International Journal of Research in Computer Engineering and Electronics. 1 ISSN 2319-376X
VOl : 2 ISSUE :1 (Feb 2013)

Data Mining through Time Series Forecasting Algorithm

G.N.K.Suresh Babu[1] , Dr.S.K.Srivatsa[2]

Abstract -In this paper we describe time series forecasting which is used for determining future predictions. Any variable that is measured over time in sequential order is called a time series. We analyze time series to detect patterns. The patterns help in forecasting future values of the time series. Many quantities in nature fluctuate in time. Examples are the stock market, the weather, seismic waves, sunspots, heartbeats, and plant and animal populations. Until recently it was assumed that such fluctuations are a consequence of random and unpredictable events. With the discovery of chaos, it has come to be understood that some of these cases may be a result of deterministic chaos and hence predictable in the short term and amenable to simple modeling. Many tests have been developed to determine whether a time series is random or chaotic, and if the latter, to quantify the chaos. If chaos is found, it may be possible to improve the short-term predictability and enhance understanding of the governing process.

Forecasting is fundamental to decision-making. There are three main methods:

- Subjective forecasting is based on experience, intuition, guesswork and a good supply of envelope-backs.
- Extrapolation is forecasting with a rule where past trends are simply projected into the future.
- Causal modeling (cause and effect) uses established relationships to predict, for example, sales on the basis of advertising or prices.

Index - Terms : Time Series, Prediction, Forecasting, Modeling

1 Introduction

A time series is a collection of observations of well-defined data items obtained through repeated measurements over time. For example, measuring the value of retail sales each month of the year would comprise a time series. This is because sales revenue is well defined, and consistently measured at equally spaced intervals. Data collected irregularly or only once are not time series.

An observed time series can be decomposed into three components: the trend (long term direction), the seasonal and the irregular (unsystematic, short term fluctuations). Time series can be classified into two different types: stock and flow. A stock series is a measure of certain attributes at a point in time and can be thought of as "stocktakes". For example, the Monthly Labour Force Survey is a stock measure because it takes stock of whether a person was employed in the reference week. Flow series are series which are a measure of activity over a given period. For example, surveys of Retail Trade activity. Manufacturing is also a flow measure because a certain amount is produced each day, and then these amounts are summed to give a total value for production for a given reporting period. The main difference between a stock and a flow series is that flow series can contain effects related to the calendar (trading day effects). Both types of series can still be seasonally adjusted using the same seasonal adjustment process. A seasonal effect is a systematic and calendar related effect. Some examples include the sharp escalation in most Retail series which occurs around December in response to the Christmas period, or an increase in water consumption in summer due to warmer weather. Other seasonal effects

- *[1] G.N.K.Suresh Babu is currently working as an Associate Professor – GKM College of Engineering and Technology, Chennai, India.*
- *[2] Dr.S.K.Srivatsa is currently working as Sr.Professor – St.Joseph's College of Engineering and Technology, Chennai, india*

include trading day effects (the number of working or trading days in a given month differs from year to year which will impact upon the level of activity in that month) and moving holidays (the timing of holidays such as Easter varies, so the effects of the holiday will be experienced in different periods each year).

2 Time Critical Decision Modeling and Analysis

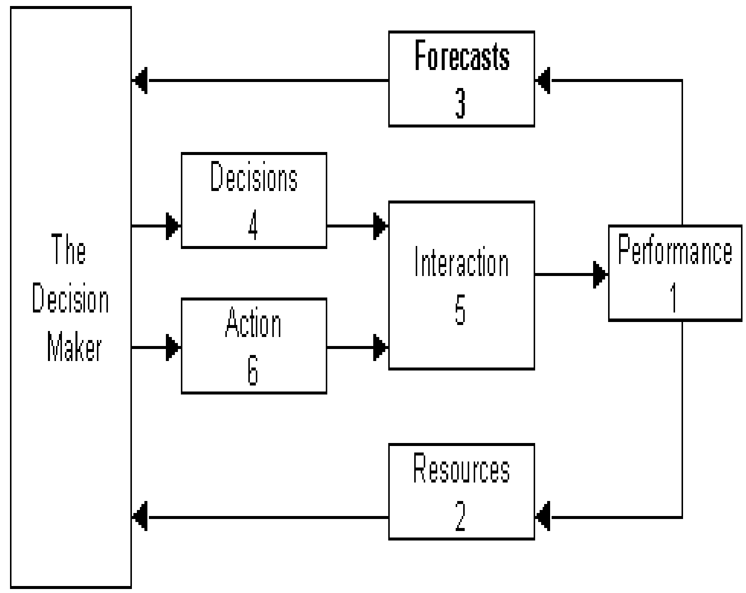
The ability to model and perform decision modeling and analysis is an essential feature of many real-world applications ranging from emergency medical treatment in intensive care units to military command and control systems. Existing formalisms and methods of inference have not been effective in real-time applications where tradeoffs between decision quality and computational tractability are essential. In practice, an effective approach to time-critical dynamic decision modeling should provide explicit support for the modeling of temporal processes and for dealing with time-critical situations.

One of the most essential elements of being a high-performing manager is the ability to lead effectively one's own life, then to model those leadership skills for employees in the organization. This site comprehensively covers theory and practice of most topics in forecasting and economics. I believe such a comprehensive approach is necessary to fully understand the subject. A central objective of the site is to unify the various forms of business topics to link them closely to each other and to the supporting fields of statistics and economics. Nevertheless, the topics and coverage do reflect choices about what is important to understand for business decision making.

Almost all managerial decisions are based on forecasts. Every decision becomes operational at some point in the future, so it should be based on forecasts of future conditions.

Forecasts are needed throughout an organization -- and they should certainly not be produced by an isolated group of forecasters. Neither is forecasting ever "finished". Forecasts are needed continually, and as time moves on, the impact of the forecasts on actual performance is measured; original forecasts are updated; and decisions are modified, and so on.

For example, many inventory systems cater for uncertain demand. The inventory parameters in these systems require estimates of the demand and forecast error distributions. The two stages of these systems, forecasting and inventory control, are often examined independently. Most studies tend to look at demand forecasting as if this were an end in itself, or at stock control models as if there were no preceding stages of computation. Nevertheless, it is important to understand the interaction between demand forecasting and inventory control since this influences the performance of the inventory system. This integrated process is shown in the following figure:



Forecasting Within an Organization:
Forecasting and Managerial Decision Making

Progressive Approach to Modeling: Modeling for decision making involves two distinct parties, one is the decision-maker and the other is the model-builder known as the analyst. The analyst is to assist the decision-maker in his/her decision-making process. Therefore, the analyst must be equipped with more than a set of analytical methods.

Integrating External Risks and Uncertainties: The mechanisms of thought are often distributed over brain, body and world. At the heart of this view is the fact that where the causal contribution of certain internal elements and the causal contribution of certain external elements are equal in governing behavior, there is no good reason to count the internal elements as proper parts of a cognitive system while denying that status to the external elements.

The time series analysis has three goals: forecasting (also called predicting), modeling, and characterization. What would be the logical order in which to tackle these three goals such that one task leads to and /or and justifies the other tasks? Clearly, it depends on what the prime objective is.

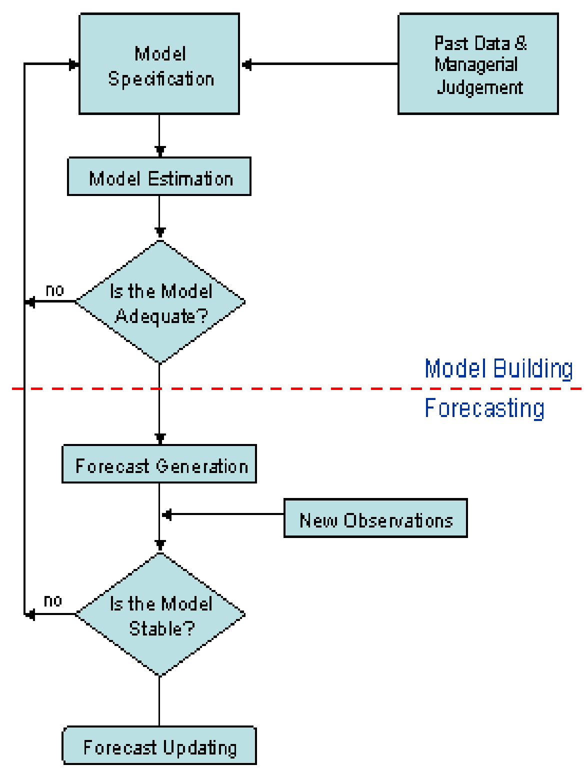
Sometimes you wish to model in order to get better forecasts. Then the order is obvious. Sometimes, you just want to understand and explain what is going on. Then modeling is again the key, though out-of-sample forecasting may be used to test any model. Often modeling and forecasting proceed in an iterative way and there is no 'logical order' in the broadest sense. You may model to get forecasts, which enable better control, but iteration is again likely to be present and there are sometimes special approaches to control problems Outliers: One cannot nor should not study time series data without being sensitive to outliers. Outliers can be one-time outliers or seasonal pulses or a sequential set of outliers with nearly the same magnitude and direction (level shift) or local time trends. A pulse is a difference of a step while a step is a difference of a time trend. In order to assess or declare "an unusual value" one must develop "the expected or usual value". Time series techniques extended for outlier detection, i.e. intervention variables like pulses, seasonal pulses, level shifts and local time trends can be useful in "data cleansing" or pre-filtering of observations.

Effective Modeling for Good Decision-Making

A Model is an external and explicit representation of a part of reality, as it is seen by individuals who wish to use this model to understand, change, manage and control that part of reality.

"Why are so many models designed and so few used?" is a question often discussed within the Quantitative Modeling (QM) community. The formulation of the question seems simple, but the concepts and theories that must be mobilized to give it an answer are far more sophisticated. Would there be a selection process from "many models designed" to "few models used" and, if so, which particular properties do the "happy few" have? This site first analyzes the various definitions of "models" presented in the QM literature and proposes a synthesis of the functions a model can handle. Then, the concept of "implementation" is defined, and we progressively shift from a traditional "design then implementation" standpoint to a more general theory of a model design/implementation, seen as a cross-construction process between the model and the organization in which it is implemented. Consequently, the organization is considered not as a simple context, but as an active component in the design of models. This leads logically to six models of model implementation: the technocratic model, the political model, the managerial model, the self-learning model, the conquest model and the experimental model. Data Gathering for Verification of Model: Data gathering is often considered "expensive". Indeed, technology "softens" the mind, in that we become reliant on devices; however, reliable data are needed to verify a quantitative model. Mathematical models, no matter how elegant, sometimes escape the appreciation of the decision-maker. In other words, some people think algebraically; others see geometrically. When the data are complex or multidimensional, there is the more reason for working with equations, though appealing to the intellect has a more down-to-earth undertone: beauty is in the eye of the other beholder - not you; yourself.

The following flowchart highlights the systematic development of the modeling and forecasting phases:



Forecasting System:
The Model-Building and The Forecasting Phases

The above modeling process is useful to:

- understand the underlying mechanism generating the time series. This includes describing and explaining any variations, seasonallity, trend, etc.
- predict the future under "business as usual" condition.
- control the system, which is to perform the "what-if" scenarios.

Statistical Forecasting: The selection and implementation of the proper forecast methodology has always been an important planning and control issue for most firms and agencies. Often, the financial well-being of the entire operation rely on the accuracy of the forecast since such information will likely be used to make interrelated budgetary and operative decisions in areas of personnel management, purchasing, marketing and advertising, capital financing, etc. For example, any significant over-or-under sales forecast error may cause the firm to be overly burdened with excess inventory carrying costs or else create lost sales revenue through unanticipated item shortages. When demand is fairly stable, e.g., unchanging or else growing or declining at a known constant rate, making an accurate forecast is less difficult. If, on the other hand, the firm has historically experienced an up-and-down sales pattern, then the complexity of the forecasting task is compounded.

3 Conclusion

In this paper, a new forecasting algorithm has been proposed to predict real-world time series. As previous step to the prediction, a clustering technique to label 24- dimensional time series has been used and the main novelty lies on the exclusive use of the labels obtained by the clustering to forecast the future behaviour of the time series, avoiding the use of the real values of the time series until the last step of the prediction process. Moreover, an automatization of the selection of the critical parameters— K and W—has been proposed. The algorithm has been successfully applied in Share changes and Buyer requirements time series of Indian Share markets providing very competitive results. The performance was accurate in all of them, showing thus the robustness and adaptability of the proposed approach for time series of different nature. This fact is specially remarkable since the approaches found in literature are usually focused on only one specific time series. Future work is focused on adjusting the model with dynamical lengths of window and on smoothing the matching sequence criterion.

4 References

[1] A. Abraham and B. Nath, “A Neuro-Fuzzy Approach for Forecasting Electricity Demand in Victoria,” Applied Soft Computing J., vol. 1, no. 2, pp. 127-138, 2001.

[2] S.K. Aggarwal, L.M. Saini, and A. Kumar, “Electricity Price Forecasting in Deregulated Markets: A Review and Evaluation,” Int’l J. Electrical Power and Energy Systems, vol. 31, no. 1, pp. 13-22, 2009.

[3] S.K. Aggarwal, L.M. Saini, and A. Kumar, “Price Forecasting Using Wavelet Transform and Lse Based Mixed Model in Australian Electricity Market,” Int’l J. Energy Sector Management, vol. 2, no. 4, pp. 521-546, 2008.

[4] N. Amjady, “Day-Ahead Price Forecasting of Electricity Markets by a New Fuzzy Neural Network,” IEEE Trans. Power Systems, vol. 21, no. 2, pp. 887-896, May 2006.

[5] J.P.S. Catalao, S.J.P.S. Mariano, V.M.F. Mendes, and L.A.F.M. Ferreira, “Short-Term Electricity Prices Forecasting in a Competitive Market: A Neural Network Approach,” Electric Power Systems Research, vol. 77, pp. 1297-1304, 2007.

[6] J. Chen, S.J. Deng, and X. Huo, “Electricity Price Curve Modeling and Forecasting by Manifold Learning,” IEEE Trans. Power Systems, vol. 23, no. 3, pp. 877-888, Aug. 2008.

[7] A.J. Conejo, M.A. Plazas, R. Espı´nola, and B. Molina, “Day-Ahead Electricity Price Forecasting Using the Wavelet Transform and ARIMA Models,” IEEE Trans. Power Systems, vol. 20, no. 2, pp. 1035-1042, May 2005.