MULTIVARIATE TIME SERIES FORECASTING WITH PARALLEL EXTRACTION OF LONG-TERM TRENDS AND SHORT-TERM FLUCTUATIONS FRAMEWORK

Yifu Zhou\textsuperscript{1}, Ziheng Duan\textsuperscript{2}\textsuperscript{*}, Haoyan Xu\textsuperscript{2}, Jie Feng\textsuperscript{2}, Anni Ren\textsuperscript{2}, Yueyang Wang\textsuperscript{3}\textsuperscript{†}, Xiaoqian Wang\textsuperscript{4}\textsuperscript{†}

\textsuperscript{1}The University of Queensland, Australia; \textsuperscript{2}Zhejiang University, China; \textsuperscript{3}Chongqing University, China; \textsuperscript{3}Purdue University, USA

ABSTRACT

Multivariate time series (MTS) forecasting has been widely used in various fields. Reasonable prediction results can help people make decisions, avoid risks, and increase profits. Generally, time series have two characteristics, namely long-term trends and short-term fluctuations. For example, stock prices will rise in the long term, but may fall slightly in the short term. These two characteristics are often relatively independent of each other. Existing forecasting methods usually do not distinguish between the two characteristics and thus cannot fully extract inherent attributes of the time series. In this paper, we propose a framework that can capture the long-term trends and short-term fluctuations of time series in parallel in order to enhance MTS forecasting performance. We formulate our model as a multi-task learning objective, the goal of which is to make predictions of long-term trends and short-term fluctuations as close as possible to the ground truth. Experiments on three real-world datasets show that the proposed method uses more supervision information and can more accurately capture the changing trends of time series, thereby improving the prediction performance.

Index Terms— Multivariate Time Series Forecasting, Long-Term Trends, Short-Term Fluctuations

1. INTRODUCTION

Time series forecasting is a key challenge in many disciplines, including finance \textsuperscript{[1]}, industry \textsuperscript{[2]}, environment \textsuperscript{[3]}, etc. Due to the diversity of the real world, a system usually contains multiple variables, which makes the research of multivariate time series (MTS) forecasting particularly important.

In most cases, time series collected from the real world have two characteristics, which are long-term trends and short-range fluctuations. As a concrete example in Fig. 1, the temperature not only has periodic long-term trends with the change of day and night, but also has fluctuations in the short term due to temporary weather reasons. Obviously, these two characteristics are relatively independent of each other, but both contain meaningful information that is important to the prediction of future values of the time series. In this case, the time series forecasting task can be decomposed into two parts, one is to predict its long-term trends, and the other part is to predict its short-term fluctuations. Together, the two constitute a complete forecasting result. Most existing time series forecasting methods take the original value of the time series as input to directly predict its future value. This confuses the two independent change rules and may reduce the prediction accuracy and interpretability of the model.

In economics, the value of the original time series is often used to represent the long-term trends, and the difference between time stamps is used as a measure of short-term fluctuations. Inspired by this representation, we design a framework to extract long-term trends and short-term fluctuations in parallel, named Parallel Forecasting Network (PFNet). PFNet is composed of three sub-modules, which are long-term trends prediction module (LTPM), short-term fluctuations prediction module (SFPM), and information fusion module (IFM). Compared with direct forecasting, PFNet integrates more supervision information and can more accurately capture the different changing patterns of the time series, thereby improving the forecasting performance. Our main contributions are:

- We first propose an MTS forecasting framework to predict long-term trends and short-term fluctuations in parallel.
- We construct a triplet loss function and integrate more supervision information to extract key features of time series.
• We empirically validate that PFNet has better performance than state-of-the-art models on various benchmark datasets.

2. FRAMEWORK

Given a matrix containing multiple observed MTS $X = [x_1, x_2, ..., x_t]$, where $x_i \in \mathbb{R}^n$ and $n$ is the number of variables. The main purpose of our forecasting model is to predict $x_{t+h}$, where $h$ is the horizon ahead of the current time stamp.

2.1. Vector Error Correction Model (VECM)

As the most common used linear MTS forecasting method, VAR \(^4\) predicts $x_{t+h}$ in the following expression: $x_{t+h} = \sum_{i=1}^{t} A_i x_i + \epsilon_t$, where $A_i (i = 1, 2, ..., t)$ are the regression coefficients when horizon is $h$, and $\epsilon_t$ is the white noise vector. Based on this, VECM \(^5\) is proposed and according to it, $\Delta x_{t+h}$ can also be written as:

$$\Delta x_{t+h} = \Pi x_t + \sum_{k=1}^{t} \Gamma_k \Delta x_{t-k} + \epsilon_t.$$ (1)

$\Pi$ and $\Gamma_k$ can be expressed by $A^1, A^2, ..., A^h$. It can be seen that the short-term fluctuations to be predicted is related to two parts: the value of MTS at time $t$ and difference between previous time stamps. In our model, we replace the linear transformation in equation \(^1\) with a neural network.

2.2. Long-Term Trends Prediction Module (LTPM)

LTPM predicts long-term trends of MTS, which takes $X$ as input and predict $x_{t+h-1}$ instead of directly predicting $x_{t+h}$. In this paper, we use Highway-CNN \(^6\) to deal with long-term trends prediction. Denote the nonlinear transformation of the multi-layer CNN as $H(X, \theta)$, where $\theta$ represents model parameters. The output of Highway-CNN can be calculated using the following equation:

$$y = H(X, \theta) \odot T(X, W_T) + W_L \odot (1 - T(X, W_T)),$$ (2)

where $\odot$ is dot product and $W_L$ is a trainable linear mapping. $W_T$ is a weight matrix. $T$ and $(1 - T)$ represent the transform gate and carry gate respectively. Here $T$ is calculated as: $T(X, W_T) = 1/(1+e^{-W_T X})$. Through LTPM, the predicted value $\hat{x}_{t+h-1}$ is obtained.

2.3. Short-Term Fluctuations Prediction Module (SFPM)

Refer to VECM, SFPM takes $\Delta X$ and $x_t$ as inputs to predict $\Delta x_{t+h}$, thereby characterizing its short-term fluctuations. Similar to LTPM, SFPM uses Highway-CNN to extract the features of $\Delta X$. In addition, a multilayer perceptron (MLP) structure is used for nonlinear transformation of $x_t$. Therefore, the expression of the output vector $\hat{x}_{t+h}$ of SFPM is:

$$\Delta \hat{x}_{t+h} = HighwayCNN(\Delta X) + MLP(x_t).$$ (3)

2.4. Information Fusion Module (IFM)

The final forecasting result is obtained by the superposition of long-term trends and short-term fluctuations. The IFM takes the output of LTPM and SFPM as input, and merges them to obtain the forecasting $\hat{x}_{t+h}$.

$$\hat{x}_{t+h} = \hat{x}_{t+h-1} + \Delta \hat{x}_{t+h}.$$ (4)

2.5. Objective Function

We use $L_1$ norm to construct a triple loss function (a weighted summation, using $c_1$ and $c_2$, of the $L_1$ loss for $x_{t+h}$, $x_{t+h-1}$ and $\Delta x_{t+h}$) and optimize the model via Adam algorithm.

$$\min_{\theta} \sum_t \left[ \|\hat{x}_{t+h} - x_{t+h}\|_1 + c_1 \|\hat{x}_{t+h-1} - x_{t+h-1}\|_1 + c_2 \|\Delta x_{t+h} - \Delta x_{t+h}\|_1 \right].$$ (5)

3. EXPERIMENTS

3.1. Experimental Setup

We use three benchmark datasets which are publicly available: Exchange-Rate \(^7\) is the exchange rates of eight foreign countries collected from 1990 to 2016, collected per day. Energy \(^8\) measurements of 26 different quantities related to appliances energy consumption in a single house for 4.5 months, collected per 10 minutes. Nasdaq \(^9\) represents the stock prices which are selected as the multivariable time series for 82 corporations, collected per minute.

We apply three conventional evaluation metrics to evaluate the performance in multivariate time series prediction: Relative Squared Error (RSE), Relative Absolute Error (RAE) and Empirical Correlation Coefficient (CORR). The comparison methods include: VAR \(^9\) stands for the well-known vector regression model. RNN \(^10\) is the Recurrent Neural Network using GRU cell with AR components. MHA \(^11\), or MultiHead Attention, stands for multihread attention components in the Transformer model. LSTMNet \(^12\) shows great performance by modeling long-term and short-term temporal patterns of MTS data. MLCNN \(^13\) is a novel multi-task deep learning framework which adopts the idea of fusing forecasting information of different future time. MTGNN \(^14\) is a joint framework for modeling multivariate time series data generally from a graph-based perspective with graph neural networks. PFNet is our proposed Parallel Forecasting Network, which predicts long-term trends and short-term fluctuations of time series in parallel.

For the training details, we conduct grid search on tunable hyper-parameters on each method over all datasets. Specifically, the same grid search range of input window size for each method is set from $\{2^0, 2^1, ..., 2^9\}$ and different hyper-parameters are chosen for each method to achieve their best

\(^1\)https://github.com/laiguokun/multivariate-time-series-data/tree/master/exchange_rate
We predict both of them in parallel through the LTPM and SFPM. Here we use Highway CNN as the network $\alpha$ and $\beta$, and MLP as the network $\gamma$. In this diagram $\Delta X_t \in \mathbb{R}^{N \times (t-1)}$ and $x_t \in \mathbb{R}^{N \times t}$. We use $x_{t-1}, x_{t+h-1}, x_{t+h}$ and $\Delta x_{t+h} \in \mathbb{R}^{N \times 1}$. As shown by the red dotted line, part of the long-term trends ($x_t$) is used with $\Delta X_t$ to predict $\Delta x_{t+h}$.

Finally, we add the predicted $x_{t+h-1}$ and $\Delta x_{t+h}$ to get the final result $x_{t+h}$.

**3.2. Main Results**

Table 1 shows the evaluation results. Following the test settings of [12], we use each model for time series predicting on future moment $\{t+3, t+6, t+12, t+24\}$. The best results for each metrics on each dataset are set bold in Table 1. We save the model that has the best performance on validation set based on RSE or RAE metric after training 1000 epochs for each method and use this model to test and record the results.

Overall, the performance of VAR is weaker than other baselines, which shows that the linear regression method is not conducive to fully extracting the variety of information in the time series. MTGNN or TEGNN achieves the best in some cases, which shows the superiority of using graph structure to exploit the relationships among different variables in multivariate time series. The results in the Table 1 indicate that our proposed PFNet outperforms these baselines in most cases.

It is worth noting that, as models for predicting the value of multiple time points in the future, PFNet performs better than MLCNN. The reason could be that LTPM and SFPM can better capture the long-term and short-term characteristics of the time series, and a single model may be easier to confuse both. PFNet uses the idea of parallel forecasting to capture the long-term trends and short-term fluctuations of multivariate time series. Thus it can break through the restriction that traditional methods and other deep learning methods cannot use both of them.

**3.3. Time Complexity Analysis**

Although the proposed PFNet architecture designs three modules, it does not require much time. Following [15], we compare the performance of all models as a function of sample size and display the results on the NASDAQ dataset in Figure 3 to prove PFNet’s efficiency. The training and testing time of our PFNet model is about twice as long as that of other simple baselines. This is in line with our expectations. Because we not only forecast the long-term trends but also forecast the short-term fluctuations and the time used for each part is similar to that of an ordinary prediction network. It is worth noting that when processing high-dimensional time series, the time required for training and prediction of MLCNN, which also adopts the idea of multi-task learning, costs 3.6 and 6.5 times compared with our model PFNet. This proves the efficiency.
of our multi-task learning framework: there is no complicated network structure, and good performance can be achieved by combining this parallel forecasting framework with ordinary prediction networks (like highway-CNN in this paper).

3.4. Ablation Study

We conduct an ablation study on exchange-rate dataset to validate the effectiveness of key components of PFNet. We name PFNet without different components as follows:

LTPM: PFNet without short-term fluctuations prediction module. Here we use multi-layer CNN with AR components [16] to perform MTS forecasting tasks. PFNet-RNN: PFNet that replaces CNN-AR module with RNN. PFNet-LSTNet: PFNet that replaces CNN-AR module with LSTNet. PFNet-xt: PFNet without MLP in short-term fluctuations prediction module. That means we only use $\Delta x_t$ to predict $\Delta \hat{x}_{t+1}$.

Table 2 shows the comparison results. The important conclusions of these results are as follows: (1) PFNet has the best performance, compared with these variants. (2) After removing the short-term fluctuations prediction module, as shown by LTPM, the performance will drop, which proves that short-term fluctuations are useful for time series forecasting and PFNet takes advantage of this. (3) The relatively weak performance after removing MLP in SFPM (PFNet-xt) also proves the contribution of introducing $x_t$ in forecasting short-term fluctuations. (4) As shown in PFNet-RNN and PFNet-LSTNet, even if the CNN-AR module is replaced by RNN or LSTNet, the results are also competitive, which proves that the framework of parallel forecasting is robust.

4. CONCLUSION

In this paper, we propose a MTS forecasting framework, PFNet, to capture the long-term trends and short-term fluctuations of time series in parallel. Inspired by the idea of multi-task learning, PFNet uses the original value and its difference between time stamps to optimize MTS forecasting.
problems. Experiments on three real-world datasets show that our model outperforms 6 baselines in terms of the three metrics (RAE, RSE and CORR). For future research, different models can be incorporated into our framework to improve prediction accuracy and make PFNet more robust.

5. REFERENCES


