

A Survey on Time Series Segmentation

Kaushik Shete, Dr. Amar Buchade
Department of Computer Engineering,
Pune Institute of Computer Technology,
Pune, India

Abstract- Time series is a set of observations of data measured over a specific interval of time. These are difficult to monitor, summarize and predict. Time series segmentation is a process organizing time series into groups that are having similar characteristics. These segments can be used for analyzing time sequence, dimensionality reduction, removing outliers and for similarity search. Analyzing multiple time series is a difficult task. By segmenting this one can easily analyze. In this paper we have taken a survey on TS segmentation based on the following categories- i) Horizontal Segmentation ii) Vertical Segmentation.

Index Terms- Cluster Analysis, Data Mining, Time Series Clustering and Segmentation.

I. INTRODUCTION

Time series are used in statistics, signal processing, pattern recognition, finance, forecasting [1]. Time series generally consists of historical data. The analysis of time series is performed in different areas like subsequent matching, indexing, clustering, classification, visualization, segmentation, trend analysis, forecasting. Data is measured over a specific time interval like a daily basis or hourly basis for a long period with multiple parameters. Because of having high dimensions and large size users cannot easily visualize time series. A time series database may have some valuable information. This information can be extracted from pattern discovery.

Time series segmentation is a process of dividing time series in the number of segments or classes. Segmentation leads to the discovery of patterns or trends. These segments will help to reduce computational power as dimensions reduce. Segmentation can be classified into-

A. Vertical Segmentation:

In this technique, segments are made on a vertical basis. This can be used to extracting patterns like trends, seasonal, cyclic, etc. and can be used to reducing time series.

B. Horizontal Segmentation:

In this technique, segments are made on a horizontal basis. This technique is used to extract a similar time series. The user gets a set of segments, where each segment is set similar time series. Also, the user can select some time series from the available graph [5].

II. RELATED WORK

Various vertical segmentation approaches are currently available like sliding window, top-down, bottom-up, piecewise linear approximation, etc. In [1] authors proposed vertical time series segmentation based on Sliding Window And Bottom-up approach. In this, authors have compared their approach with online segmentation and offline segmentation techniques. Segmentation of time series is based on a single variable. Also, the authors gave

details of existing segmentation techniques like sliding window, top-down, bottom-up. The proposed system is compared with other techniques using various datasets like Radiowaves, ExchangeRates, WaterLevel, and SpaceShuttle. In [2] authors have given the information on various segmentation techniques based on piecewise linear representation. The authors have provided algorithms for each technique. Techniques are compared with each other using various datasets like CapacityUtilization, WheatPrice, ECG and StockMarket dataset. [3] is a review of time series data mining. Authors have given various representation and indexing techniques, Similarity measures, Segmentation, Visualization and Data mining of time series. Also, authors have given information related to time series dimensionality reduction which is helpful to reduce time series data. In [4] authors have provided information related to time series analysis and forecasting. Authors have provided a classification of patterns like horizontal pattern, trend pattern, seasonal pattern, cyclic pattern, etc. It also provided information about various forecasting techniques for time series data like moving average, weighted moving average, Exponential smoothing. It also provided various decomposition methods for TS data. [5] is a software that gives segmentation functionality for multiple time series data. Segmentation is provided by 3 approaches: -

A. Graphical Query Approach

B. Hierarchical Query Approach

C. Parameter Query Approach

[6] is a review of time series clustering. Authors provided various TS clustering based on- i) Whole time series clustering ii) Sub-sequence time series clustering iii) Time point clustering. It also provided various clustering algorithms based on Hierarchical, Partitioning, Grid, Density, etc. In [7] authors have proposed the Online Divisive Agglomerative Clustering algorithm for time series clustering. In this system iteratively goes on splitting time series data until stopping criteria met. While splitting time series into 2 parts also it has the ability to aggregate clusters. The proposed system has the ability to split and aggregate the clusters. In [8] authors have provided various horizontal time series segmentation like k-means, density based, trajectory clustering with different distance and similarity measures which can be further classified into univariate and multivariate. [9] Provides various approaches for reducing time series data using R programming. These approaches help reduce time series data into fix number of points. "TSrepr" is a package used for time series representation, which contains various dimensionality reduction, preprocessing and feature extraction methods. [10] Provides various approaches for time series reduction. It also provides various feature extraction methods for time series data. Tslcust [11] is a package provided by R used for clustering TS data into K number classes. It provides various clustering types like hierarchical, partitional, fuzzy clustering with different distance measures. Proc similarity [12] is a procedure provided by SAS to compute similarity measures for time stamp data. It computes the similarity between time series. Further clustering is used for segmenting TS data into the number of classes. Tsllearn [13] is a package provided by python for analysis of TS data. Using this user can horizontally segment TS data into K classes with the help of 3 different distances: Euclidean, DTW, and softdtw. [14] Provides various methods for selecting the optimal number of clusters using R. These methods help select the optimal value of K for the clustering technique.

III. SEGMENTATION APPROACH

On the basis literature work, segmentation can be divided into 2 major categories as-
I) Vertical Segmentation II) Horizontal Segmentation

I) Vertical Segmentation-

It is mainly used to extract patterns from singular time series (time series having a single variable). Extracted patterns can be trends, cyclic, seasonal patterns, etc.

Following are the most common techniques used for vertical segmentation-

- a) Top-down Approach
- b) Bottom-up Approach
- c) Sliding Window Approach
- d) Sliding Window And Bottom-up Approach
- e) Perceptually Important Points

a) Top-down Approach-

Top-down is based on the divide and conquer process for creating segments of time series. It uses binary splits for dividing time series into a set of segments; it can be based on maximum distance or mid-point-based splits.

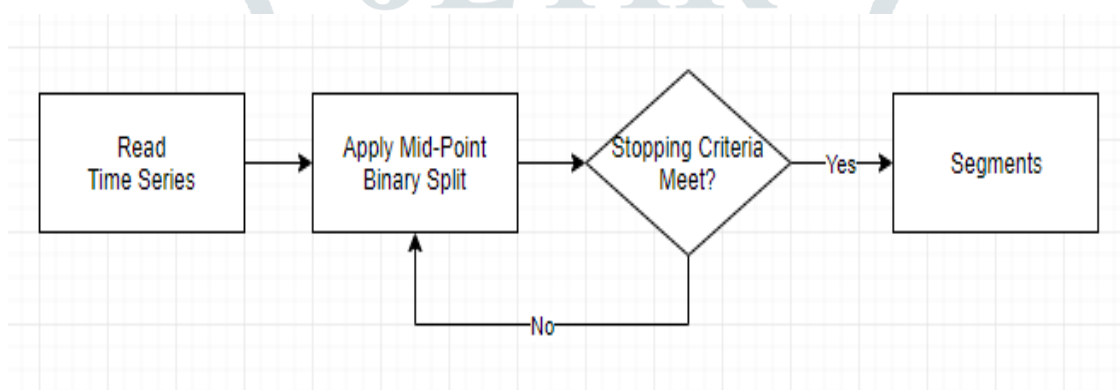


Fig.1 Flowchart Top-Down Approach

b) Bottom-up Approach-

Bottom-up approach uses merging operations. It will iteratively merge time sequences based on criterion. Merging cost can be considered as criterion for this approach.

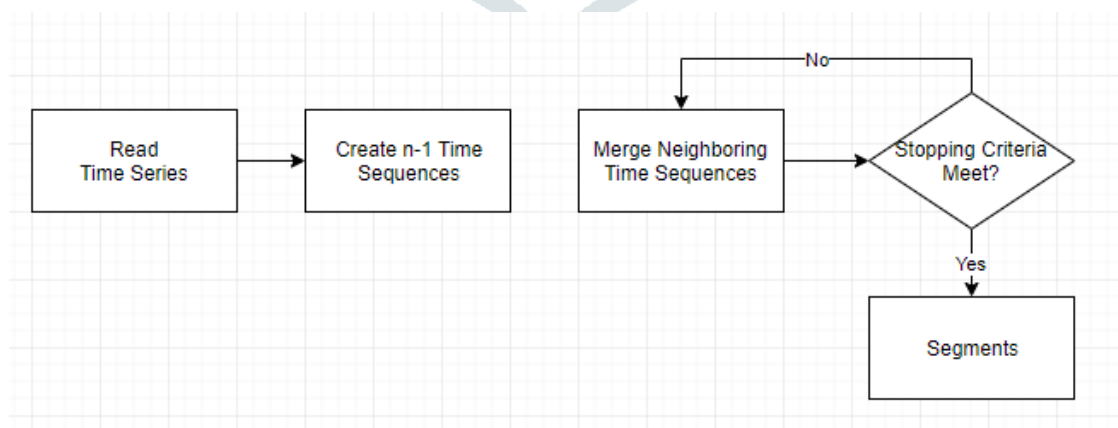


Fig.2 Flowchart Bottom-up Approach

c) Sliding Window Approach-

The sliding window approach is based on window size and distance measure. It is also known as a brute force or a one-pass algorithm. It begins by determining the first data point of a time series, which is the starting point for the window which slides in the right direction along with time series, and in that selection of segments that satisfy the user-specified threshold.

d) Sliding Window And Bottom-up Approach-

This is a combination of the sliding window and the bottom-up approach.

e) Perceptually Important Points-

It iteratively goes on extracting important points from time series data.

II) Horizontal Segmentation-

It is mainly used to extract similar time series from a given set of series.

Following are the techniques used for horizontal segmentation-

- a) Manual Segmentation
- b) Time Series Clustering
- c) Hierarchical Segmentation

a) Manual Segmentation-

Based on the user's interaction time series are divided into a set of segments. It uses the following approach for dividing time series.

- i) Graphical Query Approach- Using this user can create a segment by selecting the desired time series from a graph. For selecting the desired time series, user draws a rectangle bar using a mouse pointer around the region of interest of time series.
- ii) Hierarchical Query Approach- Using a hierarchical query user can select a leaf or level in the hierarchy. Based on the selection of the hierarchy, time series will segment.
- iii) Parameter Query Approach- Using this, based on the input level dataset or descriptive statistics parameters user able to divide time series into the segment.

b) Time Series Clustering-

It uses clustering principles for making clusters of time series. Time series clustering approach is based on-

i) Whole Time Series Clustering

It is method grouping of multiple time series based on different distance measures. Euclidean distance, dynamic time warping (DTW), minkowski distance, longest common subsequences (LCSS) are some of the distance measures that can be used for this technique. This method can be divided into the following functionality- Distance Measures Summarization and at last Clustering.

ii) Sub-sequence Time Series Clustering

This method is used to analyze similar patterns within time series on further these patterns can be used to find similar shapes in other time sequences.

iii) Time Point Clustering

This method goes on clustering points at the time point further these clusters are considered for grouping of time series.

iv) K-means Clustering- It uses Euclidean distance measure for clustering time series data. It considers series data as multivariate variables. These variables go under k-means clustering and form time series clusters.

c) Hierarchical Time Series Clustering-

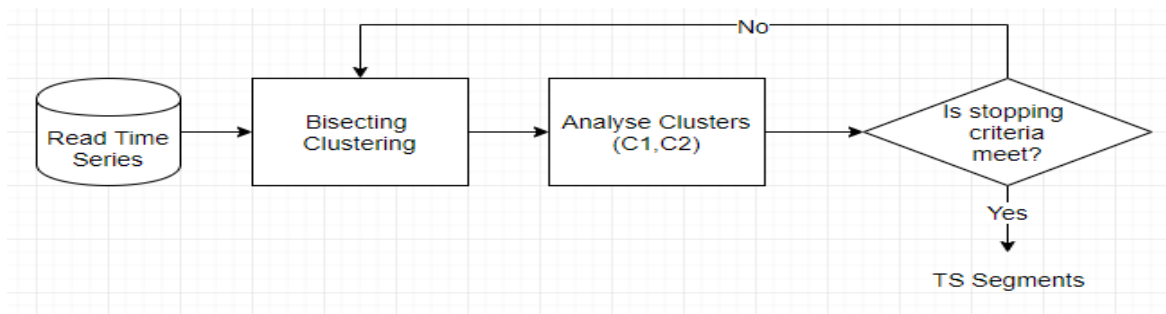


Fig.3 Hierarchical Time Series Clustering

In this system iteratively goes on splitting time series data until stopping criteria met. Time series will iteratively go on dividing into two sets. Every time a cluster which has a maximum sum square error will undergo for clustering. Stopping criteria for hierarchical clustering can be sum square error. The final set of segments is the set of similar time series.

IV. RESULTS AND ANALYSIS

Results are carried on Dow_Jones Stock Market dataset from UCI repository. Fig.4 shows time series representation where there are total 30-time series.

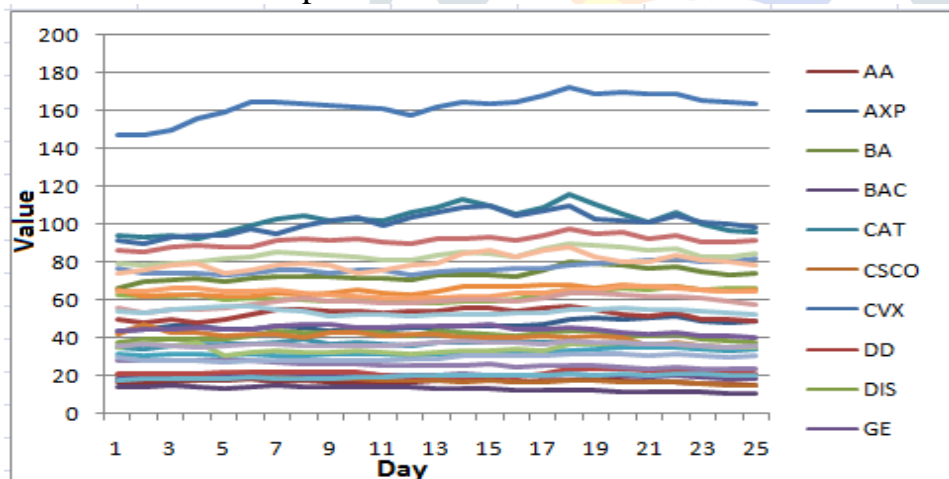


Fig.4 Time Series Representation

Fig.5 shows the results of segmentation using tsclust package using R programming. The main function of this package is to perform time series clustering. Here we have to use the hierarchical clustering approach with Euclidean distance.

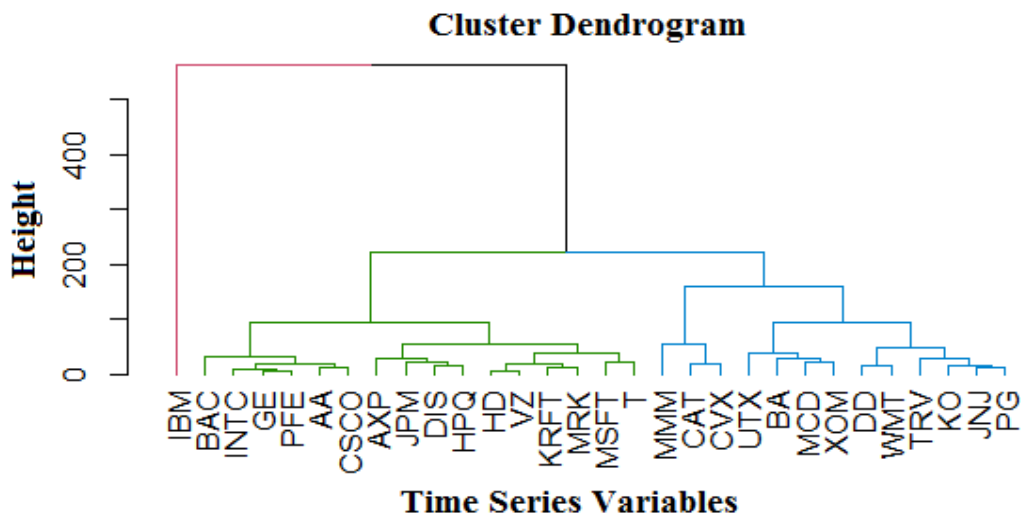


Fig.5 Time Series Segmentation

V. APPLICATIONS

1. Vertical Time Series Segmentation-
 - a. As it is used for extracting patterns within a single time series, which is helpful to identify seasonal and cyclic trends.
 - b. This can be used for feature extraction within time series also it can be used reducing time series into a fixed number of points.
2. Horizontal Time Series Segmentation-
 - a. It can be used to extract energy consumption patterns. This would be helpful to identify similar patterns within some period.
 - b. In finance finding similar TS can be used for market analysis. This would help decide which products should be kept for sale.
 - c. It also can be used in stock market analysis. This would be helpful for an investor to invest his valuable money proper company.
 - d. With the help of correlation, analysis user can identify similar time series i.e. it can be used to extract similar moving time series with the help of seasonal, cyclic and trends (upward or downward).

VI. CONCLUSION

In this paper, we analyze different approaches for time series segmentation based on vertical and horizontal segmentation. Vertical segmentation mainly applies to singular time series. Vertical segmentation can be used to extract patterns like trend, cyclic, stationary, etc. Vertical segmentation can be used for dimensionality reduction. Dimensionality reduction technique is used to reduce data points from the original time series. Horizontal segmentation is used to extract a similar time series from multiple time series data. Also using horizontal segmentation user can able to select requires set of time series.

VII. REFERENCES

- [1] Horst Bunke, Abraham Kandel, Mark Last, "Data Mining in Time Series Database", Book: Series in Machine Perception and Artificial Intelligence, Volume 57, 2004.
- [2] Miodrag Lovric, Marina Milanovic, Milan Stamenkovic, "An Algorithmic Method for Segmentation of Time Series: An Overview", JCEBI, Vol.1, pp. 31 - 53, 2014.
- [3] Tak-chung Fu, "A Review on Time Series Data Mining", Elsevier: Eng. Appl. Artif. Intell., 24 (1), pp. 164-181, 2011.
- [4] David F. Groebner, Patrick W. Shannon, Phillip C. Fry, Kent D. Smith, "Time Series Analysis and Forecasting", Business Statistics: A Decision-Making Approach 6e, chapter:15, 2005.
- [5] "SAS® Time Series Studio 14.1: User's Guide".
- [6] Saeed Aghabozorgi, Ali Seyed Shirkhorshidi, Teh Ying Wah, "Time Series Clustering: A Decade Review", Elsevier: Information Systems, Volume 53, Pages 16-38, 2015.
- [7] Pedro Pereira Rodrigues, Joao Gama, Joao Pedroso, "Hierarchical Clustering of Time-Series Data Streams", IEEE Transactions on Knowledge and Data Engineering, Volume: 20, Issue: 5, 2008.
- [8] Dimitrios Kotsakos, Goce Trajcevski et.al, "Time Series Data Clustering", Data Clustering: Algorithms and Applications, Chapter:15, 2013.
- [9] "TSrepr- Time Series Representation in R",
Website- <https://petolau.github.io/TSrepr-time-series-representations/>
- [10] Taiyeong Lee, Ruiwen Zhang, Yongqiao Xiao, and Jared Dean (SAS Institute Inc), "Feature Extraction Methods for Time Series Data in SAS R Enterprise Miner™"
- [11] "Time Series Clustering",
Website- <https://www.rdocumentation.org/packages/dtwclust/versions/3.1.1/topics/Tsclust>
- [12] "The SIMILARITY Procedure",
Website- http://support.sas.com/documentation/cdl/en/etsug/63939/HTML/default/viewer.htm#etsug_similarity_sect001.htm
- [13] "TSLEARN Clustering",
https://tslearn.readthedocs.io/en/latest/gen_modules/clustering/tslearn.clustering.TimeSeriesKMeans.html