Forecasting Seasonal Time Series Using Weighted Gradient RBF Network based Autoregressive Model

Wenjie Ruan
School of Computer Science
The University of Adelaide
Adelaide, Australia
wenjie.ruan@adelaide.edu.au

Quan Z. Sheng
School of Computer Science
The University of Adelaide
Adelaide, Australia
qsheng@cs.adelaide.edu.au

Peipei Xu
School of Electronic Engineering, UESTC
Chengdu, China
peipei.xu6@gmail.com

Nguyen Khoi Tran, Nickolas J.G. Falkner
School of Computer Science
The University of Adelaide
nguyen.tran@adelaide.edu.au

Xue Li
School of IT and EE
The University of Queensland
Brisbane, Australia
xueli@itee.uq.edu.au

Wei Emma Zhang
School of Computer Science
The University of Adelaide
Adelaide, Australia
wei.zhang01@adelaide.edu.au

ABSTRACT

How to accurately forecast seasonal time series is very important for many business area such as marketing decision, planning production and profit estimation. In this paper, we propose a weighted gradient Radial Basis Function Network based AutoRegressive (WGRBF-AR) model for modeling and predicting the nonlinear and non-stationary seasonal time series. This WGRBF-AR model is a synthesis of the weighted gradient RBF network and the functional-coefficient autoregressive (FAR) model through using the WGRBF networks to approximate varying coefficients of FAR model. It not only takes the advantages of the FAR model in nonlinear dynamics description but also inherits the capability of the WGRBF network to deal with non-stationarity. We test our model using ten-years retail sales data on five different commodity in US. The results demonstrate that the proposed WGRBF-AR model can achieve competitive prediction accuracy compared with the state-of-the-art.

1. INTRODUCTION

A time series is a sequence of data points made of successive measurements across a continuous time intervals, such as ocean tides data, counts of sunspots data, and the daily closing value of the Dow-Jones Industrial Average. The seasonal and trend patterns are usually served as basic components in many time series, especially in economic and business areas [1]. This paper proposes a novel model structure to accurately model the time series that exhibits seasonal, highly nonlinear and non-stationary features. Figure 1 shows examples of typical seasonal time series - monthly retail sale data of department, hardware and clothing in US from the year 1992 to 2001. We observe that these seasonal time series data reflects *i)* strong nonlinearity that cannot be well handled by linear models, *ii)* significant seasonal patterns which repeat a similar fluctuation pattern but with a stable increase each year.

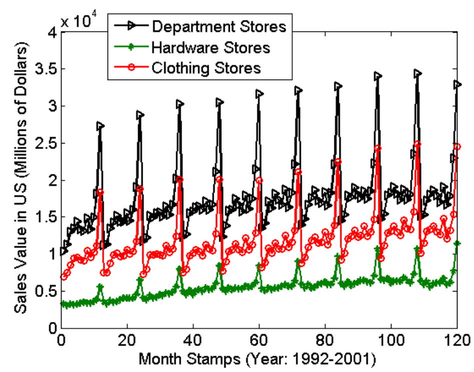

Figure 1: Three sales time-series data from USA

To accurately model this type of time-series data, many remarkable approaches are proposed over last decades [1, 20]. The most straightforward technique is to decompose the series into different components (such as seasonal part, trend part and cyclical components) by using some seasonal adjustment methods [20]. However, such decomposition may lose some valuable information among the data [12]. Another classical solution is built upon the famous Box-Jenkins model [13] that first derives a stationary time series by the time-series differencing, then the traditional Auto Regressive Moving Average (ARMA) model is applied, which is called Autoregressive Integrated Moving Average (ARIMA) [1]. However, this technique assumes that the main component of the time series is linear, thus being difficult to capture the nonlinearity in seasonal time series [2].

Recently, Artificial Neural Networks (ANNs) have attracted huge attentions and become one of the most favorable methods for time series modeling and prediction [6, 18]. Because of their remarkable universal approximation capability, various types of ANNs are also investigated to forecast the seasonal time series. For example, Zhang *et al.* [20] for the first time utilize three-layer feed-forward ANNs to model the time

CIKM'16 , October 24-28, 2016, Indianapolis, IN, USA

DOI: http://dx.doi.org/10.1145/2983323.2983899

series with seasonal and trend patterns. In [19], they further systematically investigate which factors in feed-forward ANNs have impact on the performance of quarterly time-series prediction. Nevertheless, to achieve satisfied performance, ANNs normally require a large model size (*i.e.,* many parameters) and thus need more training data.

As a result, another technique that combines both merits of ANNs and ARMA has been emerged recently for time series prediction - the neural network based functional varying-coefficient models [2, 5]. One of the modeling methods is the radial basis function network based autoregressive (RBF-AR) model [9, 16]. The RBF-AR model adopts a set of RBF networks to approximate the coefficients of a state-dependent AR model [14]. As such, RBF-AR model inherits both advances of the state-dependent AR model in the description of nonlinear dynamics and the merit of the RBF networks in functional approximation. However, applying RBF-AR into seasonal time-series prediction still requires to derive a stationary time series beforehand, leading to latent information loss thus fail to capture the fine-grained fluctuations in seasonal time series [7].

To fill this gap, this paper proposes a weighted gradient RBF network-based AR model to predict the seasonal time series. Specifically, instead of using RBF network, we design a gradient RBF (GRBF) network with regression weights to approximate the vary-coefficients in FAR. The GRBF network is a powerful ANN that can well quantify the non-stationarity in seasonal time series [4]. To further enhance its approximation capacity, we augment a linear regression function as the connection weights between the hidden layer units and the output in traditional GRBF networks instead of constants. In summary, the main contributions of this paper are as follows:

- We propose to approximate the vary-coefficients in functional autoregressive model using a wighted gradient RBF networks, called WGRBF-AR model that can model a class of nonlinear and non-stationary time series with a high accuracy. To the best of our knowledge, this paper is the first one to do so.

- We apply our model to forecast real seasonal time series. The results on five ten-year real retail sales data in US demonstrate that, with a same or smaller model size, WGRBF-AR model outperforms other similar models in terms of prediction accuracy.

2. THE PROPOSED SOLUTION

2.1 Weighted Gradient RBF Network

The traditional GRBF network is a three-layer feed-forward network, which includes an input layer, a hidden layer and an output layer [4]. The input vector of GRBF network is the gradient of the time series modeled, thus the hidden layer can respond to the gradient changes of data efficiently. Among others, the output of each hidden node is also revised with a term which is to accelerate the performance of the one-step prediction [6]. Assuming the time series under study is $\{y_t : t = 1, 2, ..., N\}$, the traditional gradient RBF network is defined as:

$$\hat{y}_i = \sum_{k=1}^{m} w_k \exp\{-\lambda \|\mathbf{x}_i - \mathbf{c}_k\|^2\} \times (y_{i-1} + \delta_k) \qquad (1)$$

where m represents the node number in the hidden layer, δ_k is a constant value associated with the center, w_k means the network weight, $\mathbf{x}_i = [y_{i-1} - y_{i-2}, y_{i-2} - y_{i-3}, \cdots, y_{i-d} - y_{i-d-1}]^T$ represents d- dimensional input vector, $\mathbf{c}_k$ is the center vector, λ is a scaling number ($\lambda > 0$) that measures the width of the symmetric response in the hidden node, and $\|\cdot\|$ is the vector 2-norm. The centers $\mathbf{c}_k$ and scalars $\delta_k (1 \leq k \leq m)$ are chosen from the training data $\{\mathbf{x}_i\}$. If $\{\mathbf{x}_j\}$ is selected as the center $\mathbf{c}_k$, the corresponding $\delta_k = y_j - y_{j-1}$. We use the Orthogonal Least Squares (OLS) algorithm [3] to select the subset $\mathbf{c}_k$.

In this paper, to further accelerate its approximation capability, we use a regression-type weight, which is a function of the input variables, to replace the constant weights in the traditional GRBF network, *i.e.,* $w_k = v_{k0} + v_{k1}(y_{i-1} - y_{i-2}) + ... + v_{kd}(y_{i-d} - y_{i-d-1})$. Fig 2 illustrates the structure of our WGRBF model. As we can see, each input node contributes to the network weights. This direct link between the weights and the input gradient variables makes the WGRBF network more efficient and thus produces a more parsimonious network.

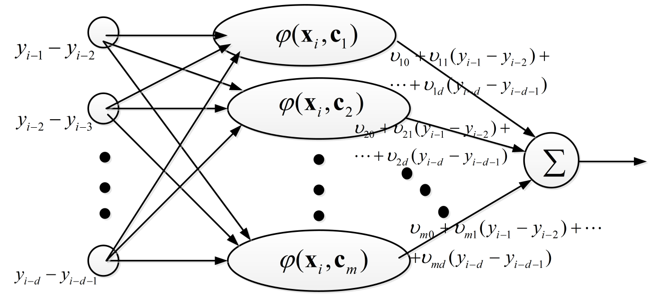

Figure 2: Structure of the WGRBF networks

2.2 WGRBF-AR Model

As aforementioned, the functional AR model is a flexible structure with strong capability of modeling nonlinear time series. We intend to enable the FAR model to model the non-stationary time series. To achieve this goal, we approximate the functional coefficients in the FAR model by using the proposed WGRBF networks. The traditional FAR model has following structures.

$$y_t = \phi_0(\mathbf{X}_t^*) + \sum_{i=1}^{p} \phi_i(\mathbf{X}_t^*) y_{t-i} + \varepsilon_t \qquad (2)$$

where $\mathbf{X}_t^*$ means the state vector consisted by lagged values of y_t, p is the model order and ε_t represents a sequence of *i.i.d.* random variables which is independent of the observations $y_{t-i} > 0$; $\phi_0(\mathbf{X}_t^*), \cdots, \phi_p(\mathbf{X}_t^*)$ indicates the functional coefficients.

In reality, most of the seasonal time series not only exhibits a trend and a local level pattern but also reveals some homogeneous features, such as some parts in the time series are very similar to other parts (see the yearly repeated pattern of sales time series in Fig. 1). We assume that the modeled time series generally have those properties. Thus we derive the WGRBF-AR model by specifying functional coefficients of the FAR model using the weighted GRBF networks. Specifically, the proposed WGRBF-AR model has a structure (without loss of generality, the first-order

WGRBFs are used) as follows.

$$\begin{cases} y_t = \phi_0(\mathbf{X}_t^*) + \sum_{i=1}^{p} \phi_i(\mathbf{X}_t^*) y_{t-i} + \varepsilon_t \\ \phi_0(\mathbf{X}_t^*) = w_0 + \sum_{k=1}^{m} w_k \exp\{-\lambda \|\mathbf{X}_t^* - \mathbf{Z}_k\|^2\} \\ \phi_i(\mathbf{X}_t^*) = w_{i,0} + \sum_{k=1}^{m} w_{i,k} \exp\{-\lambda \|\mathbf{X}_t^* - \mathbf{Z}_k\|^2\} \\ \mathbf{X}_t^* = [y_{t-1} - y_{t-2}, y_{t-2} - y_{t-3}, \cdots, y_{t-d} - y_{t-d-1}]^T \\ w_{i,k} = v_{i0,k} + v_{i1,k}(y_{t-1} - y_{t-2}) + v_{i2,k}(y_{t-2} - y_{t-3}) \\ \qquad + ... + v_{id,k}(y_{t-d} - y_{t-d-1}) \end{cases} \tag{3}$$

where $\phi_0(\mathbf{X}_t^*), \cdots, \phi_p(\mathbf{X}_t^*)$ are all composed of WGRBFs, $\mathbf{X}_t^*$ indicates input vector of WGRBF and $\mathbf{Z}_k = (z_{k,1}, ..., z_{k,d})$ means the center vector. The term $\phi_0(\mathbf{X}_t^*)$ is an offset that can grasp the time-varying local level. This term plays an important role to capture non-stationary processes with fluctuations of local mean and trend. The $\phi_0(\mathbf{X}_t^*), \cdots, \phi_p(\mathbf{X}_t^*)$ are varying coefficients that enable the model adopt to the nonlinear features [6]. In Eqn. 3, we can observe that the WGRBF-AR model actually contains the weighted GRBF network as one of its components. From the perspective of neural network, the WGRBF-AR model can be seen as a certain generalization of the WGRBF network [6]. It draws upon both advantages of the FAR model in nonlinear dynamics and the WGRBF network in dealing non-stationarity. Note that, although we only describe the first-order WGRBF-AR model as in Eqn. 3, a higher order form can be easily extended without involving any new ideas.

2.3 Identification of WGRBF-AR Model

Intuitively, we can categorize the parameters in WGRBF-AR model into a linear part and a nonlinear part, thus the model can be reformed as a linear combination of nonlinear functions. Parameter identification of our model can be viewed as a separable nonlinear least squares problem [10]. In this paper, we adopt the Variable Projection (VP) algorithm to solve the optimization problem [11]. In details, it projects the linear parameters out of the problem, leaving the nonlinear least squares problems involving only the non-linear parameters. From Eqn. 3, we observe that there are more linear parameters than nonlinear parameters. As a result, applying VP optimization method is substantially efficient since only the nonlinear parameters consume more searching resource [11]. Specifically, we represent the set of linear parameters in Eqn. 3 as:

$$\theta_L \triangleq \{w_0, w_{i,0}, v_{ij,k} \mid \\ i = 0, 1, ..., p; j = 1, 2, ..., d; k = 1, 2, ..., m\} \in \mathbb{R}^{(p+1)(m \times d+2)}. \tag{4}$$

Similarly, the set of nonlinear parameters in Eqn. 3 is formulated as

$$\theta_N \triangleq (\lambda, \mathbf{Z}_1^T, \cdots, \mathbf{Z}_m^T)^T \in \mathbb{R}^{1+m \times d}. \tag{5}$$

Thus, the model can be rewritten in the following regression form of linear parameters:

$$y_t = \varphi(\theta_N; \mathbf{X}_t^*)^T \theta_L + \varepsilon_t \tag{6}$$

As a result, the optimal parameters θ_L, θ_N can be determined by minimizing the following nonlinear function:

$$\min_{\theta_L, \theta_N} \|\mathbf{y} - \Psi(\theta_N)\theta_L\|^2 \tag{7}$$

where the component of $\mathbf{y}$ is $y(i)$; the rows of the matrix of $\Psi(\theta_N)$ correspond to the vector $\varphi(\theta_N; \mathbf{X}_t^*)^T$.

As such, we observe that Eqn. 7 has the same solution:

$$* \min_{\theta_N} \|\mathbf{y} - \Psi(\theta_N)\Psi(\theta_N)^+\mathbf{y}\|^2 = \|(\mathbf{I} - \Psi(\theta_N)\Psi(\theta_N)^+)\mathbf{y}\|^2 \tag{8}$$

where $\Psi(\theta_N)^+$ indicates the *Moore-Penrose inverse* of $\Psi(\theta_N)$ [8]. We first optimize objective function with respect to the nonlinear parameters θ_N, shown as Eqn. 8. Then we utilize very efficient linear least-squares method to optimize linear parameters $\hat{\theta_L}$ [1]. We calculate the *Jacobian matrix* of residuals analytically or using finite differences [8]. The nonlinear least squares algorithms can be applied to solve the nonlinear parameters θ_N, *e.g.,* Gauss-Newton or Levenberg-Marquardt method [9]. We choose the orders of the WGRBF-AR model by the cross-validation.

3. EXPERIMENTS

Seasonal Time-series Data: Retail sale data is a typical time series data that exhibits strong nonlinear, non-stationary and seasonal pattens. Thus many researchers utilize the monthly retail sales data to evaluate the performance of models [20, 17]. In this paper, five different ten-year retail-sales data are used to test our performance[1]. The five retail-sale datasets are retail sales amounts in department stores, book stores, clothing stores, furniture stores and hardware stores during 1992 to 2001. These series are collected in month and without seasonal adjustment, thus they all exhibit stronger seasonality (shown in Fig. 1). Similar to other works [20, 17], we adopt the *Root Mean Square Error* (RMSE) to evaluate the modeling performance.

Comparison Methods: We compare our method with the following five baseline works: *i) ANN-DSDT* [20] first utilizes trend and seasonal adjustment (*i.e.,* de-treding and de-seasonalization) to the targeted time series then adopt a 3-layer feed-forward NNs to perform the prediction; *ii) TDNN-* [17] means time-delay neural network which is well-known NN for temporal data modeling; *iii) ARIMA* [1] is the most famous time series model that can handle non-stationary time-series data; *iv) SVR* [1] means support vector regression which is an extension from *support vector machine* (SVM); *v) RBF-AR* [7] is a typical quasi-linear autoregressive model which its autoregressive coefficients are constructed by radial basis function networks. It is very successful in stationary time series modeling.

Parameter Settings: In our model, we consider 12 variables $\{y_{(t-1)} - y_{(t-2)}, y_{(t-2)} - y_{(t-3)}, ..., y_{(t-12)} - y_{(t-13)}\}$ to be the candidate input that affects y_t. We set the dimension of state vector $\mathbf{X}_t^*$ as 1 for simplicity. And the number of hidden nodes m is chosen as 0 or 1. We select the state vector $\mathbf{X}_t^*$ from input candidates. For the baseline methods, we first apply differencing to obtain a stationary time series, and we set the parameters as described in the papers [15, 17, 20]. All the experiments are conducted in MATLAB-2015b by using a 2.5GHz i7, RAM 16GB laptop with WIN10 OS.

Similar to the baseline works [20, 17, 7], for each of the series, we divide the data set into three parts: the data of the year 2000 and 2001 are used for validation and testing receptively, and the rest are adopt to train the model. We choose a model that produces the smallest RMSE in both the training and validation. We run the experiments 30 times to get the average RMSE.

Table 1 presents the prediction performance of compared

[1]https://www.usa.gov/statistics

Table 1: The RMSE of sales prediction for five commodities in 2001 using different methods

Methods	Department	Hardware	Closing	Furniture	Book
ANN-DSDT	975.55	49.17	315.43	99.45	88.74
TDNN	628.78	35.74	372.52	173.11	91.51
ARIMA	1005.41	100.71	519.6	124.44	98.17
SVR	602.35	39.47	802.79	266.47	97.98
RBF-AR	411.84	23.53	337.72	191.54	96.81
Our method	321.21	15.53	283.31	171.54	83.83

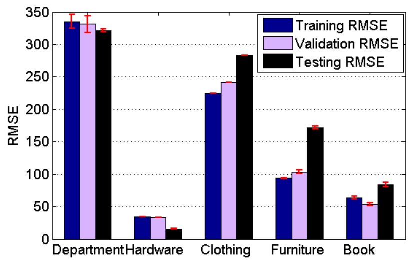

Figure 3: Mean and standard deviation of RMSE for training, validation and testing using our method

methods on five different one-year retail sales time series. The results show that our method achieves better result on four out of five datasets. Especially, for predicting the sales in the hardware and book stores, the proposed weighted gradient RBF network based AR model gets an average 15.53 and 83.83 RMSE. It is worth to mention that ANN-DSDT also reveals very competitive results on the predictions of hardware, furniture and book sales, especially for the furniture dataset, achieving 99.45 RMSE.

Fig. 3 shows the mean values and standard deviations (STD) of RMSE in training, validation and testing in the experiments by using our method. We can observed that the RMSE in training and testing is very close, which exhibits the good scalability of our model. The STD of 30 experiments are also very small, reflecting that the model identification algorithm is capable of performing a robust and stable search. In summary, the proposed WGRBF-AR model achieves competitive results to other state-of-the-art techniques, providing a promising alternative for seasonal and trend time series modeling.

4. CONCLUSION

In this paer, a weighted gradient RBF network based AR model is designed to model and predict the trend and seasonal time series. The proposed model draws upon both recent advances in gradient RBF network and functional autoregressive model that can accurately model the nonlinear, non-stationary and seasonal time series. The experiments on five retail sales datasets show that our model achieves very competitive results comparing to the state-of-the-art.

5. REFERENCES

[1] C. Chatfield. *Time-series forecasting*. CRC Press, 2000.

[2] R. Chen and R. S. Tsay. Functional-coefficient autoregressive models. *Journal of the American Statistical Association*, 88(421):298–308, 1993.

[3] S. Chen, C. F. Cowan, and P. M. Grant. Orthogonal least squares learning algorithm for radial basis function networks. *Neural Networks, IEEE Transactions on*, 2(2):302–309, 1991.

[4] E. Chng, S. Chen, and B. Mulgrew. Gradient radial basis function networks for nonlinear and nonstationary time series prediction. *Neural Networks, IEEE Transactions on*, 7(1):190–194, 1996.

[5] J. Fan and W. Zhang. Statistical methods with varying coefficient models. *Statistics and its Interface*, 1(1):179, 2008.

[6] M. Gan, C. P. Chen, H.-X. Li, and L. Chen. Gradient radial basis function based varying-coefficient autoregressive model for nonlinear and nonstationary time series. *Signal Processing Letters, IEEE*, 22(7):809–812, 2015.

[7] M. Gan, Y. Cheng, K. Liu, and G.-l. Zhang. Seasonal and trend time series forecasting based on a quasi-linear autoregressive model. *Applied Soft Computing*, 24:13–18, 2014.

[8] M. Gan, H.-X. Li, and H. Peng. A variable projection approach for efficient estimation of rbf-arx model. *Cybernetics, IEEE Transactions on*, 45(3):462–471, 2015.

[9] M. Gan, H. Peng, X. Peng, X. Chen, and G. Inoussa. A locally linear rbf network-based state-dependent ar model for nonlinear time series modeling. *Information Sciences*, 180(22):4370–4383, 2010.

[10] G. Golub and V. Pereyra. Separable nonlinear least squares: the variable projection method and its applications. *Inverse problems*, 19(2):R1, 2003.

[11] G. H. Golub and V. Pereyra. The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM Journal on numerical analysis*, 10(2):413–432, 1973.

[12] P. T. Ittig. A seasonal index for business. *Decision Sciences*, 28(2):335, 1997.

[13] A. Pankratz. *Forecasting with univariate Box-Jenkins models: Concepts and cases*, volume 224. John Wiley & Sons, 2009.

[14] M. Priestley. State-dependent models: A general approach to non-linear time series analysis. *Journal of Time Series Analysis*, 1(1):47–71, 1980.

[15] M. Qi and G. P. Zhang. Trend time–series modeling and forecasting with neural networks. *Neural Networks, IEEE Transactions on*, 19(5):808–816, 2008.

[16] Y. Qin, H. Peng, W. Ruan, J. Wu, and J. Gao. A modeling and control approach to magnetic levitation system based on state-dependent arx model. *Journal of Process Control*, 24(1):93–112, 2014.

[17] T. Taskaya-Temizel and M. C. Casey. A comparative study of autoregressive neural network hybrids. *Neural Networks*, 18(5):781–789, 2005.

[18] G. Zhang, B. E. Patuwo, and M. Y. Hu. Forecasting with artificial neural networks:: The state of the art. *International journal of forecasting*, 14(1):35–62, 1998.

[19] G. P. Zhang and D. M. Kline. Quarterly time-series forecasting with neural networks. *Neural Networks, IEEE Transactions on*, 18(6):1800–1814, 2007.

[20] G. P. Zhang and M. Qi. Neural network forecasting for seasonal and trend time series. *European journal of operational research*, 160(2):501–514, 2005.