves and the time series information of the neighbors, get a new representation of each variable, and finally through the mapping function, Map the implicit space's embedding to the predicted value corresponding to each series.

However, most of these studies are raw data from input multivariate time series and do not effectively eliminate noise in the data. In the field of digital signal processing, algorithms such as the Fourier transform wavelet transform show strong sequence analysis capabilities, can analyze the frequency domain characteristics of the sequence, and effectively propose high-frequency noise. We noticed that some of the work began to focus on the preprocessing of input data, such as ST-Norm[2] statistical analysis and regularization of raw data in time and space dimensions. This work demonstrates the tremendous value of time series preprocessing for deep learning models, but they regularize without taking into account the removal of time series data noise, and there is still room for improvement.

*Corresponding author: xinli@bit.edu.cn

Fourteenth International Conference on Digital Image Processing (ICDIP 2022), edited by Xudong Jiang, Wenbing Tao, Deze Zeng, Yi Xie, Proc. of SPIE Vol. 12342, 1234237
© 2022 SPIE · 0277-786X · doi: 10.1117/12.2644384

In this work, we propose a multivariate time series forecasting model ADWT based on wavelet transform noise reduction. The multivariate time series is transformed by the wavelet transform to obtain the wavelet coefficient, then the noise reduction is passed by the threshold function, and finally the inverse wavelet transformation is restored to the time domain, which has more pure data information. And combined with the threshold coefficient of deep learning adaptation, followed by the time prediction module based on deep learning, the end-to-end training model will be more helpful to the prediction module's capture of patterns in the signal. Innovation: 1. We are the first multivariate time series forecasting model that uses wavelet transform to denoising; 2. By comparing the basic models, we can effectively improve the prediction performance of the basic models.

2. RELATED WORK

The core assumption behind Spatial-temporal Graph Networks is that there exists a latent graph among the time series, each of which representing a node of the graph. Such a graph can contribute to the forecasting tasks by revealing fundamental insights into the system described by the time series. Therefore, many works have been devoted to combining time-series relationship inference with forecasting recently. For example, Graph Wavenet[3] uses two modules to model spatial dependencies and temporal correlations respectively, and it designs a self-adaptive matrix to capture complex spatial dependencies through node embeddings. Later on, MTGNN[4] divided the process of time series forecasting into two main components: graph learning and graph-based forecasting. Specifically, MTGNN presents a joint framework for modeling multivariate time series data, and it parameterizes the graph as a k-degree one, which is learned in an end-to-end fashion with a GNN for forecasting multivariate time series. Most recently, to improve efficiency, GTS[5] first uses the whole training dataset of multivariate time series of each node to learn the representation and then samples a discrete adjacency matrix from the edge probabilities in a differentiable way. Although the above models have achieved a certain degree of success, they all focus on mining the spatial-temporal relationships of multivariate time series in the time resolution, ignoring the frequency resolution; thus some scale information is not used which may be helpful for MST forecasting.

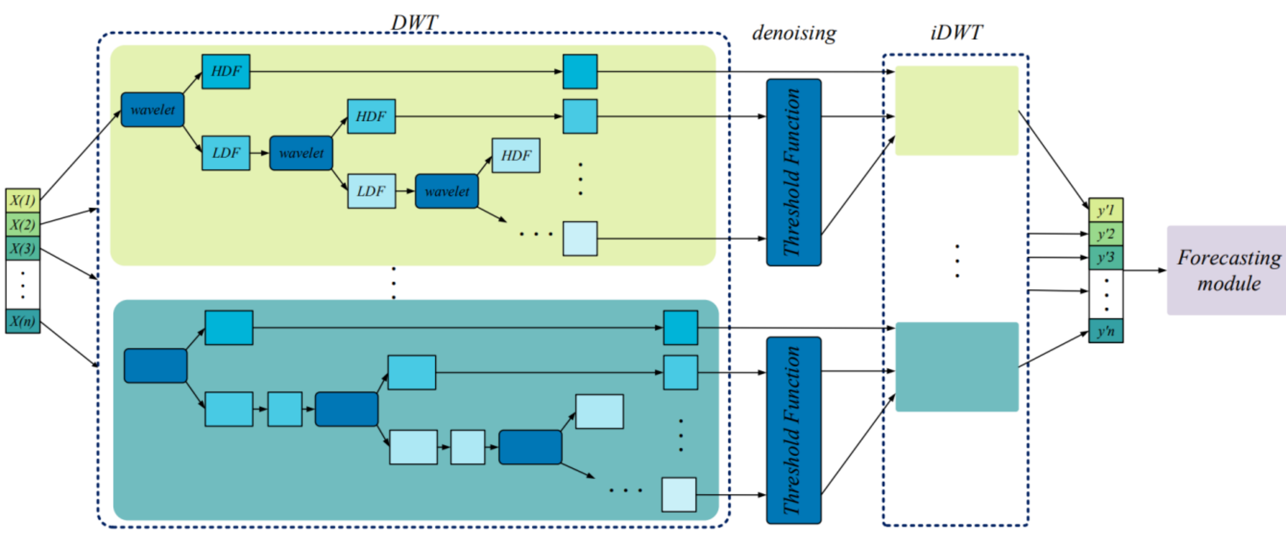

Figure 1. The framework of ADWT.

3. METHODOLOGY

The architecture of ADWT is depicted in Fig. 1. Our model is a time series forecasting model with DWT denoising, which consists of two main components: discrete wavelet transform denoising module and forecasting module. In the rest of this section, we first give the mathematical definition of our proposed model, and we present the overall framework of ADWT.

3.1 Problem Definition

Let $X \in \mathbb{R}^{N \times T}$ denotes the a multivariate time series that contain N variables and each of variable with T time steps. We defined $X = \{x_1, x_2, \cdots, x_N\}$, which contains N multivariate time series, and each time series $x_i = \{x_i^1, x_i^2, \cdots, x_i^T\}$ has a total of T time steps. At each time step t, given its previous h steps of historical observations $X^{(t-h):t}$, our goal is to learn a model f which is able to forecast its next μ steps $X^{(t+1):(t+\mu)} = f(X^{(t-h):t}, \omega)$, where ω is the model parameter. We defined the loss function between the forecasting and the ground truth as the target of optimization, that is:

$$\min \sum_t MSE(f(X^{(t-h):t}, \omega), X^{(t+1):(t+\mu)}) \tag{1}$$

The summing is because of the sliding window mechanism in model training.

3.2 Discrete Wavelet Transform Denoising

The principle is that the energy of the useful signal is concentrated on a small number of wavelet coefficients after the wavelet transformation, while the white noise is still dispersed on a large number of wavelet coefficients in the wavelet transformation domain[6][7]. Thus, relatively speaking, the wavelet coefficient values of the useful signal must be greater than the wavelet coefficient values of noise whose energy is dispersed and the amplitude is small. Therefore, in terms of the amplitude of the spectrum, the useful signal and noise can be separated. The main steps of this method are: (1) Select the appropriate orthogonal wavelet base and decomposition layer j, and decompose the noisy signal by wavelet transformation to the j layer; (2) Perform threshold processing on the wavelet coefficient obtained by decomposition. Threshold function:

$$T = \theta \frac{median(|cD_1|)}{0.6745} \sqrt{2 \ln N} / \log_2(j+1) \tag{2}$$

In the threshold function, θ is the parameter to optimize the threshold, cD_1 is the detail coefficient of the first layer of decomposition of DWT, N is the length of the data, and j is the number of decomposition layers.

3.3 The overall ADWT Framework

The model framework of ADWT is shown in the figure. Our model is a multivariate time series forecasting model, which is mainly based on the adaptive multivariate time series prediction of noise reduction by the discrete wavelet decomposition threshold function. Adaptive wavelet noise reduction is mainly based on the use of parametric threshold function to sift out the high-frequency detail coefficient of the wavelet transform of the time series, which can obtain smoother data. Finally complete the task using various popular time series forecasting modules. By using such noise reduction methods, subsequent time series forecasting models will mine patterns from the data more efficiently while retaining the data more completely.

Table 1. A Summary of Datasets.

Models	Electricity	Solar-AL	Weather
#Variables	336	137	6
#Samples	2,208	8,496	17,288
Sample rate	1 hour	10 minutes	1 hour
Input Length	128	128	128
Predict Length	24	24	24

4. EXPERIMENTS

In this section, we carry out experiments with ADWT and several commonly used baselines in three public datasets to assess the effectiveness of ADWT.

4.1 Experimental Setting

Table 1, we describe the statistics of multi-variate time series datasets. Electricity, Solar-AL and Exchange-rate datasets are shared in work[8]. The Weather dataset is released in work[9]. The datasets are split in chronological order with 60% for training, 20% for validation, and 20% for testing. Following MTGNN, we set the context length to 128 and predict length to 24 to validate the long-range predictive power of ADWT and baselines. The evaluation metrics we choose include mean absolute percentage error (MAPE), mean absolute error (MAE) and root mean squared error (RMSE).

4.2 Baseline Models

We compare ADWT with the following models:

1) LSTM: a standard encoder-decoder LSTM framework.
2) Graph Wavenet[3]: This derives a soft graph where each pair of nodes has a continuous probability of being connected
3) MTGNN[4]: This builds a graph learning module to capture the relationship between time series. MTGNN is an evolution model of Wavenet.
4) ADWT-LSTM: Use an Adaptive Discrete Wavelet Transform (ADWT) to filter the data before LSTM module.
5) ADWT-Graph Wavenet: Use an Adaptive Discrete Wavelet Transform (ADWT) to filter the data before Graph Wavenet module.
6) ADWT-MTGNN: Use an Adaptive Discrete Wavelet Transform (ADWT) to filter the data before MTGNN module.

Table 2. Baselines Comparison of 24 Points Forecasting

Models	Electricity			Solar-AL			Weather		
	MAPE(%)	MAE	RMSE	MAPE(%)	MAE	RMSE	MAPE(%)	MAE	RMSE
LSTM	22.184	31.541	67.188	22.631	1.861	4.086	8.326	2.536	3.376
ADWT-LSTM	21.844	31.090	67.701	20.701	1.940	4.042	8.027	2.463	3.293
Graph Wavenet	18.138	24.548	53.312	20.696	2.055	4.021	7.753	2.410	3.257
ADWT-Graph Wavenet	16.998	24.095	53.271	21.411	2.035	3.978	7.853	2.443	3.291
MTGNN	17.681	23.139	49.801	20.597	1.915	4.100	7.990	2.402	3.226
ADWT-MTGNN	18.642	25.884	53.365	20.854	1.960	3.988	7.537	2.395	3.258

4.3 Prediction Performance

The experimental results are presented in Table 2. We set the prediction length as the next 24 hours in Electricity and Weather datasets, and the next 6 hours in Solar-AL dataset. Table 2 is the average MAPE, MAE, RMSE of 24 points of baselines and ADWT on the three datasets, respectively. LSTM is the classical RNN-based method to handle the sequence. Graph Wavenet and MTGNN perform great in multivariate time series forecasting tasks, especially MTGNN achieving sota effects in many datasets. With the addition of the ADWT module, the prediction effect of these models on the public dataset has been improved to some extent. Because our prediction of 24 points is a long series prediction, the model with error accumulation in LSTM will perform poorly, but after noise reduction, the sequence will become smoother, so there will be a positive performance for suppressing error accumulation. From the experimental data, we can see that the prediction effect improvement on the LSTM model is more obvious. In the Graph Wavenet and MTGNN models, the input data is first convoluted in one dimension, and the data is mapped to the hidden space, which has a certain filter effect in a sense, so the improvement effect after adding ADWT is not significant enough or even be negative effects. On the whole, the noise reduction treatment of the model input data has significant significance for the patterns in the mining data.

5. CONCLUSION

In this paper, we propose a method to denoising time-series using wavelet transforms to enhance the accuracy of multivariate time series prediction tasks. Experiments have proved that our model is efficient and accurate, which can effectively improve the prediction accuracy. This proves that in the field of deep learning, the purity of the data is as important as the subtlety of the model. The integration of data noise reduction algorithms with deep learning-based prediction algorithms is conducive to more accurate mining of patterns in the data, which deserves everyone's attention, not just more complex models.

REFERENCES

[1] Flunkert, V., Salinas, D., and Gasthaus, J., "Deepar: Probabilistic forecasting with autoregressive recurrent networks," CoRR abs/1704.04110 (2017).

[2] Deng, J., Chen, X., Jiang, R., Song, X., and Tsang, I. W., "St-norm: Spatial and temporal normalization for multi-variate time series forecasting," in [KDD '21: The 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, Singapore, August 14-18, 2021], Zhu, F., Ooi, B. C., and Miao, C., eds., 269–278, ACM (2021).

[3] Wu, Z., Pan, S., Long, G., Jiang, J., and Zhang, C., "Graph wavenet for deep spatial-temporal graph modeling," in [Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019], Kraus, S., ed., 1907–1913, ijcai.org (2019).

[4] Wu, Z., Pan, S., Long, G., Jiang, J., Chang, X., and Zhang, C., "Connecting the dots: Multivariate time series forecasting with graph neural networks," in [KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, August 23-27, 2020], Gupta, R., Liu, Y., Tang, J., and Prakash, B. A., eds., 753–763, ACM (2020).

[5] Shang, C., Chen, J., and Bi, J., "Discrete graph structure learning for forecasting multiple time series," in [9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021], OpenReview.net (2021).

[6] Bracewell, R. N. and Bracewell, R. N., [The Fourier transform and its applications], vol. 31999, McGraw-Hill New York (1986).

[7] Shensa, M. J. et al., "The discrete wavelet transform: wedding the a trous and mallat algorithms," IEEE Transactions on signal processing 40(10), 2464–2482 (1992).

[8] Lai, G., Chang, W., Yang, Y., and Liu, H., "Modeling long- and short-term temporal patterns with deep neural networks," in [The 41st International ACM SIGIR Conference on Research & Development in Infor mation Retrieval, SIGIR 2018, Ann Arbor, MI, USA, July 08-12, 2018], Collins-Thompson, K., Mei, Q., Davison, B. D., Liu, Y., and Yilmaz, E., eds., 95–104, ACM (2018).

[9] Grigsby, J., Wang, Z., and Qi, Y., "Long-range transformers for dynamic spatiotemporal forecasting," CoRR abs/2109.12218 (2021).