European Journal of Statistics and Probability
Vol.9, No.1, pp, 11-16, 2021
Print ISSN: 2055-0154(Print),
Online ISSN 2055-0162(Online)

A SUGGESTED APPROACH TO ARTIFICIAL NEURAL NETWORKS MODELING OF TIME SERIES

Ismaila A. Bolarinwa and Bushirat T. Bolarinwa
Department of Statistics, The Federal Polytechnic, P.M.B. 55, Bida, Niger State, Nigeria

ABSTRACT: *This research proposed a method of selecting an optimal combination of numbers of input (number of lagged values in the model) and of hidden nodes for modeling seasonal data using Artificial Neural Networks (ANN). Three data sets (rainfall, relative humidity and solar radiation) were used in assessing the proposed procedure and the resulting ANN models were compared with two traditional models (Holt-Winter's and SARIMA). Models with large number of lagged values have shown tendency to outperform those with small number of lagged values. Selected ANN model was found to outperform the two traditional models on rainfall data; it performed better than SARIMA but worse than Holt-Winter's model on relative humidity data and performed worse than the two methods on solar radiation data. The proposed procedure has hence, performed fairly well. Oscillatory performance recorded by ANN models that resulted from the proposed procedure in relation to the other two models only attests to the fact that no particular model is best on every data set. Rather than insist on elegance or sophistication, researchers should be guided by parsimony.*

KEYWORDS: artificial neural networks, SARIMA, holt-winter, seasonal data, forecasting

INTRODUCTION

Seasonal time series data feature in virtually all areas of human endeavor. In economics, it may be average inflation rates recorded on monthly or quarterly basis, Gross Domestic Product on quarterly basis, annual volume of export and so on. In meteorology, it may be annual volume of rainfall; in health, it may be infant mortality rates computed on annual basis, annual deaths from cancer of tuberculosis and so on. Unlike most series that are characterized by mainly the trend which can be simply modeled by aa suitable trend equation, the case of seasonal data is different. The difference lies in the need to take cognizance of the seasonal nature of the data, else, poor results will result from forecasting. In response to this need, methods as decomposition model, SARIMA, Holt-Winter's among others were proposed. One of the most recent responses to this need is Artificial Neural Networks (ANN) models. Though not primarily designed for statistical analysis, ANN models have not only been applied in statistical modeling, they have in fact been shown to be superior to traditional approaches in many modeling situations. Zhang (2003), Mitria, Lee and Wu (2009), Kihoro, Otieno and Wafula (2004), Raraway and Chatfield (1998), Alon, Qi and Sadowski (2001) and Adebisi, Adewumi and Ayo (2014) are some of such comparative studies. Applications of ANN in time series analysis are too numerous to list. However, a few are Mapuwei, Bodhlyera and Mwanbi (2020), Jere, Kasesnse and Bwalya (2017), Herliansyah, Jamilatuzzahro (2017), Moreno, Pol and Gracia (2011) and Zhang and Qi (2005).

European Journal of Statistics and Probability
Vol.9, No.1, pp, 11-16, 2021
Print ISSN: 2055-0154(Print),
Online ISSN 2055-0162(Online)

This research proposes a method of selecting an optimal combination of number of input nodes (number of lagged values in the model) and number of hidden nodes in the course of modeling seasonal data using ANN models.

The remainder of the article is structured as follows: Section 2 presents the Theoretical Framework; Section 3 presents the Methodology; Section 4 presents the Results and Discussion while the last section concludes the article.

RESULTS AND DISCUSSION

Results for different model are condensed into Tables 1 to 6.

Table 1. Mean square errors of different ANN architectures for rainfall data

	H					
p	1	2	3	4	5	6
1	7040.529	6896.521	7803.026	6686.9248	7289.080	7238.430
2	7029.870	6967.096	6155.394	7013.143	6649.188	6400.125
3	5610.886	5582.189	5896.233	5623.450	6266.286	7777.794
4	4800.307	4706.601	4352.281	4613.013	4959.230	4821.573
5	3547.457	3329.952	3500.447	3482.860	3538.451	3785.012
6	2900.960	2819.320	2938.159	2854.515	4341.178	2217.388
7	2710.007	2639.053	2605.478	2693.583	2070.016	2694.137
8	26780.040	2716.042	2913.837	2719.423	2674.166	**1155.712**
9	2661.461	2671.231	2612.545	3376.564	1719.885	2569.338
10	2658.12	2538.893	2762.243	2978.847	2645.419	2667.374
11	2813.768	2161.240	1484.604	2094.161	2147.117	3082.128
12	2678.019	2388.854	2504.630	2227.165	2557.277	2279.114

Table 1 presents mean square errors obtained from the testing set for rainfall data. Typically, at a given number (h) of hidden nodes, mse reduced as p, the number of output neurons (lagged values) increased. An indication that improved accuracies were achieved with inclusion of additional lagged variables into the model. On the other hand, at fixed p, mse has not shown an identifiable pattern as h increased. The minimum observed mse of 1155.712 corresponds to p=8 and h=6, implying SANN (8, 6, 12). This implies that although higher number of lagged values is associated with higher precision, the relationship is not perfect since a model having 8 lagged values has performed best in the presence of models with higher number of lagged values.

Table 2. Mean square errors of different ANN architectures for relative humidity data

	H					
p	1	2	3	4	5	6
1	112.329	114.045	110.514	112.752	110.020	110.938
2	111.417	105.614	116.313	106.264	121.419	112.570
3	75.874	67.037	60.448	66.521	90.810	68.975
4	73.472	67.047	50.517	62.981	56.336	51.202
5	71.820	56.592	43.211	66.045	56.227	60.657
6	66.196	54.409	51.746	57.658	54.672	59.949
7	66.291	57.885	58.926	90.340	53.452	56.010
8	66.510	56.846	49.767	53.820	55.055	57.308
9	67.047	63.691	42.298	50.948	52.498	55.622
10	64.813	54.377	44.043	46.640	50.629	63.113
11	55.271	50.182	47.388	50.232	86.292	48.606
12	63.366	47.791	39.720	44.200	**45.497**	69.665

Results for relative humidity are presented in Table 2. The observed relationships in Table 1 in respect of p and h also prevail here. However, the least mse of 45.497 applies to p=12; h=5, corresponding to SANN (12, 5, 12).

Table 3. Mean square errors of different ANN architectures for relative solar radiation

	H					
p	1	2	3	4	5	6
1	2.872	2.762	2.718	3.243	2.779	2.772
2	2.491	2.376	2.358	2.517	3.540	3.932
3	2.501	2.421	2.314	2.371	2.511	2.426
4	2.504	2.917	2.214	2.313	2.567	3.696
5	2.512	2.212	2.874	2.825	1.675	2.762
6	2.453	2.439	1.909	2.139	2.574	2.001
7	2.459	2.355	2.576	2.130	1.653	2.074
8	2.278	1.454	1.120	2.024	1.783	1.937
9	2.279	2.237	1.753	1.620	1.201	1.191
10	2.300	1.557	2.509	3.411	1.586	1.373
11	2.001	1.177	0.941	1.613	0.727	1.243
12	1.342	1.177	0.941	**0.810**	0.927	1.243

Table 3 presents results for solar radiation data. The earlier observed relationships between fixed h and varying p and between fixed h and varying p still hold. The least mse (0.810) corresponds to p=12; h=4, signifying SANN (12, 4, 12). Results obtained for the three sets of data so far suggest that models with large number of lagged values tend to dominate. However, many more empirical evidences will be required to validate the claim.

Print ISSN: 2055-0154(Print),
Online ISSN 2055-0162(Online)

Tables 4 to 6 present the forecasts for January to December2019 and mean square errors.

Table 4. Comparison of forecasts of ANN to those of other models (Rainfall Data)

Month	True value (2019)	ANN	H-W	SARIMA
Jan	0	0.2855	16.3288	0.642
Feb	0.9	15.7866	18.9085	2.063
Mar	0	-12.8496	33.5181	15.166
Apr	16.1	23.2098	66.0376	46.968
May	178.6	146.1441	181.9229	170.033
Jun	312.3	218.3436	195.7313	177.178
Jul	141.5	189.8225	243.8224	220.375
Aug	420.4	209.8689	264.34	245.095
Sep	275.2	309.0187	241.8691	217.871
Oct	213.2	167.0264	117.4561	92.502
Nov	0	-0.6301	24.3943	2.894
Dec	0	9.6157	22.0262	-0.498
MSE	-	5028.899	5332.512	6193.999

On the basis of mse obtained from the forecasts and true values for January to December 2019, ANN model has performed best on rainfall data. This lays credence to the hypothesis that the approach is worthwhile.

Table 5. Comparison of the forecasts of ANN to those of other models (Relative Humidity data)

Month	True value (2019)	ANN	H-W	SARIMA
Jan	51	47.7752	53.0541	48.8851
Feb	59	50.1338	54.3138	46.108
Mar	71	58.3983	61.7672	56.0524
Apr	67	62.4925	69.3103	62.4619
May	76	75.2586	76.5816	71.3145
Jun	84	75.8231	84.6633	78.2264
Jul	85	82.3215	84.1297	78.3188
Aug	87	84.6742	89.1783	82.2804
Sep	94	86.4117	88.4162	81.6972
Oct	85	84.7006	86.4071	78.9194
Nov	75	77.3164	74.7158	67.2727
Dec	56	62.9082	60.7251	52.8346
MSE	-	38.2403	14.8841	66.2471

For the relative humidity data, ANN lagged behind Holt-Winter's model but performed better than Decomposition model. As fundamental and simple as decomposition model appears to be, it has outperformed more sophisticated models in this case. This drives home the point that model performance is not a function of model sophistication. The idea is to fit simple, yet adequate model to data being modeled as such ensures model stability and hence, reasonable out-of-sample forecasts.

Table 6. Comparison of forecasts of ANN to those of other models (Solar Radiation data)

Month	True value (2019)	ANN	H-W	SARIMA
Jan	16.7	14.7599	15.5349	16.024
Feb	17.08	18.4953	17.4661	18.2838
Mar	19.7	18.2435	18.1181	18.48
Apr	18.4	18.0647	17.6974	18.2618
May	16.9	16.1368	16.2123	16.6131
Jun	13.9	15.4324	15.5998	15.9856
Jul	13.09	14.0259	13.9049	14.4749
Aug	12.6	13.3816	12.4469	13.2888
Sep	16.7	15.3126	15.3946	16.0817
Oct	14.5	17.8234	17.0992	17.5676
Nov	18.9	18.5917	18.0874	18.6922
Dec	18.5	16.527	16.1274	16.927
MSE	-	2.4600	1.9452	1.8892

ANN performed worse than the two other models for the case of solar radiation data. That the suggested ANN approach has not performed better than the other models on all three sets of data is not a shortcoming as far as modeling is concerned. It is just normal that sometimes a model outperforms others in a particular modeling situation and other times outperformed by same models outperformed earlier. This is not in any way ruling out the need for further improvements on the proposed approach.

CONCLUSION

An approach to fitting ANN models to seasonal data has been proposed. Compared to Holt-Winter's and SARIMA models, forecasts obtained from the proposed ANN modeling procedure have competed fairly well. Models with large number of lagged values have shown tendency to outperform those with small number of lagged values. If limited time is available for generating forecasts through ANN models, search for appropriate model may be limited to models with 12 lagged values when monthly data are involved. Oscillatory performance recorded by ANN models that resulted from the proposed procedure in relation to the other

two models only attests to the fact that no particular model is best on every data set. Rather than insist on elegance or sophistication, researchers should be guided by parsimony

References

Adebiyi, A. A., Adewumi, A. O. & Ayo, C. K. (2014). Comparison of ARIMA and Artificial neural networks models for stock price prediction, *Journal of Applied Mathematics*, 2014 Article ID 614342 7pp.

Adhikari, R. & Agrawal, R. K. (2012). Forecasting strong seasonal time series with artificial neural networks, *Journal of Scientific and Industrial Research*, 71, 657-666.

Alon, I, Qi, M. & Sadowski, R. J. (2001). Forecasting aggregate retail sales: a comparison of artificial neural networks and traditional methods, *J. Retailing and Consumer Services*, 8, 147-156.

Elkateb, M. M., Solaiman, K. & Al-Turki, Y. (1998). A comparative study of medium-weather-dependent load forecasting using enhanced artificial/fuzzy neural network and statistical techniques, *Neurocomputing*, 23**,** 3-13.

Faraway, J. & Chatfield. C. (1998). Time series forecasting with neural networks: A comparative study of the airline data, *J. App. Statist*, 47, 231-250.

Gomez-Ramos, E. & Venegas-Martinez, F. Y. (2013). A review of artificial neural networks: How well do they perform in time series forecasting?, *Analitia*, 6(2), 7-15.

Gurney, K. (1997). *An Introduction to Neural Networks*. London: UCL Press.

Herliansyah, R. & Jamilatuzzahro, I. (2017). Feed forward neural networks for forecasting Indonesia exchange composite index, *GSTV Journal of Mathematics, Statistics and Operations Research* 4, 8-15.

Jere, S. Kasense, B. & Bwalya, B. B. (2017). Univariate time series analysis of second-hand car of importation in Zambia, *Open Journal Statistics*, 97, 718-730.

Kihoro, J. M. Otieno, R. O. & Wafula. C. (2004). Seasonal time series forecasting: A comparative study of ARIMA and ANN models, *African J Sc and Techno* 5, 41-49.

Mapuwei, T. W., Bodhlyera, O. & Mwambi, H. (2020). Univariate time series analysis of short-term forecasting horizons using artificial neural networks: The case of public ambulance emergency preparedness, *Journal of Applied Mathematics*, 2020.https://doi.org/10.1155/2020/2408698.

Mitrea, C. A., Lee, C. K. M. & Wu, Z. (2009). A comparison between neural networks and traditional forecasting methods: A case study, *International Journal of Engineering Business Management*, 1(2), 11-24.

Moreno, J. J. M, Pol, A. F. & Gracia, P. M. (2011). Artificial neural networks applied to forecasting time series, *Psicothema*, 23(2), 322-329.

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model, *Neurocomputing*, 50, 159-175.

Zhang, G. P. & Qi, M. (2005). Neural networks forecasting for seasonal and trend time series, *European Journal of Operations Research*, 160, 501-514.

Acknowledgement

We express appreciation to the Management of National Cereals Research Institute, Baddegi, Bida, Nigeria for making the data for the research available