SIT717 ENTERPRISE BUSINESS INTELLIGENCE

JAY TRIVEDI

218449725

Contents

Abstract 2
Introduction 2
Body of the survey 3
Models 4
How these models can be used in healthcare business intelligence? 6
Differences among various models used 7
Conclusion 7
References 8

Abstract

This paper focuses on Time series data analysis in the healthcare business industry. Stating in a layman's words time series is a sequence of numerical data that is obtained at a regular time interval. This technique has been widely used in areas such as economics, finance, environment and medicine. It can be summarised under two steps first one is to understand or describe what happened in the past (i.e. an event occurred over a month and year) and second one is to predict or forecast the future (i.e. using specific models to forecast the data). The data is represented with the help of time series graph. The aim of these graphs is to observe values on the y-axis with an increment of time against x-axis.

Objective of time series data analysis (Das 1994):

- Forecasting
- Control
- Understanding feature

Below is the visualisation of the time series data analysis

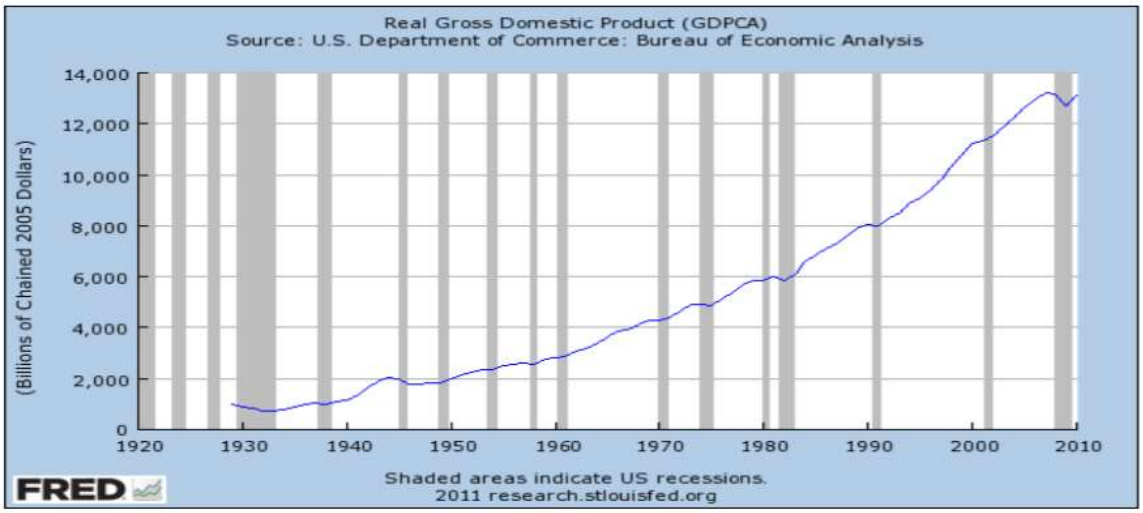

Fig.1 from *Time Series Analysis* Das (1994) page no: 4

Introduction

With the busy going and monotonous life of today's world, people now have changed their priorities and have now been focusing upon their health. The word "health" may sound small but in it there's buried a deep importance on how it affects our daily life and chores. Healthcare is itself a broad term which contains improvements of various medical services as well as its maintenance to meet the medical demands of the people. The increasing demands of the people especially the patients are widely recognised and are to be taken care of. The effects of healthcare are also seen in the modern society over the world.

One of the aims of healthcare is to improve the human health as well as to maintain it through diagnosis, treatments and by doing various disease preventions. With the number of patients increasing day by day so is the data that includes all the information needed for the hospital and its

doctors to operate. The data is a crucial part in this journey as now a day's data can also provide some insights about the future events with the help of various data analytic techniques. This survey is going to highlight it's one of the techniques also known as Time series data analysis.

With the help of time series data analysis in healthcare is that it helps you clean the data. Because of this it is possible to find the true data inside the dataset by filtering out the noise signal which is present in that dataset. Based on this the data can be forecasted for the future, which means if any incident/mishaps happen based on the data available any necessary adjustments can be made to prevent the mishap to happen or overcome it. For example, now the entire world is facing the covid-19 crisis which has shown or led to downward spiral of the economy of all the countries. Now by collecting the present data, the healthcare department can make necessary changes in their system and functionalities to be prepared to face the worst possible situation if there are any similar virus outbreaks are seen any soon.

Body of the survey

In this paper an approach is introduced in which different methods of time series data analysis is being analysed along with its advantages and disadvantages. This approach will be undertaken by keeping healthcare business intelligence as a priority. By minutely analysing each method a conclusion will be drawn that which method can be applied for the healthcare business sector to boost its business strategies which will lead to better outcome for the people and the patients.

There are four time series components that are being used for time series forecasting (Dr. Thanh Thi Nguyen, S779 Master's in IT (Professional), Deakin University, PowerPoint Slides 30 March 2020)

- Trend
- Cyclical
- Seasonal
- Irregular

Trend

In this component the data that is being used is on a persistent manner which means the data is taken under regular interval of times. It propagates in a linear motion such as an upward or downward pattern. The data that is being gathered is of a large volume such as population, technology etc. over the several years of duration.

Cyclical component

This component shows a repeating upward and downward moments over a period of time. These interactions are a result of various factors influencing the economy. The data that is being used in this component is of a 2-10 years duration. One of the important aspects of this component is that the data may vary in length.

Seasonal component

In this component a regular trend of upward and downward fluctuations is been observed. The data is mostly gathered from the weather channels, customs etc. the length of the data is usually one year.

Irregular component

In this component the data is gathered on an irregular basis which means it leads to distorted fluctuations in the graph. The reason for this is that the data is gathered in a short duration of time by the events occurred in the nature or any accidents taken place.

Models

ARIMA model

According to Chatfield (2000) the ARIMA (Auto Regressive Integrated Moving Average) model is mostly used for analysis and forecasting the time series data. This model is effective in analysing the health data in an intelligent environment. The ARIMA models fall under the category of auto-regression (AR) component where two models are applied simultaneously i.e. AR and MA, where I stand for Integrated. The ARIMA model is a formed with the help of additional seasonal terms in the ARIMA model. ARIMA can also be identified as ARIMA (p, d, q) where p stands for order of autoregressive part, d stands for degree of first differencing involved, q stands for order of moving average part. In this model we cannot apply the AR and MA processes directly, firstly we must apply differencing to make the non-stationary series into stationary. This is followed by $(X_t - X_{t-1}) = (1 - B)X_t$. in which the dth differences with be written as $(1 - B)^d X_t$. when the original data series is differentiated the d times before fitting the ARMA (p, q) process then the model with the original undifferentiated series is called the ARIMA (p, d, q). One of the challenges faced by this model is choosing the required order of differencing.

Box Jenkins Seasonal ARIMA model (SARIMA)

According to Chatfield (2000) the Box Jenkins model is also abbreviated as Box Jenkins seasonal ARIMA model. When a model follows a seasonal trend of components where s is the time periods taken over a year then it is called as seasonal ARIMA model. This model consists of two components in which (p, d, q) is the non-seasonal part of the model and $(P, D, Q)_m$ is the seasonal component of the model. This model can be written as $\emptyset(B)\varphi(B^s)(1 - b)^d(1 - B^s)^D X_t = \theta(B)\vartheta(B^s)Z_t$. another SARIMA model which can be used as is the order of $(0,1,1)x(0,1,1)_s$ which later can also be written as

$(1 - B)(1 - B^{12})X_t = (1 + \theta B)(1 + \vartheta B^{12})Z_t$. This model can also be called as airplane model because it is used to plot the airline data which is useful for forecasting purposes. One of the key things to keep in mind while deploying this model is to suitably choose the values of two orders of differencing both seasonal (D) and non-seasonal (d) to make the series stationary and remove any seasonality.

Neural network time series data analysis

Neural network time series data analysis is one of the completely different type of non-linear models that are being used now a days. Its structure is like the design of the human brain in some sense. This technique has been widely used across different scientific problems and has increasingly used in statistical applications as well. A neural net of a system can also be described as a set of input connected to a set of outputs in a non-linear way. Speaking in a time series form the “outputs” can be forecasted in a time series manner whereas the “inputs” are the values of series and also some explanatory variables. As well as the connections between the inputs and the outputs are done with the help of a number of hidden layers of neurons or nodes. The structure that is formed with this is

often referred as “architecture”. The architecture is determined on the basis of the following such as number of layers, number of neurons in each layer and how the inputs, the hidden layer and the outputs are connected. Moreover, the neural networks give more prediction accuracy as compared to other regression models. The following diagram shows the architecture of the neural networks.

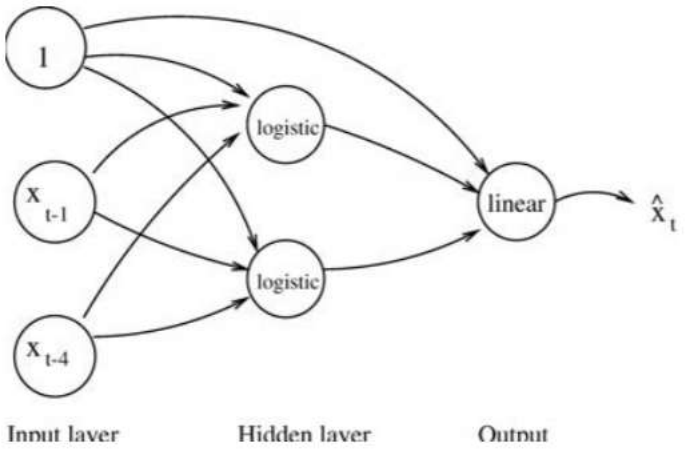

Fig 2 from *TIME-SERIES FORECASTING* Chatfield (2000) page no: 74

The figure given above is called as feed-forward type as no feedback loops are given. A suitable architecture can be determined on the basis of the content, data and perhaps also by takin external considerations using the properties of the data. When it comes to number of hidden neurons a trial-and-error method can be implemented. When we want to determine one-step-ahead forecast of the quarterly data then it is recommended to include the values of lags one and four as inputs. As well as it is also recommended to include the constant input term in the data.

Linear trend model

It is one of the simplest forecasting models used in time series. In this model the relationship between the response variable Y and its time X is a linear function. It is denoted by $Y_i = b_0 + b_1 X_{1i}$

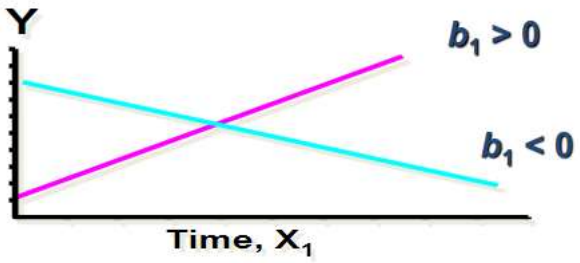

Fig 3 (Dr. Thanh Thi Nguyen, S779 Master’s in IT (Professional), Deakin University, PowerPoint Slides 30 March 2020)

Based on the above graph the value of b_1 varies, when the graph is in increasing upward direction the value of b_1 is greater than zero otherwise when the values are in decreasing downward direction the values of b_1 is less than zero. One of the benefits of using this model is that it uses slope and intercepts that gives the best average prediction of past data.

How these models can be used in healthcare business intelligence?

Healthcare is a broad sector in which there are multiple subsectors prevailing under it. Healthcare business industry is blooming in the past recent years because of different enhancements with the help of evolving technology. One of the most important aspect for the healthcare business industry is the data on which these models have been deployed and used today. These data are also taken under different parameters such as duration of time, length, event under which data is taken and many more. All these factors are being considered while deploying the above models for its forecasting purposes. The deployment of these models is based upon the type of data respected to their specific healthcare business industry.

Emergency departments

According to Kadri et al. (2014) one of the issues that has been faced by the hospitals is the overcrowding of the emergency departments, because more and more attention is being drawn towards the hospital management system to manage better flow of the patients and also to improve its management strategies. In this the data was taken by the daily attendances of the hospital emergency department. In this ARIMA model was being deployed for the forecasting of the time series analysis. As a result, a clear and concise forecasting was drawn which clearly depicts accurate forecasting of the patient attendances that has improved staff deployment and facility deployment planning. Also, one of the key things observed in it was that if there's a usage of seasonal data, then SARIMA model can be used which also provides accurate forecasting.

Smart home-based health care system.

According to Jakkula, Cook and Jain (2007) one of the branches of healthcare industry has penetrated its roots in healthcare system home. Now a days many health care appliances are directly synced with smart devices via Bluetooth that can be operated through voice commands. A technology friendly smart health care system would prove quite beneficial for the elderly people to avoid having any expensive care givers in the hospitals. In this the Box Jenkins seasonal ARIMA model have been deployed with a small amount of data keeping in mind that the data gradually increases over time. A precise forecasting of data has been provided by the SARIMA model which has led to steady performance of the healthcare system.

Wearable healthcare technology for elderly

The wearable healthcare technology has the power to improve the access of the healthcare information and also give them the freedom to take care of their health independently. As technology evolved many wearable technologies are now used for example smart watches. They are connected to your smart phones which are directly connected to the respected IoT cloud service provider. Despite the potential power of these wearable healthcare technology, their acceptance and usage have been quite low. In this study the neural network time series data analysis has been used to forecast the data, which has shown promising results that the use and infrastructural support are no longer the limitation of the wearable healthcare technology for the elderly.

Trends in childhood Asthma: Prevalence, Healthcare Utilization and Mortality.

According to Akinbami and Schoendorf (2002) The linear trend model of time series analysis has been used to analyses the trend of asthma and it was being observed that with the increasing, additional years of data collection has proved beneficial to change its trend.

Differences among various models used

Neural network time series data analysis

The analysis was carried out by Talukder et al. (2020) between Structural Equation Modelling (SEM) and Neural Networks (NN) for the wearable healthcare technology (WTH) in the elderly to identify the similarities and differences for the importance of the predicators. Where SEM results showed the most and the least significant adoptions for the social influence and functional congruence whereas NN analysis depicted that social influence was the most important factor as compared to hedonic motivation. Hence it concluded that using superior machine learning tools such as neural network analysis gave accurate predictions as compared to SEM which can only analyse linear relationships (Zhang & Qi 2005).

Box Jenkins Seasonal SARIMA time series model

This model was deployed by Jakkula, Cook and Jain (2007) under smart home-based healthcare systems. An experiment was being done between support vector machine algorithm found in Weka and seasonal ARIMA model under a data set of 40 days for support vector algorithm and of 41 days for SARIMA. The aim was to conclude that which model depicts better results of health trends. It was being keenly observed that the SARIMA model gave better accuracy as compared to the other. In conclusion, when additional data was being fetched to the SARIMA model its accuracy increased which gave better results.

ARIMA model

According to Kadri et al. (2014) the model was used in the overcrowding of emergency departments as result of more focus towards the management system and its maintenance. When the model was applied to three acuity categories as well as total patience attendance in the ED department it was being found that the results were more accurate along with its performance. On that basis better staff deployment was being deployed.

Linear trend model

The model was deployed under Trends in childhood Asthma: Prevalence, Healthcare Utilization and Mortality (Akinbami & Schoendorf 2002). The main difference that differentiates this model with the rest models is that the more data is gathered the accurate forecasting is seen in this model. This quality differentiates the linear model with others.

Conclusion

The survey has covered three major areas 1) Time series models, 2) Usage of this model in healthcare business industry and 3) Difference among each model. Firstly, when it comes to the models used it was being concluded that it solely depends on the organisation or personnel and also on the amount of data they have. At the end, each model has their differences and special functionalities that are being used in the modern world to forecast the future data. Secondly, when it comes to healthcare business industry it is pacing towards the adaption of these models deep into their roots for their advancements. With this it has led to better outputs in different scenarios as given above. With the merger of healthcare and machine learning models it will lead to better, safer and faster execution and prediction of the data that will improve the lives of the people. Thirdly,

each model is unique in their own perspective and outputs hence it depends on the goal of the user or organisation for its implementation.

References

- Akinbami, LJ & Schoendorf, KC 2002, 'Trends in childhood asthma: prevalence, health care utilization, and mortality', *Pediatrics*, vol. 110, no. 2, pp. 315-22.

- Chatfield, C 2000, *Time-series forecasting*, CRC press.

- Das, S 1994, *Time series analysis*, vol. 10, Princeton university press, Princeton, NJ.

- Jakkula, VR, Cook, DJ & Jain, G 2007, 'Prediction models for a smart home based health care system', in *21st International Conference on Advanced Information Networking and Applications Workshops (AINAW'07)*, vol. 2, pp. 761-5.

- Kadri, F, Harrou, F, Chaabane, S & Tahon, C 2014, 'Time series modelling and forecasting of emergency department overcrowding', *Journal of medical systems*, vol. 38, no. 9, p. 107.

- Talukder, MS, Sorwar, G, Bao, Y, Ahmed, JU & Palash, MAS 2020, 'Predicting antecedents of wearable healthcare technology acceptance by elderly: A combined SEM-Neural Network approach', *Technological Forecasting and Social Change*, vol. 150, p. 119793.

- Zhang, GP & Qi, M 2005, 'Neural network forecasting for seasonal and trend time series', *European journal of operational research*, vol. 160, no. 2, pp. 501-14.