Time Series Clustering and Model fitting to Time Series Clusters

Prasanta Panda
Tata Consultancy Services (TCS), Bhubaneswar, Odisha, India

***Abstract*:-In Time series forecasting, model building is the most time consuming process. Model building for all-time series is a difficult proposition in terms of time consumption when we have to deal with high volumes of time series. So, clustering the time series can help grouping similar time series in together. There are many available algorithms for creating time series clusters. We have used DTW algorithm to get DTW distance matrix and this distance matrix is used to create hierarchical clusters [4] of Time series. DTW algorithm finds the minimal distance between pairs of sequences by allowing flexible shift in alignment to compute pair wise DTW distances.**

Once time series clustering is done we need to fit the model to Time Series clusters. In this paper, a novel algorithm is proposed to fit different models to Time Series clusters. Different Time Series models are evaluated for each cluster by using this proposed algorithm. This ultimately reduces model building time and hence overall forecasting time.

***Keywords:* Time Series clustering, Model fitting to Time Series clusters, DTW distance matrix.**

I. INTRODUCTION

In business, companies seek ways to gain an edge over competitors through marketing strategies, whether they offer a product or a service. Forecasting the demand is crucial to any supplier, manufacturer, or retailer in the supply chain management system. Forecast helps determine the amount of inventory to be kept in hand, how much raw material should be purchased, and how many products should be made [1]. Also, a company can alter its business and marketing strategies to satisfy the expected demands through forecasting for future periods. For example, by monitoring consumer demand at specific prices, a business can stock items that sell well and scale back on items with poor sales. The company can also use this information to make adjustments to its pricing strategy, focusing on higher margin items or products that are in high demand. By following demand closely and making forecasts, the business gains an advantage over competitors who fail to identify a shift or change in demand. Inaccurate forecasts can lead to costly inventory buildup or stockouts. Both of these events are harmful in a business world where customer service is of the utmost importance. During this process few industries have to deal with huge volume of TS. Model building: the mandatory step, for Time Series (TS) forecasting takes a longer time when we have to deal with a large volume of TS [2].

Time series hierarchical clustering is done by taking DTW distance matrix as input. DTW distance matrix of Time series is calculated using DTW algorithm.

Once, time series clusters are done, in this paper we propose a novel method/algorithmto fit Time Series models for different Time Series clusters. Recognizing the least dissimilarity time series of a cluster, model is fitted and same model continues for the rest of the time series cluster unless user defined tolerance, ICED limit is not violated.

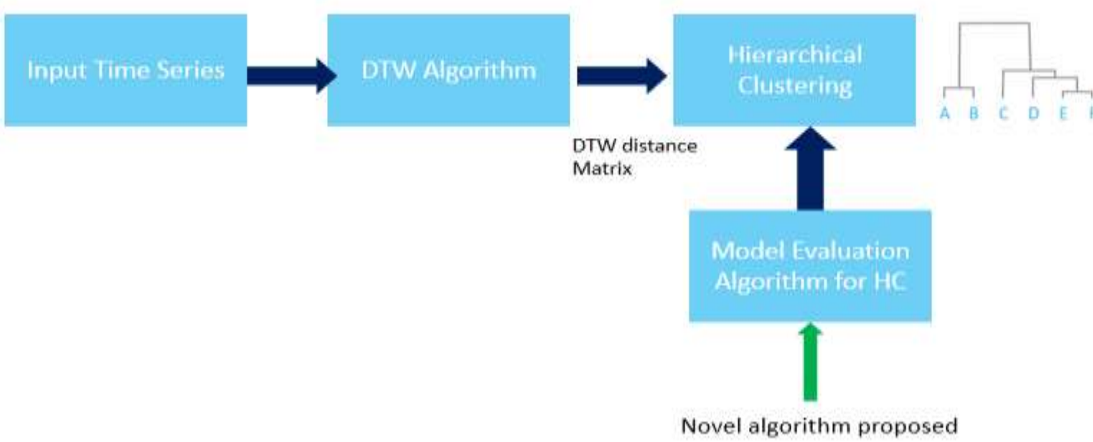


Fig-1 Time series clustering and proposed method for model fitting

International Journal of Latest Technology in Engineering, Management & Applied Science (IJLTEMAS)
Volume VIII, Issue I, January 2019 | ISSN 2278-2540

II. HIERARCHICAL CLUSTERING APPROACH FOR TIME SERIES ANALYSIS

DTW distance matrix is used as input to create Time Series hierarchical clusters. Hierarchical clusters arrange TS (Time Series) respect to its similarities index. It is an agglomerative (top down) clustering method. This builds a hierarchy of clusters, showing relations between the individual members and merging clusters of data based on similarity. In visualizing the result, a dendrogram is generated from the clustering process, representing the nested grouping of patterns and similarity levels at which groupings change [5].

2.1 Time Series (TS) Clustering

Time series clustering is to partition time series data into different groups based on similarity or distance, so that time series in the same cluster are more similar [3]. One key component in TS clustering is the function used to measure the similarity between two data being compared. These data could be in various forms including raw values of equal or unequal length, vectors of feature-value pairs, transition matrices, and so on. But DTW distance gives optimal similarity measure among the time series and hence used to create hierarchical Time Series clusters [4].

Below (Figure-2) shows data set for which Time series is prepared and then DTW distance matrix is calculated using DTW algorithm.

1	99.8	109.4	116.6	116.2	100.7	103.8	104.5	115.1	111.1	107.7	112.1	108.9	94.6	118.5
2	84.7	81.9	97.4	97.4	89.4	80.8	76.7	90.6	90.3	81.7	81.7	73.9	73.2	81
3	79.8	77.5	91.9	85.6	89.8	85.2	79.4	89.7	85.4	77.2	70.8	64.9	67.3	75.5
4	97.2	88.3	108.5	100	53.1	45.7	51.7	74.7	108.4	96.2	93.1	97.5	91.3	100.4
5	131.3	146.2	204.2	170.4	167.1	114.5	93.7	113.5	142.8	104.7	148.5	131.4	111	136.5
6	104.1	116.6	120.4	118.5	100.2	103	106.7	117.9	115.6	113.6	120.4	117.8	99.5	125.7
7	96.2	94.9	100.1	98.9	92.9	93.2	87.1	94.8	97.9	95.7	97.1	98	93.9	97.7
8	114.2	91.9	89.6	94.5	110.3	77.7	78.3	96.2	107.5	110.2	119.4	121.6	105.9	100.7
9	89.2	109.5	111.2	106.9	82.6	103.8	105.9	115.8	114.7	114.6	116.8	108.7	85.4	112.6
10	74.5	84.3	91.1	100.2	80.6	88.1	82.1	74.1	81.3	73.2	85.1	80.7	71.8	84
11	81.8	99	92.4	98.2	69.3	96.1	96.6	113.6	97.4	97	105.8	109	68.4	116.9
12	98.3	106.9	109	104.8	81	84.6	87.3	99.1	103.9	98	97.2	96.4	91.3	107.3
13	102.4	106.8	107.8	108.5	85.3	112.3	103.7	106.6	110.6	110.2	109.1	113.3	106.9	115.1
14	94	104.4	114.2	111	93	78.6	84.2	92.7	97	89.8	86.5	79	87	103.6
15	101.3	75.3	72.3	94.1	92.6	89.5	84.2	94.1	96.3	103.7	97.7	110.1	107.9	95.9
16	88.5	95.1	104.5	104.8	89.5	103.7	106.1	109.3	108.4	105.4	109.1	109.3	99.7	111.9
17	83.4	112.9	112.4	110.5	93.8	95.5	116.4	110.2	104.6	87.1	104.4	88.7	81	104.3
18	104.4	119.2	122.4	105.7	81.6	106.1	112.4	117.1	117.8	119.6	123.6	122	99.1	125.6
19	109.6	114.3	122.3	105.6	68.3	75.9	82.7	100.3	106.5	110.2	108.8	113.4	103.3	116.8
20	87.1	99.8	101.7	101.9	70	104.9	95.8	105.4	99.5	99.3	103.8	98.4	84.9	102.9
21	104.8	118.6	125.5	117.9	92.4	104.7	103.9	118.3	115.2	113.5	120.1	116	104.4	125
22	82.4	102.7	103.9	94.6	98.4	91.9	107.3	113	101.8	97.1	112.9	101.4	92.9	120.2
23	117.5	142.2	137.6	130.1	108.5	117.7	122.4	133.6	130.2	125.8	139	134.1	115.9	149.4
24	103.6	115.2	125.7	129	122.9	98.3	104.9	125.1	120	123.9	124.9	119.2	98.8	124.2
25	112.7	134.2	133.3	142.5	100.3	130.2	135.3	144.7	136.8	127.8	141.3	141.5	97.7	149.9
26	121.4	129	129.6	157.8	191.4	91.4	117	137.7	131.8	133.6	154.8	123.2	121.4	149.2
27	101.8	112	116.6	118.2	99.6	97.9	99.8	112.8	105.6	103.8	111.6	107.7	104.8	120.1
28	111.8	129.3	132.1	131.3	105.1	109.9	104.9	118.8	121.2	119.3	125.3	118	105	135.1
29	106.2	113.3	128.3	117.3	167.1	68.1	82.1	99	91.1	99.4	106.8	110.8	96.3	137.1
30	81.2	79.3	102.2	109.5	107.1	115.2	101.5	107.6	93.8	84.2	76.9	71.7	75.1	91.1
31	81.2	79.3	102.2	109.5	107.1	115.2	101.5	107.6	93.8	84.2	76.9	71.7	75.1	91.1
32	108.5	124.7	129	133.4	119.2	108.4	116	130	123	120.9	131.1	127.7	100	138.8
33	99.5	114.1	118.6	119.3	95.9	108.2	104.9	114.8	109.8	105.1	110.9	100.6	100.2	119.1
34	102.9	115.5	119.4	111.9	85.6	102.1	104.1	114.2	114.6	113.1	117.7	116.3	101	121.9
35	97.2	99.8	103.3	102.4	92.9	88.3	88.6	97.1	100.1	96.9	103.2	98.7	92.9	103.5
36	81.7	79.2	99.7	104.6	103.6	108.9	97	104	92.5	83.7	76.5	71.7	74.7	88.4
37	81.7	0	99.7	104.6	0	108.9	97	104	0	83.7	0	0	74.7	88.4
38	81.7	0	0	104.6	0	0	97	104	0	83.7	0	0	74.7	88.4
39	121.4	0	0	0	191.4	91.4	117	137.7	0	0	0	123.2	121.4	149.2

Fig 2 Time series for which clustering is done

2.2 Why Dynamic Time warping (DTW)

(DTW) is an algorithm for measuring similarity between two temporal sequences which may vary in time or speed. For instance, similarities in walking patterns can be detected using DTW, even if one person was walking faster than the other, or if there is any accelerations and deceleration during the course of an observation. DTW allows for non-linear alignments between time series not necessarily of the same length [7].In general, DTW is a method that calculates an optimal match between two given (time-dependent) sequences under certain restrictions. Intuitively, the sequences are warped in a nonlinear fashion to match each other. Here, DTW has been applied to automatically cope with time deformations and different speeds associated with time-dependent data.Given two time series, Q=q1, q2, . . . , qi, . . . , qn and R =r1, r2, . . . , rj , . . . , rm, DTW aligns the two series so that their difference is minimized. To this end, an n × m matrix where the (i, j) element of the matrix contains the distance d(qi, rj) between two points qi , and rj . The Euclidean distance is normally used. A warping path, W =w1,w2, . . . , wk, . . . , wK where max(m, n) less or equal to (K) less or equal to (m + n − 1), is a set of matrix elements that satisfies three constraints: boundary condition, continuity, and monotonically. The boundary condition constraint requires the warping path to start and finish in diagonally opposite corner cells of the matrix. The DTW algorithm computes the time axis stretch which optimally maps one time series onto another ; it outputs the remaining cumulative distance between the two.

That is w1 = (1, 1) and wK = (m, n). The continuity constraint restricts the allowable steps to adjacent cells. The monotonicity constraint forces the points in the warping path to be monotonically spaced in time. The warping path that has the minimum distance between the two series is of interest.

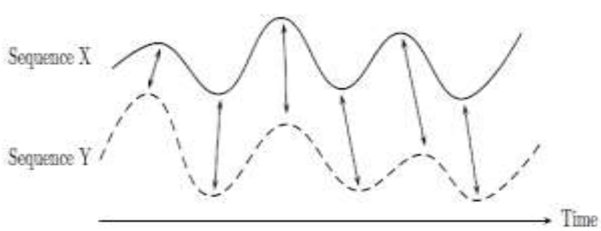


Fig 3 Time alignment of two time-dependent sequences. Aligned points are Indicated by the arrows

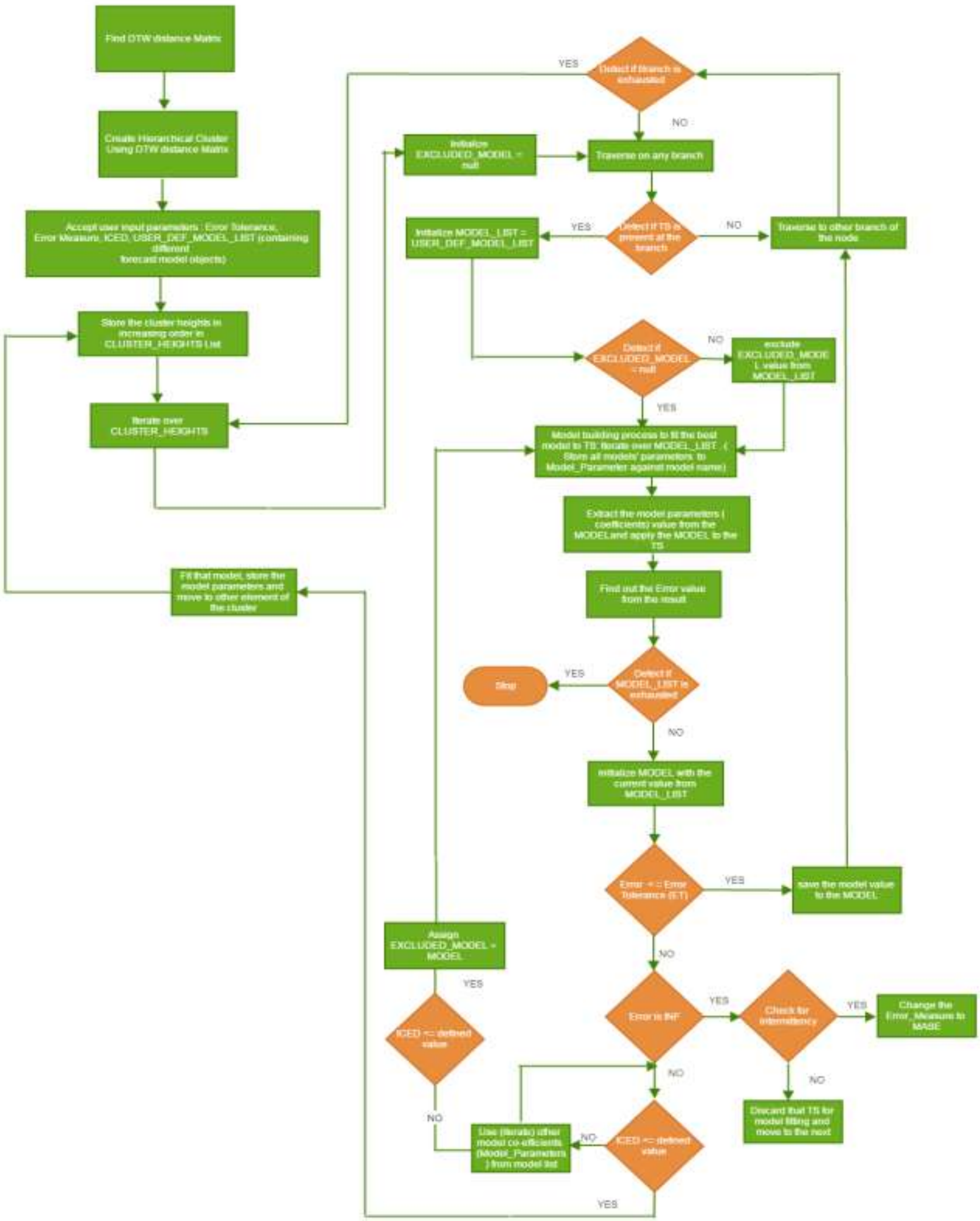
2.3 Observation from the clustering and model fitting

It is observed from figure 4 and 5 that there are many clusters formed using DTW distance. For instance, the 2nd right most cluster (in fig-4) include 6 time series. The time series 32 and 28 (the bottom most series of the cluster) has lowest level of dissimilarity (fig-5 (I)) and as we go up on the cluster the dissimilarity level increases. Algorithm would detect which time series category they belong and as per that the model would be fitted. The starting point is to fit a model to least dissimilar time series of one cluster and use the same model across the cluster until the forecasting error is within the tolerance level and ICED is satisfied Once, the forecasting error exceeds the user defined tolerance, the set of other models are evaluated and the model is picked whose tolerance level is within the defined one.

III. ALGORITHM FOR CLUSTER MODEL FITTING

1. Find the optimal distance matrix of the TS (Time Series) via DTW
2. Input the DTW distance to create Hierarchical clustering.
3. Input from the user is taken (Error Tolerance-ET, Error Measure-EM, Intra Cluster Error Difference-ICED)
4. Find out different TSs at lowest level cluster height: can be got from the output of the hierarchical cluster. (Please refer to figure 4). You can see that TS-28, TS-32, TS23, TS 25, TS26 and TS5 belong to one cluster and however TS-28 and TS-32 are most similar ones.
5. Fit a model to TS-28 and determine the Error. (say the model, in this case is is Holt Winter, Error Measure which has been selected by user is MAPE and the Error is found out)
6. If Error <- Error Tolerance (ET), fit that model to TS-32 also. If not select other model of your model list and ensure the above criteria is fulfilled.
7. If Error is not within ET for TS-32 (say Error is 30 for TS- 28 and Error is 31 for TS-32. But user given tolerance is 30. Then find out the difference of the error measure (or percentage difference of error - say intra cluster error difference-ICED) between TS 28 and TS 32.
8. If ICED is within the permissible limit (this limit is given by the user), use the same model's parameter (TS 28) to TS 30
9. If not, iterate other models' parameter (of your current model list)
10. If not, try to evaluate all models for TS32 except HW.

Below is the Figure showing the algorithm for model fitting to TS cluster.


Find DTW distance Matrix
Create Hierarchical Cluster Using DTW distance Matrix
Accept user input parameters : Error Tolerance, Error Measure, ICED, USER_DEF_MODEL_LIST (containing different forecast model objects)
Store the cluster heights in increasing order in CLUSTER_HEIGHTS List
Iterate over CLUSTER_HEIGHTS
Fit that model, store the model parameters and move to other element of the cluster
Detect if branch is exhausted
Initialize EXCLUDED_MODEL = null
Traverse on any branch
Detect if TS is present at the branch
Initialize MODEL_LIST = USER_DEF_MODEL_LIST
Traverse to other branch of the node
Detect if EXCLUDED_MODEL = null
exclude EXCLUDED_MODEL value from MODEL_LIST
Model building process to fit the best model to TS: Iterate over MODEL_LIST . (Store all models' parameters to Model_Parameter against model name)
Extract the model parameters (coefficients) value from the MODEL and apply the MODEL to the TS
Find out the Error value from the result
Detect if MODEL_LIST is exhausted
Stop
Initialize MODEL with the current value from MODEL_LIST
Error <= Error Tolerance (ET)
save the model value to the MODEL
Assign EXCLUDED_MODEL = MODEL
ICED <= defined value
Error is INF
Check for intermittency
Change the Error_Measure to MASE
Discard that TS for model fitting and move to the next
Use (iterate) other model co-efficients (Model_Parameters) from model list
ICED <= defined value

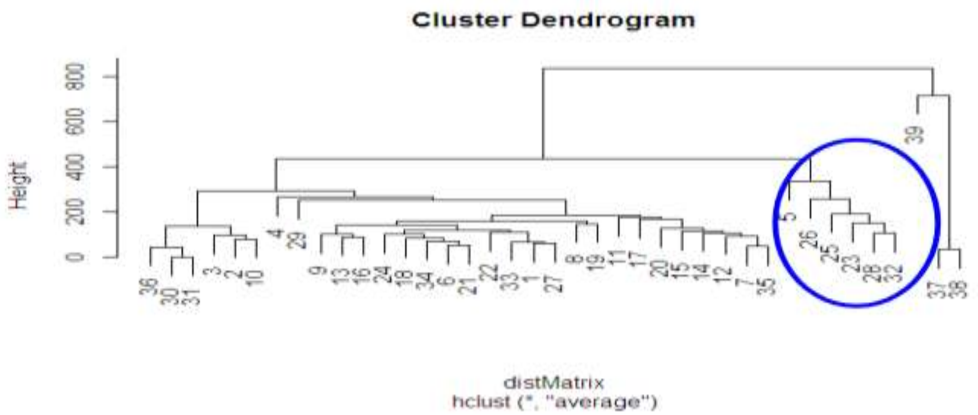


Fig 4 Dendogram for the time series

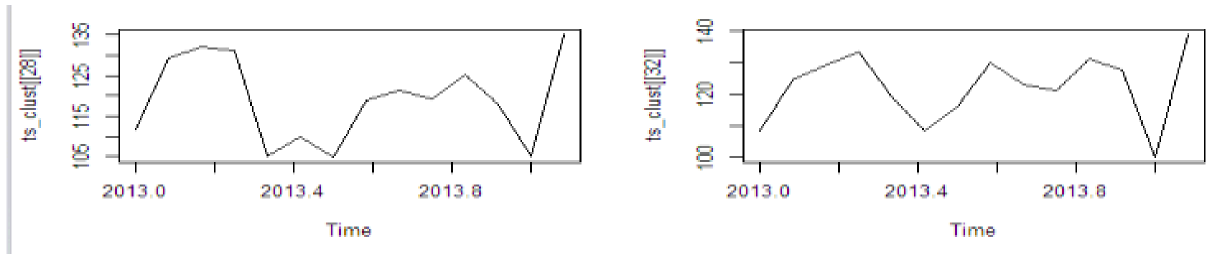


Fig-5 (i) Plot for time series(28 and 32)

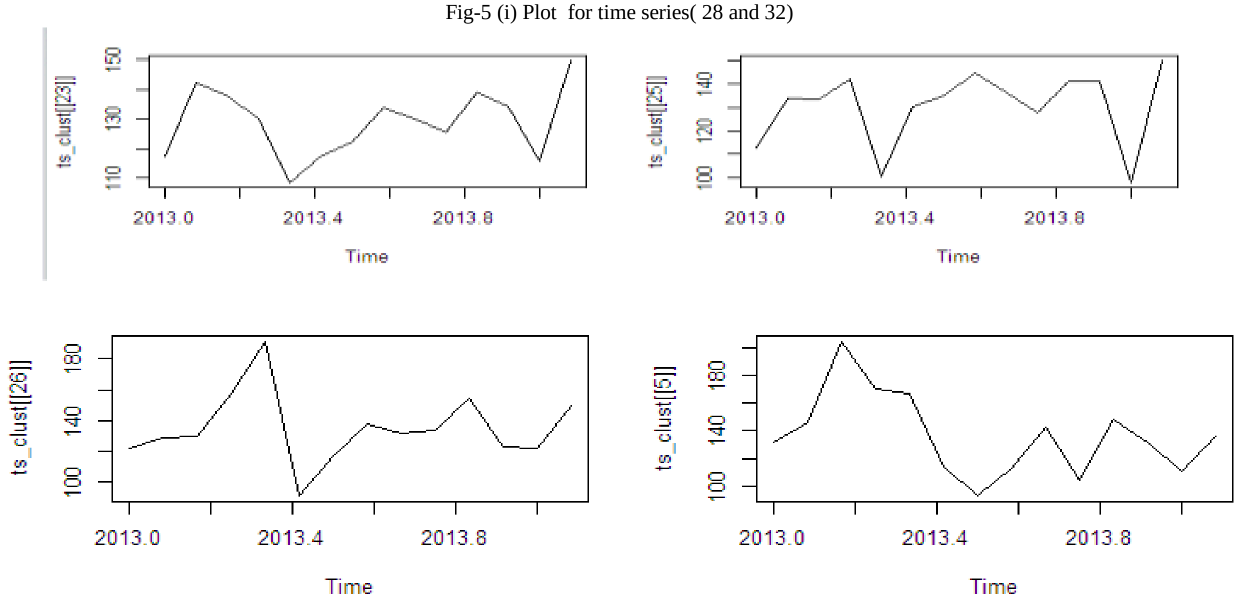


Fig-5 (ii) Plot for time series(23,25,26 and 5)

IV. TIME OPTIMIZATION AFTER MODEL FITTING TO TS CLUSTER

Below are the optimizing parameters on which we can do time optimization.

No of clusters remaining constant: C

Each cluster has t Time Series

Model repository has K TS models

Average time to fit one model for one TS =T_{avm}

4.1 Assumption:

Average time to fit one model for one TS =T_{avm} =0.8min

For each cluster, x % of TSs require model fitting x =0.4

4.2 Overall time taken for model fitting in traditional way

The total time to fit the model to all time series = No of TS * average time to fit one model* No of models = 100*0.8*5 =400min

$$T_{Old} = \sum_{j=1}^{j=p}(C \times t_j) \times \sum_{i=1}^{i=n}(T_{avm} \times K_i) \text{ ---(1)}$$

4.3 Overall time taken for model fitting in new approach

For one cluster, for first TS (TS at lowest level) all models are evaluated.

$$T_{New} = \sum_{j=1}^{j=p}(x/100) \times (C \times t_j) \times \sum_{i=1}^{i=n}(T_{avm} \times K_i) \text{ --(2)}$$

Saving in the time compared to traditional approach in percentage = (T_{old}– T_{new})/ T_{old}* 100 – Time for creating clusters (using DTW distance matrix) = 60 – (5 to 10) = **50 to 55%**

Usually 5-10% of the time (in terms of T_{avm})

V. CONCLUSION

Time series analysis and forecasting is done at different level of supply chain management system. With the historical data, model is fitted and using that model prediction is done for future period. When the time series increase, this model building activities takes a longer time. Instead of creating model for individual time series, clustering of time series is performed. Optimal distance (DTW distance matrix) is found out by using DTW algorithm. Hierarchical clustering is done taking the DTW distance matrix into consideration. Our algorithm finds which model can be a fit (within the tolerance level) to a cluster. The lowest dissimilar time series (TS at the lowest cluster height of one cluster) is the starting point to fit a model to a cluster. Once model is decided w.r.t the least dissimilar time series, same model is evaluated for the other TSs of the cluster unless the defined tolerance is not exceeded. In case, the forecasting error exceeds the defined tolerance value, the next level of model is tried out. The whole process saves time compared to individual model fitting to volume of time series.

REFERENCES

[1]. Time Series Analysis, by James Douglas Hamilton, Princeton University Press- 1994, Chapter-4, pp-72-115
[2]. Time series clustering and classification by the autoregressive metric, June 2007, Elsevier Publication,by Marcella Corduas, Domenico Piccolo
[3]. Characteristic-Based Clustering for Time Series Data, Data Mining and Knowledge Discovery, November 2006, Volume 13, Issue 3, pp 335–364, by Xiaozhe Wang, Kate Smith, Rob Hyndman
[4]. The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances, Data Mining and Knowledge Discovery, May 2017, Volume 31, Issue 3, pp 606–660, by Anthony Bagnal,Jason Lines,AaronBostrom, James Large, Eamonn Keogh
[5]. Clustering of time series data—a survey, Pattern Recognition, Volume 38, Issue 11, Pages 1857-1874 , May 2005, T. Warren Liao
[6]. https://cran.r-project.org/web/packages/dtw/dtw.pdf
[7]. https://en.wikipedia.org/wiki/Dynamic_time_warping

ABBREVIATIONS AND DEFINITIONS

1. ET- Error Tolerance: is the user defined value beyond which user cannot accept the error.
2. EM-Error Measure: is the error definition: MAPE, MASE, MAE etc
3. ICED-Intra Cluster Error Difference: The difference of the error value between two time series at the time of model building within a cluster.
4. Cluster: Group of Time Series arranged within on the basis of their similarity distance.
5. Cluster Height: The height on the basis of which hierarchical cluster is done.
6. Model List: the list of Models (like Holt Winter, ARIMA etc) which are going to be used for model fitting to TS.