A Forecasting Decision Support System

Hanaa E.Sayed [a, *], Hossam A.Gabbar [b], Soheir A. Fouad[c], Khalil M. Ahmed[c], Shigeji Miyazaki [a]

[a] *Department of Systems Engineering, Division of Industrial Innovation Science, Okayama University, Japan*

[b] *Faculty of Energy Systems and Nuclear Science, UOIT, ON L1H 7K4 Ontario, Canada*

[c] *Computer Science & Systems Engineering Department, Alexandria University, Egypt*

Abstract— **Nowadays forecasting is needed in many fields such as weather forecasting, population estimation, industry demand forecasting, and many others. As complexity and factors increase, it becomes impossible for a human being to do the prediction operation without support of computer system. A Decision support system is needed to model all demand factors and combine with expert opinions to enhance forecasting accuracy. In this research work, we present a decision support system using winters', simple exponential smoothing, and regression statistical analysis with a new proposed genetic algorithm to generate operational forecast. A case study is presented using real industrial demand data from different products types to show the improved demand forecasting accuracy for the proposed system over individual statistical techniques for all time series types.**

Keywords: Forecasting; Regression Analysis; Genetic Algorithm; Support System

I. Introduction

Demand forecasting is very important to support supply chain processes as it is considered the start point for them. Many industries such as FMCG (Fast Moving Consumer Goods), oil, and media are seeking more accurate demand forecasting. Focus on FMCG industry is still not that strong from researchers. Many reasons cause poor forecast accuracy: unspecialized forecasters, unsuitable forecasting techniques used for each case, highly dynamic markets, short-term activity planning, etc. Still, research is active in this area to provide better forecasting accuracy.

Improving forecasting models is considered a vital part in the overall supply chain process. And building a decision support system for forecasting is not that easy task, due to the complexity and the multiple relations between factors affecting the demand. Kuo [4] proposed a decision support system for the stock market via fuzzy Delphi and ANN (Artificial Neural Networks). In another studies, [3] developed an intelligent sales forecasting system which considered quantitative and qualitative factors by integrating ANN and FNN (Fuzzy Neural Network). Combining Forecasts is being proofed to outperform than using individual methods [1]; [6]. They proofed that combining forecasts is decreasing the risk of wrong forecasts than using individual methods. Many techniques have been used to combine methods such as moving average, rule based techniques, and Genetic Algorithms. Genetic Algorithms is known by their superior capabilities to search in a big size of population space [2].

In this research work, we propose a decision support system which utilizes statistical techniques and genetic algorithm with input from expert people in order to guide and tune the outcome final forecast. A modified genetic algorithm is proposed to enhance the searching capabilities by Modified Gaussian Mutation Operator. We have chosen simple statistical methods in order to show that we don't need here complex ones, just the use of the proposed genetic algorithm with the proper understanding and modeling for demand factors through the decision support system will give better forecast accuracy. Furthermore, they are chosen to cover all time series components: level, trend, and seasonality. Three traditional statistical methods are used: (1) Simple Exponential Smoothing, (2) Additive Winters', and (3) Multiplicative Winters' methods [7]. A real case study is presented by applying this model in FMCG products. Results of this model exhibited overall improvement in the forecast accuracy over individual methods for all types of time series. Sec. (2) is presenting an overall for the proposed model framework with model flowchart. The detailed model is explained in Sec. (3). In Sec. (4), the case study is presented with the results. And finally conclusion is in Sec. (5).

II. Proposed Model Framework

We call it Baseline Activities & Forecasting Integration (BAFI) Model. BAFI inputs include controllable and uncontrollable factors.

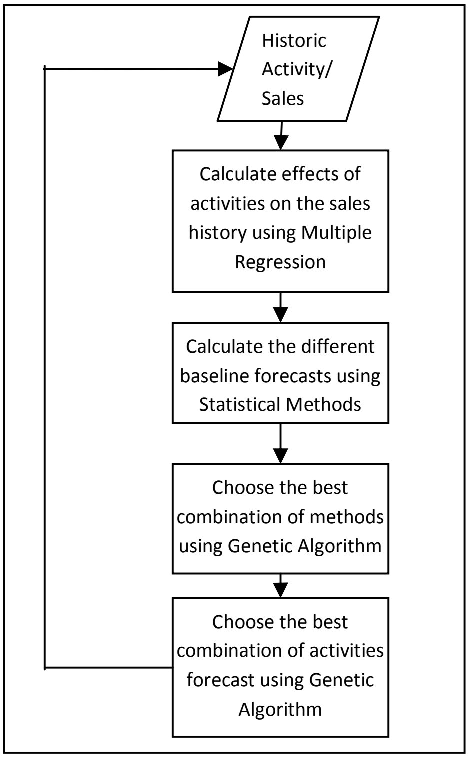

Figure 1: BAFI Flow Chart

Controllable factors are those such as historic sales and historic demand factors such as promotions and prices. Uncontrollable factors are those such as temperature and number of working days. Number and types of factors are flexible, based on the industry type and the situation we can change them. Multiple regression is used to calculate the factors effect of the sales series. Subsequently, using different statistical methods, BAFI calculates different forecasts for the baseline series. Then it chooses the best combination of methods using the modified genetic algorithm. Weights between methods can be changed to obtain the best combination, which gives the least MSE. Uncontrollable factors are forecasted using statistical methods; then the best combinations of the controllable factors in the future are calculated using a genetic algorithm to maximize sales within the cost limit. A flow chart of BAFI is shown in Figure 1. The final outcome of the model is operational forecast which is total forecast number including baseline forecast and activities forecast.

III. Forecasting Decision Support System

The framework of the proposed demand forecasting decision support system comprises three sub-models, as portrayed in Figure 2.

The first sub-model is called *Cause & Effect Model*. Its inputs are history demand series and demand factors. The model outputs are cleaned history demand and RFM (Regular Factors Matrix) of possible factors on the demand series. The first model is responsible for the following consecutive steps: (1) evaluating real demand by eliminating the effects of unavailability; (2) calculating the effects of demand factors on the cleaned sales using multiple regression method; and (3) establishing the knowledge base, which is updated after each run of the model.

The second sub-model is called *Intelligent Time Series Model*. Its inputs are the cleaned history demand series, RFM. It is responsible for calculating time series components (trend, seasonality, and cycles) of the real demand and calculating a raw or baseline forecast, which represents demand without the effects of demand factors. The baseline is calculated by combining selected statistical methods: SES (Simple Exponential Smoothing), winters' additive and winters' multiplicative methods. These methods are chosen so that all the times series components (Level, Trend, and seasonality) are being forecasted, then combined using the following module. Also they are chosen so simple to show that the model is improving the total forecast accuracy although the statistical methods used are most simple ones. The best combination is obtained by calculating the best set of weights for each method using a genetic algorithm that minimizes MSE (Mean Square Error).

Finally, the last sub-model is *Intelligent Combining Model*. Its inputs are the forecast, RFM, generated baseline, which the same input is used to the previous module. Its output is the final forecast including factors forecast. The model compares factors timing using a genetic algorithm, which minimizes the cost and increases the profit of the forecast. The final outcome of the model is the final demand forecast (Operational Forecast) and activities which maximizes profit. The operational forecast is the summation of the baseline forecast and activities forecast.

The model can be further tuned using opinions of experienced people, which can add or remove any of the activities followed by recalculation of the forecast model based on the new parameters suggested by experts. Proposed model is considered a decision support system as it is the main input for the demand planning meeting. Top management is using the forecast results as starting point and then add the market intelligence and other expert opinions which are not included in the model, or shown suddenly. And by that way we eliminate the error occurs due to human bias.

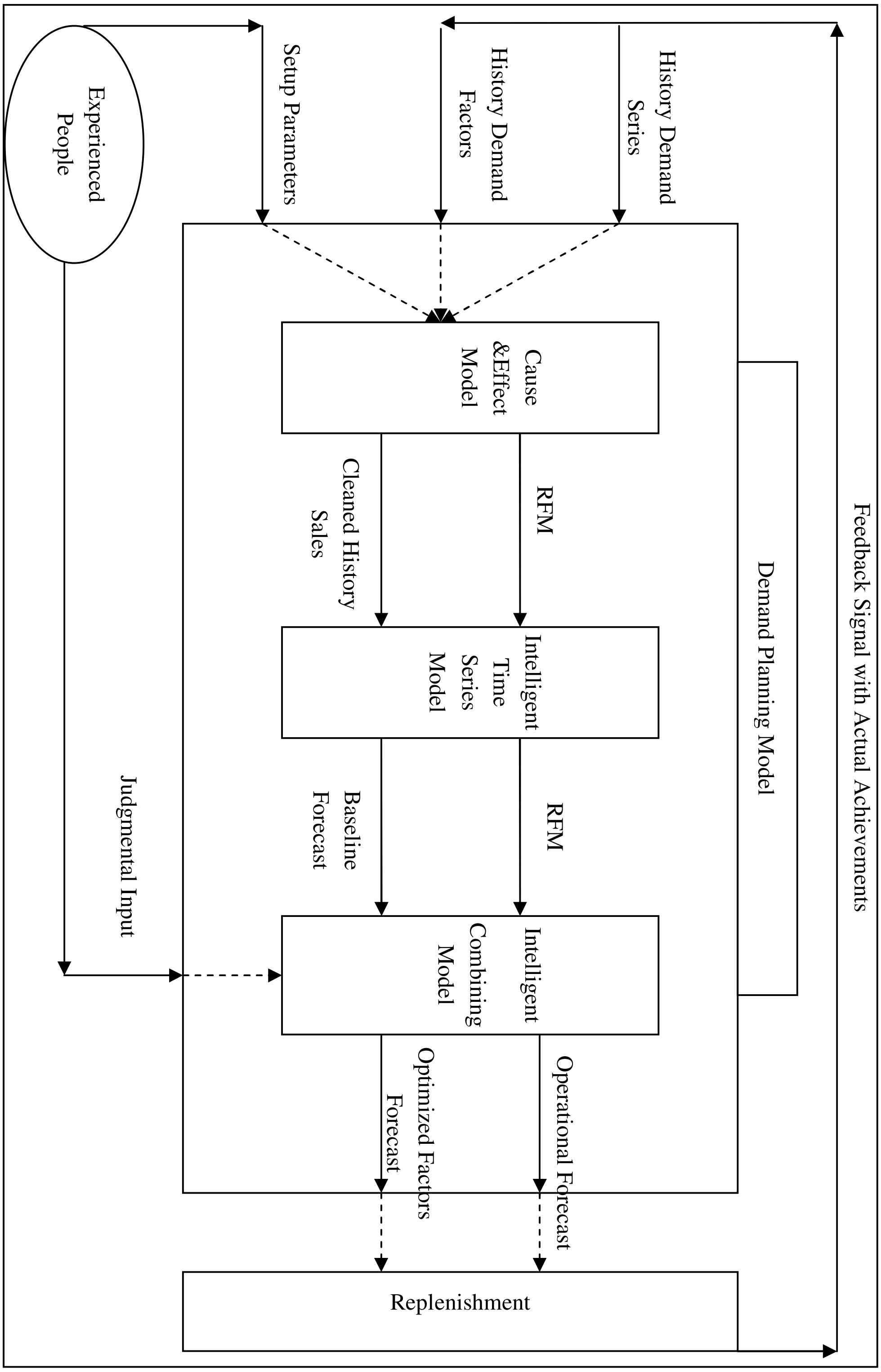

Figure2: BAFI Frame Work

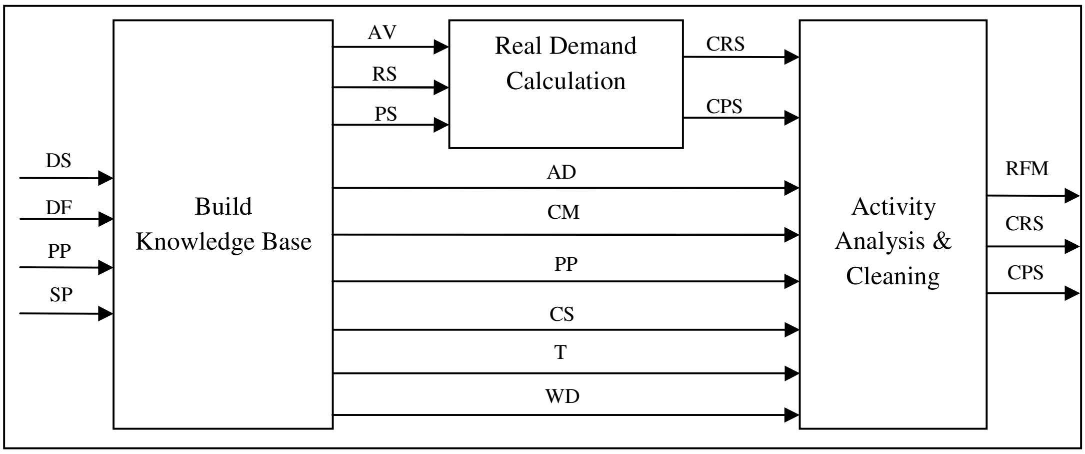

Figure3: Cause and Effect Model

A. Cause and Effect Model

Cause and effect model is the first sub-model in the solution framework. Inputs to it are history demand series, which is inferred from sales history, demand factors, product prices, and setup parameters. The outcomes of that model are RFM and cleaned history demand series (Raw History Demand).

The first sub-module, *Build Knowledge Base*, uses the domain knowledge rule. Inputs, DS (Demand Series), DF (Demand Factors), PP (Product Prices), and SP (Setup Parameters), are categorized and inserted into predefined factors parameters, which are used throughout the model, AV (Availability), RS (Regular Series), PS (Promoted Series), AD (Advertising), CM (Commission), CS (Cost Series), T (Temperature), and WD (Working Days). This division is useful because it facilitates utilization of rules and constraints. It is easy to add additional factors in the future, where the system can adapt itself automatically without changing model structure.

The second sub-module, *real demand calculation*, is used to eliminate the effects of availability factor from regular series and promotion series to generate real history regular demand and real history promotion series.

The final sub-module is *activity analysis and cleaning*. Its inputs are real history demand series and real history factors series. The outcome of cause and effect model is CRS (Cleaned Regular Series), CPS (Cleaned Promotion Series), RFM, and PFM (Promotion Factors Matrix). Multiple Regression method is used to calculate demand activities that affect percentages on demand series. The term cleaned means that it represents the real demand with no effect of demand factors. Multiple regression method is used to calculate the effects of demand factors on demand series. In Figure 3, the framework of cause and effect model is presented.

B. Intelligent Time Series Model

After cleaned regular sales are calculated, using previous module, history sales will be analyzed using statistical and computational methods. At *Intelligent Times Series Model*, methods of three types are used, as shown in Figure 4: SES), additive winters' method and multiplicative winters' method [7]. Then the genetic algorithm is used to calculate best weights among the three methods to give the least MSE and generate baseline forecast using optimum weights.

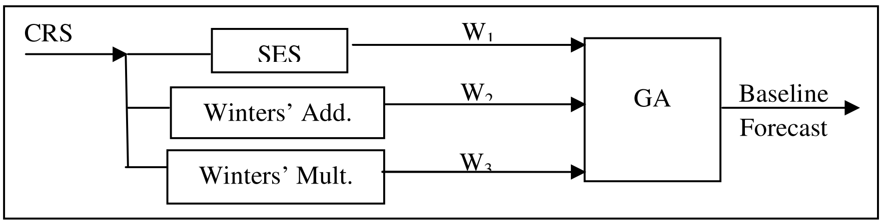

Figure 4: Intelligent Times Series Model

GA Mutation

Genetic Algorithms are famous of their global search capability [5]. A new evolutionary algorithm of mutation is presented in this paper. We called it "Modified genetic algorithm". It uses mutation operator, Modified Gaussian mutation. And it is defined as follows:

$$\sigma_j = k.Min\{x_j^{t-1} - a_j, b_j - x_j^{t-1}\}\left(1 - \frac{t}{M_g}\right)^s \left(1 + \frac{t}{M_g}\right),$$

j=1,2,…,M (1)

where k is a constant within the closed interval [0, 1]; t is the generation; x_j^{t-1} is the jth variable to be optimized in the (t-1)th generation; $[a_j, b_j]$ is the jth variable's scope; M_g is the maximum generation; s is a shape parameter; and M is the number of variables in our case they are three.

The mutation of the jth variable x_j is given by:

$$x_j' = x_j + \varepsilon_j, \quad j= 1, 2, \dots, M \tag{2}$$

$$\varepsilon_j \sim N(0, \sigma_j), \tag{3}$$

where ε_j is distributed as a Gaussian random number variable with zero mean and σ_j standard deviation.

C. Intelligent Combining Model

The objective of this module is to calculate the optimum choices of the activities in the future periods which maximize the profit. Therefore, using the introduced genetic algorithm module, we can choose the best time we can do the activity and the best timing.

The first step is to forecast the uncontrollable factors for the required periods. Both T and WD are forecast using the model intelligent time series model at Figure 5. Then to calculate the best combination of the controllable factors (AD, CP, and CM), timings using the genetic algorithm and the values are calculated using the RFM from the values of forecasted T and WD corresponding to the time point.

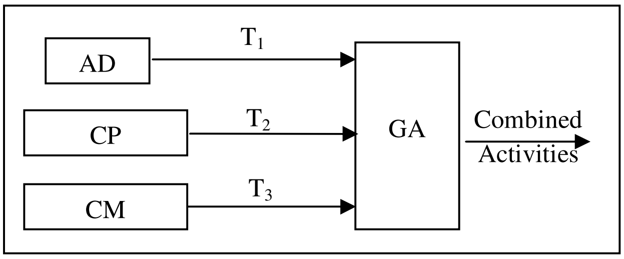

Figure 5: Intelligent Combining Model

IV. Case Study & Results

Data used in the selected case study are from real FMCG industry with different products and different time series types. Data used are history actual demand, history activities, along with products price. They are collected as part of improving forecasting accuracy project in an international industrial company. We collected and cleaned the data for all its products, 1500 products. Products are coming from foods division and humans. The Human care division includes creams, oils, shampoos, soap, tea, and tomato paste. After collecting the times series of the sales history, data cleaning is performed to ensure correctness of sales and data continuity in the history. The cleaned time series data are fed to the model where a knowledgebase is constructed. After running each time series, the last optimum weights obtained are stored in the knowledgebase for use in subsequent runs. History is divided into a training set and test set. Consequently, model results can be compared with the test set, which is useful to calculate the MSE and to choose optimum results.

Table 1: MSE Comparison of methods

MSE	SES	Additive Winters'	Multiplicative Winters'	BAFI
Static Trend & Static Seasonality	30%	40%	70%	18%
Static Trend & Linear Seasonality	60%	40%	75%	20%
Static Trend & Exp. Seasonality	55%	40%	68%	16%
Linear Trend & Static Seasonality	50%	40%	85%	19%
Linear Trend & Linear Seasonality	65%	40%	77%	17%
Linear Trend & Exp. Seasonality	70%	55%	60%	16%
Exp. Trend & Static Seasonality	72%	50%	55%	20%
Exp. Trend & Linear Seasonality	77%	50%	52%	19%
Exp. Trend & Exp. Seasonality	80%	60%	45%	18%

Minitab Tool is used for the same time series to prove that the model is calculating better MSE results. Similarly, the MSE is calculated using the individual statistical methods and results are compared with the proposed model MSE. Table (1) is showing the improvement MSE for the proposed model over individual methods for all types of times series.

Conclusion

Forecasting using single statistical method showed limitations in providing accurate forecasting. The use of a decision support system is needed for providing accurate forecasting specially for complex environments which have different forecasting factors. As described in this paper, the use of a modified genetic algorithm is proposed to combine statistical methods and for combining activities with the baseline forecast, which suggests forecasts that are more accurate. First, genetic algorithm searches for the best combination of weights between the methods, which minimizes MSE error. Then genetic algorithm searches between the activities timing to choose the best timing.

Selected case study is based on FMCG products, which have different types of times series. In future studies, this model can be tested for other types of industrial disciplines. Also a future work plan is to add neural networks techniques. Comparison of the obtained results shows better accuracy than that obtained using traditional methods. Other combinations of forecast methods can be included in the proposed solution for better forecast accuracy. This model is proposing a new way to model the factors which are affecting the demand with combination of expert people opinions.

REFERENCES

[1] Timmermann. *Forecast Combinations*, UCSD, 2005.

[2] D.E. Goldberg. Genetic Algorithms in Search Optimization, and Machine Learning, Addison- Wesley, 1989.

[3] Kuo, R. J., Wu, P., & Wang, C. P. An intelligent sales forecasting system through integration of artificial neural networks and fuzzy neural networks with fuzzy weight elimination. Neural Networks, 15, 909–925, 2002.

[4] Kuo, R. J. A sales forecasting system based on fuzzy neural network with initial weights generated by genetic algorithm. European Journal of Operational Research, 129, 496–517, 2001.

[5] M. Gen, R. Cheng. *Generic Algorithms and Engineering Design,* Wiley, 1997.

[6] M. Hibon, T. Evgeniou. To combine or not to combine: selecting among forecasts and their combinations, *International Journal of Forecasting*; 21; 15–24, 2005.

[7] S. Makridakis, R. J. Hyndman, S. C. Wheelwright. *Forecasting Methods and Applications*, Wiley, 1998.