# Final Project of Pattern Recognition

## Deep Learning for Multivariate Time Series Forecasting

**1811018 Guan Song Wang**

Dept. Software Engineering, Tongji University, Shanghai, CN

## Introduction

The topic we choose for my final course project is multivariate time series forecasting. Multivariate time series (MTS) data are common in daily life ranging from household electricity consumption, meteorological observation, currency exchange rate, solar power production, and even music notes can all be considered as time series data. There may exist complex dynamic interdependencies between different series that are significant but difficult to capture and analyze. As for forecasting, the task is to predict future time series based on observed data in the past. For example, precipitation in the next days, weeks or months can be forecast according to historical measurement. The further ahead we try to forecast, the harder it is.

When it comes to MTS forecasting using deep learning, it often faces a major research challenge, that is, how to capture and leverage the dynamics dependencies among multiple variables as well as the mixture of short-term and long-term repeating patterns. It is commonly recognized that deep learning is suitable for capturing complex dependencies. And it has also received an increasing attention in time series analysis. For instance, naive RNN models [1], ARIMA models combining with Multilayer Perceptron (MLP) [2,3,4], vanilla RNN with Dynamic Boltzmann Machines [5] and CNN models [6] have been studied in this direction.

To the best of our knowledge, Long- and Short-term Time-series Network (LSTNet, 2018) [7] is the first model that is designed specifically for MTS forecasting with up to hundreds of evolving variables. In LSTNet, CNN is utilized to capture short-term patterns, whereas LSTM or GRU is responsible for memorizing relatively long-term patterns. LSTNet adds a recurrent-skip layer or a typical attention to deal with gradient vanishing problem coming with long-term RNN. And traditional autoregression is also part of the entire model that helps to tackle the scale insensitive problem of neural networks.

Finally, the main contributions of this project are summarized as follows:

* We reproduce the LSTNet by implementing the model with pytorch and repeating the main experiments.
* Several original deep CNN models and RNN models are designed to compare with LSTNet.
* Based on the multi-head attention mechanism in [8], I introduce a temporal attention usage only model with autoregression part for MTS forecasting.
* We rigorously evaluate my approaches and LSTNet on 4 benchmark tasks and 2 of them achieve state-of-the-art prediction performance.

## Related Work

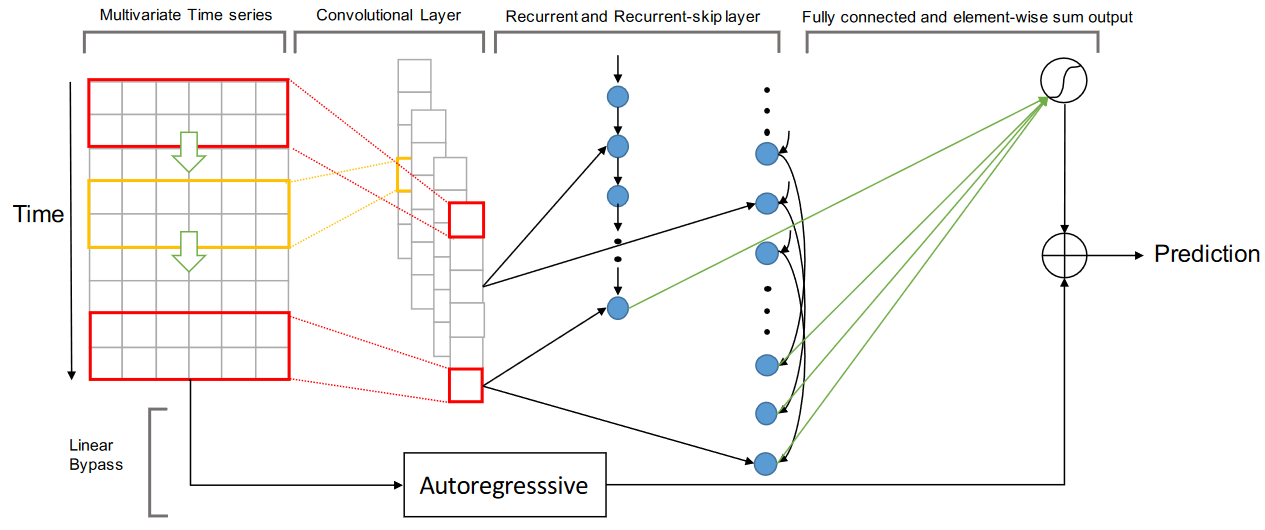
The most renowned model for linear univariate time series forecasting is the autoregressive integrated moving average (ARIMA) [9], which encompasses other autoregressive time series models, including autoregression (AR), moving average (MA), and autoregressive moving average (ARMA). Additionally, linear support vector regression (SVR) [10] treats forecasting problem as typical regression problem with time-varying parameters. However, these models are rarely used in high dimensional multivariate time series forecasting due to their high computational cost.

To forecast MTS data, vector autoregression (VAR), which is a generalization of AR-based models, is proposed [11,12]. VAR models extend AR models to the multivariate setting but ignores the dependencies between output variables. Nevertheless, neither AR-based models nor VAR-based models are capable of capturing non-linearity. For that reason, substantial effort has been made for non-linear models for time series forecasting based on kernel methods [13], ensembles [14] and Gaussian Processes (GP) [15].

GP is a non-parametric method for modeling distributions over a continuous domain of functions. This contrasts with models defined by a parameterized class of functions such as VARs and SVRs. It can be applied to multivariate time series forecasting task as suggested in [16], However, the power of GP comes with the price of high computation complexity. A straightforward implementation of Gaussian Process for multivariate time-series forecasting has cubic complexity over the number of observations. What's more, GP and VAR apply predetermined non-linearity and may fail to recognize different forms of non-linearity for different MTS.

Recently, deep neural networks have received great amount of attention due to their adaptable abilities in capturing non-linear interdependencies. Two variants of RNN, namely long short-term memory (LSTM) [17] and gated recurrent unit (GRU) [18], have shown promising results in several NLP tasks and have also be employed on MTS forecasting. As mentioned before, previous work in this area starts from using naive RNN [1], to hybrid models that combined ARIMA and Multilayer Perceptron [2,3,4], and to the latest Dynamic Boltzmann Machine with RNN [5]. Although these models can be applied to MTS, they mainly target univariate or bivariate time series.

## LSTNet



## Multi-Head Attention

## Evaluation

The four benchmark datasets we use are all public available, which are published by Lai et al.. Table 1 summarizes the corpus statistics. The descriptions of these datasets are as follows:

* Solar-Energy: the solar power production data from photovoltaic plants in Alabama State in 2006.
* Traffic: two years (2015-2016) of data provided by the California Department of Transportation that describes the road occupancy rate (between 0 and 1) on San Francisco Bay area freeways.
* Electricity: a collection of electricity consumption of 321 clients in kWh.
* Exchange Rate: the exchange rates of eight foreign countries (Australia, British, Canada, China, Japan, New Zealand, Singapore and Switzerland) from 1990 to 2016.

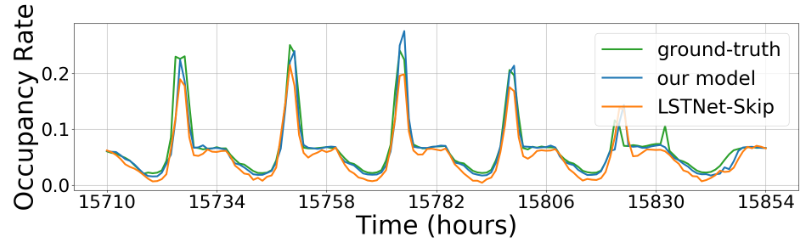
|  |  |  |  |
| --- | --- | --- | --- |
| **Dataset** | **L** | **D** | **S** |
| **Solar-Energy** | 52560 | 137 | 10 min |
| **Traffic** | 17544 | 862 | 1 hour |
| **Electricity** | 26304 | 321 | 1 hour |
| **Exchange Rate** | 7588 | 8 | 1 day |

Table 1: Dataset Statistics, where **T** is length of time series, **D** is number of variables, **S** is the sample rate.

On these benchmark MTS datasets, since we make comparison of our models with LSTNet, we follow the same evaluation metrics. The first metric is the root relative squared error (RSE), which is defined as:



## Performance



## Conclusion

## Reference

1. J. Connor, L. E. Atlas, and D. R. Martin. Recurrent networks and narma modeling. *Advances in Neural Information Processing Systems*, pages 301–308, 1991.
2. Jain and A. M. Kumar. Hybrid neural network models for hydrologic time series forecasting. *Applied Soft Computing*, 7(2):585–592, 2007.

.

1. G. P. Zhang. Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50:159–175, 2003.
2. G. E. Box and D. A. Pierce. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association*, 65(332):1509–1526, 1970.
3. S. Dasgupta and T. Osogami. Nonlinear dynamic boltzmann machines for timeseries prediction. *AAAI-17*. Extended research report available at goo. gl/Vd0wna, 2016.
4. J. B. Yang, M. N. Nguyen, P. P. San, X. L. Li, and S. Krishnaswamy. Deep convolutional neural networks on multichannel time series for human activity recognition. In Proceedings of the 24th *International Joint Conference on Artificial Intelligence (IJCAI)*, Buenos Aires, Argentina, pages 25–31, 2015.
5. Lai, G.; Chang, W.-C.; Yang, Y.; and Liu, H. Modeling long- and short-term temporal patterns with deep neural networks. *SIGIR*, 95–104, 2018.
6. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *Advances in Neural Information Processing Systems,* 5998-6008, 2017.
7. G. E. Box, G. M. Jenkins, G. C. R., and Ljung, G. M. Time series analysis: forecasting and control. *John Wiley & Sons*, 2015.
8. Cao, L.-J., and Tay., F. E. H. Support vector machine with adaptive parameters in financial time series forecasting. I*EEE Transactions on neural networks,* 1506–1518, 2003.
9. J. D. Hamilton. Time series analysis, volume 2. *Princeton university press Princeton*, 1994.
10. H. Lütkepohl. New introduction to multiple time series analysis. *Springer Science & Business Media*, 2005.
11. Chen, S.; Wang, X. X.; and Harris, C. J. Narxbased nonlinear system identification using orthogonal least squares basis hunting. *IEEE Transactions on Control Systems*, 78–84, 2008.
12. Bouchachia, A., and Bouchachia, S. Ensemble learning for time series prediction. *Proceedings of the 1st International Workshop on Nonlinear Dynamics and Synchronization*, 2008.
13. Frigola, R., and Rasmussen, C. E. Integrated preprocessing for bayesian nonlinear system identification with gaussian processes. *IEEE Conference on Decision and Control*, 552–560, 2014.
14. S. Roberts, M. Osborne, M. Ebden, S. Reece, N. Gibson, and S. Aigrain. Gaussian processes for time-series modelling. *Phil. Trans. R. Soc. A*, 371(1984):20110550, 2013
15. Hochreiter, S., and Schmidhuber, J. Long short-term memory. *Neural Computation,* 9(8):1735–1780, 1997.
16. Cho, K., van Merrienboer, B., G ¨ ulc¸ehre, C, Bahdanau, D., ¨ Bougares, F., Schwenk, H., and Bengio, Y. Learning phrase representations using rnn encoder–decoder for statistical machine translation. *In Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1724–1734, 2014.