



Fuzzified Crisp Relative Cure Hierarchical Clustering for Temporal Relational Data

L. Jaya Singh Dhas, B. Mukunthan

Abstract - A bitemporal data clustering is a significant solution to the diverse problems for finding the intrinsic structure and compact information over temporal data. The temporal data are collected in the series of particular time periods. The various data mining methods have been developed in the temporal relational data analysis. But the accurate analysis was not performed with minimum time. An efficient technique called Fuzzy Crisp Relative Spherical CURE Hierarchy Clustering (FCRSCHC) is introduced for improving the temporal relational data analysis by partitioning the total dataset into different clusters with minimum time as well as space complexity. The CURE hierarchical structure takes the number of scattered temporal data points in the spherical surface for the clustering. After that, 'k' number of clusters and the representative points (i.e. cluster centroid) are initialized. Then the distance between the representative point and the temporal data point are calculated using spherical coordinates. The minimum distance between the data points are grouped into a particular cluster. Then the fuzzy memberships between the two cluster representative points are calculated based on the distance metric. The CURE hierarchical structure merges the two clusters based on the crisp relation between the representative points. Then, the newly obtained clusters are validated using the silhouette coefficient to identify the data points are close to its own cluster or their neighboring clusters. Finally, the optimal numbers of clusters are obtained and minimize the incorrect data clustering which improves the accuracy. The experimental evaluation is performed using a bitemporal dataset with various parameters such as clustering accuracy, false alarm rate, clustering time and space complexity. The results show that FCRSCHC technique improves the clustering accuracy and minimize the time as well as space complexity as compared to the state-of-the-art-works.

Keywords - Bitemporal data analysis, CURE hierarchical clustering, fuzzy membership, crisp relation, silhouette coefficient

I. INTRODUCTION

Bitemporal data analysis is a specific case of temporal database information to handle the sequential data along with two different timelines. Time series analysis is a statistical technique that describes the data in the series of particular time periods for tracking the behavior of a particular phenomenon. Clustering is an essential data mining technique that partitioning the data points into different groups.

Manuscript published on 30 September 2019

* Correspondence Author

L. Jaya Singh Dhas*, Research Scholar, Department of Computer Science, Jairams Arts and Science College, Karur, Tamilnadu, India. <https://orcid.org/0000-0002-0136-3941> Email: jayasinghdhas@yahoo.com

B. Mukunthan, Research Supervisor & Assistant Professor, Department of Computer Science, Jairams Arts and Science College, Karur, Tamilnadu, India. <https://orcid.org/0000-0001-8452-3164> Email: dr.mukunthan.bmk@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

A variety of clustering algorithms has been developed for a large bitemporal dataset with very various attribute counts. Since the dimensionality of temporal data is considerably larger and more complicated. The research work is concentrated to involve the relevant data object clustering algorithm to overcome the computational problems.

A novel density-based clustering algorithm called chronoClust was developed in [1] for grouping a time-series data. The designed clustering algorithm failed to extend the performance of clustering accuracy. A bi-weighted ensemble approach was developed in [2] for grouping the temporal data using a hidden Markov model. The approach failed to minimize the temporal data clustering time as well as space complexity. An Expectation-Maximization algorithm was designed in [3] for grouping the temporal data. But the clustering accuracy was not improved and also the optimal number of clusters was not obtained.

Clustering a temporal network (ClueNet) was introduced in [4] to cluster the data into different groups. The ClueNet provides higher computational complexity while increasing the temporal network data. A hybrid clustering algorithm was developed in [5] to group the time series data based on similarity. The designed algorithm failed to use any split or merge the clustering algorithm for obtaining the optimal number of clusters. A Fuzzy clustering algorithm was designed in [6] for Spatial-Temporal data based on the autocorrelation. The clustering time was not minimized using the designed algorithm. Clustering of time series data was presented in [7] to measure variation between time series data using copula-based distance measure. But the cluster validation was not performed to achieve the higher accuracy.

A stepwise spatial and temporal clustering strategy was developed in [8] based on the similarity between the data. The designed strategy was not minimized the false alarm rate during the clustering process. A spatiotemporal (ST-OPTICS) clustering technique was developed in [9]. The clustering technique minimizes the time complexity but the performance of space complexity remained unaddressed. A Generalized k-means based clustering technique was designed in [10] for temporal data. Though the technique minimizes the time consumption, the clustering accuracy was not minimized.

A dynamic fuzzy cluster (DFC) algorithm was introduced in [11] for dynamically grouping the time series data. But the algorithm failed to perform the cluster validation. Two incremental fuzzy clustering algorithms were developed in [12] based on a Dynamic Time Warping distance. The clustering algorithms failed to minimize the incorrect clustering of time series data. An HMM-based hybrid meta-clustering algorithm was developed in [13]. The algorithm failed to minimize clustering time.

An Evolutionary Clustering based on Graph regularized Nonnegative Matrix Factorization was developed in [14] to analyzing the temporal networks. But the technique failed to solve the time-consuming problems during clustering analysis. A mixed fuzzy clustering (MFC) algorithm was designed in [15] for grouping the time series data.

But the MFC algorithm failed to perform the cluster validation to minimize the false alarm rate. A Correlation-based clustering of big spatiotemporal data was presented in [16] with less memory capacity. The clustering error was not minimized using the correlation based clustering technique. An automatic similarity measure selection was developed in [17] for clustering the time series data. But the accurate clustering results was not provided since it failed to use the distance measures for grouping the time series data into different clusters. A novel clustering technique was designed in [18] for grouping the spatiotemporal sequence data with respect to density. The designed method minimized the time complexity but the algorithm failed to describe how the total dataset was partitioned into smaller ones and the clusters were merged together. A cross clustering algorithm was designed in [19] for grouping the time series data and improving the clustering performance. The algorithm minimizes the time complexity, but the setting of the distance parameter in the cross clustering application was not performed to further improving the clustering accuracy. A unified framework was introduced in [20] for clustering the spatiotemporal data with minimum space and time complexity. The framework failed to validate the clustering technique.

The major issues are identified from the above said literature are overcome by introducing a novel clustering technique called FCRSCHC. The proposal contribution is summarized in below subsection on the contrary to existing techniques.

1.1 Proposal contribution

The major limitations of the proposed FCRSCHC technique on the contrary to the existing literature are summarized as follows,

- The FCRSCHC technique is designed to improve the clustering accuracy of temporal relation data and minimize the time complexity using CURE hierarchical clustering algorithm. The algorithm groups similar scattered data points in the spherical surface through the distance measure. The CURE hierarchical structure merges the two clusters by measuring the crisp relation between the two representative points. The crisp relation is defined by the fuzzy membership function.
- To minimize the false alarm rate in the clustering process, the silhouette coefficient based cluster validation is performed. The coefficient effectively measures how each data point is similar in one cluster and the neighboring clusters and thus provides a way to obtain an optimal number of clusters. This helps to correctly group all the similar data points into the particular clusters resulting in minimizes the space complexity.

1.2 Structure of the paper

This paper is organized into five different sections in the following manner. Section 2 provides a detailed description of the clustering with neat diagram. Section 3 illustrates the experimental evaluation using a temporal dataset with different parameters settings. Followed by, the statistical results analysis is presented under the different parameters using various techniques in section 4. Finally, section 5 concludes the paper.

II. METHODOLOGY

The bitemporal time series datasets comprise a large volume of data and it leads to higher computational vulnerability. In order to solve the issues in the bitemporal data analysis, Fuzzy Crisp Relative Spherical CURE Hierarchy Clustering (FCRSCHC) technique is developed. The time series data clustering is performed by using the following system model. Given the bitemporal datasets D_b , the 'n' number of time series data points $DP_1, DP_2, DP_3, \dots, DP_n$ collected at any time 't'. The total dataset D_b partitioned into the number of clusters i.e. $C_1, C_2, C_3, \dots, C_m$ where $C_i \neq C_j$. The clusters are formed by grouping the similar time series data together depends on the distance similarity. The Hierarchy Clustering technique merges the two clusters based on the system similarity. Based on the system model, the FCRSCHC technique is designed.

2.1 Fuzzy Crisp Relative Spherical CURE Hierarchy clustering method

A Fuzzy Crisp Relative Spherical CURE Hierarchy clustering (FCRSCHC) method is developed for partitioning the temporal data objects into different groups. The clustering is the data mining model helps to minimize the similarity between the clusters and maximize the similarity within the cluster. The similar temporal data points are assigned into one cluster whereas the different data points are grouped into other clusters. The conventional clustering algorithm did not work well with clusters of different size and different density. Therefore, the CURE (Clustering Using REpresentatives) is a hierarchical clustering technique used for partitioning the large databases into different subsets. The block diagram of the FCRSCHC method is shown in figure 1.

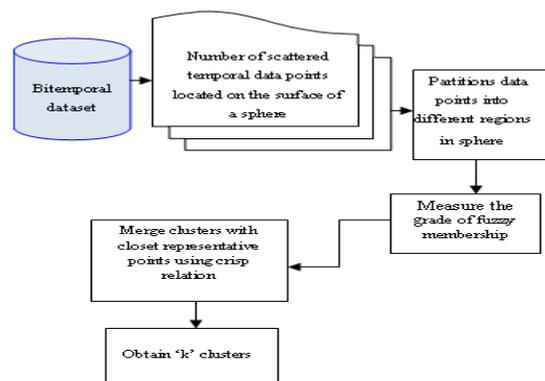


Figure 1 Block diagram of the FCRSCHC method

As shown in figure 1, a block diagram of the FCRSCHC method is illustrated to obtain the optimal 'k' clusters. The bitemporal data points are collected from the dataset. Initially, numbers of scattered temporal data points are located on the surface of a sphere for clustering. CURE hierarchical structure identifies the well scattered bitemporal data points $DP_1, DP_2, DP_3, \dots, DP_n$ in the sphere 'S'. The total sphere (S) is partitioned into different regions (i.e. clusters). By applying a FCRSCHC method, initially 'k' number of clusters and representative points (i.e. centroid) are chosen randomly in surface of the sphere. Each cluster has one representative point. After that, the scattered data points are assigned to the cluster through the distance measure. The distance between the data points in the sphere and the cluster representative points are calculated using the spherical coordinates (r, θ, φ) .

Let us consider the data point DP_i in the spherical coordinates is $(\theta_{D_i}, \varphi_{D_i})$ and the representative point R_j of the cluster in the spherical coordinates denoted as $(\theta_{R_j}, \varphi_{R_j})$. Therefore the given two spherical coordinates, the distance between them is mathematically calculated using the below equation,

$$d(DP_i, R_j) = \text{arc cos}(\sin \theta_{D_i} \cos \varphi_{D_i} \sin \theta_{R_j} \cos \varphi_{R_j} + \sin \theta_{D_i} \sin \varphi_{D_i} \sin \theta_{R_j} \sin \varphi_{R_j} + \cos \theta_{D_i} \cos \theta_{R_j}) \quad (1)$$

In (1), $d(DP_i, R_j)$ denotes a distance between the data point and the representative point of the cluster. In the above equation (1), the coordinates (θ, φ) are mathematically calculated as follows,

$$\theta = \text{arc cos}\left(\frac{x}{\sqrt{x^2+y^2}}\right) \quad (2)$$

$$\varphi = \text{arc cos } z \quad (3)$$

In (2) (3), x, y, z denotes a Cartesian coordinates. Based on the distance measure, the FCRSCHC technique groups the data points into the particular cluster with minimum distance between them. In this way, the numbers of regions are formed on the sphere surface with similar data points over the time periods. After partitioning the data points on the sphere surface, the nearest clusters are merged for minimizing the dimensionality of the space which estimates the level of similarity of data points. The interconnections between the two clusters are identified by the fuzzy membership function. The fuzzy membership function finds the neighboring representative points for merging the two clustering set. The membership is computed by the following equation,

$$U_{ij} = \left(\sum_{j=1}^m \frac{R_{ij}}{R_{ik}}\right)^{-\left(\frac{2}{p-1}\right)} \quad (4)$$

In (4), U_{ij} fuzzy membership function, R_{ij} denotes a distance between i^{th} and j^{th} cluster representative points, R_{ik} denotes a distance between i^{th} and k^{th} cluster representative points, p denotes a fuzziness parameter $p > 1$. The membership grade is generally represented by a real number in the closed interval $[0, 1]$ and it indicates the strength of the relation between the two cluster representative points. If the distance between the two representative points are minimum, then the membership value is higher. Otherwise, the membership value is minimized. Based on the membership value, the nearest representative points are combined to form a one cluster through the crisp relation. The following operation is

performed to find the crisp relation between the cluster representative points.

$$C_1 \cup C_2 = \text{arg max } [U_{ij}] \quad (5)$$

In (5), $C_1 \cup C_2$ represents the union (i.e. combine) of two clustering sets that contains all of the data points. arg max denotes a argument of the maximum function, U_{ij} denotes a membership function. The maximum membership functions between the cluster representative points are merged into one cluster in the spherical surface. After merging the clusters, the new representative point is defined i.e. updated for the newly combined cluster. The newly obtained cluster includes the all data points in the two clusters C_1 and C_2 that were merged. Finally, the well-partitioned data points are shrunken to the cluster centroid. As a result, the hierarchical clustering technique merges the nearest pairs of representative points to form a single cluster which results in minimizes the dimensionality of the cluster in sphere space.

2.2 Silhouette coefficient based cluster validation

After obtaining the 'k' number of clusters, the proposed FCRSCHC technique uses the Silhouette coefficient for validating the data points within the cluster and between the cluster to minimize the incorrect temporal data point clustering. The Silhouette coefficient is mathematically expressed as follows,

$$\rho_s = \frac{(D_n(dp) - D_w(dp))}{\max\{D_w(dp), D_n(dp)\}} \quad (6)$$

In (6), ρ_s denotes a Silhouette coefficient, $D_w(dp)$ denotes an average distance of data points with respect to all the other data points in the cluster it's assigned. $D_n(dp)$ represents the average distance of data points with respect to all the other data points to the neighboring clusters. The Silhouette coefficient provides the value in the range of -1 to +1. The value '+1' indicates that the data point is very close to the cluster it's assigned and far away from its neighboring cluster. The value '-1' indicates that the data point is close to the neighboring cluster than the cluster it's assigned. As a result, the validation effectively improves the clustering accuracy and minimizes the false positive rate. The algorithmic process of the proposed FCRSCHC technique is described as follows,

Input:	Bitemporal dataset D_b , number of time series data points $DP_1, DP_2, DP_3, \dots, DP_n$
Output:	Improves the bitemporal data clustering accuracy
Begin	
1.	Initialize 'k' clusters
2.	for each cluster
3.	Select one representative point R_j
4.	Measure the distance $d(DP_i, R_j)$
5.	Partitions dataset into 'k' clusters with minimum distance
6.	Calculate the degree of membership U_{ij} between the cluster representative points
7.	Find the crisp relation $C_1 \cup C_2$
8.	Merge two clusters based on $\text{arg max } [U_{ij}]$
9.	Update the representative point R_j'
10.	Shrink the data points (DP) to the representative point R_j'
11.	Calculate the Silhouette coefficient ' ρ_s '
12.	if ($\rho_s = +1$) then
13.	Data points are correctly grouped into the cluster
14.	else
15.	Data points are close to the neighboring cluster
16.	end if
17.	end for
18.	Obtain the optimal k' clusters
end	

Algorithm 1 Fuzzy Crisp Relative Spherical CURE Hierarchy Clustering

Algorithm 1 clearly describes the bitemporal data clustering and minimizing the dimensionality as well as the clustering time. Initially, the total dataset is divided into a number of clusters on the spherical surface based on the distance between the data points and cluster representative point. After clustering, the closest representative points in the surface are merged by measuring the crisp relation based on the fuzzy membership grades. Followed by, the new clusters are formed and update the representative points. Finally, the proposed FCRSCHC technique validates the data points within and between the clusters through the Silhouette coefficient. If the coefficient returns +1, then the data points are correctly grouped into the particular cluster. Otherwise, the data points are close to the neighboring cluster. As a result, all the bitemporal data points are grouped into the cluster resulting in improves the clustering accuracy and minimize the false positive rate.

The above-explained processes are implemented in the experimental using time series dataset in the following section.

III. Experimental evaluation and parameter settings

An experimental assessment of proposed FCRSCHC technique and existing methods ChronoClust [1] and Bi-weighted ensemble approach [2] is implemented using Java language with Activity Recognition from Single Chest-Mounted Accelerometer Data Set. The dataset is collected from the UCI machine learning repository. This dataset comprises the temporal data points from a wearable accelerometer fixed on the chest. The characteristics of the dataset are univariate, sequential and time-Series. The attributes characteristics are real and the association tasks performed by the dataset are clustering and classification. The data points are separated by each participant and each file comprises the sequential number, x acceleration, y acceleration, z acceleration, and labels. The labels are used to provide the clustering outcomes which are represented by the seven numbers. The label 1 to 7 represents the human Activity Recognition such as Working at Computer, Standing up Walking and Going updown stairs, Standing, Walking, Going UpDown Stairs, Walking and Talking with Someone and Talking while Standing.

The performance results of FCRSCHC technique and existing methods ChronoClust [1] and Bi-weighted ensemble approach [2] are evaluated with the different parameters such as,

- clustering accuracy
- false alarm rate
- clustering time
- space complexity

IV. Results and Discussions

The experimental evaluation results of proposed FCRSCHC technique and existing methods ChronoClust [1], Bi-weighted ensemble approach [2] are discussed and the results are compared in this section. The different performance metrics are used in this section such as clustering accuracy, false alarm rate, clustering time and space complexity. The obtained results are discussed with the help of graphical representation. For each section, the statistical calculation is

given to show the performance of the proposed technique and conventional techniques.

4.1 Impact of clustering accuracy

Clustering accuracy is referred to the ratio of a number of (no. of) time series data points are correctly clustered to the total number of data points. The mathematical formula for calculating the clustering accuracy is given below,

$$\text{clustering accuracy} = \left[\frac{\text{No. of DP correctly clustered}}{n} \right] * 100 \quad (7)$$

From (7), 'n' represents the number of data points *i.e* DP. The clustering accuracy is measured in percentage (%).

Sample calculation:

- **Proposed FCRSCHC technique:** No. of data points correctly clustered is 880 and the total number of data points is 1000. The clustering accuracy is calculated as,

$$\text{clustering accuracy} = \frac{880}{1000} * 100 = 88\%$$

- **Existing ChronoClust:** No. of data points correctly clustered is 810 and the total number of data points is 1000. The clustering accuracy is calculated as,

$$\text{clustering accuracy} = \frac{810}{1000} * 100 = 81\%$$

- **Existing Bi-weighted ensemble approach:** No. of data points correctly clustered is 780 and the total number of data points is 1000. The clustering accuracy is calculated as,

$$\text{clustering accuracy} = \frac{780}{1000} * 100 = 78\%$$

Case 1: Let us consider, the number of bitemporal data taken as input from the Activity Recognition dataset. The ten runs are considered with various input data points. 1000 data points are taken in the first run to calculate the clustering accuracy. The proposed FCRSCHC technique correctly grouped 880 data points into seven different clusters among the 1000 data points. Then the classification accuracy is 88% using FCRSCHC technique. Whereas, the clustering accuracy of existing ChronoClust [1] and Bi-weighted ensemble approach [2] are 81% and 78% respectively. The statistical analysis shows that the proposed FCRSCHC technique improves the clustering accuracy than the other two existing methods. The various results of clustering accuracy are shown in the following graph.

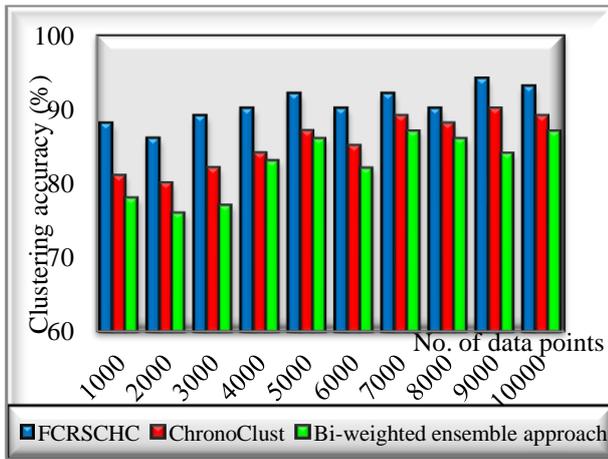


Figure 2 experimental results of clustering accuracy

Figure 2 shows the experimental results of clustering accuracy with respect to the number of temporal data points. The temporal data points are taken in the range of 1000 to 10000. The graphical results confirm that the clustering accuracy of FCRSCHC technique is said to be increased. This improvement is achieved by performing the hierarchical clustering of temporal data points. The proposed FCRSCHC technique initializes the clusters and the representative points in the sphere surface. The minimum distance between each data points and the representative point are grouped into the clusters. Similarly, all data points are grouped into the different clusters. Then the nearest representative points between the clusters are combined. The crisp relation between the representative's points is calculated by the membership grade for merging the two clusters. After merging, the optimal numbers of clusters are obtained. Based on the clustering results, human activities are correctly recognized. As a result, the hierarchical clustering technique correctly groups all the temporal data points into the different clusters.

The observed results of the FCRSCHC technique are compared to the accuracy of the ChronoClust [1] and Bi-weighted ensemble approach [2]. The comparison results confirm that the FCRSCHC technique improves the clustering accuracy by 6% and 10% than the existing methods.

4.2 Impact of false alarm rate

The false alarm rate is measured as the ratio of a number of time series data points are incorrectly clustered to the total number of data points. The mathematical formula for the false alarm rate is given below,

$$FAR = \left(\frac{\text{No. of DPs incorrectly clustered}}{n} \right) * 100 \quad (8)$$

From (8) FAR represents the false alarm rate, DPs denotes a data points, n represents the number of data points. The false alarm rate is measured in the unit of percentage (%).

Sample calculation:

- **Proposed FCRSCHC technique:** No. of data points incorrectly clustered is 120 and the total number of data points is 1000. Then the false alarm rate is calculated as,

$$FAR = \frac{120}{1000} * 100 = 12 \%$$

- **Existing ChronoClust:** No. of data points incorrectly clustered is 190 and the total number of data points is 1000. Then the false alarm rate is calculated as,

$$FAR = \frac{190}{1000} * 100 = 19\%$$

- **Existing Bi-weighted ensemble approach:** No. of data points incorrectly clustered is 220 and the total number of data points is 1000. Then the false alarm rate is calculated as,

$$FAR = \frac{220}{1000} * 100 = 22 \%$$

Figure 3 illustrates the experimental results of false alarm rate versus a number of temporal data points taken in the range from 1000 to 10000. The data point is given to the 'x' axis and the results of the false alarm rate are obtained at the 'y' axis. The above graph shows that the proposed FCRSCHC technique outperforms well and provides accurate clustering results with minimum false alarm rate. The incorrect classification of the FCRSCHC technique is minimized by performing the cluster

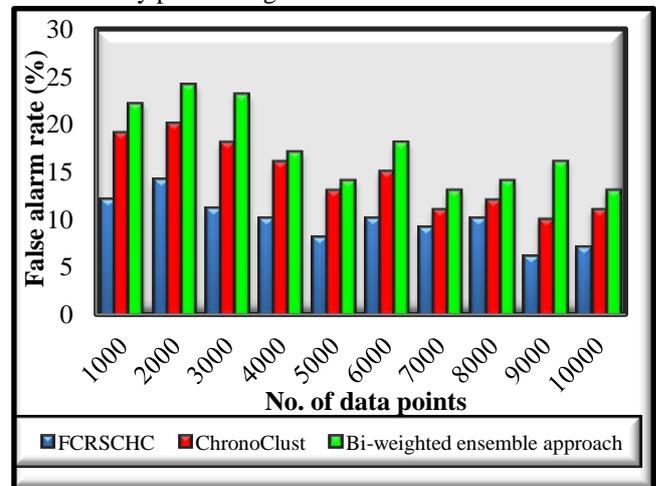


Figure 3 Experimental results of false alarm rate

validation. After obtaining the 'k' clusters, the cluster centroid gets updated for the newly obtained cluster. In this case, the cluster validation is performed through the silhouette coefficient for correctly identifying the cluster members within the clusters to minimize the error rate. The silhouette coefficient measures the average distance of the data points and its cluster center along with their neighboring cluster center. Based on the coefficient values, the cluster members are correctly identified. This helps to minimize the incorrect data point clustering. Let us consider the 1000 data points as input for grouping the temporal data points to recognize human activities. The false alarm rate of FCRSCHC technique is 12% whereas the false alarm rate of ChronoClust [1] and Bi-weighted ensemble approach [2] are 19% and 22% respectively.

Similarly, the nine different results are obtained with various input data points. The observed results of the proposed clustering technique are compared to the existing results. The average is taken for the comparison results and shows that the false alarm rate is considerably minimized by 33% and 44% using FCRSCHC technique as compared to the existing clustering techniques.

4.3 Impact of clustering time

Clustering time is measured as an amount of time taken by the algorithms to group the similar data point into the cluster. The clustering time is mathematically calculated as follows,

$$C_T = \text{No. of data points} * \text{time (grouping single DP)} \quad (9)$$

From (9), C_T indicates the clustering time, DP denotes a data point. The clustering time is measured in the unit of milliseconds (ms).

Sample calculation

- **Proposed FCRSCHC technique:** No. of data points is 1000 and the time taken for grouping single data point is 0.022ms, then the overall clustering time is calculated as follows,

$$C_T = 1000 * 0.022ms = 22ms$$

- **Existing ChronoClust:** No. of data points is 1000 and the time taken for grouping single data point is 0.025ms, then the overall clustering time is calculated as follows,

$$C_T = 1000 * 0.025ms = 25ms$$

- **Existing Bi-weighted ensemble approach:** No. of data points is 1000 and the time taken for grouping single data point is 0.028ms, then the overall clustering time is calculated as follows,

$$C_T = 1000 * 0.028ms = 28ms$$

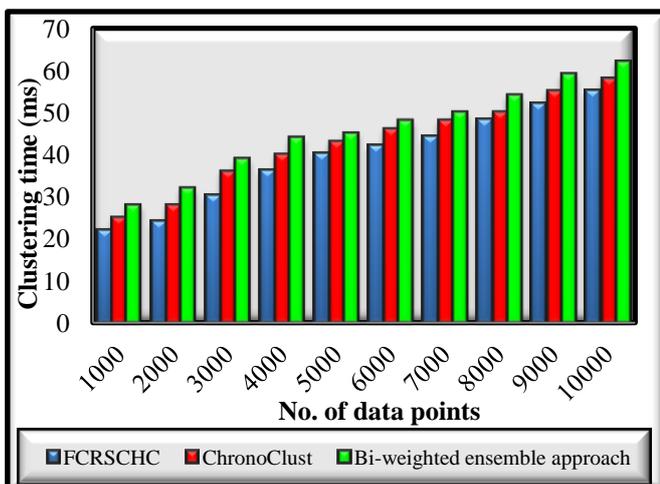


Figure 4 experimental results of clustering time

The experimental results of clustering time versus a number of temporal data points are shown in figure 4. The above figure clearly shows that with the increase in the number of

data points, the clustering time is also found to be in the increasing trend. This is due to the application of Fuzzy Crisp Relative Spherical CURE Hierarchy Clustering along with the cluster validation. By applying this, with the temporal clusters, the data points are partitioned into different groups based on time series. The cure clustering technique accurately groups the temporal data points with minimum time. For example, with 1000 data points, clustering time for single data point is '0.022ms' using FCRSCHC technique, where as the clustering time for data points is '0.025ms' using ChronoClust [1] and clustering time for single data point being '0.028ms' using Bi-weighted ensemble approach [2]. Therefore, the overall clustering time was found to be '22ms', '25ms' and '28ms' using FCRSCHC technique, ChronoClust [1] and Bi-weighted ensemble approach [2] respectively. The above statistical analysis proves that the clustering time is minimized using FCRSCHC technique. The average of ten various results shows that the FCRSCHC technique minimizes the clustering time by 9% when compared to ChronoClust [1] and 16% as compared to Bi-weighted ensemble approach [2].

4.4 Impact of space complexity

Space complexity is measured to the amount of memory space taken by the algorithms to store similar data points into the cluster. The space complexity is mathematically calculated as follows,

$$\text{space complexity} = \text{No. of DP} * \text{memory space (storing single DP)} \quad (10)$$

From (10), DP denotes a data point. The space complexity is measured in terms of mega bytes (MB).

Sample calculation:

- **Proposed FCRSCHC technique:** No. of data points is 1000 and the memory space taken for storing single data point is 0.01MB, then the overall space complexity is calculated as follows,

$$\text{Space complexity} = 1000 * 0.01MB = 10MB$$

- **Existing ChronoClust:** No. of data points is 1000 and the memory space taken for storing single data point is 0.012MB, then the overall space complexity is calculated as follows,

$$\text{Space complexity} = 1000 * 0.012MB = 12MB$$

- **Existing Bi-weighted ensemble approach:** No. of data points is 1000 and the memory space taken for storing single data point is 0.013MB, then the overall space complexity is calculated as follows,

$$\text{Space complexity} = 1000 * 0.013MB = 13MB$$

Figure 5 illustrates the experimental results of space complexity with respect to a number of data points using three methods namely FCRSCHC technique, ChronoClust [1] and Bi-weighted ensemble approach [2].



As shown in the above figure, the proposed FCRSCHC technique utilizes the less memory space for storing the temporal data points as compared to the existing clustering technique. This is owing to the application of the Spherical hierarchical structure based clustering for large dimensional data sets. The CURE hierarchical structure identifies the well scattered bitemporal data points in the spherical surface. In addition, the silhouette

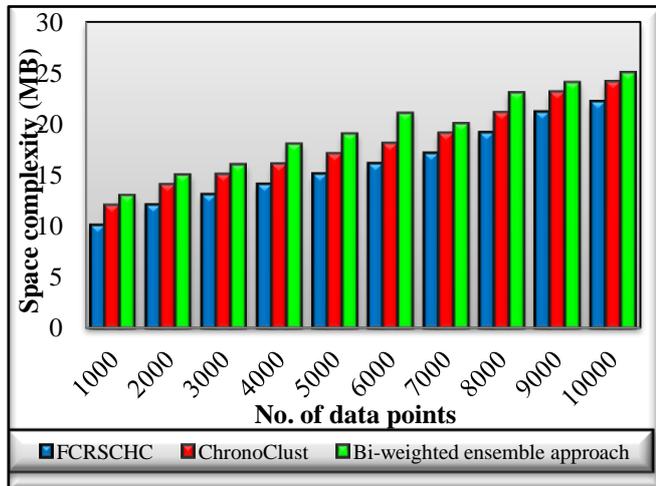


Figure 5 Experimental results of space complexity

coefficient is used to find similar data points within the cluster. This in turn minimizes space complexity. For example, 1000 data points are considered for calculating the space complexity. The FCRSCHC technique consumes 10MB for storing the 1000 temporal data points whereas the 12MB and 13MB consumed by the ChronoClust [1] and Bi-weighted ensemble approach [2]. The result shows that the FCRSCHC technique minimizes the storage space. Totally ten results are obtained for three methods and the results of the proposed technique are compared with the existing results. Then the average value is taken for ten various results. Finally, the average result confirms that FCRSCHC technique minimizes the space complexity by 12% when compared to existing ChronoClust [1]. Similarly, the performance results of space complexity also minimized by 19% when compared to existing Bi-weighted ensemble approach [2]. The above discussion clearly shows that the FCRSCHC technique improves the clustering accuracy and minimizes the time, false alarm rate as well as space complexity when compared to the state-of-the-art methods.

V. CONCLUSION

An efficient technique called FCRSCHC is designed with the aim of temporal relational data analysis through the clustering process. By applying the FCRSCHC technique, the clustering accuracy is improved and minimized the clustering time. The fuzzy crisp relation based hierarchical clustering algorithm groups the similar temporal data points into the different clusters. The hierarchical structure takes the well scattered bitemporal data points in the spherical surface and groups the data point which is close to the representative point of that particular cluster. After that, the nearby cluster representative points are merged to minimize the dimensionality. Finally, the cluster validation is performed to identify the weather data points are well

matched to its own cluster and poorly matched to neighboring clusters. This helps to minimize the false alarm rate and improve the clustering accuracy. The experimental assessment is done using the temporal dataset for recognizing human activities. The various observed results show that the FCRSCHC technique achieves better clustering accuracy and minimum clustering time, false positive rate as well as space complexity than the state-of-the-art methods.

REFERENCES

- Givanna H. Putri, Mark N. Read, Irena Koprinska, Deeksha Singh, Uwe Röhm, Thomas M. Ashurst, Nicholas J.C. King, "ChronoClust: Density-based clustering and cluster tracking in high-dimensional time-series data", Knowledge-Based Systems, Elsevier, Volume 174, 15 June 2019, Pages 9-26
- Yun Yang and Jianmin Jiang, "Bi-weighted ensemble via HMM-based approaches for temporal data clustering", Pattern Recognition, Elsevier, Volume 76, April 2018, Pages 391-403
- Hani El Assaad, Allou Samé, Gérard Govaert, Patrice Aknin, "A variational Expectation-Maximization algorithm for temporal data clustering", Computational Statistics & Data Analysis, Volume 103, 2016, Pages 206-228
- Joseph Crawford, Tijana Milenkovic, "ClueNet: Clustering a temporal network based on topological similarity rather than denseness", PLoS ONE, Volume 13, Issue 5, 2018, Pages 1-25
- Saeed Aghabozorgi, Teh Ying Wah, Tutut Herawan, Hamid A. Jalab, Mohammad Amin Shaygan, and Alireza Jalali, "A Hybrid Algorithm for Clustering of Time Series Data Based on Affinity Search Technique" The Scientific World Journal, Hindawi Publishing Corporation, Volume 2014, March 2014, Pages 1-12
- Pierpaolo D'Urso, Livia De Giovanni, Marta Disegna, Riccardo Massari, "Fuzzy clustering with spatial-temporal information", Spatial Statistics, Elsevier, Volume 30, April 2019, Pages 71-102
- Beibei Zhang and, Baiguo An, "Clustering time series based on dependence structure", PLoS ONE, Volume 13, Issue 11, 2018, Pages 1-22
- Xin Yao, Di Zhu, Yong Gao, Lun Wu, Pengcheng Zhang, Yu Liu, "A Stepwise Spatio-Temporal Flow Clustering Method for Discovering Mobility Trends", IEEE Access, Volume 6, 2018, Pages 44666 - 44675
- K.P. Agrawal, Sanjay Garga, Shashikant Sharma, Pinkal Patel, "Development and validation of OPTICS based spatio-temporal clustering technique", Information Sciences, Elsevier, Volume 369, 2016, Pages 388-401
- Saeid Soheily-Khah, Ahlame Douzal-Chouakria, Eric Gaussier, "Generalize d k -means-base d clustering for temporal data under weighted and kernel time warp", Pattern Recognition Letters, Elsevier, Volume 75, 2016, Pages 63-69
- Min Ji, Fuding Xie, and Yu Ping, "A Dynamic Fuzzy Cluster Algorithm for Time Series", Abstract and Applied Analysis, Hindawi Publishing Corporation, Volume 2013, March 2013, Pages 1-7
- Yongli Liu, Jingli Chen, Shuai Wu, Zhizhong Liu, Hao Chao, "Incremental fuzzy C medoids clustering of time series data using dynamic time warping distance", PLoS ONE, Volume 13, Issue 5, 2017, Pages 1-25
- Yun Yang and Jianmin Jiang, "HMM-based hybrid meta-clustering ensemble for temporal data", Knowledge-Based Systems, Elsevier, Volume 56, 2014, Pages 299-310
- Wei Yu, Wenjun Wang, Pengfei Jiao, Xuewei Li, "Evolutionary clustering via graph regularized nonnegative matrix factorization for exploring temporal networks", Knowledge-Based Systems, Elsevier, Volume 167, 2019, Pages 1-10
- Cátia M. Salgado, Marta C. Ferreira, Susana M. Vieira "Mixed Fuzzy Clustering for Misaligned Time Series", IEEE Transactions on Fuzzy Systems, Volume 25, Issue 6, 2017, Pages 1777 - 1794
- Marc Hüsch, Bruno U. Schyska, Lueder von Bremen, "CorClustST - correlation-based clustering of big spatio-temporal datasets", Future Generation Computer Systems, Elsevier, 2018, Pages 1-33
- Usue Mori, Alexander Mendiburu, Jose A. Lozano, "Similarity Measure Selection for Clustering Time Series Databases", IEEE Transactions on Knowledge and Data Engineering (Volume 28, Issue 1, 2016, Pages 181 - 195

18. Keon Myung Lee, Sang Yeon Lee, Kyung Mi Lee and Sang Ho Lee, "Density and Frequency-Aware Cluster Identification for Spatio-Temporal Sequence Data", *Wireless Personal Communications*, Springer, Volume 93, Issue No.1, March 2017, Pages 47–65.
19. Wenshuo Zhou, Kuangrong Hao, Chunli Jiang, Lei Chen, Xue-Song Tang, Xin Cai, "A New Cross Clustering Algorithm for Improving Performance of Supervised Learning", *IEEE Access*, Volume 7, Pages 56713 – 56723
20. Min Deng, QiLiang Liu, JiaQiu Wang, Yan Shi, "A general method of spatio-temporal clustering analysis", *Science China Information Sciences*, Springer, Volume 56, Issue 10, 2013, Pages 1–14

AUTHORS PROFILE



L. Jaya Singh Dhas, received his Bachelor of Science in Computer Science from Madurai Kamaraj University - Madurai, India in 1991 and Master of Computer Applications from Bharathidasan University - Tiruchirapalli in the year 1996 and M.Phil from Alagappa University - Karikudi in the year 1998. He is currently working as a Assistant Professor, Department

of Computer Science, Scott Christian College (Autonomous), Nagercoil since 1998 and his research work focuses on Algorithms, Big Data Analytics, Data Mining.



B. Mukunthan pursued Bachelor of Science in Computer Science from Bharathiar University, India in 2004 and Master of Computer Applications from Bharathiar University in year 2007 and Ph.D from Anna University - Chennai in 2013. He is currently working as Research Advisor in Department of Computer Science, Jairams Arts & Science College, Affiliated to Bharathidasan

University, Tiruchirapalli since 2016. He is a member of IEEE & IEEE computer society since 2009, a life member of the MISTE since 2010. He has published more than 10 research papers in reputed international journals. He is also Microsoft Certified Solution Developer. His main research work focuses on Algorithms, Bioinformatics, Big Data Analytics, Data Mining, IOT and Neural Networks. He also invented a Novel and Efficient online Bioinformatics Tool and filed for patent. He has 12 years of teaching experience and 10 years of Research Experience.