UACEE International Journal of Advancements in Electronics and Electrical Engineering – IJAEEE
Volume 2 : Issue 2 [ISSN 2319 – 7498]
Publication Date : 05 June 2013

Decomposition of time series data, in discrete non-linear time series data systems

[R.K Singh]
[Sunil Bhaskaran]

***Abstract*—Towards the end of the 20th century, we have seen an improved interest among Statisticians and Computer engineers to explore data from any data source with respect to the change in time. However, most of the techniques used remains the same as that used in conventional data mining. Capturing, indexing, representing and storing the data remains the key issue in time series data mining. Indexing is a very critical under job under noise conditions. The indexing system exploded the database volume. In time series data ming a statistical models which provides descriptions for the sampling of data, (data collected on global warming, flood and flood forecasting pattern etc) are deviced. In order to provide a statistical arrangement for describing the nature of a continues stream of data that fluctuate in a random fashion with respect to the time, we assume a time series can be defined as a collection of random variables indexed according to the order they are obtained. Here we are assuming a time series data as a sequence of (linear or non-linear) random variables, which can be represented as t1,t2,t3,t4 …., where the random variables t1,t2 etc. are the observed values with respect to the change in time. Trend detection and recording is the most important activity in time series data mining. In practice, it is accomplished using linear and nonlinear regression technique that satisfactorily helps in identifying non-monotonous trend component in the time series. It is already been proved that statistical methods such as moving average can be effectively used in smoothening data flow.**

Keywords—Time Series Data ming - TSDM, regression, segmentation, Time series decomposition, discrete time series, sampled series.

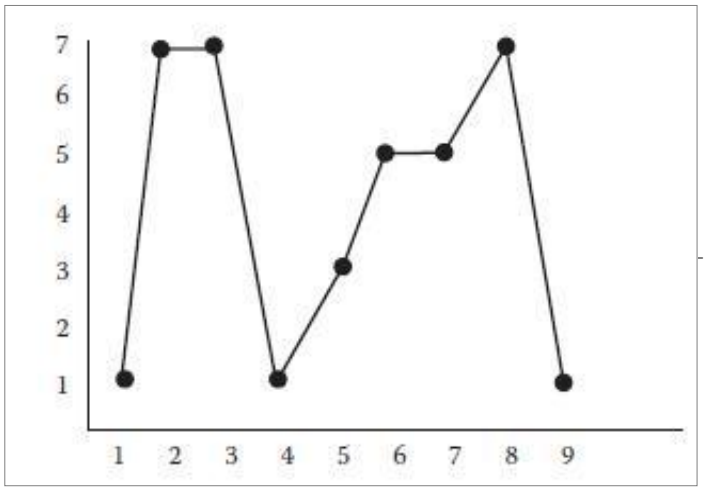
Introduction

"Prediction is very difficult, especially if it's about the future" – Nils Bohr. The difficulty of the analysts is that, there is no immediate way available for calibrating these forecasts. He is required to waiting for future observations to come. It is sometimes helpful to check the forecasting ability of the model by employing data readily available with the analyst[1]. It is achieved in different ways.

R.K Singh, Controller of Examination
Uttarakhand Technical University, Dehradun
India
rksinghkec12@rediffmail.com

Sunil Bhaskaran, Assistant Professor
Army Cadet College Wing, IMA/ Dehradun
India
bsunil2003@gmail.com

A time series is said to be continuous when observations



are made without any break in time as in Figure: 1.

Figure 1 : A time series data representation model

1. The term continuous, is used for series of this type even when the measured variable can only take a discrete set of values. A time series can be treated as discrete when observations are taken at a specific time interval. Usually they are equally spanned. The term discrete, is used for time series of this type even though the measured variables are continuous in nature[2,3].

We are mainly concentrating on discrete time series, where the observations are taken at equal intervals for this research activity. We are also considering some continuous time series briefly, while giving some references regarding the analysis of discrete time series taken at unequal intervals.

Discrete time series can arise in several ways[3]. Given a continuous time series, we could record the values at equal intervals of time to give a discrete time series, sometimes called a sampled series. To avoid any data lose, the sampling interval between successive readings must be carefully decided. A different type of discrete series is obtained when a variable does not have an immediate value, but can generate the values over some equal intervals of time. Monthly exports and daily rainfalls are examples of such time series. Some time series are discrete by its origin. An example for such category is the dividend paid by a company to its shareholders over a continues period of time or up side and down side movement of the index of an equity market as shown in figure : 2.

Figure 2: Performance of (NIFTY-India), National stock exchange 100 share index. Courtesy National Stock Exchange

(Source, http://www.nseindia.com) – accessed on 19/02/2013.

Many statistical concepts and theories are available for random samples of independent and continues observations. The peculiarity feature of time-series data analysis is the fact that, the successive observations are not necessarily be independent. In this case, the analysis must take into account the time order of the observations. If the are observations are continues and independent, future values sets may be predicted from previously observed data sets. When a time series is predicted 100 % accurately, it is said to be deterministic. In actual practice, most time series are stochastic in nature. In stochastic model the future is predicted only partly. It is determined by past values, so that exact predictions are impossible. It must be replaced by the idea that, the future values have a probability distribution[4,5,6]. It is determined on the basis of a knowledge of past values.

As the main partnership functions of soft computing, the following technologies have been used:

Fuzzy logic - which has to deal with the imprecisions in computing and to perform the approximation on the basis of reasoning.

Neurocomputing (ANN) - which is required for learning and self recognition purposes.

Probabilistic reasoning - which is essential in dealing with uncertainty and belief propagation.

Industrial applications

In Japan, during 1980s, the earliest use of fuzzy logic was recorded. It was used in the process industry. Fuzzy logic facilities are capable of solving complex nonlinear and uncertainty problems of a chemical reactors. Such advanced techniques started replacing the highly skilled plant operator. During the same period, neural networks were also applied in statistical analysis of huge data sets acquired through sensor by time series analysis and forecasting techniques. This techniques was later extended and applied in data mining for managing very large amounts of complex. However such techniques of soft computing are based on pattern recognition and multi sensor data fusion. It was found successful in geting a better understanding on process behavior through the analysis and identification of essential process features which are hiding in data piles. Scientist were successful in unearthing some accompanying problems related to plant monitoring, diagnosis, quality control, production monitoring and forecasting, plant logistics and various other services.

Decomposition process in time series data

Data transformations is useful in many statistical activities such as stabilization of the variations of data. The decomposition activity is treated as an important technique in time series analysis, particularly for making adjustments in seasonality. It is constructing, a number of component series from an observed time series, which could be further used to reconstruct the original by additions or multiplications, where each of these has a certain characteristic or exhibiting a particular type of behavior. Non constant variance is common phenomena in time series data. There are also several adjustments which are useful in developing time series models for forecasting. The most widely used adjustments are trend adjustments and seasonal adjustments. This adjustments are sometimes called trend and seasonal decomposition[9,10].

To make a proper forecast in the context of a multi-component time series, it is required to know to what extent the particular components are present in the time series data. It emphasis needs for the decomposition of time series data. Such decomposition will help us to identify and extract the partial data superimposed to the main time series data.

A time series that shows a strict trend is called a non stationary time series. Modeling and forecasting of such time series are simplifying the job of the analyst.

One way to implement this is to fit a regression model describing the trend component and then subtracting it out of the original observations, leaving a set of residuals which are free of trends[6]. The time series decomposition process can be represented dramatically, as shown in figure : 3

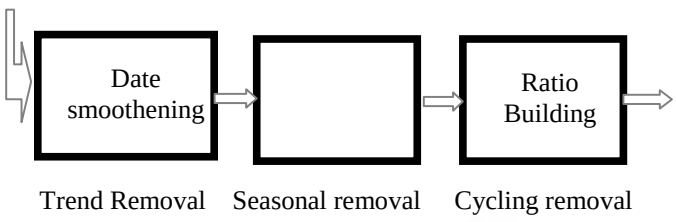


Figure 3: Time series decomposition process

Relation-Seeking

It is defined as tasks that imply a search for occurrences of specified relations between characteristics or between references[7]. Here are some example tasks of this type:

When did the stock price increase or decreased by more than 15% in comparison with the previous day?

I there any earthquakes recorded in last 72 hours before a given earthquake?

Which all states in India recorded a sharp decline in gender ratio for the period from 1981 to 1991?

In which metro city of India and in which year the crime against women exceed the burglary rate?

Find pairs of earthquakes which happened in the time interval between them is not exceeding more than 48 hours and what was the distance between their epicenters in km.

In an attempt to illustrate such a task graphically, we arrived at the idea of a metaphorical representation of a specified relation by something like a stencil, or mask. This stencil is to be moved over a set to find elements that fit in its holes and, hence, are related in the way specified. This metaphor is presented in Fig. 3.6. The shape on the left depicts a stencil, which represents symbolically a specified relation, denoted by . This is assumed to be a binary relation, i.e. a relation between two elements[8]. The goal is to find pairs of elements of the characteristic set C linked by the relation .

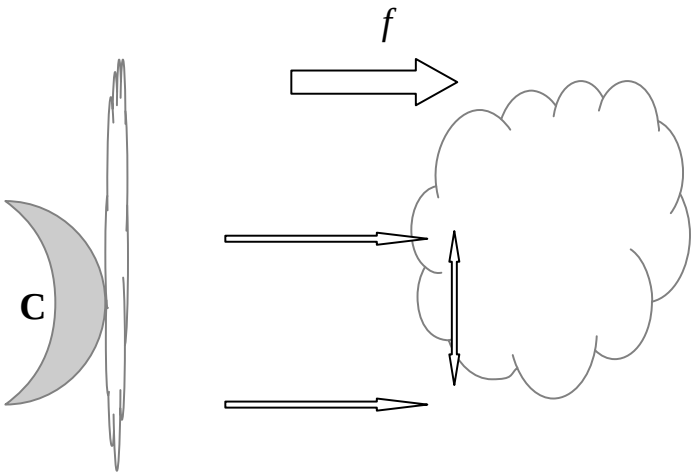


Figure 3: A graphical illustration of a relation-seeking task. A specified relation

between characteristics, denoted by, is represented as a stencil, or mask.

The ultimate goal is to find what references correspond to the characteristics linked by the relation

Related Work

Analysis of link structure with an idea of getting informational organization remains as an issue in a number of research field which are interconnected. Here we are discussing some of the established approaches and try to establish the connections between our work and them.

Let us consider the case of bibliometrics. It is the study of written documents and their citation in various international journals including there structure. Research in this field has long been concerned with the use of citation with the purpose of identifying core sets or clusters of articles, authors, or journals from a particular fields of a topic. Small and Griffith designed co-citation analysis as a principle method for quantifying the common intellectual areas among pairs of documents[11]. Researchers have successfully identified and applied the co-citation analysis to the mining of the semantic structure of the World Wide Web. After this study, a large number of applications of bibliometrics were made in the hyper linked conditions with the aim of extracting the most authenticated pages connected to the query of the user.

Conclusion

Decomposition is an important statistical method that is applied on time series data used for reconstructing a time series under analysis, into small sub functional units. On the basis of the fundamental activity involved in the decomposition process can be divided in to two categories, which are.

- Based on rates of the amplitude change of the data.
- Based on predictability of the data.

Developing and implementing procedures with the idea of monitor the performance of the model developed for forecasting, is an essential activity in the designing of the end forecasting system design. It is not a matter on record, how much effort has been used in developing the forecasting model. Regardless of how perfectly the model works in the initial days, it is likely that its performance will deteriorate, over time. However the pattern of the time series will change, either because the inertial forces that drive the process may evolve through time. It may due to the facts such as external events like new customers entering the market. A change in level or a slope may occur in variable that is being forecasted. Some times, it is also possible for the inherent variability in the data to increase or decrease. Subsequently the performance monitoring is important.

Acknowledgment

I take this option to express my sincere gratitude to Prof R.K Singh, Controller of Examination, Uttarakhand Technical University, Dehradun for encouraging me to write this research paper and get it published. He is always an inspiration for me. His active guidance and research

aptitude is helping me to compete my thesis for the award of doctorate degree in Computer Science and Engineering from Uttarakhand Technical University, Dehradun. Also, I am thankful to my wife Ms Bisini, for providing me a conducive atmosphere and moral support, without which, it was not possible for me to complete this work.

References

[1] Sheng Chang, Wynne Hsu and Mong Li Lee. (2006). *Mining Dense Periodic Patterns in Time Series Data, Proceedings of the 22nd International conference on data engineering (ICDE'06)*, 8-7695-2570-9/06, IEEE.

[2] Jose Zubcoff , Jesús Pardillo and Juan Trujillo. (2009). *A UML profile for the conceptual modeling of data-mining with time-series in data warehouse*. Information and Software Technology 51(2009) 977–992, Science Direct, ELSEVIER.

[3] Juan Trujillo. (2011). *A review on time series data mining - Engineering Applications of Artificial Intelligence*, 24 (2011) 164–181. ELSEVIER.

[4] Xiao Hu, Peng Xu, Shaozhi Wu, Shadnaz Asgari and Marvin Bergsneider. (2010). *A data mining framework for time series estimatio*n. Journal of Biomedical Informatics, 43 (2010) 190–199. ELSEVIER.

[5] Das P.K, Maya Nayak, Senapati. M.R and Lee I.W.C. 2007. *Mining for similarities in time series data using wavelet-based feature vectors and neural networks*. Engineering Applications of Artificial Intelligence, 20 (2007) 185–201. Science Direct, ELSEVIER.

[6] Zhe Song, Xiulin Geng, Andrew Kusiak, and Chang Xu. 2011. *Mining Markov chain transition matrix from wind speed time series data. Expert Systems with Applications* xxx(2011)xxx–xxx. Science Direct, ELSEVIER.

[7] Chun-Hao Chen, Tzung-Pei Hong and Vincent S. Tseng. 2009. *Mining fuzzy frequent trends from time series. Expert Systems with Applications*, 36 (2009) 4147–4153. Science Direct, ELSEVIER.

[8] Huei-Wen Wu and Anthony J.T. Lee. 2009. *Mining closed Knowledge Engineering*, 68 (2009) 1071–1090, Science Systems, 24 (2011) 492–500. IEEE, Science Direct, ELSEVIER. ELSEVIER.

[9] Hailin Li and Chonghui Guo . 2011. *Piecewise cloud patterns in multi-sequence time-series databases. Data & approximation for time series mining*. Knowledge-Based systems 24(2001) 202–215. Science Direct, ELSEVIER.

[10] Dash P.K, Behera H.S and Lee I.W.C 2009. *Time sequence data mining using time–frequency analysis and soft computing techniques*. Applied Soft Computing, 8 (2008) 202–215. Science Direct, ELSEVIER.

[11] W. N. Venables, D. M. Smith and the R Development Core sequence data mining using time–frequency analysis and Team, *"An Introduction to R. Manual of R language"*, *soft computing techniques. Applied Soft Computing*, Institute for Statistics and Mathematics, 2007.

About the authors

Prof R.K Singh

The first author, Prof R.K Singh, is the supervisor of the second author in research at Uttarakhand Technical University Dehradun, Uttarakhand. Prof R.K Singh is the controller of examination at Uttarakhand Technical University, Dehradun, India.

Sunil Bhaskaran

The second author is presently working as Assistant Professor in Computer Science at Army Cadet College Wing of Indian Military Academy, Dehradun. ACC Wing, IMA is a recoganised study center of JNU, New Delhi. I am also actively pursuing my PhD in Computer Science and Engineering under the guidance of Professor R.K Singh, Controller of Examination, Uttarakhand Technical University, Dehradun.