

Deep Learning-based Time Series Forecasting: Models and Applications

by Bin Wang

Thesis submitted in fulfilment of the requirements for
the degree of

Doctor of Philosophy

under the supervision of
Associate Professor Guangquan Zhang
and Distinguished Professor Jie Lu

University of Technology Sydney
Faculty of Engineering and Information Technology
Australian Artificial Intelligence Institute

September 2022

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, *Bin Wang*, declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Computer Science at the Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of the requirements for a degree at any other academic institution except as fully acknowledged within the text. This thesis is the result of a Collaborative Doctoral Research Degree program with Southwest Jiaotong University.

This research is supported by the Australian Government Research Training Program.

Signature: Production Note:
Signature removed prior to publication.

Date: *9, September, 2022*

Acknowledgements

It is an extremely remarkable and unforgettable journey for pursuing my Ph.D. degree at the University of Technology Sydney. I am indebted to the people for all their help and support.

I would like to express sincere gratitude to my principal supervisor, A./Professor Guangquan Zhang. Without his encouragement and advice, I would not have been able to accomplish this Ph.D program. His vision and wisdom influence me deeply, which point out the directions of my research. It always warms me to be invited as a member of his home party. I also would like to express thankfulness to my co-supervisor, Distinguished Professor Jie Lu. Her optimism, enthusiasm and confidence guided me to the right direction, especially when I become puzzled in my research progress. She initiated the dual doctoral degree program between UTS and SWJTU, provided me scholarship and offered me an opportunity to join DeSI Lab in the Australian Artificial Intelligence Institute. In addition, I would like to express my respect to my supervisor at SWJTU, Professor Tianrui Li. His encouragement and high standards of academic pursuit motivated me to face challenges positively. It is fortunate to be his postgraduate in my life.

I would like to thank every member of the DeSI Lab. It was a wonderful experience to learn from these dedicated researchers. I especially thank Dr. Dianshuang Wu, Dr. Anjin Liu, Dr. Hang Yu, Dr. Ximeng Wang, Dr. Ruiping Yin,

Dr. Qian Zhang, Dr. Yiliao Song, Dr. Shan Xue, Dr. Feng Liu, Dr. Feng Gu and Dr. Zhen Fang who have shared their opinions and comments with me. Furthermore, I want to thank for the kind help from Dr. Mingshan Jia, Dr. Yan Zhang, Dr. Mengxi Jiang, Dr. Joakim Skarding and other friends, may our friendship last forever.

I would like to express my heartfelt appreciation to my parents for their love and support.

Abstract

With the advent of the era of big data and artificial intelligence, people can obtain various data and extract valuable information and knowledge through artificial intelligence technology. As a typical representative of complex data, time series modeling and forecasting have always been hot topics. In the big data environment, time series often have the characteristics of multi-source complexity, dynamic heterogeneity, uncertainty, and nonlinearity, which brings tremendous challenges to the processing and prediction of time series data. As a cutting-edge approach of artificial intelligence, deep learning has efficient automatic feature extraction and robust representation learning capabilities. Using deep learning to enhance time series forecasting performance has become an important research direction. This dissertation studies the point estimation and uncertainty quantification of time series forecasting based on deep learning methodology. This thesis achieves research contributions as follows:

- (1) It proposes a point estimation network based on noisy residual connection and snapshot ensemble forecasting methods. The model reasonably simplifies the input data and introduces Gaussian noise to the residual connection to achieve regularization so that the model can significantly reduce the training time while ensuring generalization accuracy. The method of forecasting using snapshot ensemble for point estimation has also been explored. It can significantly

improve forecasting accuracy without much difference from the training time of a single model. The experiment is evaluated based on two realistic traffic flow datasets and confirms the effectiveness of the proposed methods.

- (2) It develops a deep uncertainty quantification method based on the assumption of Gaussian distribution. The method employs a new likelihood loss function based on the Gaussian distribution as the training loss function of the encoder-decoder model, allowing the model to simultaneously predict the point estimation and prediction interval of multiple variables at multiple time steps. The research also reported the generalization improvement introduced by training the deep forecasting model with the likelihood loss function. This research applies the proposed method to solve a real-world weather forecasting problem. By designing an effective multi-source information fusion mechanism, it can improve the forecasting accuracy of various meteorological variables significantly.
- (3) It further presents a deep uncertainty quantification method without distribution assumption. The method constructs a novel loss function without a distribution hypothesis, hence the so-called distribution-free loss function, as the objective loss function for training the encoder-decoder to model the data distribution adaptively at the training stage and then predict the point estimation and the prediction interval at the forecasting stage. The effectiveness of the proposed method is verified on three public datasets.
- (4) It constructs a novel deep quantile fusion network for robust point estimation. By designing a novel loss function, the network can transform the hidden layer into a quantile layer with semantic information and take these quantiles as

the input features of the following layer to predict the target variable, thereby constructing a deep forecasting model with high accuracy and great robustness. The experiments are evaluated on eight UCI regression datasets and a time series dataset and demonstrate the effectiveness of the proposed methods.

This study focuses on time series data with characteristics of complexity, heterogeneity, abnormality, nonlinearity, and uncertainty. It discusses data preprocessing and feature fusion, proposes effective deep point estimation and uncertainty quantification methods, and applies these methods to handle practical forecasting tasks. The research works of this dissertation can promote the related research of deep learning in time series forecasting, promote the development of related industrial applications, and have both theoretical significance and application value.

Table of Contents

CERTIFICATE OF ORIGINAL AUTHORSHIP	iii
Acknowledgements	v
Abstract	vii
List of Figures	xv
List of Tables	xix
1 Introduction	1
1.1 Background	1
1.2 Research Challenges and Methodologies	3
1.2.1 Research Challenges	3
1.2.2 Research Methodologies	4
1.3 Research Questions and Objectives	5
1.4 Research Significance	8
1.5 Thesis Structure	10
1.6 Publications Related to This Thesis	13
2 Literature Review	15

Table of Contents

2.1	Deep forecasting	16
2.1.1	Model architectures	16
2.1.2	Model loss functions	18
2.2	Uncertainty quantification	20
2.2.1	Concept of uncertainty quantification	20
2.2.2	Deep uncertainty quantification	21
2.3	Applications of deep forecasting	24
2.3.1	Flow prediction	24
2.3.2	Weather forecasting	25
3	Noise Residual Network for Spatio-temporal Traffic Flow Forecasting	29
3.1	Introduction	30
3.2	Problem Statement	32
3.3	Proposed Methodology	35
3.3.1	Preliminary	35
3.3.2	Model framework	35
3.3.3	Forecasting using snapshot ensemble	36
3.4	Experiments and Analysis	41
3.4.1	Datasets and preprocessing	41
3.4.2	Experimental settings and baselines	42
3.4.3	Experimental evaluations and analysis	45
3.5	Summary	51
4	Deep Uncertainty Quantification for Weather Forecasting with Distribution Assumptions	53
4.1	Introduction	54

4.2	Problem Statement	56
4.3	Proposed Methodology	62
4.3.1	Data preprocessing	62
4.3.2	Model framework	63
4.3.3	Likelihood loss function	66
4.4	Experiments and Analysis	71
4.4.1	Baselines	71
4.4.2	Experimental environments	72
4.4.3	Parameter settings	72
4.4.4	Performance analysis	74
4.5	Summary	82
5	Deep Uncertainty Quantification without Distribution Assumptions	83
5.1	Introduction	84
5.2	Proposed Methodology	87
5.2.1	Model framework	87
5.2.2	Hybrid loss function	91
5.2.3	Training algorithm	93
5.3	Experiments and Analysis	93
5.3.1	Datasets	95
5.3.2	Experimental settings and baselines	97
5.3.3	Experimental evaluations	100
5.3.4	Performance analysis	100
5.4	Summary	106
6	Deep Forecasting via Quantile Fusion	109

Table of Contents

6.1	Introduction	110
6.2	Deep Quantile Fusion Methods	113
6.2.1	Preliminary	113
6.2.2	Model framework	113
6.2.3	Loss function	114
6.2.4	End-to-end V.S. two-stage training algorithms	116
6.3	Experiments and Analysis	118
6.3.1	Datasets	118
6.3.2	Experimental settings and baselines	118
6.3.3	Experimental results and analysis	128
6.4	Model Extension	131
6.4.1	Deep auto-encoder with quantile fusion	132
6.4.2	Dataset	134
6.4.3	Experimental analysis	134
6.5	Summary	135
7	Conclusion and Future Research	137
7.1	Conclusions	137
7.2	Future Study	139
	Bibliography	143
	Abbreviations	166

List of Figures

1.1	Thesis structure	12
3.1	Spatio-temporal grids.	33
3.2	Spatio-temporal flow matrix	34
3.3	(a) shows a residual unit consisting of 4 stacked BN+ReLU+Conv in ST-ResNet. (b) displays the residual unit with Gaussian regularization in Noise-ResNet.	34
3.4	Model architecture of Noise-ResNet.	37
3.5	Comparison between two learning rate schedules (Huang et al., 2017)	39
3.6	Illustration of dynamic schedule of the learning rate E times in one snapshot period	41
3.7	Loss changes of the last 20 epochs during training and testing . . .	48
3.8	Sensitivity analysis of Noise-ResNet on BikeNYC dataset	51
4.1	Three historical target variable series from 03/01/2015-05/31/2018. They show a strong seasonal variation of t2m, a weak seasonal variation of 2-meter rh2m , and almost no seasonal variation of w10m.	59

List of Figures

4.2	The variation of mean (solid line) and 90% confidence interval (shaded area) for 10 stations during the target forecasting time zone (3:00 intraday to 15:00 of the next day) from 03/01/2015-05/31/2018. Values of Y-axis are following normalization.	61
4.3	DUQ for sequential point estimation and prediction interval.	65
4.4	A test sample at one station is chosen to visualize the forecasting of 3 target variables in the future 37 hours. We can see that all predicted points fall into the prediction interval given by $1 - z = 90\%$	78
5.1	Framework of DeepPIPE. The encoder $E(\cdot)$ first encodes input series $\mathbf{X}_{1:t}$ to hidden states \mathbf{c} which is transferred to form the initial state of decoder $D(\cdot)$. $D(\cdot)$ then decodes \mathbf{c} and consequently generates the sequential prediction interval and point estimation simultaneously. The proposed hybrid loss function based on related assumptions is adopted to train the whole network.	90
5.2	Examples of three time-series datasets. Each dataset illustrates the first 800-step time series.	96
5.3	Predictive examples of DeepPIPE on appliances energy dataset. DeepPIPE achieved better MPIW significantly than DeepMVE.	101
5.4	Forecasting examples of DeepPIPE on air quality dataset. DeepPIPE achieved better MPIW significantly than DeepMVE.	102
6.1	The framework of deep quantile fusion.	115
6.2	DNN(50-11)-based and DeepJMQR models.	120
6.3	DNN(50-11-5)-based models.	121
6.4	The output quantiles of two models trained o/w CL.	131

6.5	10 output samples on Concrete dataset. For each sample, the yellow points represent 11 quantiles, the green points are the ground truth and the red points are the regressed/predictive values.	132
6.6	The framework of quantile fusion-based AutoEncoder.	133
6.7	Examples of reconstructing 10-step time series.	135

List of Tables

3.1	Datasets TaxiBJ and BikeNYC	42
3.2	Comparison among different baselines on BikeNYC	47
3.3	Comparison among different baselines on TaxiBJ	47
4.1	The <i>SS</i> performance of different methods on 9 days. The column P-value compare the best performing method DUQ_{Esb10} with other methods, using one-tail paired T-test.	79
4.2	The <i>RMSE</i> performance of different methods on 9 days. Since <i>RMSE</i> and <i>SS</i> are not fully linear relationship, the counterpart assessment does not reach the optimal at the same time.	80
4.3	PICP on every day	81
4.4	T-test among single models for $DUQ_{300-300}$	81
4.5	Iterations	81
5.1	Summarized key differences between DeepPIPE and previous methods. PE and PI represents point estimation and prediction interval, respectively.	87
5.2	Information of three real-world datasets. 1/3-step prediction indicate that datasets are formatted for <i>n</i> -step-ahead prediction.	98

List of Tables

5.3	Experimental results on Brent oil price dataset. DeepPIPE achieved the best on MAE and MPIW.	105
5.4	Experimental results on appliances energy dataset. DeepPIPE achieved the best on MAE and MPIW.	105
5.5	Experimental results on air quality dataset. DeepPIPE achieved the best on MAE and MPIW.	106
6.1	RMSE evaluation. The ranking of five competitive methods, i.e., DNN (50-11-5)-based [QL+MSE]-E2E:DQF, [QL+CL+MSE]-E2E:DQF, [QL+MSE]-2S:DQF, [MSE] and DeepJMQR, are indicated at upper-right corner. The first-place one is also displayed in bold. The $\pm X$ reported is standard error on 20 testing folds (not standard deviation).	125
6.2	MAE evaluation. Note that the ranking of methods are not consistent as RMSE evaluation since both two evaluation are not linearly correlated.	126
6.3	Top-n hits of five competitive methods on eight datasets.	127
6.4	An example statistic of quantile crossing on one testing data fold.	127
6.5	The evaluation metric is RMSE, N is the number of testing data and D is the number of input feature. From left to right, it is MC-Dropout (Keras+Theano) (Gal and Ghahramani, 2016), MC-Dropout (Keras+Tensorflow) (our reproduction), DeepEnsemble (Lakshminarayanan et al., 2017), and PBP (Hernández-Lobato and Adams, 2015).	127
6.6	Reconstruction error evaluation.	134