Integrating Neural Networks with Time Series Forecasting: Improving Sales

Shailendra Singh Kathait, Dr Aankita Kaur, Shubhrita Tiwari, Anubha Varshney

***Abstract*— Sales forecast helps in achieving the sales goals.This further affects the sales revenue, efficiency, customer retention, cost etc. of an organization. It needs effective and accurate decision making in forecasting for revenue and demand planning. Time series analysis is widely used in quantitative methods of forecasting. Integrating self-adaptive, nonlinear, fault tolerance, data driven feed forward neural networks will help in improving the accuracy level of sales.**

***Index Terms*— Sales forecasting, revenue, customer retention, fault tolerance, feed forward neural networks**

I. INTRODUCTION

Neural Networks are one of the popular data mining tools used to predict accurate results in practical ap- plications. Neural Networks are analytical techniques inspired by the biological nervous systems of brain for processing of information and cognitive system for pro- cess of learning. They are capable of learning from ex- isting data using the learning algorithm. They can also predict new observations using previous observations based on their learning ability. The processor of neural network consists of simple processing units which are massively parallel distributed. Neural Networks process information in parallel mode with high speed in a dis- tributed manner and is robust, fault tolerant and recall patterns from partial or noisy pattern and then finally make predictions. So, they are widely used to deal with real world complex problems in the fields of business, industry and science etc.

According to Simon Haykin, the definition of neural network as an adaptive machine and its learning part resemblance with human brain is as stated:

"A neural network is a massively parallel distributed processor made up of simple processing units which has a natural propensity for storing experiential knowledge and making it available for use.".

It resembles brain in the following aspects:

1) Knowledge is acquired by the network from its environment through a learning process.

2) Inter-neuron connection strengths, known as synaptic weights are used to store the acquired knowledge. [1][2][3]

Forecasting is a discipline of predictive analytics. Forecasting in general means predicting the future. To eval- uate predictive research various forecasting techniques are used as in what would happen under different cir- cumstances. Importance is also given to understanding the application of qualitative and quantitative methods of forecasting, understanding of underlying stochastic processes, deriving generalized estimators of seasonal factors etc.

Statistics is quantitative while predictive analysis is both quantitative and qualitative. Forecasting is about pre-dicting the future. Logistic predictive analysis uses both quantitative and qualitative methods to improve the design and competitiveness of a business problem by estimating the past and future behavior of a business process among organizations for the improvement in the application of business problem.

Sales data is huge and unique as it comprises of vol- ume, variety and velocity of data. Volume is in terms of sales and price details, quantity, items sold, time of day, date, customer data etc. Velocity is defined as col- lection of data in terms of yearly, monthly to weekly, daily to hourly data. Variety is in terms of direct sales, sales distribution, online sales, competitor sales etc. Our research problem deals with improvement in man- agement of sales data in terms of forecasting based on some specific requirements. Integration of sales data with detailed customer data is used for efficient and effective operations. Sales data contains information re- garding the sales done for a time span of 5 years and customer data contains all the relevant information re- lated to customers (Name, Contact details etc.).

In this paper, we have used time series forecasting to predict future (sales value/observations) by using some function of past observations. Also the neural network that we are using here has non-linear function so we also implemented constructing the non-linear auto re- gression model.

II. LITERATURE REVIEW

Previously forecasting was generally based upon statsit- ical methods like ARIMA modelling etc. that provided lower accuracy rates in the predictions. The approach for forecasting used by us is based on deep neural net- works, that learns through the data provided at the time of training and then make predictions on the test- data based upon its learning. This highly increased the accuracy of forecasting.

III. METHODOLOGY

Forecasting is a time specified experimentation. Mostly ARIMA modelling is used in forecasting time series data like sales, call volume, inventory etc. Here we will give comparative analysis of using various popular tech- niques that can help in solving forecasting problem in predicting sales and analyzing the impact of promotions on sales activities:

Manuscript received Jan 24, 2017

A. ARIMA Modelling:Time Series Prediction
B. Linear Regression and segment based Forecasting
C. Neural Networks: Machine Learning for Forecasting.

A. Moving Average and ARIMA Modelling

ARIMA modelling technique with non-seasonal effect is classified as ARIMA (p,q,d) where p is the number of autoregressive terms, d is the number of non-seasonal differences and q is the number of lagged forecast errors in the prediction equation.

ARIMA modelling uses lags of the differential series called auto regressive terms, lags of the forecast errors called moving average terms and a time series which needs to be differentiated to be made stationary to con- vert into an integrated version of a stationary series. To capture ARIMA model for a time series first we have to identify the order of differencing needed to stationarize the series and also for removing the seasonality compo- nent, in conjunction with a variance stabilizing trans- formations such as logging and deflating. ARIMA mod- elling technique is capable of capturing trend, cyclic and seasonal patterns in the data but fails to capture the ef- fect of other independent factors which are non-seasonal and non-cyclic in nature.

The sales series showed an increasing trend but had very noise in it. There was no seasonality pattern in sales. Predicting such a series accurately was a vital challenge. The sales series showed an increasing trend but had very noise in it. There was no seasonality pattern in sales. Predicting such a series accurately was a vital challenge. As first step, the series was smoothened us- ing up to 6 moving average and exponential smooth- ing with alpha 0.1 to 0.9. The series was forecasted using both moving average and exponential smoothen- ing method, error rate of the series was too high from both the techniques, i.e.47 percent. As a second step, we used ARIMA modeling method for forecasting. The result from ARIMA models were not satisfactory as the average error rate was very high i,e. 32 percent with a range of 0 to 75 percent.

As a third step, we segmented agents into various seg- ments using CART analysis on the basis of various key performance indicators, demographics and other exter- nal variables. After identification of key segments, we forecasted the sales of each segment but the average er- ror rate was still substantial, i.e. 27 percent with the range of 2 percent to 65 percent.

B. Linear Regression and Segment-based forecasting

For minimizing error rate, we built a linear regression model for forecasting agent's sales using: up to 3 lagged value of sales , error component obtained from ARIMA modelling i.e. difference between model fit and actual values, key performance indicators and monthly promo- tions and offer variables. Quantitative (impact on sales) and qualitative (nature of offer) analysis of monthly promotions were done to create variables as an input to regression model. Using it we were able to drop down the average error rate to 20 percent with error range of 4 percent to 35 percent but this was again on a very higher side.

Linear Regression is often used to solve time series fore- casting problems. Here the equation is formed with the help of lag variables (calculated manually) for the data- set (up to 4 lag variables) and regressing the lag vari- ables on the dependent variable. It can also incorporate independent variables, which may further improve the accuracy of time series.

$$\text{Sales}_t = A+\alpha\text{Lag1}+\beta\text{Lag2}+\gamma\text{Lag3}+\delta\text{Lag4}+ \text{Promo}(0,1)+\text{OtherEffectVariables}(1)$$

C. Machine Learning for forecasting

We used artificial feed-forward neural networks with multiple hidden layers. In our study, the function is nonlinear. The strength of each connection is measured by a quantity called weight. Here weights were adjusted iteratively.Many neural network fitting problems may or may not converge to local minimum since a neu- ral network model that fits well may give poor out of sample forecast. To fit the neural network models to obtain the best forecasts of the test data rather than the best fit to the training data. Also it is not wise to use neural network model in black box, so the wise analyst needs to use traditional modelling statistics to select a good neural network model, i.e. to select ap- propriate lagged variables as input. Therefore we used ARIMA modelling and Linear Regression initially then proceeded to neural networks.

Deep learning neural networks were used for training the models. Using neural networks, the forecasting was done. The model used consisted of multiple hidden lay- ers that increased the computation and accuracy.

Our objective is to forecast monthly sales retailer. For this we have used sales data for four years from Jan 2009 to Dec 2012 in building our forecasting model. Af- ter this we have created 12 lags (lag1-lag12) so our data gets reduced to by 12 months and our new data is avail- able from Jan 2010 to Dec 2012. This data was used to train the network. For validation we have used data for six months from Jan 2013 to June 2013. Values from Jan 2013 to Dec 2013 were predicted for forecasting. Besides the 12 lags, we also generated exponential of the lags as well as their transforms using different mathe- matical algorithms and used them as features for train- ing the neural network that increased the accuracy of forecasting.

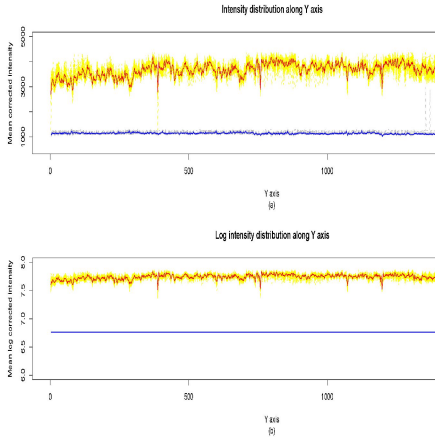

Fig. 1 Intensity and Log Distribution

Here the value was forecasted using the values at lags 1 and 12. In the beginning we have from lag 1 to lag 13, they are the input provided called as the input node. Forecast is the

output called as the output node. Hid- den layer of 100 neurons are the processing units. Also a constant value 1 was introduced as the input term as the residual of input lags was added. Feed-forward neural networks with densely connected multiple hid- den layers provided various combinations of neurons to train neural network. We got the optimal output with
24 neurons. After training neural network using var- ious combinations of neurons in the hidden layer, we achieved a great success in forecasting sales series. The average error rate was brought down to 4 percent with a range of 0 percent to 4 percent.

IV. OBSERVATIONS

From all the methods used, the minimum error was obtained in neural network approach. So this was used in forecasting.

CONCLUSION

The multi-layered neural network approach gave the best results as compared to the statistical methods. Dif- ferent type of neural networks with different configura- tion of hidden layers can be used for obtaining more accurate forecasting.

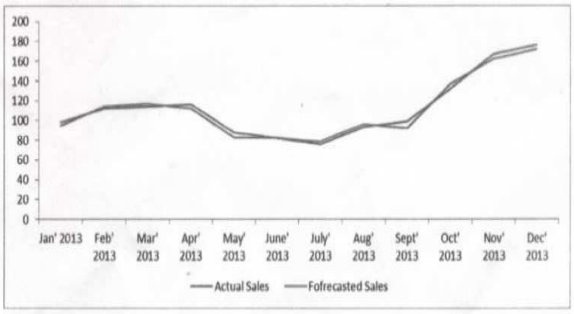

Fig. 2 Actual Sales Vs Forecasted Sales for the year 2013

Model Techniques	Error
Moving Average And Exp Smoothening	47%
ARIMA	32%
Linear Regression	20%
Multi-layered Neural Networks	4%

Fig. 3 Observation Table

REFERENCES

[1] B. Noble, G. Eason, I. N. Sneddon, "On certain inte- grals of Lipschitz-Hankel type involving products of Bessel functions", Phil. Trans. Roy. Soc. London, vol. A247, pp.529551, (1955).
[2] J. Clerk Maxwell,"A Treatise on Electricity and Mag- netism", 3rd ed., vol. 2. Oxford: Clarendon, pp.6873, (1892).
[3] C. P. Bean, I.S.Jacobs,"Fine particles, thin films and ex- change anisotropy", Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, pp. 271350, (1963).