Forecasting ensemble empirical mode decomposition

Introduction

This chapter introduces the background of time series and the importance of forecasting. The motivation behind the project is elaborated and finally the aims and objectives are given.

1. 1 Background

Time series can be defined as a sequence of observations or measurements that are taken at equally spaced timed interval (Xu, 2012). Hence, it is a stochastic process and can be expressed as (Xu, 2012):

x(t) = xi; i = 1; 2; :::; N: (1. 1)

Some examples of time series data include yearly profit, monthly recorded temperature, hourly electrical consumption.

Time series are classified into two categories mainly the stationary time series and non stationary time series. Stationary time series consist of data which remain fixed irrespective of the whereabouts. A stationary process is one where the mean, variance and autocorrelation do not vary with time (Nau, 2014). For example, the financial stock change of Mauritius remains constant in Mauritius as well as in any other place in the world. Non stationary time series on the contrary involve data that keeps changing over time. For instance, if we consider meteorological data of Mauritius, the data collected are varied considerably from region to region as well as accordingly throughout the year. For example, we have more rainfall over regions on the Central Plateau compared with the coastal regions as demonstrated by Figure (1. 1) which illustrates the variation of

rainfall collected for Mauritius over distinct regions from 1960 – 1990. while figure 1. 2 shows the difference in signal data between the two classes of time series. All meteorological data including temperature, wind speed, solar irradiance irradiance, sea pressure and many more weather parameters similar to rainfall have variations both in time and location.

Hence, we can conclude that meteorological data are non stationary in nature.

Figure 1. 1: Distribution of rainfall for Mauritius for the year 1961-1990

Source: http://unfccc. int/resource/docs/natc/maunc1/chap1/chapter1. htm

Figure 1. 2: Difference between stationary and non stationary series ,

Source: http://en. wikipedia. org/wiki/Stationaryprocess

Time series modeling is a vast field of research. The analysis of time series signals can be extrapolated to meet demands of analytical results and predicting results in various fields, such as :

1. Economical
2. Climatological
3. Biological
4. Financial and others

Due to its implementation in various fields, continuous research are been done in order to design model for forecasting with better accuracy and efficiency. The behaviour of time series is governed by four main aspects namely trend, seasonal variation, cyclic variation and random variation (Xu,

2012). Trend of time series can be pictured as the evolution of the series over time and hence gives the forthcoming pathway of the data. Hence, trend analysis is very efficient in predicting extensive behaviour of data. Phonetically, a general assumption in most time series techniques is that the data are stationary. Transformation of non stationary to stationary is often done to manipulate the data for analysis.

Forecasting is of high precedence in application of time series as it can predict future events based on past events, specially when using in the field of limited resources. Forecasting may be classified as a prediction, a projection or estimate of a future activity. In fact, we have two types of forecasting methods namely qualitatively and quantitatively.

Qualitative methods are non mathematical computations whereas quantitative methods are rather objective methods based on mathematical computations.

1. 2 Motivation

We belong to a world of success in which one of the leading factor to success is our ability to predict the result of our choices making all of us in a way or another forecasters.

Climate consists of one of the major applications of forecasting. Over years, newer and better models are been investigated so as to improve forecasting accuracy as much as possible. Investigating weather parameters is highly necessary so as to be able to predict weather situations which are required in various fields such as aviation, shipping, oceanography and agriculture.

Moreover, it is helps to evade weather hazards. Mauritius has being confronted to drastic changes in weather conditions recently. We have already a weather station which is deploying its best methods for weather forecasting but is unable to predict accurately unexpected changes in weather, for example the recent flash flood in March 2013 or one of the most worst drought that stroke Mauritius in 2002. Therefore, in order to prevent further incidents or life taking calamities, it is of high importance to have accurate and early predictive models in order to take preventive measures to make sure that the population is safe well before such events occur. This project comprises of investigating a different method for forecasting meteorological data.

Throughout this project we will be dealing with time series models based of data which has been collected over years and try to foresee future events based on the fundamentals patterns confined within those data.

The most commonly used forecasting model for time series was the Box – Jenkins models (ARIMA and ARMA models) (Peel et al., 2014). They are non-static models that are beneficial in forecasting changes in a process. Many models have further been developed among which is listed the Hilbert Huang Transform (Huang and Shen, 2005).

Since climate data are of nonlinear and non-stationary nature, Hilbert Huang Transform is capable of improving accuracy of forecast since most previous traditional methods are designed for stationary data while this method is efficient in both cases. On the other hand, recognizing all the advantages of Artificial Neural Network, it is of no surprise that this methodology has

gained so much interest in the this field of application. ANN have proven to be more effective, compared to other traditional methods such as Box-Jenkins, regression models or any other models (Khashei and Bijari, 2009) as a tool for forecasting.

Both successful models mentioned however carries their own associated percentage error. As a means to minimize error, both models can be combined to give rise to a new hybrid model with better performance capabilities.

1. 3 Aims And Objectives

1. In this project, the aim is to develop a combined model from two completely different computational models for forecasting namely Ensemble Empirical Mode Decomposition and Artificial Neural Network so as to improve accuracy of future predictions of time series data.

2. EEMD will be adopted as the decomposition technique to obtain a set of Intrinsic Mode Functions (IMF) and residual for meteorological time series data for Mauritius signal while ANN will be the forecasting tool which will take as input parameters the non obsolete IMFs. The results obtained will be compared with real data in order evaluate the performance of the model. The idea is to reduce error associated with each model when employed separately as both models possess their own skill in determining trend in complex data.

3. Eventually, the model will be applied to forecast meteorological data mainly rainfall from MMS and wind speed from studies conducted by fellow colleagues.

1. 4 Structure of Report

1. Chapter 2 consists of a literature review on the models and their applications

2. Chapter 3 introduces Ensemble Empirical Mode Decomposition and validate the EMD model.

3. Chapter 4 introduces the Artificial Neural Network and validate the network.

4. Chapter 5 present the results from application of EEMD to meteorological data. The EEMD-ANN hybrid model is also introduced and validate. Finally the following is applied to rainfall and wind speed data.

5. Chapter 6 presents the conclusion and the future work.