Comparison of Strategies for Multi-step-ahead Prediction of Time Series using Neural Network

Nguyen Hoang An
Faculty of Management Information Systems
Banking University of Ho Chi Minh City

Duong Tuan Anh
Faculty of Computer Science and Engineering
Ho Chi Minh City University of Technology
dtanh@cse.hcmut.edu.vn

Abstract— **If the one-step forecasting of a time series is already a challenging task, performing multi-step ahead forecasting is more difficult. Several approaches that deal with this complex problem have been proposed in literature: recursive (or iterated) strategy, direct strategy, combination of both the recursive and direct strategies, called DirREC, the Multi-Input Multi-Output (MIMO) strategy, and the last strategy, called DirMO which aims to preserve the advantageous aspects of both the Direct and MIMO strategies. This paper aims to review existing strategies for multi-step ahead forecasting using neural networks and compare their performances empirically. To attain such an objective, we performed several experiments of these different strategies on three datasets: NN3 competition dataset, the Vietnam composite stock price index (VNINDEX) and the closing prices of the FPT stock. The most consistent findings are that the DirREC strategy is better than all the other strategies for multi-step ahead forecasting using neural network.**

Keywords—time series; multi-step-ahead prediction; neural networks; strategies of prediction.

I. Introduction

There are a few methods commonly used for time series prediction. The Box-Jenkins method, which gives quite good results with linear time series, is one of them. The Box-Jenkins method is considerably effective especially with linear and stationary datasets and with series that are non-stationary but turned into stationary via transformations. One of the techniques that has been used widely for time series forecasting since the late 1980s is artificial neural networks (ANN). ANN provides linear and nonlinear modeling without the assumptions as to the relation between input and output variables. Therefore, ANN is more flexible and applicable than other methods (Zhang et al., 1998, [17]).

Time series forecasting can be performed for both single and multiple periods. Unlike one-step ahead forecasting, multi-step ahead forecasting tasks are more difficult (Tiao & Tsay, 1994 [16]), since this kind of forecasting has to deal with additional complications, like accumulation of errors, reduced accuracy, and increased uncertainty (Sorjamaa et al., 2007 [15]). Consequently, accurate prediction of time series over long future horizons recently has become a hot topic which attracts a lot of attention from researchers.

According to Taieb et al., 2012 [4], five strategies have been proposed in the literature to tackle multi-step ahead forecasting tasks. Three of them are Recursive, Direct and MIMO (Multi-input Multi-output). The fourth one which combines Recursive and Direct is called DirREC. The fifth one which combines MIMO and Direct is called DirMO.

Despite the fact that many studies have compared between some of the five multi-step ahead approaches, these studies bring out contradictory findings and it is not clear which approach give better results. For example, in comparing between Recursive and Direct strategies with ANN as underlying predictor, Hill et al., 1994 [11] and Chang et al., 2007 [7] concluded that Recursive forecast yields better results than the Direct forecast while Zhang, 1994 [19] found that Direct forecast yields better results than the Recursive forecast. The Zhang's finding is supported by Hamzacebi et al., 2009 [10] in their study to compare direct and iterative ANN forecast approaches in multi-step-ahead time series prediction. These confusing findings bring out a strong need for further efforts of comparative evaluation for strategies of multi-step ahead time series prediction.

In this work, for the first time, we aim to compare empirically the performance of all the five multi-step-ahead forecasting strategies using feed-forward neural networks model. To attain such an objective, we performed several experiments of these different strategies on one benchmark dataset: NN3 competition and two real-life datasets: the daily Vietnam composite stock price index (VNINDEX) and the closing prices of the FPT stock. The most consistent findings are that the DirREC strategy is better than all the other strategies for multi-step ahead forecasting using ANN.

The rest of the paper is organized as follows. In Section 2 we give a definition of time series and explain some basic ideas of one-step-ahead prediction of time series using ANN. Section 3 introduces the five different multi-step ahead forecasting strategies for time series. Section 4 reports the experiments to compare the performance of all the five multi-step ahead forecasting strategies with ANN. Section 5 gives some discussion, conclusions and remarks for future work.

II. Background

A. *Time Series and Forecasting Methods*

A time series is a sequence of historical measurements y_t of an observable variable y at equal time intervals.

978-1-4673-8234-2/15 $31.00 © 2015 IEEE
DOI 10.1109/ACOMP.2015.24

The time series forecasting domain has been influenced, for a long time, by linear statistical methods such as ARIMA models. However, in the late 1970's and early 1980s, it became increasingly clear that linear models can not adapt to many real applications. In the same period, several useful nonlinear time series models were proposed such as the threshold autoregressive model and the autoregressive conditional heteroscedastic (ARCH) model ([9]). However, the analytical study of nonlinear time series analysis and forecasting is still in its infancy compared to linear models.

In the last two decades, machine learning models have drawn attention and have established themselves as serious contenders to classical models in the forecasting field. For instance, Artificial Neural Networks (ANNs) can outperform the classical statistical models such as linear regression and Box-Jenkins approaches [18]). Later, other approaches appeared such as decision trees, support vector machines and k-nearest neighbors ([1], [9]).

B. *One-step-ahead Forecasting with ANN*

Many different ANN models have been proposed since 1980s. Perhaps the most influential models are the multi-layer perceptrons (MLP), Hopfield networks, and Kohonen's self organizing networks. In this work, we focus on the multi-layer perceptrons.

An MLP is typically composed of several layers of nodes. The first layer is an input layer where external information is received. The last layer is an output layer where the problem solution is obtained. The input layer and output layer are separated by one or more intermediate layers called hidden layers. The nodes in adjacent layers are usually connected acyclic arcs from a lower layer to a higher layer.

For time series prediction, the inputs are typically the past observations of the time series and the output is the future value. The ANN perform the following function mapping

$$X_{t+1} = f(X_t, X_{t-1}, \ldots, X_{t-p}) \qquad (1)$$

where X_t is the observation at time t.

Before an ANN can be used to perform any desired task, it must be trained to do so. Basically, training is the process of determining the connection weights which are the key elements of an ANN. The knowledge learned by a network is stored in the links and nodes in the form of link weights and node biases. It is through the links that an ANN can carry out complex nonlinear mappings from its input nodes to its output nodes. An ANN training is a supervised one in that the desired response of the network (target value) for each input pattern (example) is always available.

The training input data is in the form of vectors of input variables or training patterns. Corresponding to each element in an input vector is an input node in the network input layer. Hence the number of input nodes is equal to the dimension of input vectors. Whatever the dimension, the input vector for a time series forecasting system will be almost always composed of a moving window of fixed length along the series (see Fig. 1).

The training process is usually as follows. First, examples of the training set are entered into the input nodes. The activation values of the input values are weighted and accumulated at each node in the first hidden layer. The total is then transformed by an activation function, for example the sigmoid function $f(x) = 1/(1+e^{-x})$, into the node's activation value. It in turn becomes an input into the nodes in the next layer, until eventually the output values are found. The training algorithm is used to find the weights that minimize some overall error measure such as the sum of squared errors (SSE) or mean squared errors (MSE). Hence, the network training actually is an unconstrained nonlinear minimization problem.

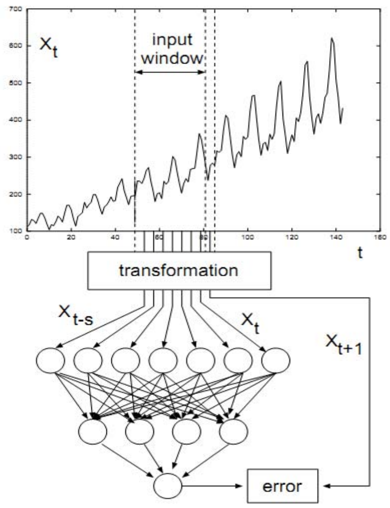

Fig. 1. Training an ANN for time series forecasting ([13])

As for the number of hidden layers, in our work, we use one single hidden layer. Many previous works show that a single hidden layer is sufficient for ANNs to approximate any complex nonlinear function. The use of an MLP for forecasting time series implies that the input nodes are connected to a number of past observed values supposed sufficient to identify the process at future time steps. As far as the number of input nodes is concerned, there is no theory yet to tell how many nodes are needed. As for the number of output nodes, MLP should have one output node in the case of *single step* prediction and might have k output nodes in the case of *k-step-ahead* prediction. The number of hidden nodes in the single hidden layer can be determined by experimental design.

III. Strategies for Multi-step-ahead-prediction of Time Series

A multi-step-ahead (also called long-term) time series forecasting task consists of predicting the next H values $[y_{N+1}, \ldots, y_{N+H}]$ of a historical time series $[y_1, \ldots, y_N]$ composed of N observations, where $H > 1$ denotes the *forecasting horizon*.

We will use a common notation where f and F denote the functional dependency between the past and future observations, d refers to the embedding dimension of the time series and w represents the term that includes modeling error, disturbances and/or noise.

According to Taieb et al., 2012 [4], there exist five strategies that have been proposed in the literature to tackle a multi-step ahead forecasting tasks. These five strategies can be applied to any underlying prediction model, such as ARIMA, ANN, support vector machine, k-nearest-neighbors or Lazy Learning algorithm. The five strategies will be explained in this section.

A. Iterative Strategy

The oldest and most intuitive forecasting strategy is the Iterated strategy (also called Recursive strategy) ([10], [12], [14], [15]). In this strategy, a single model f is trained to perform a *one-step-ahead* forecast, i.e.

$$y_{t+1} = f(y_t, \ldots, y_{t-d+1}) + w$$

with $t \in \{d, \ldots, N - 1\}$

When forecasting H steps ahead, we first forecast the first step by applying the model. Subsequently, we use the value just forecasted as part of the input variables for forecasting the next step (use the same one-step-ahead model). We continue in this manner until we have forecasted the entire horizon.

Let the trained one-step-ahead model be $\hat{f}$. Then the forecasts are given by:

$$\hat{y}_{N+h} = \begin{cases} \hat{f}(y_N, \ldots, y_{N-d+1}) & if\ h = 1 \\ \hat{f}(\hat{y}_{N+h-1}, \ldots, \hat{y}_{N+1}, y_N, \ldots, y_{N-d+h}) & if\ h \in \{2, \ldots, d\} \\ \hat{f}(\hat{y}_{N+h-1}, \ldots, \hat{y}_{N+h-d}) & if\ h \in \{d+1, \ldots, H\} \end{cases} \quad (2)$$

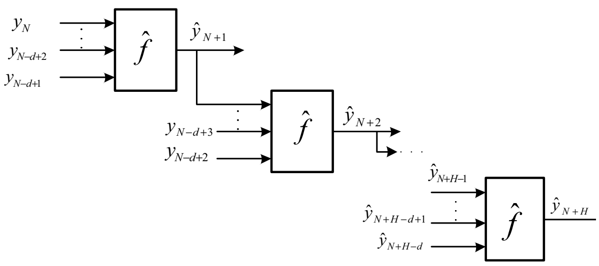

Fig. 2. The architecture of Iterated strategy for muti-step-ahead forecasting

The main concern of the Iterated strategy is whether it can produce a sequence of successful predictions to prevent error for propagating. Fig. 2 shows the architecture of the Iterated strategy for multi-step-ahead forecasting.

B. Direct Strategy

The Direct (also called Independent) strategy ([10], [12], [14], [15]) consists of forecasting each horizon independently from the others. In other words, H models f_h are learned (one for each horizon) from the time series $[y_1, \ldots, y_N]$ where

$$y_{t+h} = f_h(y_t, \ldots, y_{t-d+1}) + w$$

with $t \in \{d, \ldots, N - H\}$ and $h \in [1, \ldots, H]$

The forecasts are obtained by using the H learned models $\hat{f}$ as follows:

$$\hat{y}_{N+h} = \hat{f}_h(y_N, \ldots, y_{N-d+1}) \quad (3)$$

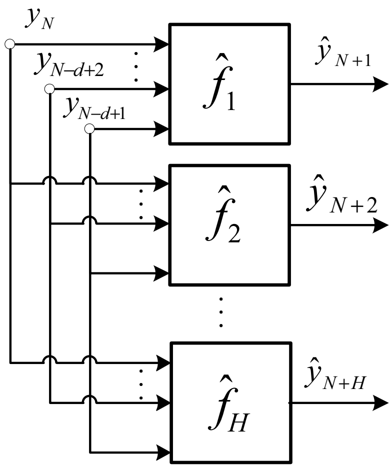

Fig. 3. The architecture of Direct strategy for multi-step-ahead forecasting

The Direct strategy does not use any approximated values to compute the forecasts, being then immune to the accumulation of errors. However, the H models are learned independently inducing a conditional independence of the H forecasts. Fig. 3 shows the architecture of Direct strategy for multi-step-ahead forecasting.

C. DirREC Strategy

The DirREC strategy (Sarjamaa & Lendasse, 2006 [14]) combines the architectures and the principles underlying the Direct and the Iterated strategies. DirREC computes the forecasts with different models for each horizon (like the Direct strategy) and at each time step, it enlarges the set of inputs by adding variables corresponding to the forecasts of the previous step (like the Recursive strategy). However, note that unlike the two previous strategies, the embedding size d is not the same for all the horizons. In other words, the DirREC strategy learns H models f_h from the time series $[y_1, \ldots, y_N]$ where

$$y_{t+h} = f_h(y_{t+h-1}, \ldots, y_{t-d+1}) + w$$

with $t \in \{d, \ldots, N - H\}$ and $h \in [1, \ldots, H]$

To obtain the forecasts, the H learned models are used as follows:

$$\hat{y}_{N+h} = \begin{cases} \hat{f}_h(y_N, \ldots, y_{N-d+1}) & if\ h = 1 \\ \hat{f}_h(\hat{y}_{N+h-1}, \ldots, \hat{y}_{N+1}, y_N, \ldots, y_{N-d+1}) & if\ h \in \{2, \ldots, H\} \end{cases} \quad (4)$$

Fig. 4 shows the architecture of DirREC strategy for multi-step-ahead forecasting.

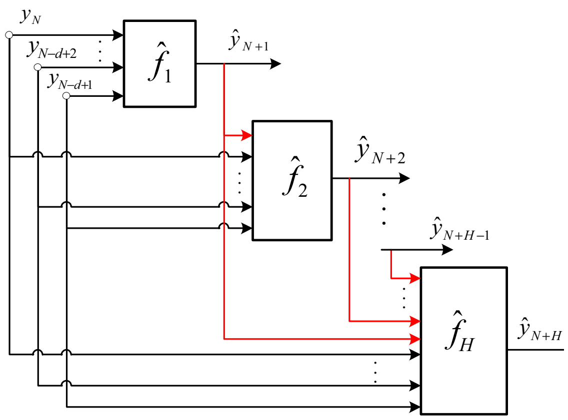

Fig. 4. The architecture of DirREC strategy for multi-step-ahead forecasting

D. MIMO strategy

The three previous strategies (Iterated, Direct and DirRec) may be considered as Single Output strategies ([3]) since they model the data as a (multiple-input) Single Output function (see Eqs. (2), (3) and (4)).

The MIMO strategy (Bontempi, 2008 [5]) learns one Multiple-Output model F from the time series $[y_1,\ldots,y_N]$ where

$$[y_{t+H},\ldots, y_{t+1}] = F(y_t,\ldots, y_{t-d+1}) + \boldsymbol{w}$$

with $t \in \{d, \ldots, N\text{-}H\}$, F: $\Re^d \rightarrow \Re^H$ is a vector-valued function, and $\boldsymbol{w} \in \Re^H$ is a noise vector.

The forecasts are returned in one step by a Multiple-Output model $\hat{F}$ where:

$$[\hat{y}_{t+H},\ldots,\hat{y}_{t+1}] = \hat{F}(y_N,\ldots,y_{N-d+1}) \tag{5}$$

The main idea of the MIMO strategy is to preserve, between the predicted values, the stochastic dependency characterizing the time series. This strategy avoids the conditional independence assumption made by the Direct strategy as well as the accumulation of errors in the Iterated strategy.

However, the need to preserve the stochastic dependencies by using only one model has a drawback as it constrains all the horizons to be forecasted with the same model structure. This constraint could reduce the flexibility of the forecasting strategy (Ben Taieb et al., 2009 [2]).

Fig. 5 shows the architecture of MIMO strategy for multi-step-ahead forecasting.

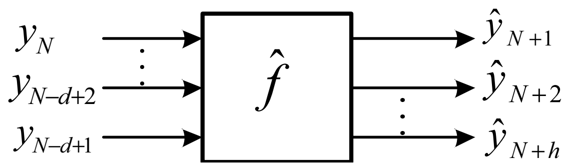

Fig. 5. The architecture of MIMO strategy for multi-step-ahead forecasting

E. DirMO strategy

The DirMO strategy (Ben Taieb et al., 2009 [3]) aims to preserve the most appealing aspects of both the Direct and MIMO strategies. DirMO forecasts the horizon H in blocks, where each block is forecasted in a MIMO fashion. Thus, the H-step-ahead forecasting task is decomposed into n Multiple-Output forecasting tasks ($n = H / s$), each with an output of size s ($s \in \{1,\ldots, H\}$).

The DirMO strategy learns n models F_p from the time series $[y_1,\ldots,y_N]$ where

$$[y_{t+p*s},\ldots\, y_{t+(p-1)*s+1}] = F_p(y_t,\ldots, y_{t-d+1}) + \boldsymbol{w}$$

with $t \in \{d, \ldots, N\text{-}H\}$, $p \in \{1,\ldots, n\}$ and F: $\Re^d \rightarrow \Re^s$ is a vector-valued function if $s > 1$, and $\boldsymbol{w} \in \Re^H$ is a noise vector.

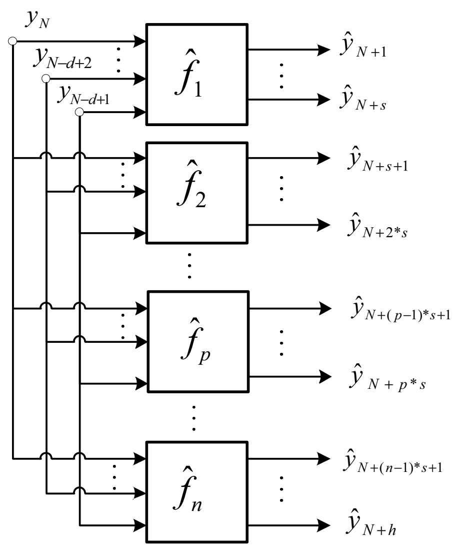

Fig. 6. The architecture of MIMO strategy for multi-step-ahead forecasting

The H forecasts are returned by the n learned models as follows:

$$[\hat{y}_{N+p*s},\ldots,\hat{y}_{N+(p-1)*s+1}] = \hat{F}_p(y_N,\ldots,y_{N-d+1}) \tag{6}$$

Fig. 6 shows the architecture of MIMO strategy for multi-step-ahead forecasting.

IV. EXPERIMENTAL EVALUATION

We implemented all the five multi-step ahead forecasting strategies using ANN in Microsoft C#. We conducted all the experiments on Intel Core2 Duo CPU T9300 @ 2.50GHz, 4.00 GB RAM, NVIDIA GeForce 8600M GT, HDD 320 GB.

A. *The Datasets*

In this work, we experimented the five different strategies over three datasets: NN3 competition dataset, the daily Vietnam composite stock price index (VNINDEX) and the daily closing prices of the stock FPT.

NN3 is a competition on time series forecasting organized in 2007 [8]. The dataset for this competition consists of 111 monthly time series. All time series are drawn from many empirical business time series. Some of time series are strongly seasonal, some are both trended and seasonal. Each time series has different length varying from 50 to 126 data points. In each time series, the competition required to forecast the values of the next 18 months, using the given historical data points. Fig. 7 shows two time series (11th and 55th series among 111 time series of the NN3 dataset).

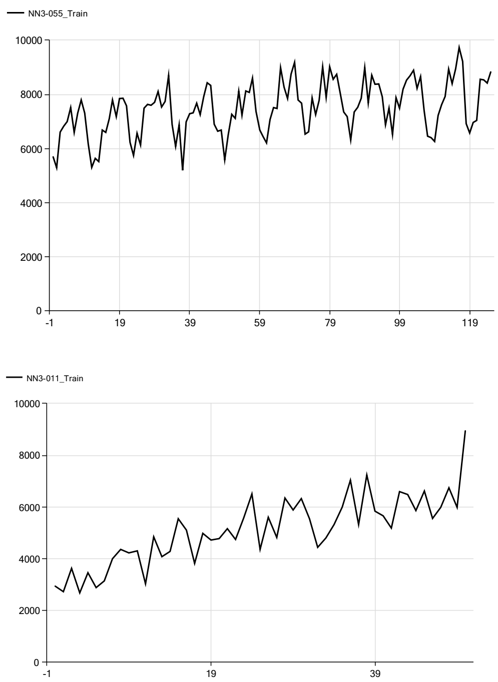

Fig. 7. Two time series from NN3 time series forecasting competition: (top) the 55th time series and (bottom) the 11th time series

For experiment, our work also uses two more real-world datasets from Vietnam stock market, downloaded from the website of Bao Viet Securities Company (http://www.bvsc.com.vn/DownloadMSData.aspx)

The VNINDEX dataset records the composite stock index of Ho Chi Minh City Stock Exchange. This dataset consists of 1000 daily data points taken from 30/12/2009 to 31/12/2013.

The FPT dataset records the stock close prices of the FPT Company. This dataset consists of 385 daily data points from 28/05/2012 to 04/12/2013.

Each dataset is divided into two parts: 80% for training and 20% for testing. The experimental results are used to compare the performance of the different strategies for multi-step-ahead prediction of time series.

Fig. 8 and Fig. 9 show the VNINDEX dataset and the FPT dataset, respectively.

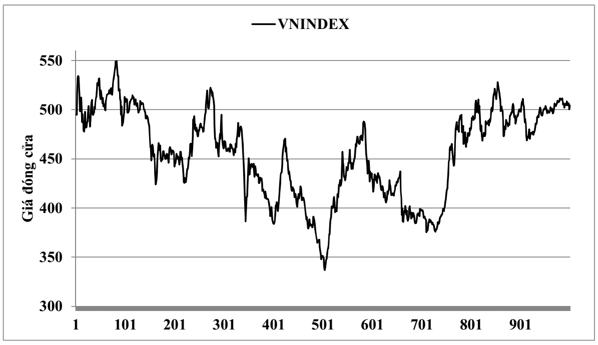

Fig. 8. VNINDEX dataset

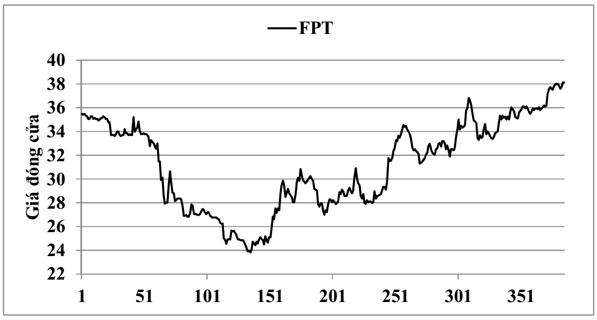

Fig. 9. FPT dataset

B. *Evaluation Criteria*

In this study, the mean squared error (MSE), the mean absolute percentage error (MAPE) and the symmetric mean absolute percentage error (SMAPE) are used as evaluation criteria. The formula for MSE, MAPE and SMAPE are given as follows:

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(\hat{y}_t - y_t)^2$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|\hat{y}_t - y_t|}{y_t}$$

$$SMAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|\hat{y}_t - y_t|}{(y_t + \hat{y}_t)/2}$$

where *n* is the number of observations, y_t is the actual value in time period *t*, and $\hat{y}$ is the forecast value for time period *t*.

C. Experimental Results

• *NN3 Dataset*

In the experiment for NN3 dataset, we used the three-layer feed forward neural network. For the selection of activation function, a sigmoid function and a linear function are adopted in the hidden and output layer, respectively. The number of input neurons is 10. The number of hidden neurons is 8. The number of output neurons depends on the forecasting strategy (e.g., for DirMO, the number of output neurons is 3). The learning rate is 0.01. The momentum constant is 0.9. The forecasting horizon is 18. The time series in the dataset are transformed to the range [0.01, 0.99]. The backpropagation algorithm is used as the learning algorithm and the network is trained less than 2000 epochs and termination condition is based on training error $E \leq 1.0\text{-}E5$.

Table I shows the average test errors over the NN3 dataset with all the five forecasting strategies. The MAPE, SMAPE and MSE of each time series are averaged over 111 time series in the datasets to obtain the mean MAPE*, SMAPE*, MSE* defined as:

$$MAPE^* = \frac{1}{111}\sum_{i=1}^{111} MAPE(i)$$

$$SMAPE^* = \frac{1}{111}\sum_{i=1}^{111} SMAPE(i)$$

$$MSE^* = \frac{1}{111}\sum_{i=1}^{111} MSE(i)$$

where $MAPE(i)$, $SMAPE(i)$, $MSE(i)$ denote the errors collected over the i-th time series in the NN3 dataset.

TABLE I. AVERAGE TEST ERRORS OVER NN3 DATASET WITH ALL THE STRATEGIES

Criterion	Direct	DirMO	DirREC	Iterated	MIMO
MAPE*	23.146	22.372	23.114	24.381	23.371
SMAPE*	19.215	18.964	20.265	19.420	19.361
MSE*	4.07E+08	3.74E+08	4.30E+08	3.49E+08	3.52E+08

The average test errors of all the five forecasting strategies over the NN3 dataset are shown in Table I. From the experimental results in Table I, DirMO < DirREC < Direct < MIMO <Iterated ranking is obtained in terms of MAPE; DirMO < Direct < MIMO < Iterated < DirREC ranking is obtained in terms of SMAPE; and Iterated < MIMO < DirMO < Direct < DirREC ranking is obtained in terms of MSE. It is also noticed that for NN3 dataset, there is not a distinct difference between the strategies in terms of test errors. However, in summary, it appears that DirMO is the best performer for NN3 dataset.

• *VNINDEX Dataset*

In the experiments for both VNINDEX and FPT datasets, we used the three-layer feed forward neural network. For the selection of activation function, a sigmoid function and a linear function are adopted in the hidden and output layer, respectively. The number of input neurons is 10. The number of hidden neurons is 10. The number of output neurons depends on the forecasting strategy (e.g., for DirMO, the number of output neurons s is 3). The learning rate is varied from 0.01 to 0.46. The momentum constant is 0.9. The backpropagation algorithm is used as the learning algorithm and the network is trained less than 2000 epochs and termination condition is based on training error $E \leq 1.0\text{-}E5$.

The test errors (in terms of MAPE) over the VNINDEX dataset are shown in Table II. Here we experiment each strategy with the forecasting horizon varying from 2 to 14. The motivation is to test the stability of each strategy when we forecast further into the future. Note that for the two horizon values 2 and 3, Table II does not report the measurements since in these cases, DirMO is equivalent to MIMO. The test errors (in terms of MSE) over the VNINDEX dataset are given in Table VI (Appendix).

TABLE II. TEST ERRORS OVER THE VNINDEX DATASET WITH ALL THE STRATEGIES (IN TERMS OF $MAPE_{MEAN}$)

Horizon	Direct	DirMO	DirREC	Iterated	MIMO
2	5.851	N/A	5.208	7.170	8.288
3	7.377	N/A	5.178	8.717	6.297
4	8.987	12.614	5.071	10.533	15.145
6	13.284	17.131	5.191	13.443	25.622
8	17.411	20.923	5.966	17.149	26.632
10	18.774	21.315	7.471	20.689	27.724
12	20.934	23.883	9.016	24.689	26.513
14	22.384	24.628	10.411	26.831	25.398
Average	**14.375**	**20.082**	**6.689**	**16.153**	**20.202**

According to the experimental results in Table II and Table VI, it can be seen that DirREC strategy gives the best performance in terms of the average test errors. The DirREC strategy brings out less than fourth of the error of the Direct strategy and gives less than sixth of the error of the Iterated strategy.

From the experimental results in Table II, we can derive the ranking for the five multi-step ahead forecasting strategies as reported in Table III.

For all the five forecasting strategies over the VNINDEX dataset, the forecasting error increases when the forecasting horizon increases. However, regarding the rate of this increasing, DirREC strategy is found as the most stable in comparison with all the other strategies.

• *FPT Dataset*

The test errors (in terms of MAPE) over the FPT dataset are shown in Table IV. Here we experiment each strategy with the forecasting horizon varying from 2 to 14 ($H \in \{2, 4, 6, 8, 10, 12, 14\}$. For DirMO, the number of output neurons s is 3.

According to the experimental results in Table IV, it can be seen that DirREC strategy gives the best performance in terms of the average test errors. The DirREC strategy brings out about haft of the error of the Direct strategy and gives less than eighth of the error of the Iterated strategy. The test errors (in terms of MSE) over the FPT dataset are given in Table VII (Appendix).

TABLE III. THE RANKING OF FORECASTING PERFORMANCE OF ALL THE STRATEGIES OVER VNINDEX DATASET (IN TERMS OF $MAPE_{MEAN}$)

Horizon	Ranking				
	1	2	3	4	5
2	DirREC	Direct	Iterated	MIMO	DirMO
3	DirREC	MIMO	DirMO	Direct	Iterated
4	DirREC	Direct	Iterated	DirMO	MIMO
6	DirREC	Direct	Iterated	DirMO	MIMO
8	DirREC	Iterated	Direct	DirMO	MIMO
10	DirREC	Direct	Iterated	DirMO	MIMO
12	DirREC	Direct	DirMO	Iterated	MIMO
14	DirREC	Direct	DirMO	MIMO	Iterated

TABLE IV. TEST ERRORS OVER THE FPT DATASET WITH ALL THE STRATEGIES (IN TERMS OF $MAPE_{MEAN}$)

Horizon	Direct	DirMO	DirREC	Iterated	MIMO
2	14.649	N/A	13.514	18.202	19.934
3	15.740	N/A	13.851	23.710	26.873
4	16.539	26.301	14.016	29.525	27.118
6	18.571	29.495	14.401	40.590	33.892
8	21.123	29.539	15.042	46.454	34.248
10	23.861	31.170	15.531	49.105	36.411
12	25.806	31.064	15.784	51.682	32.962
14	27.294	31.258	15.841	53.848	29.699
Average	**20.448**	**29.804**	**14.748**	**39.140**	**30.142**

From the experimental results in Table IV, we can derive the ranking for the five multi-step ahead forecasting strategies as in Table V.

Again, for all the five forecasting strategies over the FTP dataset, the forecasting error increases when the forecasting horizon increases. However, regarding the rate of this increasing, DirREC strategy is found as the most stable in comparison with all the other strategies.

Although we experimented on the three datasets NN3, VNINDEX and FPT, we pay more attention to the experimental results from the two stock datasets. The reason is that NN3 datasets contains many time series with trended and seasonal patterns which are not suitable for prediction using ANN. ANN can not capture well trended and seasonal variations in time series if the time series have not been preprocessed by detrending and deseasonalization (Zhang et al., 2005 [18]). The two stock datasets VNINDEX and FPT are the two time series without trend and seasonality. Therefore, in this experiment, we mainly use the experimental results on the two stock datasets to rank the prediction performances of the five strategies for multi-step-ahead prediction. According to the experimental results from the two datasets VNINDEX and FPT, we can see that DirREC is the best strategy for multi-step-ahead prediction since it has the lowest prediction error.

TABLE V. THE RANKING OF FORECASTING PERFORMANCE OF ALL THE STRATEGIES OVER FPT DATASET (IN TERMS OF $MAPE_{MEAN}$)

Horizon	Ranking				
	1	2	3	4	5
2	DirREC	Direct	Iterated	MIMO	DirMO
3	DirREC	Direct	Iterated	MIMO	DirMO
4	DirREC	Direct	DirMO	MIMO	Iterated
6	DirREC	Direct	DirMO	MIMO	Iterated
8	DirREC	Direct	DirMO	MIMO	Iterated
10	DirREC	Direct	DirMO	MIMO	Iterated
12	DirREC	Direct	DirMO	MIMO	Iterated
14	DirREC	Direct	MIMO	DirMO	Iterated

The different results from two different kinds of datasets (NN3 and stock datasets) imply that the intrinsic properties of datasets have a critical impact on the prediction performance of strategies of multi-step ahead time series forecasting.

V. DISCUSSION AND CONCLUSION

Multi-step-ahead forecasting is still an open challenge in time series forecasting. In this work, comparisons of all the five strategies for multi-step-ahead forecasting: Iterated, Direct, DirREC, MIMO and DirMO using feed-forward neural networks as underlying model were presented. Based on the performed comparison with the five forecasting strategies over the three datasets NN3, VNINDEX and FPT, superiority of the DirREC strategy is pointed out. Next to DirREC are the strategies Direct and DirMO. The two strategies Iterated and MIMO give the worst prediction performance.

We attribute the best performance of DirREC strategy to the fact that the DirREC strategy combines the advantages of the Iterated and Direct strategies and this strong point is illustrated in the experimental section.

The findings in this work support two previous studies which claim the superiority of the direct strategy against iterative strategy (Zhang, 1994 [19], Hamzaçebi et al., 2009 [10]) when these authors use ANN as underlying forecasting model for multi-step ahead prediction. The results of our work seem not consistent with one previous work which reported that Multiple-Output strategies (MIMO and DIRMO) are the best strategies (Ben Taieb et al., 2012 [4]). However, we should notice that Lazy Learning algorithm (a local model) had been used as underlying predictor in [4] rather than ANN (a global model) and the authors in [4] tested on only one dataset: NN5. All the different findings from our work and [4] indicate that the underlying prediction model and the characteristics of the tested datasets have critical effects on the accuracy of strategies for multi-step ahead time series forecasting.

Therefore, from out work, it is not possible to draw the conclusion that the DirREC gives best results for all kind of time series or with all kind of underlying forecasting models. Further studies performed with more real-world datasets and different underlying forecasting models are necessary to generalize the conclusion presented in this work.

A possible direction for future work could be applying some other better training algorithms for neural networks to improve its prediction performance. Also, possibly tailoring some detrending and deseasonalization methods for the five strategies could also be a promising research point.

References

[1] E. Alpaydin, Introduction to Machine Learning, The MIT Press, 2004.

[2] S. Ben Taieb, G. Bontempi, A. Sorjamaa, and A. Lendasse, "Long-term prediction of time series by combining direct and MIMO strategies," in *2009 International Joint Conference on Neural Networks*, 2009, pp. 3054–3061.

[3] S. Ben Taieb, A. Sorjamaa, and G. Bontempi, "Multiple-output modeling for multi-step-ahead time series forecasting," *Neurocomputing*, vol. 73, no. 10–12, Jun. 2010, pp. 1950–1957.

[4] S. Ben Taieb, G. Bontempi, A. F. Atiya, and A. Sorjamaa, "A review and comparison of strategies for multi-step ahead time series forecasting based on the {NN5} forecasting competition," *Expert Syst. Appl.*, vol. 39, no. 8, 2012, pp. 7067–7083.

[5] G. Bontempi, "Long term time series prediction with multi-input multi-output local learning," in *Proceedings of the 2nd European Symposium on Time Series Prediction (TSP), ESTSP08*, 2008, pp. 145–154.

[6] G. Bontempi and S. Ben Taieb, "Conditionally dependent strategies for multiple-step-ahead prediction in local learning," *Int. J. Forecast.*, vol. 27, no. 3, 2011, pp. 689–699.

[7] F.-J. Chang, Y.-M. Chiang, and L.-C. Chang, "Multi-step-ahead neural networks for flood forecasting," *Hydrol. Sci. J.*, vol. 52, no. 1, 2007, pp. 114–130.

[8] S. Crone, "Forecasting Competition for Neural Networks & Computational Intelligence," 2008. [Online]. Available: http://www.neural-forecasting-competition.com/NN3/index.htm. [Accessed: 15-Mar-2014].

[9] J. G. De Gooijer, R. J. Hyndman, "25 years of Time Series Forecasting", International Journal of Forecasting 22(3), 2006, pp. 443-473.

[10] C. Hamzaçebi, D. Akay, and F. Kutay, "Comparison of direct and iterative artificial neural network forecast approaches in multi-periodic time series forecasting," Expert Syst. Appl., vol. 36, no. 2, Part 2, 2009, pp. 3839–3844.

[11] T. Hill, L. Marquez, M. O'Connor and W. Remus, "Artificial neural networks for forecasting and decision making", International Journal of Forecasting, Vol. 10, 1994, pp.5-15.

[12] D. M. Kline, "Methods for multi-step time series forecasting neural networks," in G.P. Zhang (Ed.) *Neural Networks in Business Forecasting*, Hershey, PA, USA, IGI Global, 2004, pp. 226–250.

[13] T. Kolarik and G. Rudorfer, "Time series forecasting using neural networks," vol. 25, no. 1, 1994, pp. 86–94.

[14] A. Sorjamaa, and A. Lendasse, "Time series prediction using DirRec strategy", In M. Verleysen (Ed.), ESANN06, European Symposium on Artificial Neural Networks, Bruges, Belgium, 2006, pp. 143- 148.

[15] A. Sorjamaa, J. Hao, N. Reyhani, Y. Ji, and A. Lendasse, "Methodology for long-term prediction of time series," *Neurocomputing*, vol. 70, no. 16–18, 2007, pp. 2861–2869.

[16] G.C. Tiao and R. S. Tsay, "Some advances in nonlinear and adaptive modeling in Time Series", Journal of Forecasting, Vol. 13 (2), 1994, pp. 109-131.

[17] G. Zhang, B. E. Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks:: The state of the art," Int. J. Forecast., vol. 14, no. 1, 1998, pp. 35–62.

[18] G. Zhang, M. Qi, "Neural network forecasting for seasonal and trend time series", European Journal of Operational Research 160, 2005, pp. 501-514.

[19] X. Zhang, "Time series analysis and prediction by neural networks", Optimization Methods and Software, vol. 4, 1994, pp. 151-170.

Appendix

TABLE VI. THE RANKING OF FORECASTING PERFORMANCE OF ALL THE STRATEGIES OVER VNINDEX DATASET (IN TERMS OF MSE_{MEAN})

Horizon	Direct	DirMO	DirREC	Iterated	MIMO
2	0.002772	N/A	0.002164	0.004501	0.008571
3	0.004956	N/A	0.002138	0.006967	0.003227
4	0.007161	0.015797	0.002067	0.010508	0.022231
6	0.015390	0.026177	0.002190	0.017566	0.046062
8	0.024324	0.034188	0.003134	0.027757	0.047535
10	0.027338	0.034193	0.005126	0.038279	0.051389
12	0.032256	0.040522	0.007175	0.049705	0.047665
14	0.035895	0.041966	0.009133	0.056337	0.045832
Average	**0.018761**	**0.032141**	**0.004141**	**0.026453**	**0.034064**

TABLE VII. TEST ERRORS OVER THE FPT DATASET WITH ALL THE STRATEGIES (IN TERMS OF MSE_{MEAN})

Horizon	Direct	DirMO	DirREC	Iterated	MIMO
2	0.020133	N/A	0.017258	0.032910	0.040027
3	0.022687	N/A	0.017742	0.059879	0.067303
4	0.024666	0.063557	0.017871	0.093604	0.067588
6	0.030549	0.076228	0.018476	0.168277	0.093741
8	0.039686	0.075121	0.019717	0.201792	0.094006
10	0.050203	0.080736	0.020635	0.211808	0.101949
12	0.058201	0.079743	0.021063	0.231060	0.087219
14	0.064364	0.080249	0.021081	0.247044	0.072876
Average	**0.038811**	**0.075939**	**0.019231**	**0.155797**	**0.078089**