Fitting Conventional Neural Network Time Series Models on Sand Price Indices Dataset

Nor Azura Md Ghani1*, Saadi Bin Ahmad Kamaruddin1, Ismail Musirin2, Hishamuddin Hashim3

1Center for Statistical Studies and Decision Sciences, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
2Faculty of Electrical Engineering, Universiti Teknologi MARA, Shah Alam, Selangor Darul Ehsan, Malaysia
3Pilihsepak Sdn. Bhd., Lot 5431, Tingkat 1(B), Jalan J9, Fasa 6, Taman Melawati, Kuala Lumpur, Malaysia
*Corresponding author E-mail: azura@tmsk.uitm.edu.my

Abstract

This result-based paper discusses on the best aftereffects of both fitted BPNN-NAR and BPNN-NARMA on MCCI Sand dataset regarding distinctive error measures. This exploration examine the outcomes as far as the execution of the fitted forecasting models by every arrangement of input lags and error lags utilized, the execution of the fitted anticipating models by various hidden nodes utilized, the execution of the fitted estimating models when joining both inputs and hidden nodes, the consistency of error measures utilized for the fitted determining models, and in addition the general best fitted estimating models for Malaysian sand price indices dataset. In this examination, Malaysian sand price indices monthly data from January 1980 to December 2013 were adapted. The examination of BPNN-NAR on Malaysian sand data shows that insufficient or inadequate combination of input and error lags lead to greater RMSE. Correspondingly, the number of input lags for BPNN-NAR, as well as the number of input and error lags for BPNN-NARMA really have direct effect to the models’ performances. The higher or the lesser hidden nodes to the input lags, the higher the network’s RMSE. On the other hand, the higher the input lags lead to the higher network’s RMSE.

Keywords: Neural Network; Conventional Nonlinear Autoregressive (NAR); Conventional Nonlinear Autoregressive Moving Average (NARMA); Sand Price Indices

1. Introduction

As indicated by [1], input slacks alone may not be satisfactory to absolutely harsh the fundamental time series process. He included that one slack can’t catch the fundamental relationship or not adequate to show a procedure, in this manner there is a need of satisfactory number of shrouded neurons for ideal system execution. The two sources of info and shrouded hubs can fundamentally influence the learning of a neural system when anticipating a period arrangement. This examination is keen on exactness of the models rather than stingy time arrangement models, as recommended by [2]. For time arrangement, inputs likewise incorporate slacks. As the information dimensionality builds, display intricacy increments and learning turns out to be more troublesome, prompting poor joining. The test here is to choose or locate the best blends of information slacks with sufficient number of concealed hubs that will prompt a predominant model, particularly to a specific dataset of a specific situation. On the off chance that the preparation set is adequately vast, ANN will sum up precisely and will deliver exact yields for inputs not in the preparation set. At the point when neural systems are legitimately prepared, they can sum up and extrapolate extra subtle elements of the capacity mapping the contributions to yields. All things considered, no earlier presumptions of needy and autonomous factors should be made since the neural system is prepared on watched information [3, 4]. This construes neural frameworks are usually utilized as “discovery” gadgets that is no prior data about the strategy was accepted; the objective was to build up a procedure demonstrate construct just with respect to perceptions of its information yield conduct [5, 6, 7]. By definition, the black-box is a demonstrating strategy that builds the model simply in light of the data picked up from the structure, and it doesn’t rely upon other data about the framework [8]. No early assumptions about the model structure are made, and rather, the demonstrating method’s worry is to make a non specific model that maps the data yield relationship of the dataset [9, 10]. Disclosure models are known to be suitable and adaptable, as the aide data about the framework may not be speedily open. From this examination, one deficiency lies in the way that that the model multifaceted nature grows [7] Also, another great position of the discovery distinguishing proof is that it can show stream that are definitely gotten by numerical models [12, 13].

2. Methodology

In this examination, Malaysian sand price indices monthly data from January 1980 to December 2013 were adapted. These datasets are auxiliary information gathered from three unique sources, UKAS, from Malaysian Prime Minister's Office, Malaysian CIDB, and Malaysian Statistical Department. In this examination, just Malaysian focal locale total value records were adjusted which involve 0% exceptions. The aggregate N=408 (a year x 34 years) from January 1980 to December 2013 (base 1980=100). As a matter of first importance, this examination talks about the execution of the fitted estimating models on Malaysian total Price lists information by each arrangement of info slacks and mistake slacks utilized. Table 1 indicates execution aftereffects of fitted
anticipating models of two-layer tansig/direct exchange works on Malaysian sand Price records information. For examination purposes, this exploration utilized information slacks from 5 to 40 and concealed hubs from 5 to 45 for BPNN-NARMA show, while for the BPNN-NARMA display, this exploration utilized the mistake slacks going from 5 to 40, together with input slacks from 5 to 40 and concealed hubs from 5 to 45. The most widely recognized way to deal with choose this property is experimentation or experimentation [14, 15, 16], though extraordinary systems (dependable guideline) and calculations (pruning and developing) are also available [17, 18, 19]. It is to be seen that BPNN-NAR is a feed-forward neural system write demonstrate, while BPNN-NARMA is an intermittent neural system compose display [18]. This investigation reports the blunder measures on the test dataset, which is the most key trademark, mirroring ANN's speculation capacity [19, 20]. Table 2 demonstrates the information parceling for organize pre-handling. Figure 1 demonstrates the flowchart of the BPNN-NAR and BPNN-NARMA.

3. Research Findings

As a matter of first importance, this examination talk about the execution of the fitted determining models on Malaysian sand Price indices data by every arrangement of input lags and error lags utilized. Figure 2 demonstrates the execution of BPNN-NAR as for input lags and hidden nodes on Malaysian sand Price indices data in view of RMSE. Then again, Figure 3 demonstrates the execution of BPNN-NARMA as for input lags and hidden nodes on Malaysian sand Price indices data in light of RMSE. Not the same as the outcomes with Malaysian total data, it can be watched that the general lines of BPNN-NARMA are more towards the focal point of zero qualities, which means zero error contrasted with BPNN-NAR. In view of RMSE aftereffects of Malaysian Sand data, this examination can see that both BPNN-NAR and BPNN-NARMA models on Malaysian Sand data create much littler errors.

In view of input and error lags 5, the ideal number of hidden nodes was 20. In light of input and error lags 10, the ideal number of hidden nodes was 10. In view of input and error lags 15, the ideal number of hidden nodes was 15. In light of input and error lags 20, the ideal number of hidden nodes was 20. In light of input and error lags 25, the ideal number of hidden nodes was 25. In view of input and error lags 30, the ideal number of hidden nodes was 30. In view of input and error lags 35, the ideal number of hidden nodes was 35. In view of input and error lags 40, the ideal number of hidden nodes was 45.

In view of these outcomes, BPNN-NARMA display performed superior to the BPNN-NAR show. This is parallel with the discoveries of [21], even many works have demonstrated that the BPNN-NARMA show beat BPNN-NAR display [22-24].

4. Conclusions

The examination of BPNN-NAR on Malaysian sand data shows that insufficient or inadequate combination of input and error lags lead to greater RMSE. Correspondingly, the number of input and error lags really have direct effect to the network’s RMSEs. The higher or the lesser hidden nodes to the input lags, the higher the RMSE. On the other hand, the higher the input lags lead to the higher network’s RMSE. Among the earliest attempts are [26-27]. From the results, obviously that there is no need to adjust NAR and NARMA models when handling data with zero outliers problem. However, there is a need to modify NAR and NARMA models so that they may handle outliers issue effectively, or else the development of NAR model to NARMA demonstrate should not be proceeded since the NARMA model will tend convey greater mistakes, and the results are not strong for additionally use [28-36].

It is immovably trusted that the reason of this finding because of non-presence of anomaly issue in the data. In a book entitled ‘Nonlinear framework distinguishing proof: from traditional ways to deal with neural systems and fluffy models’ by [25], it is said that the neural system perform well when the data don’t comprise of anomalies.

By and large, as the quantity of input and error lags expanded, the errors are likewise expanded. From this exploration, if the system is given sufficient hidden nodes, it might surmised the data well.

Table 1: Combinations of Input Lags, Error Lags and Percentage of Outliers for BPNN - NAR and BPNN - NARMA Models

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Notation</th>
<th>Outliers</th>
<th>Input Lags</th>
<th>Error Lags</th>
<th>Hidden Nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monthly Malaysian Sand Price Indices Data</td>
<td>Sand</td>
<td>0%</td>
<td>5, 10, 15, 20, 25, 30, 35, 40</td>
<td>5, 10, 15, 20, 25, 30, 35, 40</td>
<td>5, 10, 15, 20, 25, 30, 35, 40</td>
</tr>
</tbody>
</table>

Table 2: Size of Data Partitioning for Training, Validation and Testing Sets for Each Data Used in This Research

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Total Sample Size (N)</th>
<th>In-sample Data</th>
<th>Out-of-sample Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training (70%)</td>
<td>Validation (15%)</td>
<td>Testing (15%)</td>
</tr>
<tr>
<td>Sand</td>
<td>408</td>
<td>286</td>
<td>61</td>
</tr>
</tbody>
</table>

Table 3: Best Results of Ordinary BPNN-NAR and BPNN-NARMA Models on Malaysian Aggregate Price Indices Data based on Different Lags

<table>
<thead>
<tr>
<th>Input Lags</th>
<th>Error Lags</th>
<th>Hidden Nodes</th>
<th>RMSE ANN Models</th>
<th>RMSE BPNN-NARMA Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>5</td>
<td>20</td>
<td>0.146</td>
<td>0.120</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>10</td>
<td>0.123</td>
<td>0.119</td>
</tr>
<tr>
<td>15</td>
<td>15</td>
<td>15</td>
<td>0.152</td>
<td>0.124</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>20</td>
<td>0.161</td>
<td>0.130</td>
</tr>
<tr>
<td>25</td>
<td>25</td>
<td>20</td>
<td>0.168</td>
<td>0.135</td>
</tr>
<tr>
<td>30</td>
<td>30</td>
<td>35</td>
<td>0.175</td>
<td>0.139</td>
</tr>
<tr>
<td>35</td>
<td>35</td>
<td>35</td>
<td>0.165</td>
<td>0.135</td>
</tr>
<tr>
<td>40</td>
<td>40</td>
<td>45</td>
<td>0.141</td>
<td>0.124</td>
</tr>
</tbody>
</table>
Fig. 1: Flowchart of Backpropagation Conventional Neural Network 
BPNN-NAR and BPNN-NARMA Mechanism
Fig. 2: Performance of BPNN-NAR with Respect to Input Lags and Hidden Nodes on Malaysian Sand Price Indices Data based on RMSE

Fig. 3: Performance of BPNN-NARMA with Respect to Input Lags, Error Lags and Hidden Nodes on Malaysian Sand Price Indices Data based on RMSE

Acknowledgement

We want to express our gratefulness and appreciation to Unit Kerjasama Awam Swasta (UKAS) over Prime Minister's Office, Construction Industry Development Board (CIDB) then Malaysian Statistics Department. Appreciation to Universiti Teknologi MARA for financial support under the Grant Scheme No. 600-RMI/DANA 5/3/CIFI (65/2013).
References


