International Journal of Hospital Research 2013, **2**(3):133-142
www.ijhr.iums.ac.ir
RESEARCH ARTICLE

IJHR

Open Access

A Three-phase Hybrid Times Series Modeling Framework for Improved Hospital Inventory Demand Forecast

Nima Riahi [1], Seyyed-Mahdi Hosseini-Motlagh [2*], Babak Teimourpour [3]

[1] Department of Information Technology, Faculty of Engineering, Payam Noor University, Tehran, Iran [2] Department of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran [3] Department of Industrial Engineering, Faculty of Engineering, Tarbiat Modares University, Tehran, Iran

Abstract

Background and Objectives: Efficient cost management in hospitals' pharmaceutical inventories have the potential to remarkably contribute to optimization of overall hospital expenditures. To this end, reliable forecasting models for accurate prediction of future pharmaceutical demands are instrumental. While the linear methods are frequently used for forecasting purposes chiefly due to their simplicity, they have serious deficiencies in capturing nonlinearities in real-world problems. On the other hand, real world time series data are rarely pure linear or nonlinear, calling for development of forecasting models accounting for both these features of the data. To help meeting this need in the health/healthcare domain, this study undertook development of a hybrid framework consisting of a linear and a nonlinear component to improve forecasting of operating rooms' pharmaceutical demand.

Methods: A hybrid modeling framework combining AutorRgressive Integrated Moving average (ARIMA) as the linear component, and Artificial Neural Network (ANN) as the nonlinear component was developed. The method encompasses three phases: 1) Fitting a linear ARIMA model to the targeted time series, (2) Building an ANN model based on the residuals of the ARIMA model, and (3) Build the hybrid model by combining ARIMA and ANN models for the final forecast. Using the pharmaceutical inventory database of the Iranian Mohem Hospital for fitting AMIRA model and training ANN model, the forecast performance of all three models was compared by calculating the corresponding mean squared error and mean absolute error values, and by superimposing the time series patterns of the operating rooms' drug demand independently predicted by each model to the corresponding observed pattern.

Findings: Both quantitative and intuitive comparisons demonstrated that our hybrid ARIMA-ANN framework outperforms forecasting capability of either ARIMA or ANN models. In particular, the hybrid model showed remarkably superior capability in capturing the nonlinear behavior of the operating rooms' pharmaceutical demand time series.

Conclusions: Our proposed framework sets a ground for developing mathematical and computational forecasting models with ever higher predictive accuracy and supports the promotion of using such forecasting models in practical cost optimization in health facilities.

Keywords: Hospital, Operating Room, Pharmaceutical Demand, Forecasting, Time Series Modeling, Data Mining, Artificial Neural Networks

Background and Objectives

In the contemporary era, the health systems, in particular in developing countries, are struggling with increasing costs which limit their capabilities in rapid development of a health community. This fact, require development of effective cost optimization strategies, at various levels of the health system, in order to save resources for more vital appropriations. Hospitals, as the major subsystem of the health system, consume the majority of overall health budgeted. On the other hand, the sophisticated structural and functional nature of these organization, escalate the likelihood of inefficient utilization of resources in these institutes, thereby posing an unnecessary burden on the health system. Overcoming this challenge will require implication of scientific approaches in hospital management, in particular in the financial aspects.

These costs in a hospital can be divided into different categories: human resources costs, liabilities and capital costs, and inventory costs, the third of which

*Corresponding author: Seyyed-Mahdi Hosseini-Motlagh, Department of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran, P.O.Box: 13114-16841, Tel: +98 21 73225070, Fax: +98 21 73021653, Email: motlagh@iust.ac.ir

being the focus of this study. Because the pharmaceutical inventories are distributed in several parts of a hospital (partly for of supplying the drugs needed in the operating rooms), their efficient management can lead to a potentially remarkable cost saving in the hospital. An integral part of an optimal inventory management is the availability of an accurate demand forecasting system. The more accurate is the demand forecasting, the more efficient the cost saving in the inventory and the less inventory shortage will become.

Given the need for continued research for developing ever more accurate forecasting systems for efficient hospital inventory management, this study attempted to develop a new forecasting framework, by combining two previously established methods, as described in the following.

One of the popular and pervasive techniques for developing forecasting systems is the time series modeling [1]. Generally, there are two main approaches to time series modeling, including the linear approach and the nonlinear approach. Several of linear models, including regression models, exponential smoothing models, and moving average models have developed so far, for use in prediction of time series data. One of the most common linear models of the kind is the autoregressive integrated moving average (ARIMA), which is frequently used in different forecasting domains [1, 2]. While the linear methods are frequently used for forecasting purposes chiefly due to their simplicity, the inherent linear assumption of these models limit their capability to provide accurate forecasts for nonlinear phenomena. On the other hand, while nonlinear models are thought to be more suitable for the majority of real-world problems, the nonlinear territory is usually highly complicated, posing difficulties in developing accurate mathematical/computational model capable of capturing the nonlinearities. To partially address this issue, several efforts have been devoted to developing nonparametric models, which do need empirical determination of parameters required for developing rigorous nonlinear mathematical models. A tractable kind of nonlinear non-parametric models is the Artificial Neural Networks (ANNs), which have recently gained remarkable of attention [3]. The major strength of ANNs is their data-based nature allowing them to by-pass the confining presumptions about the structure of the data. This feature, together with the inherently nonlinear structure of these models, renders ANNs particularly suitable for capturing the underlying complicated nonlinear behavior of the real-world phenomena.

Although ANNs have proven powerful in the forecasting domain, the natural time series data such as operating rooms' pharmaceutical demand data are rarely pure linear or nonlinear. This fact, limits the predictive capability of the solely nonlinear models [8, 12], calling for development of forecasting models accounting for both linearity and nonlinearity in the real-world phenomena. Indeed, previous attempts in this way has demonstrated the superior forecasting performance of the combined linear-nonlinear models as compared with the single models [4-12] and the higher robustness of the hybrid models to the noisy situation [1].

Considering the need for tailoring such efforts to the health/healthcare domain, this study undertook development of a hybrid framework consisting of a linear and a nonlinear component to improve forecasting of operating rooms' pharmaceutical demand.

Methods

Forecasting Framework Development Process

Figure 1 hieratically illustrates the steps for developing a hybrid forecasting model according to our method.

Data Collection

We collected the secondary data from Moheb Hospital, which is one of the major subspeciality centres in Iran. This hospital offers about 560 types of surgical operations, in which over 779 codes of drugs are used. The structure of Moheb Hospital' inventory database accounts for operation's name, operation's code, operation's date, patient admission code, drugs and supplies used in every surgery, drug's code, and number of drugs used.

Preprocessing the data

The collected data for forecasting often include noisy, missing, and inconsistent datasets. Hence, for an accurate data-mining-based forecasting, the data usually have to be cleaned, reduced, transformed, and integrated.

Data Cleaning and Reduction

By reviewing the drug inventory database, it was observed that some of the drugs were not used in any operations throughout a year; hence these drugs were excluded from the study.

In addition, because of the limited research time, it was impossible for us to include all 779 drug codes in our study. However, instead of selecting a random

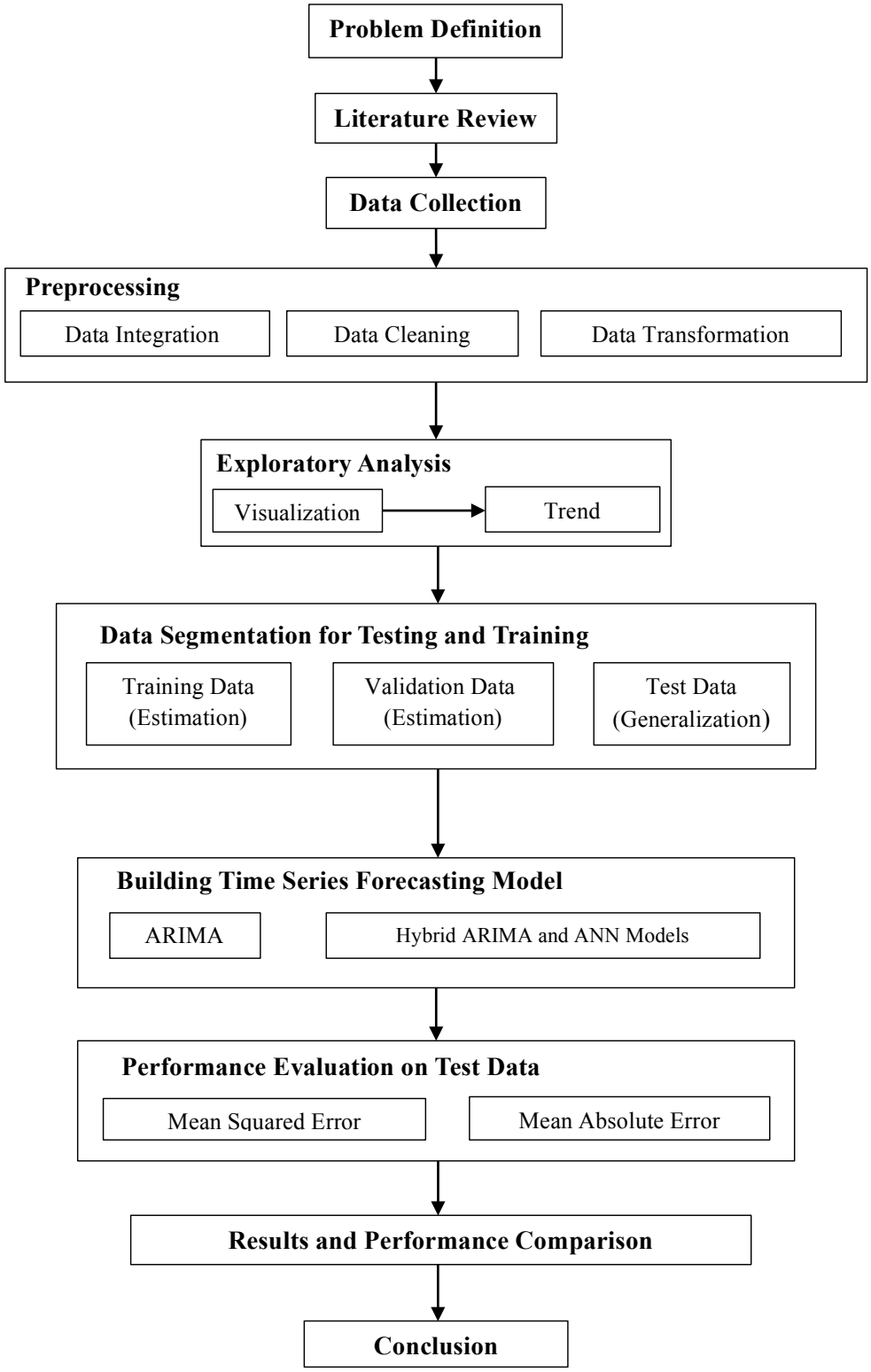

Figure 1 The steps for developing a hybrid model for hospital inventory demand forecasting

sample among the drug codes' list, we rationally selected a list of 118 drug codes, which were consumed permanently in all 53 weeks of a year. This way of selection allows us to focus on the drugs with adequate consumption history data.

On the other hand, during review of the drug inventory database, we found a zero drug consumption rate at the 1st, 2nd, and 52nd weeks, which correspond

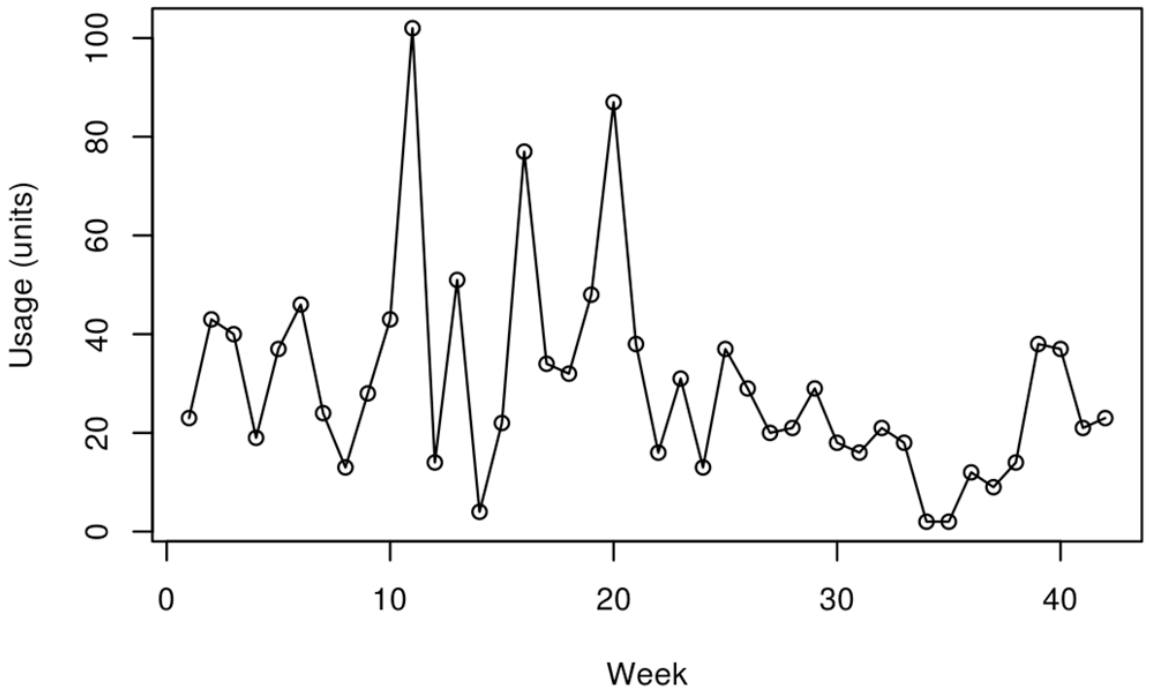

Figure 2 Time series of consumption of the sample drug code during weeks 3 – 44

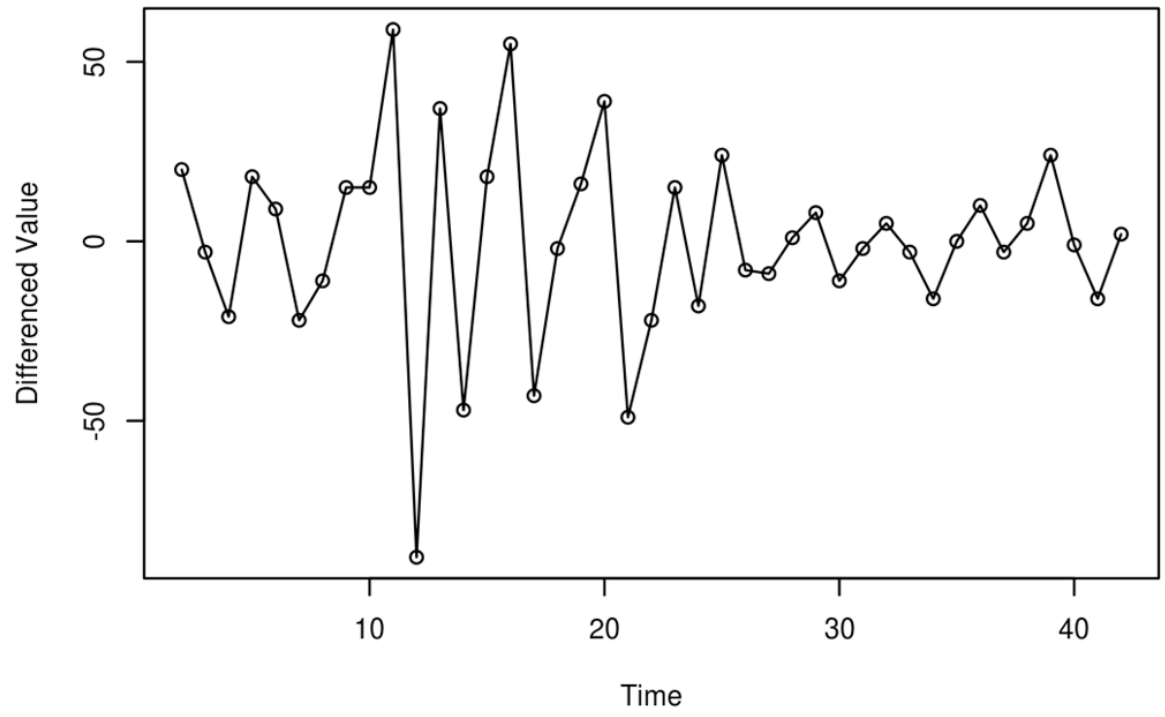

Figure 3 The once-differenced time series of consumption of the sample drug code during weeks 3 – 44

to the New Year holidays. By excluding these noisy data from the dataset, a matrix of Drug-Week with 49 rows, related to 3rd to 51st weeks of the year, and 118 columns related to the selected drug codes was created, with each element related to the consumption of j^{th} drug in i^{th} week. After the data cleaning, a significant reduction in data dimensions was achieved.

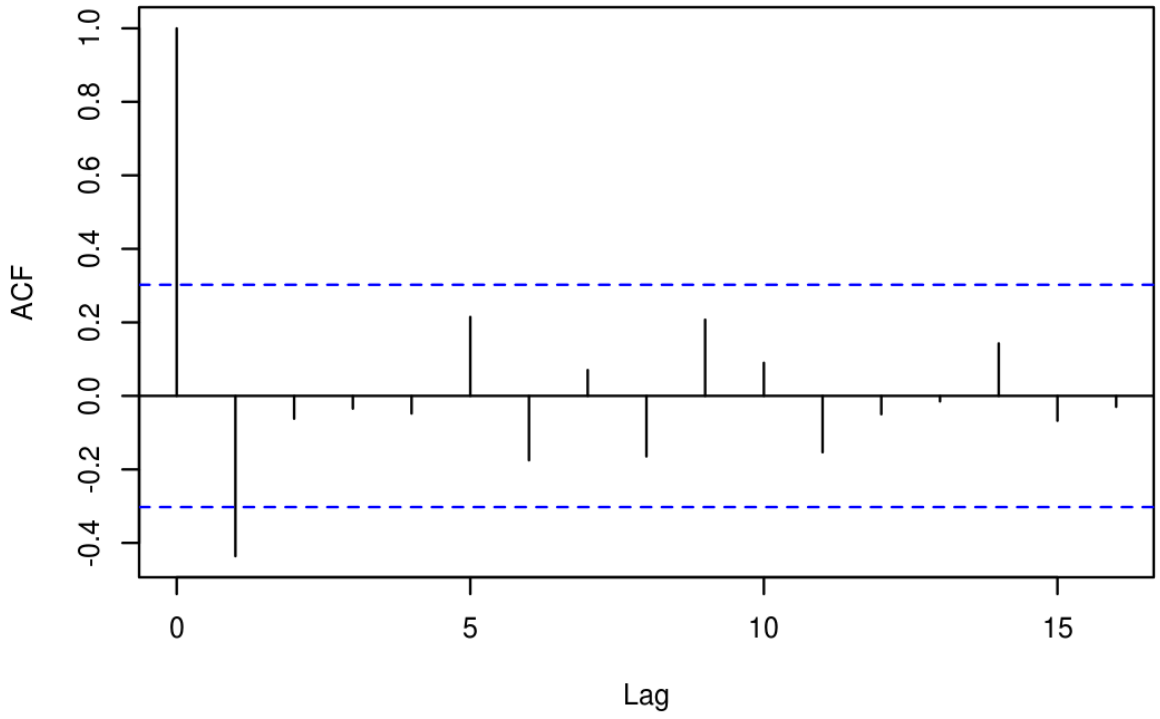

Figure 4 The autocorrelogram of the once-differenced time series of consumption of the sample drug code during weeks 3 – 44

Data Transformation

In ANNs, the sigmoidal or logistic functions are used as the activation function for the hidden and the output nodes. These functions produce outputs in the (0, 1) bracket. Therefore, we have to transform the ANN data to the scale of (0.1, 0.9).

Three-phase Forecasting Method

We assumed that a time series is consisted of two fundamental components: linear and nonlinear components. A general model describing a particular time series, thus, can be formulated as the following:

$$o_t = Linear\ (t) + Nonlinear\ (t) + e_t \quad (1)$$

where, o_t denotes the time series observation at time *t*, *Linear (t)* and *Nonleinear (t)* represent the linear and nonlinear components of the time series, respectively, and e_t denotes the random error term.

Table 1 Comparison of the forecasts by ARIMA and Hybrid models for the consumption of the sample drug code during weeks 45 – 51

Week	ARIMA Forecast	ANN Forecast	Hybrid Model Forecast	Observed Value	Error (ARIMA)	Error (Hybrid Model)
45	18	3	21	11	-7	-10
46	17	3	20	24	6	3
47	17	5	22	24	7	2
48	16	5	21	30	14	9
49	15	5	20	4	-11	-16
50	15	1	16	18	3	2
51	14	4	18	20	6	2

The three-phase model building process comprises the following sequential steps:

Phase 1: Building an ARIMA model based on the raw data as the linear component

Let the estimated value from the ARIMA model at time *t* be *Linear (t)*. We can then calculate the model residual at time using of the following relation:

$$r_t = o_t - \widehat{Linear}\,(t) \quad (2)$$

Since real-world data have nonlinear pattern, these residuals should have nonlinear behaviors that need to be captured by nonlinear models (Phase 2).

Phase 2: Building an ANNs by using the residuals of ARIMA model as the nonlinear component

In our modeling, the nonlinear behavior of the target time series is to be captured by an ANN model. For developing the neural network model, we chose, experimentally, two relatively similar Multi-Layer Perceptron (MLP) structures: 1) a three-layer feed forward perceptron neural network (MLP with one hidden layer) and 2) a four-layer feed forward perceptron neural network (MLP with two hidden layers) with one-step-ahead forecasting method.

Table 2 Comparison of ARIMA and Hybrid models with respect to MSE and MEA of the forecast values

	ARIMA	Hybrid
MSE	70.85	65.42
MAE	7.71	6.29

The model is described by the following equation:

$$\widehat{Nonlinear}\,(t) = f(r_{t-1}, r_{t-2}, \ldots, r_{t-k}) + \delta_t \quad (3)$$

where represents a nonlinear function modelled by ANN, denotes random error, and denotes the width of moving window.

Phase 3: Build the hybrid forecasting model by combining the models from Phase 1 and Phase 2

The ANN model developed in Phase 2 which accounts for nonlinear residuals, will then be combined the linear ARIMA model developed in Phase 1 to yield the final forecasting model. The hybrid forecast model can, hence, be represented as the following:

$$\widehat{o_t} = \widehat{Linear}\,(t) + \widehat{Nonlinear}\,(t) \quad (4)$$

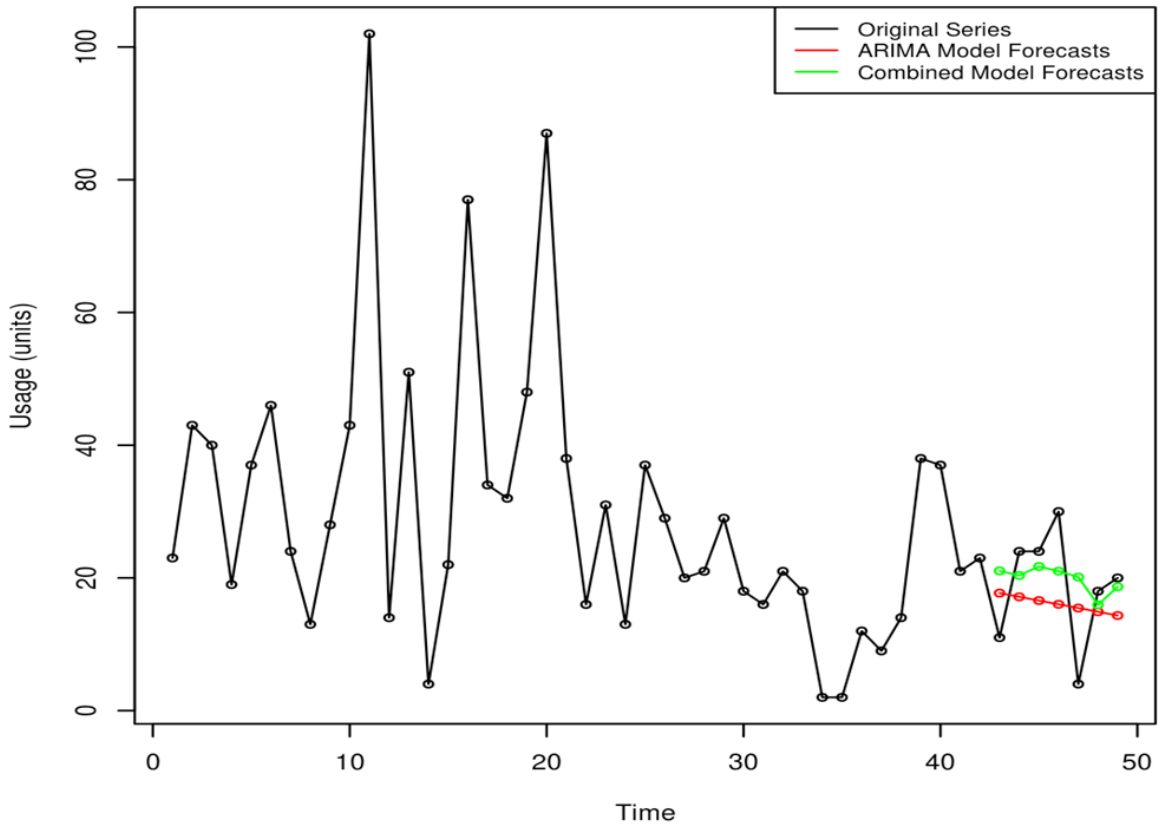

Figure 5 Forecasted vs. observed times series of consumption of the sample drug code during weeks 45 – 51

Building Time Series Forecasting Model

Developing ARIMA moded based on time series data

As stated before, the general ARIMA model [2] features an autoregressive parameter (p) and a moving average parameters (q). The model formulation also explicitly accounts for a differencing parameter (d). An ARIMA model can thus be abstracted as ARIMA (p, d, q), where for instance, a model described as ARIMA (2,1,1) includes two autoregressive parameters (p) and one moving average parameter (q), which will be calculated based on the time series data after the time series was differenced for once (d = 1; see below). Here, we select a sample drug code 2146 to describe the model building method.

Step 1: Transforming the data

As we know, the input series for ARIMA needs to be stationary, that is it should have a constant mean, variance, and autocorrelation through time. Therefore, usually, the series first needs to be differenced until it becomes stationary (it also often requires log transforming to stabilize the variance). The number of times the series needed to be differenced to achieve stationarity is reflected in parameter d. In order to determine the adequate level of differencing, the plot of data and autocorrelogram should be examined [13].

The time series plot of the sample drug code is shown in Figure 2. As seen the use of the drug is not stationary, as indicated by different mean and variance through time, suggesting the need for differencing. After differencing for once (d = 1), the series became approximately stationary (Figure 3).

Step 2: Identification

At this stage, we need to decide how many autoregressive (p) and moving average (q) parameters are needed to yield an effective, but still parsimonious model of the process (a parsimonious model has the fewest parameters and greatest degrees of freedom among all models that fit the data). By examining the autocorrelogram of the differenced time series of the sample drug code (Figure 4), the best fitted model to its consumption data is found to be ARIMA (0, 1, 1).

Step 3: Forecasting

After estimation of the parameters of the ARIMA model, the future demand and associated confidence intervals can be predicted and compare with the test data, which were excluded from the training stage. The estimation process is performed on the transformed (differenced) data. For doing so, before the forecasts are generated, the series is integrated in the inverse of differencing so that the forecast values become compatible with the input data.

Developing an ANNs model by using ARIMA residuals

We apply ANNs to the residuals of ARIMA model (that bear some nonlinearities). We next choose the three-layer feed forward perceptron neural network with one-step-ahead forecasting method. In the experiment, we chose 2-6 input and 1-10 hidden nodes. Doing so, we examined 50 different MLP structures and selected the best one with the minimum Mean Squared Error (MSE). Since we have generated the initial weights of arrows randomly (using random function), we will get different MSEs in each iteration. To solve this problem, we repeat the above-mentioned procedure for 100 times for each drug code and calculate the mean of these 100 MSEs. We then select the best model with the minimum mean MSE.

After identifying the model parameters, we used the *nnet* package of R software for modeling and forecasting [14]. An inverse data transformation was performed to produce values in the original scale. Ultimately, cost function was optimized using BFGS method [15–19]. After completing these steps, an optimized network with three input nodes and four hidden nodes was achieved.

Building a hybrid forecasting model by combining the linear and nonlinear models

A hybrid model was developed by combining the ARIMA and ANNs models, obtained from Phase 1 and Phase 2, and used for final forecasting.

Results and Discussion

Table 2 shows the results of forecasting of the test data using the ARIMA, ANN and the hybrid models. Forecast values of the hybrid model are calculated by the summation of the corresponding values of ARIMA and ANN models. In addition, the error values are calculated as the difference of observed drug consumption values with their point forecasts. These error values were then used for calculation of MSE and Mean Absolute Error (MAE).

Although the best ARIMA and ANN models has been chosen among all, the forecasting results of the single models are not satisfactory enough according to the Table 2.

In addition, while at most weeks (weeks 46-48, 50, and 51), our proposed model has yielded better forecasting results (lower absolute errors) as compared

with the ARIMA model, at weeks 45, the AMIRA model, and at week 49, the ANN model have performed much better. This observation can be explained by the likely linear nature of the data from 45th week and nonlinear nature of the data from 49th week.

While comparison of the results based on single error values could be misleading, using MAE and MSE can enable more robust conclusions. Table 3 represents the MAE and MSE values. The noticeably lower MSE and MAE values associated with our three-phase modeling method as compared with those of the AMIRA method (7.66% and 18.41%, respectively), indicates the higher performance of the hybrid ANN-ARIMA model, in forecasting the pattern of consumption of the sample drug code.

Figure 5 superimposes the observed time series of the sample drug code (black line), and the forecasted time series produced by ARIMA model (red line), and the hybrid model (green line). The figure shows that the hybrid model clearly outperforms the ARIMA model in capturing the inherent nonlinearity in the time series' behavior.

Overall, it can be concluded that combining linear models with nonlinear models for forecasting purposes can alleviate the limitations of both pure linear or nonlinear models, resulting in higher forecasting accuracy. In addition, our proposed framework sets a ground for developing mathematical and computational forecast models with ever higher predictive accuracy and supports promotion of using such forecast models in practical cost optimization in health facilities.

Future studies can further explore the potential of hybrid linear-nonlinear models in accurate forecasting of temporal behaviors, by comparing the prediction results of hybrid and single models, when applied to dynamic time series data.

Conclusions

The main purpose of this study was to compare the performance of a hybrid linear-nonlinear forecasting model with models characterized by either the single features. A novel three-phase method for developing hybrid models was developed, in which AMIRA and ANN were used as the linear and nonlinear model components, respectively. The method was implemented for modeling the behavior of the time series of operating rooms' pharmaceutical demand. Comparison of the forecast results between the three models demonstrated that the hybrid AMIRA-ANN framework outperforms either of the two, particularly in capturing the nonlinear behavior of the pharmaceutical demand time series. Our proposed framework sets a ground for developing mathematical and computational forecast models with ever higher predictive accuracy and supports promotion of using such forecast models in practical cost optimization in health facilities.

Abbreviations

(ARIMA): autoregressive integrated moving average; (ANN): artificial neural network; (MLP): multilayer perceptron; (MSE): mean squared error; (MAE): mean absolute error

Competing Interests

The authors declare no competing interests.

Authors' Contributions

SMHM designed the study, coordinated the study procedure and provision of the requirements, guided development of hybrid modeling framework, and contributed to revision of the manuscript. NR develoed the idea, and made the major contribution to the development of the hybrid modeling framework, collecting, refining, and analyzing the data, building the forecast models, comparing the forecast and observed data, interpreting the results, and drafting and revising the manuscript. BT was involved in developing the forecast models, analyzing the data, interpretation of the results, and drafting the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to appreciate the Administrator and staff in Moheb Hospital, for their kind cooperation in this study.

Received: 8 March 2013 Revised: 19 May 2013 Accepted: 24 May 2013

References

1. Zhang GP. Neural Networks in Business Forecasting. IGI Global; 2003. [http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-59140-176-6].
2. Box GE, Jenkins GM, Reinsel GC. Time series analysis: forecasting and control. Holden-Day; 1976.
3. Zhang G, Eddy Patuwo B, Y Hu M. Forecasting with artificial neural networks: The state of the art. *Int J Forecast* 1998, 14(1):35-62.
4. Bishop CM. Mixture density networks. Technical report. Birmingham Aston university, 1994. [http://eprints.aston.ac.uk/373/1/NCRG_94_004.pdf]
5. Clemen RT. Combining forecasts: A review and annotated bibliography. *Int J Forecast* 1989, 5(4):559-83.
6. Hansen JV, Nelson RD. Time-series analysis with neural networks and ARIMA-neural network hybrids. *J Exp Theor Artif Intell* 2003, 15(3):315-30.
7. Hippert HS, Pedreira CE, Souza RC, editors. Combining neural networks and ARIMA models for hourly temperature

forecast. Neural Networks, IEEE-INNS-ENNS International Joint Conference on; 2000: IEEE Computer Society.
8. Terui N, Van Dijk HK. Combined forecasts from linear and nonlinear time series models. *Int J Forecast* 2002, 18(3):421-38.
9. 9. Tseng FM, Yu HC, Tzeng GH. Combining neural network model with seasonal time series ARIMA model. *Technol Forecast Soc Change* 2002, 69(1):71-87.
10. Weigend AS, Mangeas M, Srivastava AN. Nonlinear gated experts for time series: Discovering regimes and avoiding overfitting. *Int J Neural Syst* 1995, 6(04):373-99.
11. Zhang GP. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* 2003, 50:159-75.
12. Zhang GP, Qi M. Neural network forecasting for seasonal and trend time series. *Eur J Oper Res* 2005, 160(2):501-14.
13. Hill T, Lewicki P. *STATISTICS: methods and applications*. Electronic version: StatSoft, Inc. Electronic Statistics Textbook. Tulsa, OK: StatSoft; 2013 [http://www.statsoft.com/textbook/]
14. Venables WN, Ripley BD. Modern applied statistics with S. 4th edition. Springer; 2002.
15. Broyden CG. A new double-rank minimization algorithm. *AMS Notices*. 1969, 16:670.
16. Broyden CG. The convergence of a class of double-rank minimization algorithms 1. general considerations. *IMA Journal of Applied Mathematics* 1970, 6(1):76-90.
17. Fletcher R. A new approach to variable metric algorithms. *Comput J* 1970, 13(3):317-22.
18. Goldfarb D. A family of variable-metric methods derived by variational means. *Math Comput* 1970, 24(109):23-6.
19. Shanno DF. Conditioning of quasi-Newton methods for function minimization. *Math Comput* 1970, 24(111):647-56.

Please cite this article as:
Nima Riahi, Seyyed-Mahdi Hosseini-Motlagh, Babak Teimourpour. A Three-phase Hybrid Times Series Modeling Framework for Improved Hospital Inventory Demand Forecast. *International Journal of Hospital Research* 2013, **2**(3):133-142.