Journal of Computer Sciences

Original Research Paper

A Fuzzy Clustering Based Method for the Spatio-Temporal Data Analysis

Anahita Zolghadr and Afsaneh Jalalian

Department of Computer Engineering, Raja University, Qazvin, Iran

Article history
Received: 03-10-2021
Revised: 08-05-2022
Accepted: 14-05-2022

Corresponding Author:
Anahita Zolghadr
Department of Computer Software, Raja University, Iran
Emai: anahitazolghadr@gmail.com

Abstract: Spatiotemporal data is a type of data that is collected by the sensors. This type of data has two spatial and temporal dimensions. There are many challenges in analyzing spatiotemporal big data. Common evaluation metrics of clustering methods are not appropriate for spatiotemporal data. Previous clustering methods and the conventional evaluation metrics are efficient for data like time series with only one segment. Therefore, other metrics are required to evaluate the clustering of such data. In this study, energy function, reconstruction, and prediction metrics are used to evaluate the quality of spatiotemporal data clustering. The purpose of this study is to minimize the energy function using the Fuzzy C-Mean method on spatiotemporal data. The obtained results are compared with those obtained using k-medoid, DBSCAN, COBWEB, X-means, and TLBO. Also, the energy function, reconstruction, and prediction metrics are used to evaluate the quality of the clusters. The clustering methods are implemented on the dataset of parking located in the CBD area of Australia.

Keywords: Spatiotemporal Data, Clustering, Energy Reduction, FCM Method

Introduction

Today, with the development of novel technologies in the context of social networks, the Web of Things (WOT) is extended, and big data is generated. The emergence of big data related to spatial location knowledge, called geographical big data, provides opportunities for recognizing the urban area. The available database processing methods are not sufficient for providing fast reliable results in the context of geographical big data because it needs to define approximate "metrics" and increase the run-time of the queries. Big data obtained from devices like smart and portable phones, and Geographical Positioning System (GPS) devices, have penetrated our daily life and they have shown great potential for practical applications like climatology, natural disaster management, public health, product protection, smart cities, emergency management, and environment monitoring. Geographical big data is a subset of big data including spatial location data (Deng *et al*., 2019; Madbouly *et al*., 2022). Using various sensors to collect data over time is one application of the Web of Things (WoT). In WoT applications, there are many various sorts of data, such as temperature, light, and sound and they come in many different forms. Data quality might vary over time or from one device to another, therefore it depends on time and location (Barnaghi *et al*., 2013). Sensor data typically includes two features: Time and location This is referred to as spatiotemporal data. Spatiotemporal data is similar to time-series data since it comprises temporal information of an object, but it contains a vector of temporal information. While each object in time series data has a unique timestamp, the differences between time series and spatiotemporal data are (Shao *et al*., 2016):

- Time series data can measure time intervals, while spatiotemporal data measure both temporal and spatial intervals
- Time series data is in one dimension space, but the spatiotemporal data is in multiple Euclidean spaces
- The distances among data in a time series are the same, but the spatiotemporal data does not meet this condition

Because of the differences between the two types of data, common clustering algorithms such as k-means, DBSCAN, and COBWEB, which are usually used to cluster time series data, are not proper for spatiotemporal data.

There are several approaches to analyzing spatiotemporal big data (Shao *et al*., 2016; Sheng *et al*., 2010;

Yao and Sheng, 2012; Yao *et al*., 2013). When dealing with spatiotemporal big data, we confront numerous challenges. Real-time data, network density, massive volumes of spatiotemporal data and how the sensors link, heterogeneity in location and time, and reliable data sampling are some of these challenges (Shao *et al*., 2016; Sheng *et al*., 2010; Yao and Sheng, 2012; Yao *et al*., 2013; Atluri *et al*., 2018).

One of the existing issues is selecting acceptable methods for clustering this type of data, as well as evaluating the quality of spatiotemporal data clustering. There are numerous methods for evaluating clustering quality, including Silhouette, Davis Boldin, and index C (Shao *et al*., 2016; Thrun,2018; Nerurkar *et al*., 2018; Rousseeuw, 1987; Sun *et al*., 2010), but these approaches are not appropriate for spatiotemporal data since they use only one criterion to identify the best partition. While analyzing and evaluating spatiotemporal data require considering the distance between the two spatial and temporal domains (Shao *et al*., 2016). The Fuzzy C-Means (FCM) technique is used in this study for clustering. In addition, two evaluation metrics, including prediction and reconstruction metrics, which are employed for fuzzy clustering of spatiotemporal data, have been introduced. This research is based on a large dataset of parking data from the CBD area in Australia (Shao *et al*., 2016).

The objectives of this study are as follows:

- Providing an FCM-based clustering method on spatiotemporal data
- Providing an optimal distance function and examining the results on spatiotemporal data
- Introducing two optimization metrics including reconstruction and prediction error to optimize the efficiency of the proposed clustering method
- Comparing the proposed method with other clustering methods on a parking dataset

In the following, the properties of spatiotemporal data and time-series data and their problems are described in detail. Clustering of the spatiotemporal and time-series data is introduced and the previous clustering methods are discussed in the next section

Related Work

In this section, the spatiotemporal data is described.

Spatiotemporal Data

This type of data includes different types of data with different characteristics and different approaches for knowledge extraction (Kisilevich *et al*., 2009). At least one Spatio-temporal property is present in spatiotemporal objects (Atluri *et al*., 2018; Dhundale and Takalikar, 2015). The spatial feature of objects is related to their geographical location, while the temporal feature refers to the time intervals for each valid object (Sheng *et al*., 2010).

There are three major dimensions to spatiotemporal data: Spatial dimension, temporal dimension, and data dimension.

Temporal Dimension

Temporal events are divided into time sections. It is critical to select the best method for measuring the intervals between time segments. The temporal dimension is defined below.

It is assumed that we have a temporal dataset like $[t_s, t_e]$ such that the data vector, relates to a dependent variable like $f(t)$, $t \in [t_s. t_e]$. The vector includes time-related features (Shao *et al*., 2016).

Spatial Dimension

The spatial dimension of an object reveals whether it has a fixed or variable location. This refers to whether or not their location is dynamic and can vary over time. This property indicates an object's location in space at various moments (over a time interval) (Sheng *et al*., 2010). The data in this dimension solely contain information on the object's location, such as length and width (Shao *et al*., 2016).

Data Dimension

The data dimension of each spatiotemporal data has a period but can have multiple data points. This dimension can reveal vital information such as the number of parking violations. The data dimension is time-dependent, which means that data evolves through time. The complexity and assessment of similarity between two spatiotemporal events increase with distance uncertainty in data dimensions (Shao *et al*., 2016).

In addition, two general characteristics of autocorrelation and heterogeneity are defined for this type of data. Auto-correlation indicates that adjacent spatiotemporal data are interdependent and this autocorrelation in the dataset contributes to the transparency and integrity of observations. Each sample belongs to a different group or population, which is referred to as heterogeneity. This leads to a uniform distribution (Atluri *et al*., 2018).

There are various types of ST data that are used in many cases. There are distinct methods for collecting and representing each sort of ST data in two domains. As a result, STDM problems are classified into various groups. ST data is classified into four general groups. 1-Event data 2-Trajectory data 3-point Reference data 4-Raster data:

- Event data are distinct events in time and space. The set of Spatiotemporal event data (ST) is seen as a spatiotemporal pattern. Each event may have

additional properties that provide us with further information about the event. A forest fire, for example, can be viewed as a spatial polygon event

- Trajectory data depicts the movements of objects over time. The animal migration pattern is one example. Sensors installed on moving objects can be used to acquire this data. Trajectory data is utilized in a variety of applications, including transportation and the environment
- Point reference data take a continuous ST field measurement, such as temperature, vegetation, or a set of moving points in space and time. Meteorological factors such as temperature and humidity, for example, are measured by balloons floating in space, and weather observations are recorded continuously at various locations and times
- A discrete or continuous ST field can be evaluated when dealing with Raster data. This data can have a definite location in space as well as fixed points in time. This type of data is employed in a variety of applications, including remote control and brain imaging (Atluri *et al*., 2018)

Problems of Spatiotemporal Data Clustering

Due to the complexity of spatiotemporal data, clustering this type of data has various problems and challenges, including the following (Sheng *et al*., 2010):

- The spatial and temporal features of the objects vary continuously
- Adjacent spatiotemporal objects affect each other. For example, rain and wind affect fire intensity
- Spatiotemporal data represent various multidimensional data like time, location, and non-spatial features of spatiotemporal objects. Thus, processing, analyzing and data mining should be carried out at different levels for all features
- Because spatiotemporal data comprise at least two spatial and temporal features, the effect of spatial and temporal components on each other must be investigated independently in each object while clustering (Sheng *et al*., 2010)

Clustering Spatiotemporal Data

Clustering is one of the most essential strategies for exploring and analyzing data. (Shao *et al*., 2016; Bouguettaya *et al*., 2015; Liu *et al*., 2010; Yu and Rege, 2010; Birant and Kut, 2007). Clustering is a technique in which data with common features are clustered inside a cluster. Due to the presence of spatial and temporal dimensions, clustering ST data presents several issues. For instance, clustering locations based on their temporal information. Spatial clusters in raster ST data clustering must be continuous. Ignoring this will result in spatial cluster fragmentation. As a result, errors will occur in the interpretation of cluster information (Yao and Sheng, 2012).

Also, the spatial and temporal dimensions include various types of data with various features, providing various methods for knowledge extraction (Yao *et al*., 2013).

Clustering is divided into 4 groups hierarchical, partitioning, hard and fuzzy. Hierarchical clustering connects an input of a hierarchy to its corresponding output. The partitioning clustering employs an objective function to convert the input partition to a fixed set of clusters. In hard clustering, the patterns are in specific Clusters. But due to the overlap of the clusters, some patterns are located in a single cluster or different groups of data. This feature challenges using hard clustering in real-world applications. Fuzzy clustering was developed to overcome such limitations and it provides more information about the membership of the patterns. After the introduction of the Fuzzy theory by Zadeh, the scientists employed fuzzy theory for clustering (Nayak *et al*., 2015; Madbouly *et al*., 2022; Theodoridis and Koutroumbas, 2008; Iglesias and Kastner, 2013; Jain *et al*., 1999; Legany *et al*., 2006; Guha *et al*., 1998).

The result of a clustering algorithm in the same dataset might be different from each other because other input parameters of an algorithm might change the behavior and execution of the algorithm significantly. The purpose of cluster validity is to find a partition that best fits the principle data. Usually, two dimensions datasets are used to evaluate the clustering algorithms, because the reader can easily verify the result. But visual validation and visualization of high-dimensional data are not easy, thus official methods are required (Legany *et al*., 2006).

Materials and Methods

Here, three algorithms of fuzzy c mean, k-medoid, and TLBO are presented and their performance is compared. The Fuzzy C-Means (FCM) algorithm is a fuzzy clustering method, which is based on the minimum second-order criterion in each cluster. The membership degree of the data in a cluster is near one another. FCM has the advantage of generating new point clusters where the degree of membership of points in one cluster is close to each other. The FCM approach typically employs three operators. The fuzzy membership function, the partitioning matrix, and the objective function (Nayak *et al*., 2015; Ben Ayed *et al*., 2014; Alomoush *et al*., 2018; Rao *et al*., 2011; Gopala Krishna and Lalitha Bhaskari, 2016; Yang *et al*., 2018).

The Fuzzy C-means Algorithm in Temporal and Spatial Dimensions

For clustering the spatiotemporal data, it is assumed that there are n data as $X_1. X_2. \ldots . X_n$ such that each one contains temporal and spatial components. The data x_i, the

link between spatial and temporal components of data, is represented as $X_i = [X_i(S)|X_i(t)]^T$ such that $X_i(S)$ is the spatial component of X_i and $X_i(t)$ is the temporal component of the data. Assuming that there are r attributes in the spatial component and q attributes in the temporal component (Shi and Pun-Cheng, 2019; Izakian *et al*., 2012):

$$X_i = \left[X_i(S) \middle| X_i(t) \right]^T = \left[x_{i1}(s) ..... x_{ir}(s) \middle| x_{i1}(t) ..... x_{iq}(t) \right]^T \quad (1)$$

The purpose of the FCM method is to construct a set of c clusters including a set of initial samples $v_1. v_2. \ldots . v_c$ and the fuzzy partition matrix $U = [u_{ik}]$, c, $k = 1.2. \ldots . n$ such that (Shi and Pun-Cheng, 2019; Izakian *et al*., 2012; Zhou *et al*., 2008):

$$u_{ik} \in [0.1]. \quad (2)$$

$$\sum_{i=1}^{c} u_{ik} = 1 \cdot \forall k. \quad (3)$$

and:

$$0 < \sum_{k=1}^{n} u_{ik} < n \cdot \forall n \quad (4)$$

The following equation is obtained by minimizing the objective function (Shi and Pun-Cheng, 2019; Izakian *et al*., 2012; Zhou *et al*., 2008):

$$J = \sum_{i=1}^{C} \sum_{k=1}^{n} u_{ik}^{m} d^2 \left(v_i \cdot x_k \right) \quad (5)$$

If m>1, it is considered the fuzzification coefficient. The distance d is also known as the Euclidean or relative distance. Because spatiotemporal data is divided into two components of space and time, the distance calculation function must be defined such that the space and time components can be calculated individually. The distance function can be generalized to include two components to accomplish this (Shi and Pun-Cheng, 2019; Zhou *et al*., 2008; Yang *et al*., 2018):

$$d_{\lambda}^{2} \left(v_i \cdot x_k \right) = \left\| v_i(s) - x_k(s) \right\|^2 + \lambda \left\| v_i(t) - x_k(t) \right\|^2 \cdot \lambda \geq |0 \quad (6)$$

Equation 6 is a generalized function of distance and it can be used to control the effect of different parts of space and time on each other. When $\lambda = 0$, the data has no temporal attributes and the spatial attribute effect is taken into account. The greater the value of λ, the larger the time effect. The following equation is constructed by using the distance function given in Eq. 6 in the objective function (Shi and Pun-Cheng, 2019; Zhou *et al*., 2008; Yang *et al*., 2018):

$$J = \sum_{i=1}^{C} \sum_{k=1}^{n} u_{ik}^{m} d_{\lambda}^{2} \left(v_i \cdot x_k \right) \quad (7)$$

Equations 8 and 9 represent the initial sample and the partition matrix that is obtained through optimizing the objective function (Shi and Pun-Cheng, 2019; Zhou *et al*., 2008; Yang *et al*., 2018):

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^{m} X_k}{\sum_{k=1}^{n} u_{ik}^{m}} \quad (8)$$

$$u_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\frac{d_{\lambda}(v_j \cdot x_k)}{d_{\lambda}(v_j \cdot x_k)} \right)^{2/m-1}} \quad (9)$$

To obtain optimal clusters, the partition matrix and the initial sample should be updated continuously. It should be considered that the weight λ should be highly flexible.

In the following, two evaluation metrics are introduced (Shi and Pun-Cheng, 2019):

The k-medoid method is also known as the object-based technique representative. This is an unsupervised learning algorithm that is comparable to the K-means algorithm. This approach computes the mean point by inventing a hypothetical point, whereas the k-medoid computes the mean point by computing the closest real point in the dataset. This method's scalability can be extended by using algorithms like PAM, CLARA, and CLARANS (Han *et al*., 2012).

The TLBO technique is an algorithm inspired by the teaching-learning process presented by (Rao *et al*., 2011) It is one of the most recently developed algorithms. This idea was inspired by a classroom learning experience. A crucial idea of TLBO is the teacher's influence on the learner. This algorithm does not necessitate any control parameter. It is a population-based technique, similar to other nature-inspired algorithms. Learners (students) are regarded as the study population, or as candidate solutions. There are numerous applications of TLBO for various optimization problems (Gopala Krishna and Lalitha Bhaskari, 2016; Rao and Savsani, 2012; Satapathy and Naik, 2011; Satapathy *et al*., 2012a,b; Parvathi *et al*., 2012; Naik *et al*., 2012; Nayak *et al*., 2012; Ren *et al*., 2022).

Various characteristics of an optimization problem are analogous to various courses taught in the classroom. The scores acquired in each of these subjects, as well as the total scores obtained, are regarded as "fitness." The teacher is regarded as the best solution since he or she is regarded as the best person in the community. The TLBO is broken into two parts: Teacher and student (Gopala Krishna and Lalitha Bhaskari, 2016).

The TLBO algorithm is devoid of any algorithmic parameters. As a result, no parameters need to be tweaked to optimize the performance of the FCM-TLBO algorithm

(Gopala Krishna and Lalitha Bhaskari, 2016). In the C-means fuzzy method, the termination condition is attained when the generated solution is not improved further. The number of iterations in the spatial domain is 8, 20, and 50, whereas the number of iterations in the spatiotemporal domain is 10, 20, and 50. Furthermore, in all circumstances, the weight power equals 2 (m = 2).

Evaluation Metrics

It can be said that the validation of the results obtained by a clustering algorithm tries to provide a measure of the algorithm's success and accuracy. Here, we identify two strategies for investigating clustering solutions.

On the one hand, there are cluster or clustering methods, which attempt to evaluate findings by mathematical analysis and direct observation of solutions based on the intrinsic qualities of the input dataset. In other words, it comprises idealistic analysis approaches because they focus on the definition given to a cluster, regardless of the rationale for establishing clustering (i.e., end application).

Clustering solutions, on the other hand, can occasionally be tested directly by the program or the environment that simulates the application (clustering evaluation). This is a practical (or engineering) approach to testing that focuses on program-based testing. Generalizations are riskier in this context. It is worth noting that the program has introduced corruption and deformations, as well as boundary constraints and unique data utilized for testing.

The value of such quantitative metrics is always relative in both cases, implying that they are the only tools accessible to specialists to evaluate clustering (Iglesias and Kastner, 2013; Han *et al*., 2012).

In this case, two metrics are employed: Reconstruction and prediction. The nature of these two metrics is depicted in Fig. 1 (Shi and Pun-Cheng, 2019; Adhikari *et al*., 2015).

Reconstruction Metric

This evaluation metric reconstructs the main data and the partition matrix by minimizing the sum of distances (Shi and Pun-Cheng, 2019):

$$F = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m} \left\| v_i - x_k \right\|^2 \tag{10}$$

Such, $\hat{x}_k$ is the reconstructed version of x_k. If the gradient of F is set to zero, considering $\hat{x}_k$, we have (Shi and Pun-Cheng, 2019):

$$\hat{X}_k = \frac{\sum_{i=1}^{c} u_{ik}^{m} V_i}{\sum_{i=1}^{c} u_{ik}^{m}} \tag{11}$$

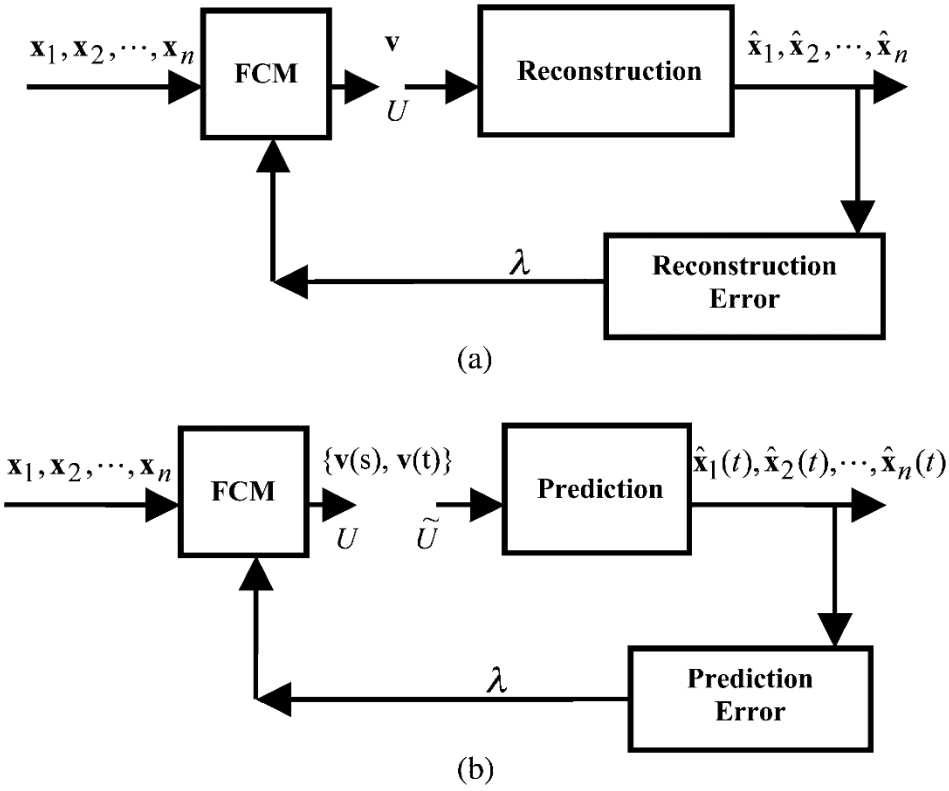

Fig. 1: Evaluation metrics of the FCM algorithm (Shi and Pun-Cheng, 2019)

When reconstruction is complete, $x_1.x_2....x_n$ is the new dataset is constructed using Eq. 11. The reconstruction quality as the λ function is defined as follows: (Shi and Pun-Cheng, 2019):

$$E(\lambda)=\sum_{k=1}^{n}\left\|X_k-\hat{X}_k\right\|^2 =\sum_{k=1}^{n}\left\|x_k(s)-\hat{x}_k(s)\right\|^2+\sum_{k=1}^{n}\left\|x_k(t)-\hat{x}_k(t)\right\|^2 \quad (12)$$

Such that:

$$\left\|x_k(s)-x_k(s)\right\|^2=\frac{1}{r}\sum_{j=1}^{r}\frac{\left(x_{kj}(s)-x_{kj}(s)\right)^2}{\sigma_j^2} \quad (13)$$

and:

$$\left\|x_k(t)-x_k(t)\right\|^2=\frac{1}{q}\sum_{j=1}^{q}\frac{\left(x_{kj}(t)-x_{kj}(t)\right)^2}{\sigma_j^2} \quad (14)$$

σ_j^2 is the variance of the j^{th} attribute. In this study, the Euclidean distance is used. The reconstruction error $E(\lambda)$ is a function of λ and its minimum is determined inside a specific set of values of λ (Shi and Pun-Cheng, 2019).

Prediction Metric

Since the spatiotemporal data is comprised of spatial and temporal components, the centers of the cluster of this type of data also have spatial (v(s)) and temporal (v(t)) components. If the spatial component of each data and the spatial component of the cluster centers are known, a new partition matrix can be comprised, called $\tilde{U}$: (Shi and Pun-Cheng, 2019):

$$u_{ik}=\frac{1}{\sum_{j=1}^{c}\left(\frac{\left\|v_i(s)-x_k(s)\right\|}{v_j(s)-x_k(s)}\right)^{2/(m-1)}} \quad (15)$$

Also, the new partition matrix and the temporal component of the cluster centers (v(t)), can be used to minimize the sum of distances:

$$F=\sum_{i=1}^{c}\sum_{k=1}^{n}\tilde{x}_{ik}^{m}\left\|v_i(t)-\hat{x}_k(t)\right\|^2 \quad (16)$$

Such that $\hat{x}(t)$ is the predicted temporal component of the k[th] data. If the gradient of F is set to zero, considering $\hat{x}(t)$, we have:

$$\hat{X}_k(t)=\frac{\sum_{i=1}^{c}\tilde{u}_{ik}^{m}\,v_i(t)}{\sum_{i=1}^{c}\tilde{u}_{ik}^{m}} \quad (17)$$

The prediction quality is evaluated using the following prediction error:

$$E(\lambda)=\sum_{k=1}^{n}\left\|x_k(t)-\hat{x}_k(t)\right\|^2=\frac{1}{q}\sum_{k=1}^{n}\sum_{j=1}^{q}\frac{\left(x_{kj}(t)-\hat{x}_{kj}(t)\right)^2}{\sigma_j^2} \quad (18)$$

The prediction quality is calculated using the sum of normalized Euclidean distances between the temporal component and the predicted temporal component. Similar to the previous metric, the purpose is to minimize E(λ) by adjusting λ (Shi and Pun-Cheng, 2019).

Organization and Implementation

This study employs data acquired by parking sensors in the CBD area of Australia, which totaled one gigabyte in 2012. Each record contains information such as the district name, street name, arrival and departure times, and so on.

Only spatial information was used to cluster the data in the first experiment. That is, each data simply has the latitude and longitude properties. The second experiment employs all spatiotemporal features. Spatial and temporal event information are both included in spatiotemporal information. The goal is to investigate the impact of temporal information on energy production. The results suggest that clustering spatiotemporal data reduces energy production when both spatial and temporal components are present. In this experiment, information for one day was chosen at random from a one-year total dataset.

Clustering evaluation metrics were employed in the last experiment, including both the reconstruction and prediction metrics. One day is chosen at random. The energy function, as well as the evaluation metrics, are calculated as a result of the fuzzy clustering.

The number of clusters in c-means has been chosen as an internal index by the Bayesian information criteria (Shao *et al*., 2016; Nerurkar *et al*., 2018). This suggests that the number of clusters chosen by c-means is nearly optimal. The R X-means clustering program is utilized, which chooses the best value of k automatically (Shao *et al*., 2016; Pelleg and Moore, 2000). After clustering with the methods described above, the energy functions *E*(*t*), *E*(*s*), density, and *E*(*f*) are computed.

Data clustering with the Fuzzy c-means method will be done and the distance of each point from its cluster center, partition matrix, and the objective function will be calculated. Energy function, *E*(*t*), *E*(*s*), Density, and *E*(*f*) will be calculated.

As mentioned in the introduction, common metrics for evaluating clustering methods cannot be applied to spatiotemporal data. Various evaluation methods are carried

out using cluster validation approaches. Cluster validation depicts the process of evaluating the clustering algorithm's results. Metrics such as energy function ($E(f)$), prediction, and reconstruction can be used to analyze this type of data.

The purpose of the energy function is to increase the intra-cluster similarity while decreasing the inter-cluster similarity. It should be noticed that this improves the cluster quality. Equation 19 specifies this function (Shao *et al*., 2016):

$$E(f)=E_{spatial}(f)+E_{temporal}(f) \tag{19}$$

$E_{temporal}(f)$ is used to measure the dissimilarity between f and the data observed in the time domain, whilst $E_{spatial}(f)$ is used to measure the similarity in the spatial domain:

$$E(f)=\left(\frac{1}{\alpha}\right)*\sum_{f_p=f_q\in L} Dist_s(p.q)+\sum_{f_p=f_q\in L} Dist_T(p\cdot q) \tag{20}$$

Fp is the label for the point p. $Dist_T$ is used to measure distance in the time domain. α is a constant parameter that is greater than one ($\alpha>>1$), implying that the relationship between the first and second parameters is linear.

Measuring Distances in the Spatial-Temporal Domain

As discussed above, in this study, the Euclidean distance is used to calculate spatiotemporally. $Dist_S$ is determined using the following equation (Shao *et al*., 2016):

$$Dist_S(p\cdot q)=\sqrt{\left(p_x-p_x\right)^2+\left(p_y-q_y\right)^2} \tag{21}$$

The length and width of the points p and q are represented by x and y.

The distance in the time domain is determined as a density function, hence F and G are regarded as probability density functions in the following form: (Shao *et al*., 2016):

$$\int_{-\infty}^{+\infty}\left(F(t)-G(t)\right)^2dt \tag{22}$$

Similarity and Balance

Both metrics are used to measure distance in space and time and measure intracluster similarity. Another point that should be considered is the balance of the clusters. This means that the distinction between each pair of clusters is negligible. Variance is a critical parameter. According to Eq. 5, variance is measured for this purpose: (Shao *et al*., 2016):

$$E_{balance}=Var(X)=\sum_{i=1}^{k}\left(x_i-\mu\right)^2 \tag{23}$$

μ is the average size of all clusters, x_i is the density of the i^{th} cluster and k is the number of clusters.

Density

Here, a general definition of density in spatiotemporal data features space is given based on Eq. 6 (Shao *et al*., 2016). X_i is a point in space. P_{xi} is the ratio of time events at this point to all events in the time domain. For example, the high incidence and length of parking violations at a specific point suggest that high-volume points can increase the burden of officers in these areas. As a result, P_{xi} has a significant impact on intra-cluster density in spatiotemporal dimensions (Shao *et al*., 2016):

$$Density=\sum_{x_i\in C_i} P_{xi}\sum_{x_j\in C_i,i\neq j} P_{xj}\times Dist_s\left(x_i.x_j\right) \tag{24}$$

Results

The simulation results are divided into two sections: Spatial and spatial-temporal. The spatial section of the data provides information about the location of each data, such as the length and width of the parking lot and the spatiotemporal section of the data contains information about the start and end times of each event.

The clustering results of x-means, DBSCAN, and COBWEB methods are compared with the methods provided in this study, namely FCM, k-medoid, and TLBO, according to (Shao *et al*., 2016).

Implementation Results of the Spatial Domain

As can be seen in Table 1, the maximum energy in the FCM method is produced by cluster 7. Also, according to Table 2, the maximum energy in the k-medoid method is produced by cluster 2.

It should be noted that the difference of the maximum and minimum energy by FCM (min = 0.087 and max = 2.0087) is less that x-means (min = 1,366.57 and max = 430,130) and DBSCAN (min = 0.9866 and max = 1,235,190).

The number of iterations for implementation of TLBO is 8, 15, and 20 and the number of variables is 4.

Implementation Results of the Spatiotemporal Domain

In this section, information regarding the time and location of the events is used.

As can be seen in Table 4, cluster 3 produces maximum spatial energy, while Cluster 9 produces maximum temporal energy. However, using the FCM approach, cluster 9 produces maximum energy. Cluster 6 also produces minimum energy.

As in the spatial domain, the difference between the maximum and minimum energy produced by the FCM technique with min = 0.96 and max = 19.50 is less than that produced by the x-means methods with min = 2.71 and max = 26,281.5 and DBSCAN with min = 0.264 and max = 59,118.6.

Also, for implementing the TLBO method, the number of iterations is 10, 15, and 20 and the number of variables is 4.

Figure 10 shows the temporal distribution of data for 10 clusters in the FCM method. As can be seen in Fig. 5, the clusters are balanced and maximum energy is produced by cluster 9.

Tables 4 and 5 compare spatiotemporal clustering approaches. The energy obtained from clustering using the aforementioned approaches is given in the $E_{spatial}$ and $E_{temporal}$ columns. Equation 5 is also used to calculate $E_{balance}$. The total $E_{spatial}$, $E_{temporal}$ and $E_{balance}$ of the x-means, DBSCAN, and COBWEB techniques are extracted and compared (Shao *et al*., 2016).

The reconstruction and prediction metrics are used to measure the quality of the clusters. The results of these two evaluation metrics are given in Table 9.

Table 1: Results of FCM clustering in spatial domain

Clusters	$E_{spatial}$	$E_{temporal}$	Density
Cluster 1	0.27291	0.330500	2.5000
Cluster 2	1.91800	0.533000	7.0000
Cluster 3	0.97402	1.294400	3.9000
Cluster 4	0.25930	0.433500	2.1000
Cluster 5	0.62090	1.144700	3.3000
Cluster 6	0.46820	0.545400	4.4000
Cluster 7	2.00780	2.553100	6.3000
Cluster 8	0.08790	0.021000	1.4000
Cluster 15	0.01790	0.064600	0.0218
Cluster 20	0.00643	0.009321	0.0238
Cluster 50	0.00254	0.004197	0.0293

Table 2: The energy produced in k-medoid method in spatial domain

Clusters	$E_{spatial}$	$E_{temporal}$	Density
Cluster 1	846.1400	15.3000	30.060
Cluster 2	2041.0200	26.7400	28.980
Cluster 3	1236.7900	16.1251	31.420
Cluster 4	0.1698	0.0500	764.009
Cluster 5	398.0100	67.8500	14.360
Cluster 6	148.6700	23.1830	12.080
Cluster 7	9.0040	2.8500	55.566
Cluster 8	75.1100	10.5900	2.540
Cluster 15	49.3590	9.8800	2.190
Cluster 20	40.1290	9.0010	1.980
Cluster 50	20.6500	5.3320	1.530

Table 3: Implementation of TLBO in spatial domain

Clusters	Worst score	Best score
Cluster 8	83.3901	1.5767000
Cluster 15	65.7681	0.0378560
Cluster 20	55.2142	0.0029530
Cluster 50	37.3420	1.021e-09

Table 4: Results of clustering using FCM in Spatiotemporal domain

Clusters	$E_{spatial}$	$E_{temporal}$	Density
Cluster 1	75.67000	13.9400	1.10
Cluster 2	38.41000	9.6619	0.77
Cluster 3	124.20000	16.2586	1.38
Cluster 4	102.20000	15.3155	1.20
Cluster 5	53.21000	13.9541	0.66
Cluster 6	2.23000	0.9655	0.57
Cluster 7	13.06000	11.1258	0.44
Cluster 8	69.46000	15.5955	1.32
Cluster 9	123.14000	19.5025	1.02
Cluster 10	40.08000	16.3894	1.33
Cluster 15	0.04300	8.1040	0.08
Cluster 20	0.00340	2.0780	0.17
Cluster 50	0.00046	1.6763	0.16

Table 5: Results of clustering using k-medoid in spatiotemporal domain

Clusters	$E_{spatial}$	$E_{temporal}$	Density
Cluster 1	4326.0202	127.140	0.39
Cluster 2	1087.7100	117.142	0.21
Cluster 3	821.5500	92.453	0.14
Cluster 4	2852.1900	51670.000	0.40
Cluster 5	111.8270	522.879	0.11
Cluster 6	31.3330	17.580	0.06
Cluster 7	8.2000	3.240	0.05
Cluster 8	16314.0800	1244.001	0.16
Cluster 9	7.4600	4.303	0.12
Cluster 10	1.4400	2.521	0.09
Cluster 15	10.2423	6.713	0.05
Cluster 20	5.9070	2.308	0.02
Cluster 50	0.1100	1.790	0.01

Table 6: Implementation using TLBO in the spatiotemporal domain

Clusters	Worst score	Best score
Cluster 10	65.7314	0.298920
Cluster 15	53.5759	0.042091
Cluster 20	63.9475	0.001570
Cluster 50	43.6619	1.387e-09

Table 7: Comparing the FCM, k-medoid and TLBO clustering methods

	FCM			k-medoid			
Clusters	$E_{spatial}$ (Total energy)	$E_{temporal}$ (Total energy)	$E_{balance}$	$E_{spatial}$ (Total energy)	$E_{temporal}$ (Total energy)	$E_{balance}$	TLBO Best score
C = 8	6.60903	6.855600	1481.600	4754.913	162.6881	3,285.912	1.5767000
C = 15	0.01790	0.064600	257.190	4804.272	93.8402	2562.927	0.0378560
C = 20	0.00643	0.009321	4.631	4849.401	70.8302	1803.471	0.0029530
C = 50	0.00254	0.004197	0.198	729.441	25.5600	584.301	1.021e-09
Clustering in spatiotemporal domain							
C = 10	641.66000	132.708000	0.080	35,560.810	2,648.1290	0.190	0.2989200
C = 15	640.70300	129.812800	0.050	25572.050	2690.5140	0.110	0.0420910
C = 20	530.95000	98.147000	0.010	10539.250	536.7800	0.090	0.0015700
C = 50	266.40000	53.180000	0.005	729.441	25.5600	0.080	1.387e-09

Table 8: Comparing the energy generated by clustering methods (Shao *et al*., 2016)

$E_{balance}$	$E_{temporal}$ (Total energy)	$E_{spatial}$ (Total energy)	Clustering method
Clustering in spatial domain			
121،979.9	22’132.23	465’965.18	x-means
22’717’033	43’606.79	1’241’492	DBSCAN
Clustering in the spatiotemporal domain			
0.21	33’312.99	230’380.94	x-means
0.95	59’425.45	1’107’414.5	DBSCAN
8.18	51’247’869.73	46’194.95	COBWEB

Table 9: Evaluation metrics of clustering

Clusters	Reconstruction metric	Prediction metric
Cluster 2	0.6802	1.4790
Cluster 3	0.5130	1.2636
Cluster 4	0.4197	1.1877
Cluster 5	0.3615	1.1394
Cluster 6	0.3564	5.3811
Cluster 7	0.3487	5.0185
Cluster 8	0.3446	4.8336
Cluster 9	0.3369	4.2127
Cluster 10	0.3227	3.9133
Cluster 20	0.1100	1.7000
Cluster 50	0.0500	0.0900

Fig. 2: Clustering using the k-medoid method in the spatial domain

Fig. 3: Clustering using the FCM method in the spatial domain

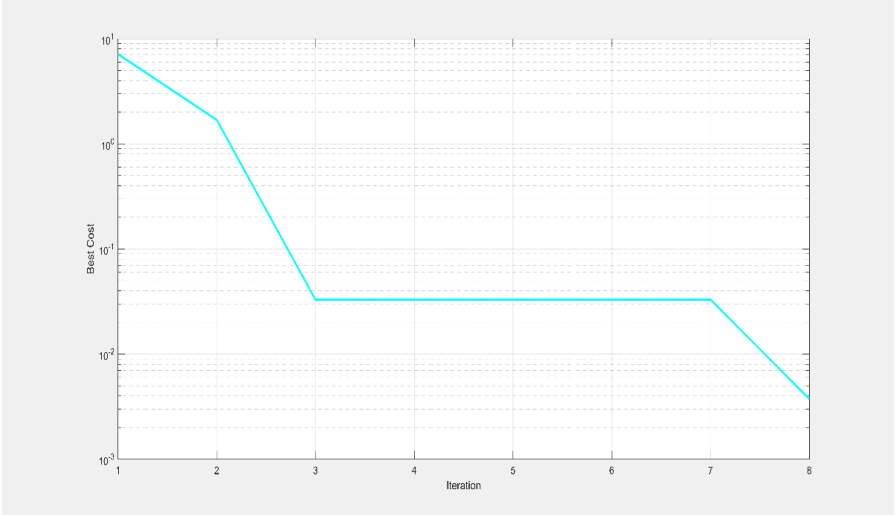

Fig. 4: Best scores for 8 iterations in the spatial domain

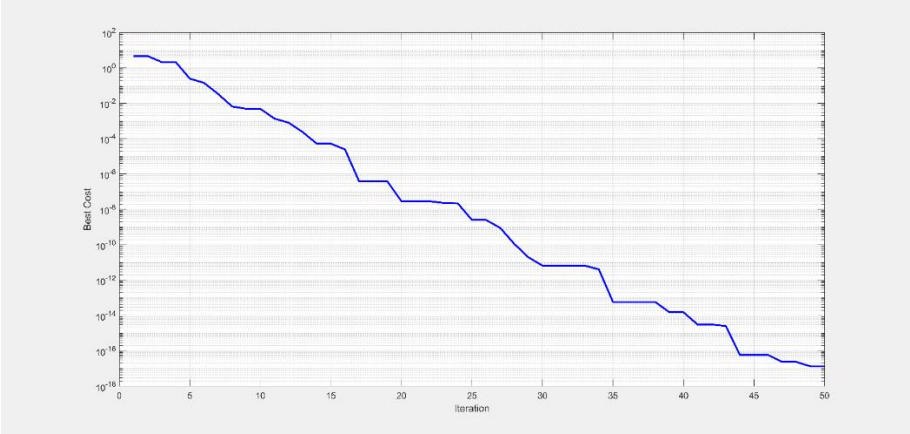

Fig. 5: Best scores for 50 iterations in the spatial domain

Fig. 6: Clustering using FCM in the spatiotemporal domain

Fig. 7: Clustering using k-medoid in the spatiotemporal domain

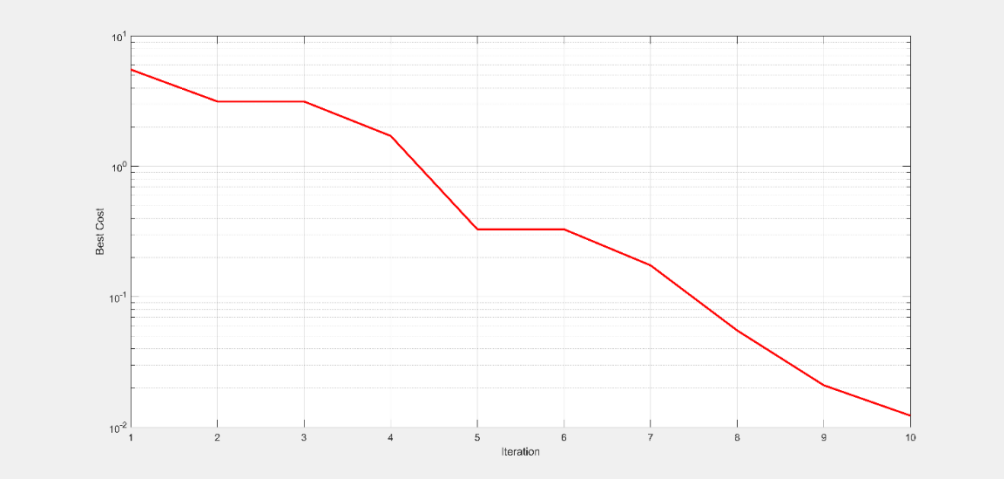

Fig. 8: Best scores for 8 iterations in the spatial domain

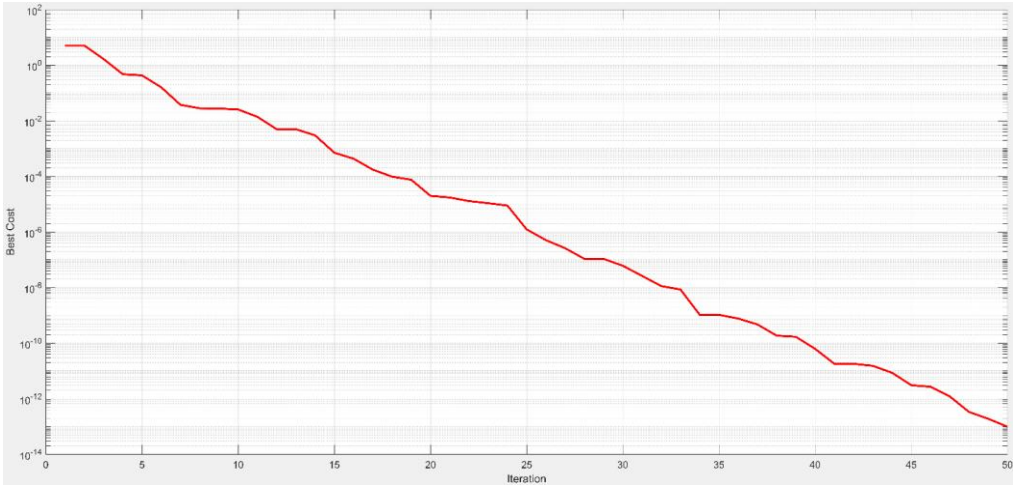

Fig. 9: Best scores for 50 iterations in the spatiotemporal domain

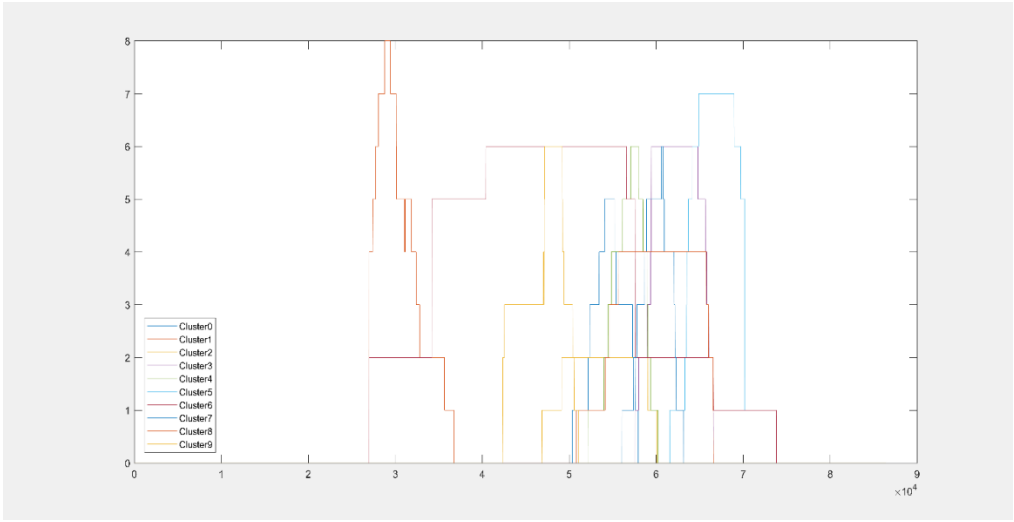

Fig. 10: Temporal distribution of data for 10 clusters in the FCM method

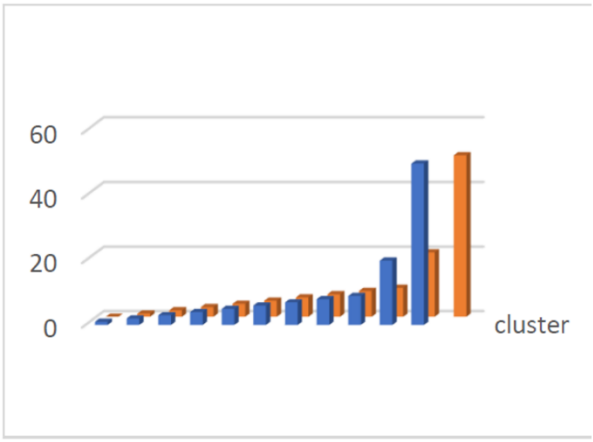

Fig. 11: Results of evaluation criteria

Discussion

In the spatial domain, as shown in Fig. 2 and 3, the FCM method is more balanced than the k-medoid method and the small value of density represents that intra-cluster distribution is normal.

Also, according to (Shao *et al*., 2016), maximum energy E(s) and E(t) are produced by the last cluster in the x-means method and the clusters in this method are unbalanced. In general, the x-means method is more balanced compared to the DBSCAN method and the FCM approach uses less energy than the two methods discussed (Shao *et al*., 2016), x-means and DBSCAN and the proposed k-medoid method is more balanced, and efficient than x-means. This suggests a normal intra-cluster data distribution, which improves the FCM method's similarity and balance. This method, in general, is more balanced and uses less energy than the x-means and DBSCAN methods.

As can be seen in Table 3, Fig. 4 and 5, by increasing the number of clusters, the obtained score decreases, which indicates that the performance of this algorithm is improved.

In the Spatiotemporal domain temporal component is used and the purpose is to examine the effect of temporal information on energy reduction. The number of clusters is 10.

Finally, the results of the mentioned methods are compared with those of the methods mentioned in (Shao *et al*., 2016).

The spatial and temporal energy generated by the FCM method is lower in the spatiotemporal domain than the energy generated by the k-medoid approach, as shown in Tables 4 and 5. This method uses less energy than the x-means method except for clusters 8 and 9. Thus, it has enhanced the energy production process compared to the x-means technique, demonstrating the superior performance of the proposed methods in spatiotemporal data clustering.

Also, Fig. 6 and 7 show more equilibrium than the methods mentioned in (Shao *et al*., 2016).

As in the spatial domain, in the spatiotemporal domain, this reflects the normal intra-cluster distribution, which increases similarity and balance in the FCM approach.

Furthermore, because of using the time parameter, the energy $E(t)$ and $E(s)$ produced in the spatiotemporal domain are less than the energy produced in the spatial domain. In this case, the x-means method outperforms the DBSCAN method. Although the COBWEB approach is more balanced in the spatiotemporal domain than the DBSCAN method, it has not improved compared to the x-means method, according to (Shao *et al*., 2016).

In general, the results of the FCM approach are not much better than the DBSCAN method because the DBSCAN method has less energy in half of the clusters, according to (Shao *et al*., 2016). In terms of energy production, it can be determined that this approach is more efficient than the FCM method. However, it produces substantially less energy than the x-means technique.

According to Table 6, Fig. 8 and 9, by increasing the number of clusters, the obtained score decreases.

When diverse time intervals are included in a cluster, the cluster produces less energy. DBSCAN, which has maximum energy, cannot partition the time domain into different time intervals. As a result, clustering algorithms that can divide the temporal domain into various intervals are likely to use less energy (Shao *et al*., 2016). The FCM approach generates a smaller $E_{balance}$ and is more balanced in both spatial and temporal domains than previous methods. This is because FCM uses a new distance function. Because effective regulation of the time section makes the data more balanced.

According to Table 7, when comparing $E_{balance}$, the FCM technique outperforms the k-medoid method in both spatial and spatial-temporal domains, owing to the usage of the fuzzy method. Because of its independence from the control parameter, the TLBO approach outperforms the proposed methods significantly in the spatial domain. However, when clustering with 10 clusters in the spatiotemporal domain, the TLBO approach improves performance less than the other methods, but this problem is resolved by increasing the number of clusters.

By comparing the values of $E_{balance}$ in Tables 7 and 8, it can be concluded that all three methods outperform the x-means and DBSCAN clustering methods in the spatial domain, and the FCM and Kmedoid methods outperform the x-means, DBSCAN and COBWEB in the temporal domain.

When using TLBO with 10 clusters, no improvement is achieved, but by increasing the number of clusters, the performance of this method improves compared to the methods mentioned in Table 8.

As can be seen, in Fig. 11, as the number of clusters increases, both metrics decrease. The smaller the values of the evaluation metrics, the quality of the clusters is higher.

Conclusion

As previously stated, significant amounts of spatiotemporal data are generated and recorded by systems that record sequential remote sensing, mobility, and social media data (sensors). These intricate and implicit relationships are very dynamic. Finding solutions to ensure real-time data, latency, network congestion, and recognizing links between sensors and spatial heterogeneity are all challenges when dealing with this type of data.

Because of the complexity of spatiotemporal data, clustering this type of data presents several problems and challenges, including 1-continuous and discrete

changes in the spatial and non-spatial features of spatiotemporal objects and 2-the effect of neighboring spatiotemporal objects on each other.

Because of the challenges highlighted in this study, three strategies were utilized to cluster spatiotemporal data. The results of the proposed method were compared to those of other approaches. The following can be mentioned according to the findings.

The distance function in the FCM clustering technique is employed in this research, in which its time domain is discriminated from the spatial domain with the weight parameterλ. When data contains both space and time parameters, $\lambda = 1$ is used, whereas when data just contains spatial information, $\lambda = 0$ is used. This enables the data's temporal component to be calculated efficiently.

It can be concluded that clustering spatiotemporal data with a fuzzy method in which the temporal component's effect is regulated by a parameter such as weight can boost intra-cluster similarity and balance while reducing energy generation. In spatiotemporal data clustering, the FCM approach uses minimum energy in both the spatial and temporal domains.

The k-medoid approach uses less energy than the x-means method and it has been enhanced since it calculates the mean point by finding the nearest real point in the dataset. As a result, it outperforms x-means. Also, according to the obtained data, the TLBO approach outperforms the other two methods because it does not rely on any algorithm parameters other than population size and the maximum number of iterations.

The focus of this study is on spatiotemporal data and it can be extended to other types of data like time series data.

As mentioned, the parameterλ is used in calculating the distance of the FCM algorithm. In future works, the effect of using this parameter on the energy function can be studied.

Acknowledgment

I would like to take this opportunity to appreciate my dear brother and colleague Mr. Behshad Benvidi, who helped me in this research.

Author's Contributions

Anahita Zolghadr: Participated in all experiments, coordinated the data analysis, and contributed to the writing of the manuscript.

Afsaneh Jalalian: Contribute to drafting the article, Planning the experimental setup, Organizing the study, give final approval for submission.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and that no ethical issues are involved.

References

Adhikari, S. K., Sing, J. K., Basu, D. K., & Nasipuri, M. (2015). Conditional spatial fuzzy C-means clustering algorithm for segmentation of MRI images. Applied soft computing, 34, 758-769. doi.org/10.1016/j.asoc.2015.05.038

Alomoush, W., Alrosan, A., Norwawi, N., Alomari, Y., Albashish, D., Almomani, A., & Alqahtani, M. (2018). A survey: Challenges of image segmentation based fuzzy C-means clustering algorithm. Journal of Theoretical and Applied Information Technology. https://www.researchgate.net/publication/327338071

Atluri, G., Karpatne, A., & Kumar, V. (2018). Spatio-temporal data mining: A survey of problems and methods. ACM Computing Surveys (CSUR), 51(4), 1-41., doi.org/10.1145/3161602

Ayed, A. B., Halima, M. B., & Alimi, A. M. (2014, August). Survey on clustering methods: Towards fuzzy clustering for big data. In 2014 6th International conference of soft computing and pattern recognition (SoCPaR) (pp. 331-336). IEEE. doi.org/10.1109/SOCPAR.2014.7008028

Barnaghi, P., Sheth, A., & Henson, C. (2013). From data to actionable knowledge: Big data challenges in the web of things [Guest Editors' Introduction]. IEEE Intelligent Systems, 28(6), 6-11. doi.org/10.1109/MIS.2013.142

Birant, D., & Kut, A. (2007). ST-DBSCAN: An algorithm for clustering spatial-temporal data. Data and knowledge engineering, 60(1), 208-221. doi.org/10.1016/j.datak.2006.01.013

Bouguettaya, A., Yu, Q., Liu, X., Zhou, X., & Song, A. (2015). Efficient agglomerative hierarchical clustering. Expert Systems with Applications, 42(5), 2785-2797. doi.org/10.1016/j.eswa.2014.09.054

Deng, X., Liu, P., Liu, X., Wang, R., Zhang, Y., He, J., & Yao, Y. (2019). Geospatial big data: New paradigm of remote sensing applications. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12(10), 3841-3851. doi.org/10.1109/JSTARS.2019.2944952

Dhundale, V. V., & Takalikar, M. (2015), "Survey on Spatio-Temporal Clustering". International Journal of Science and Research (IJSR), https://www.ijsr.net/get_abstract.php?paper_id=ART2016667, Volume 5 Issue 7, July 2016, 1998-2001

Gopala Krishna, P., & Lalitha Bhaskari, D. (2016). Fuzzy C-means and fuzzy TLBO for fuzzy clustering. In Proceedings of the Second International Conference on Computer and Communication Technologies (pp. 479-486). Springer, New Delhi. doi.org/10.1007/978-81-322-2517-1_46

Guha, S., Rastogi, R., & Shim, K. (1998). CURE: An efficient clustering algorithm for large databases, Proc. of ACM SIGMOD International Conference on Management of Data, pp, 73-84. https://doi.org/10.1145/276305.276312

Han, J., Kamber, M. & Pei, J. (2012). Data mining concepts and techniques, third edition Morgan Kaufmann Publishers., ISBN: 978-9380931913

Iglesias, F., & Kastner, W. (2013). Analysis of similarity measures in times series clustering for the discovery of building energy patterns. Energies, 6(2), 579-597. doi.org/10.3390/en6020579

Izakian, H., Pedrycz, W., & Jamal, I. (2012). Clustering spatiotemporal data: An augmented fuzzy C-means. IEEE transactions on fuzzy systems, 21(5), 855-868**.** doi.org/10.1109/TFUZZ.2012.2233479.

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. ACM computing surveys (CSUR), 31(3), 264-323. doi.org/10.1145/331499.331504

Kisilevich, S., Mansmann, F., Nanni, M., & Rinzivillo, S. (2009). Spatio-temporal clustering. In Data mining and knowledge discovery handbook (pp. 855-874). Springer, Boston, MA. doi.org/10.1007/978-0-387-09823-4_44

Legany, C., Juhasz, S., & Babos, A. (2006) Cluster Validity Measurement Techniques, Proceedings of the 5th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases. 388-393.

Liu, Y., Xiong, N., Zhao, Y., Vasilakos, A. V., Gao, J., & Jia, Y. (2010). A multi-layer clustering routing algorithm for wireless vehicular sensor networks. IET Communications, 4(7), 810-816. doi.org/10.1049/iet-com.2009.0164.

Madbouly, M. M., Darwish, S. M., Bagi, N. A., & Osman, M. A. (2022). Clustering Big Data Based on Distributed Fuzzy K-Medoids: An Application to Geospatial Informatics. IEEE Access, 10, 20926-20936. https://ieeexplore.ieee.org/abstract/document/9706220

Naik, A., Satapathy, S. C., & Parvathi, K. (2012). Improvement of initial cluster center of c-means using teaching-learning-based optimization. Procedia Technology, 6, 428-435. doi.org/10.1016/j.protcy.2012.10.051

Nayak, J., Naik, B., & Behera, H. (2015). Fuzzy C-means (FCM) clustering algorithm: A decade review from 2000 to 2014. Computational intelligence in data mining-volume 2, 133-149. doi.org/10.1007/978-81-322-2208-8_14.

Nayak, R.**,** Naik, A., Parvathi, K., Satapathy, S. C., Panda, B. S., (2012). QoS Multicast Routing Using Teaching Learning Based Optimization, pp. 49-55. Springer, Berlin (2012), Online ISBN: 978-81-322-0740-5, doi.org/10.1007/978-81-322-0740-5_6.

Nerurkar, P., Shirke, A., Chandane, M., & Bhirud, S. (2018). A novel heuristic for evolutionary clustering. Procedia Computer Science, 125, 780-789., doi.org/10.1016/j.procs.2017.12.100

Pelleg, D., & Moore, A. W. (2000, June). X-means: Extending k-means with efficient estimation of the number of clusters. In Icml (Vol. 1, pp. 727-734). https://web.cs.dal.ca/~shepherd/courses/csci6403/clustering/xmeans.pdf

Rao, R. V., & Savsani, V. J. (2012). Mechanical design optimization using advanced optimization techniques. doi.org/10.1007/978-1-4471-2748-2

Rao, R. V., Savsani, V. J., & Vakharia, D. P. (2011). Teaching–learning-based optimization: A novel method for constrained mechanical design optimization problems. Computer-aided design, 43(3), 303-315. doi.org/10.1016/j.cad.2010.12.015

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics, 20, 53-65. doi.org/10.1016/0377-0427(87)90125-7

Satapathy, S. C., & Naik, A. (2011, December). Data clustering based on teaching-learning-based optimization. In International conference on swarm, evolutionary and memetic computing (pp. 148-156). Springer, Berlin, Heidelberg. doi.org/10.1007/978-3-642-27242-4_18

Satapathy, S. C., Naik, A., & Parvathi, K. (2012a, August). 0-1 integer programming for generation maintenance scheduling in power systems based on teaching learning-based optimization (TLBO). In International Conference on Contemporary Computing (pp. 53-63). Springer, Berlin, Heidelberg.

Satapathy, S., Naik, A., & Parvathi, K. (2012b). High dimensional real parameter optimization with teaching learning-based optimization. International Journal of Industrial Engineering Computations, 3(5), 807-816. doi.org/0.5267/j.ijiec.2012.06.001

Shao, W., Salim, F. D., Song, A., & Bouguettaya, A. (2016). Clustering big spatiotemporal-interval data. IEEE Transactions on Big Data, 2(3), 190-203. doi.org/ 10.1109/TBDATA.2016.2599923

Sheng, Q. Z., Zeadally, S., Luo, Z., Chung, J. Y., & Maamar, Z. (2010). Ubiquitous RFID: Where are we? Information Systems Frontiers, 12(5), 485-490. doi.org/10.1007/s10796-009-9212-x

Shi, Z., & Pun-Cheng, L. S. (2019). Spatiotemporal data clustering: A survey of methods. ISPRS international journal of geo-information, 8(3), 112. doi.org/10.3390/ijgi8030112

Sun, L., Cheng, R., Cheung, D. W., & Cheng, J. (2010, July). Mining uncertain data with probabilistic guarantees. In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 273-282). doi.org/10.1145/1835804.1835841

Theodoridis, S., & Koutroumbas, K. (2008). Pattern Recognition, Fourth Edition; 2008, Academic Press, ISBN-10: 1597492728

Thrun, M. C. (2018). Approaches to cluster analysis. In Projection-Based Clustering through Self-Organization and Swarm Intelligence (pp. 21-31). Springer Vieweg, Wiesbaden. doi.org/10.1007/978-3-658-20540-9_3

Yang, X., Xie, Z., Ling, F., Li, X., Zhang, Y., & Zhong, M. (2018). Spatio-temporal super-resolution land cover mapping based on fuzzy C-means clustering. Remote Sensing, 10(8), 1212. doi.org/10.3390/rs10081212

Yao, L., & Sheng, Q. Z. (2012, October). Exploiting latent relevance for relational learning of ubiquitous things. In Proceedings of the 21st ACM international conference on Information and knowledge management (pp. 1547-1551). doi.org/10.1145/2396761.2398470

Yao, L., Sheng, Q. Z., Gao, B. J., Ngu, A. H., & Li, X. (2013, December). A model for discovering correlations of ubiquitous things. In 2013 IEEE 13th International Conference on Data Mining (pp. 1253-1258). IEEE. doi.org/10.1109/ICDM.2013.87

Yu, Q., & Rege, M. (2010, July). On service community learning: A co-clustering approach. In 2010 IEEE International Conference on Web Services (pp. 283-290). IEEE. doi.org/0.1109/ICWS.2010.47

Zhou, X. C., Shen, Q. T., & Liu, L. M. (2008). New two-dimensional fuzzy C-means clustering algorithm for image segmentation. Journal of the Central South University of Technology, 15(6), 882-887. doi.org/10.1007/s11771−008−0161−1