Improving the Trajectory Clustering using Meta-Heuristic Algorithms

Haiyang Li*, Xinliu Diao
Department of Basic Courses, Henan Polytechnic Institute, Nanyang Henan 473000, China

***Abstract*—The rapid growth of GPS trajectories obscures valuable information regarding urban road infrastructure, urban traffic patterns, and population mobility. An innovative method termed trajectory regression clustering is introduced to improve the extraction of hidden data and generate more precise clustering results. This approach belongs to the unsupervised trajectory clustering category and has the objective of minimizing the loss of local information inside the trajectory. It also seeks to prevent the algorithm from getting stuck in a suboptimal solution. The methodology we employ consists of three primary stages. To begin with, we present the notion of trajectory clustering and devise a distinctive approach known as angle-based partitioning to segment line segments. The evaluation results indicate a significant improvement in the clustering accuracy of the proposed method compared to existing methodologies, especially for a high number of clusters. The HCMGA and HCMMOPSO algorithms have improved clustering accuracy for MBP values by 0.61% and 0.64%, respectively, as compared to previous approaches. Moreover, based on the implementation findings, the ant colony approach demonstrates superior accuracy compared to alternative methods, while the particle swarm method exhibits faster convergence.**

Keywords*—Ant colony method; particle swarm algorithm; HCM clustering; and trajectory lines

I. Introduction

An analysis of the travel characteristics of moving objects, such as automobiles and individuals, can provide insights into travel patterns. This analysis reveals information about people's frequent travel habits, patterns of traffic congestion, and social activity patterns [1]. Travel patterns have been utilized in various domains, such as furnishing decision-making data for urban planning and emergency situations [2], analyzing and optimizing routes to offer personalized travel suggestions for residents, dispatching vehicles [3], and optimizing and selecting stations [4]. These applications can provide valuable insights for urban construction and growth. The authors in Reference [5] proposed a method that takes into account the road network while clustering road segment spatial trajectories. This method was designed to replace density-based clustering and Euclidean-based distance calculations, with the goal of achieving faster and more efficient clustering. The algorithm described in reference [6] is a scalable and efficient density clustering method that utilizes big data computing. In addition, the authors of Reference [7] introduced an enhanced density-based technique specifically designed for clustering stops in trajectories. The study described in Reference [8] introduced a clustering approach based on anisotropic density (angle-based standard deviation). This algorithm was utilized to identify spatial point patterns along with any accompanying noise.

Typically, clustering techniques can be classified into many categories such as density-based, partitioning-based, grid-based, hierarchical-based, and graph-based. These algorithms are extensively used in spatial data processing for a variety of purposes. In addition, each of these categories encompasses other renowned clustering algorithms, such as partitioning-based K-means, K-median, and fuzzy C-means (HCM), each with its own distinct advantages and disadvantages. Specifically, density-based clustering algorithms are commonly employed to extract concealed information from a given dataset and process any GPS datasets, since they are especially well-suited for identifying clusters with irregular forms and identifying clusters that do not overlap [9, 10, 11]. Nevertheless, managing the intersecting clusters (such as trajectory crossover) becomes challenging when dealing with fuzzy clusters and the absence of localized trajectory information. Furthermore, it is responsive to both the specified neighborhood and the density, as determined by the value of MinPts. This research specifically examines partitioning-based techniques, such as HCM. Nevertheless, they exhibit certain limitations, such as being susceptible to the choice of initial cluster centers, delayed convergence, and a propensity to get trapped in local optima. In this paper, a new trajectory regression clustering technique is introduced. The technique is based on partition clustering and combines the AngPart method, which generates line segments based on angles, with the HCML algorithm, a Lagrange-based fuzzy C-means clustering algorithm, and the LSR model, which is used to create an unsupervised trajectory clustering method. This method is an alternative to using a map-based knowledge base. Specifically, HCML is an innovative approach for clustering regression data without the need for supervision. Initially, a line segment partitioning technique is developed to generate line segments using three GPS data points. This method efficiently preserves the local information of trajectories. The novel clustering algorithm presented in this study combines a unique fuzzy C-means (NHCM) with the Lagrange operator [12] and Hausdorff-based K-means++ [13]. The NHCM is employed to cluster line segments, while K-means++ is utilized to generate the initial cluster centers of the line segments. This combination aims to capture the global optimum and prevent the algorithm from getting trapped in local optima. The original fuzzy C-means (HCM) technique is a clustering approach based on partitioning [14]. In this algorithm, the computation of distances between line segments requires the use of Hausdorff distances instead of Euclidean distance [15]. Once the concealed GPS data is extracted and acquired, the LSR (Least Squares Regression) method is utilized to perform trajectory regression. The objective is to regress and generate trajectories based on the clustering results, without relying on a map-based knowledge base. These trajectories can then be used to analyze and describe various urban aspects such as people, vehicles, roads, traffic flow, and serve as a reference for road planning.

Background: HCML can enhance line segment partitioning and maintain the local information of trajectory by employing the angle-based technique prior to the clustering process. For instance, when two GPS data points are produced as line segments, it becomes challenging to articulate the local information between GPS data points and to comprehend the connection between consecutive GPS points, such as the alteration in steering and intersecting angle.

Furthermore, the method being discussed, HCML, is a form of unsupervised learning. Hence, in cases when a map-based knowledge base is not required, the least squares regression model (LSR) is employed to generate the trajectories of the clustering outcomes.

Problem: In order to assess the performance and efficacy of HCML, an actual GPS dataset from Beijing, China is employed as an experimental test. The experiments involve comparing HCML with K-median, K-means, and HCM clustering methods using the MBP (Pakhira-Bandyopadhyay-Maulik)-index [26] cluster evaluation criteria. This is discussed in Section 2. The MBP-index is an effective unsupervised evaluation tool [27,28]. However, it is important to be aware that distances in the MBP-index necessitate the utilization of the Hausdorff method for calculating the distance between the center of line segments and other line segments. Furthermore, LSR is employed to accomplish the regression of the clustering outcomes. The experimental findings demonstrate that HCML outperforms K-means, K-median, and HCM algorithms in terms of trajectory regression quality (refer to Section 5).

Proposed Solution: Thus, the primary content of the study is succinctly outlined as follows:

1) The angle-based partitioning method (AngPart) is offered as a strategy for generating unique line segments.

2) This study proposes a novel clustering technique called fuzzy C-means (NHCM), which integrates the Lagrange operator with AngPart and K-means++.

3) This paper introduces a trajectory regression technique that utilizes least squares regression (LSR) to analyze population movement patterns along trajectories. The technique provides insights into the state of population migration and can serve as a valuable reference for urban road planning.

4) HCML has been demonstrated to be effective when used to actual taxi GPS data in Beijing, China.

The subsequent sections of the paper are arranged in the following manner. Section II provides a description of taxi GPS data in Beijing, China. Section III presents the angle-based normalization method employed for the taxi GPS data. Section IV introduces a trajectory regression technique that combines HCML with LSR. Section V outlines the tests and offers the data for evaluating the effectiveness of the recommended procedures. Section VI serves as the final part of the report, providing a conclusion and proposing potential future research.

II. Related Works

The integration of information technology into transportation systems is currently a prominent trend. This is because it effectively addresses key issues faced by traffic operators, such as traffic congestion and accidents. Therefore, observing the traffic situation is essential for traffic operators, particularly at intersections [16]. The traffic data obtained from the monitoring system is frequently extensive, necessitating diligent attempts to identify noteworthy trends within it. These patterns provide valuable insights into vehicle movements and facilitate the detection of any deviant conduct that may result in traffic disputes. Nevertheless, it will be an arduous task for traffic operators to manually monitor the movement of vehicles at a crossroads, especially when there are thousands of vehicles passing through. Therefore, the process of grouping vehicle trajectory data to identify comparable patterns is carried out using the k-means and fuzzy c-means (HCM) clustering algorithms. Since various clustering techniques necessitate the input parameter of the number of clusters, this research focuses on studying the appropriate number of clusters for the clustering process [17].

Analyzing urban travel patterns helps assess the regularity of inhabitants' mobility, offering information for urban traffic planning and emergency decision-making. Clustering techniques have been extensively utilized to uncover concealed insights from extensive trajectory data concerning trip patterns. Implementing soft constraints in the clustering process and statistically evaluating their performance remains a challenging task. This paper introduces a refined trajectory clustering approach, known as TC-FDBSCAN, which utilizes fuzzy density-based spatial clustering of applications with noise to perform classification on trajectory data. Initially, we establish the trajectory distance by taking into account various features and determining the corresponding weight factors to quantify the similarity between trajectories [18]. In the HCM clustering method, membership degrees and membership functions are created to extend the standard DBSCAN method.

A modified version of the fuzzy c-means algorithm is proposed to solve the inverse kinematics and trajectory planning problem for a redundant manipulator, while taking into account performance criteria. A novel HCM clustering approach is introduced, which utilizes a newly developed generalized validity index. This index is built on weighted within-scatter metrics and between-cluster scatter metrics specifically designed for the manipulator. The issue of redundant manipulator, which refers to a nonlinear system with several inputs and outputs, has not been previously addressed using the clustering method. The trajectory planning algorithm for the manipulator is simulated using Matlab. The entire process, starting from gathering data to verifying the model, is demonstrated using a robot manipulator with four degrees of freedom. The simulated results are being compared to the numerical approaches used for trajectory planning. The findings are visually depicted. The proposed method offers the benefits of being straightforward, adaptable, and exhibiting excellent tracking capabilities [19].

The mining of trajectory databases (TD) has garnered significant attention as a result of the widespread use of tracking devices. However, the mining procedure has not yet included the presence of uncertainty in TD, such as GPS inaccuracies. This work examines the impact of uncertainty in TD clustering and presents a three-step methodology to address it. Initially, we give a conceptual framework for representing trajectories using intuitionistic point vectors, which captures the inherent uncertainty. Additionally, we introduce a robust distance metric to handle this uncertainty. Furthermore, we provide CenTra, an innovative approach that addresses the challenge of identifying the centroid trajectory of a set of moves. Furthermore, we provide a modified version of the fuzzy C-means (HCM) clustering algorithm that incorporates CenTra during its updating step. The empirical assessment on real-world TD substantiates the efficiency and efficacy of our methodology [20].

This research examines the application of HCM clustering to identify probable spatial patterns by incorporating rough set and fuzzy set theory. Initially, we suggest a rapid technique for measuring similarity by utilizing the approximate distances between trajectories. Significant reductions in processing time would be achieved, particularly for lengthy trajectory sequences. In addition, we present a summarization strategy that minimizes the number of distance calculations needed for similarity measurement. Furthermore, the membership degree is modified in order to enhance the clustering quality and performance. Furthermore, we enhance the fuzzy C-means algorithm by using a novel similarity measure and the membership degree function. The efficacy of our methods is demonstrated by experimental findings obtained from two actual datasets of trajectories. These results involve the assessment of clustering validity and the computation performance for big datasets. The computing performance of the proposed HCM clustering algorithm exhibits a clear improvement as the size of the trajectory dataset rises [21].

Clustering trajectory data is a method used to identify and display the underlying structure in the movement patterns of mobile objects. This technique has a wide range of possible applications in fields such as traffic control, urban planning, astronomy, and animal research. This study introduces a novel method for grouping trajectory data using a Particle Swarm Optimization (PSO) methodology. The strategy takes into account the Dynamic Time Warping (DTW) distance, which is a widely-used measure for comparing trajectory data [22]. The suggested technique may identify the (almost) optimal number of clusters and the (almost) ideal cluster centers during the clustering process. In order to enhance the performance of the suggested method, a Discrete Cosine Transform (DCT) representation of cluster centers is utilized. This approach helps to minimize the dimensionality of the search space and improve the method's performance in relation to a certain performance index. The suggested method can incorporate different cluster validity indices as the objective function for optimization. The experimental findings, obtained from both synthetic and real-world datasets, demonstrate the improved performance of the proposed technique compared to fuzzy C-means, fuzzy K-medoids, and two evolutionary-based clustering techniques previously suggested in the literature [23].

Next, the NT algorithm utilizes the characteristics of noise in order to actively reduce the impact of noise [24]. In 2022, a method was provided for grouping ship trajectories at sea. This approach utilized the Douglas-Peucker compression technique and the DBSCAN algorithm. The arrangement of this data defines the dispersion of traffic volume and the customary path for ship passage. Initially, the appropriate parameters are derived for the Douglas Poker method by analyzing the alterations in the ships' trajectories. This is done with the aim of enhancing the clustering of the ships' trajectories. The DTW distance matrix is calculated using these parameters to compress the data obtained from the traffic lines. Next, the enhanced DBSCAN algorithm is utilized to accomplish density-based clustering. The DBSCAN algorithm selects its optimal settings by considering the statistical properties of the distribution of ships' routes [25].

III. Proposed Method

The suggested approach will present the path utilizing an innovative pre-processing technique. This method utilizes the angle formed between the lines as a reference to determine the linear areas in a specific order based on the sequence of points. Subsequently, an attempt is made to employ evolutionary approaches to better the performance of particle swarms and ant colonies in order to address the limitations of the HCM clustering technique. The overall framework of the suggested technique is illustrated in Fig. 1.

A. *Using GPS Coordinates to Determine Starting and Ending Places*

The trajectory in the proposed approach comprises the GPS coordinates specified by the user for both people and moving vehicles. These figures represent data about a motion that starts at one point and concludes at another. During this process, sampling is conducted at regular intervals, often less than two minutes, and the spatial data of the moving item is recorded [26]. This enables the collection of data related to a sequence of locations for the mobile entity (refer to Fig. 2). The suggested technique utilizes route information represented as Ti = {(p, a)} = {(p1, a1), (p2, a2), (pi, ai)}, where pi is a pair consisting of latitude and longitude coordinates, and a represents the angle formed by the line from the current point to the next point.

B. *Tracing the Pathways of GPS Data*

The results pertaining to the clustering which involves the regionalization of the trajectories are comparable. Fig. 3 illustrates an instance of the clustering of garlic lines.

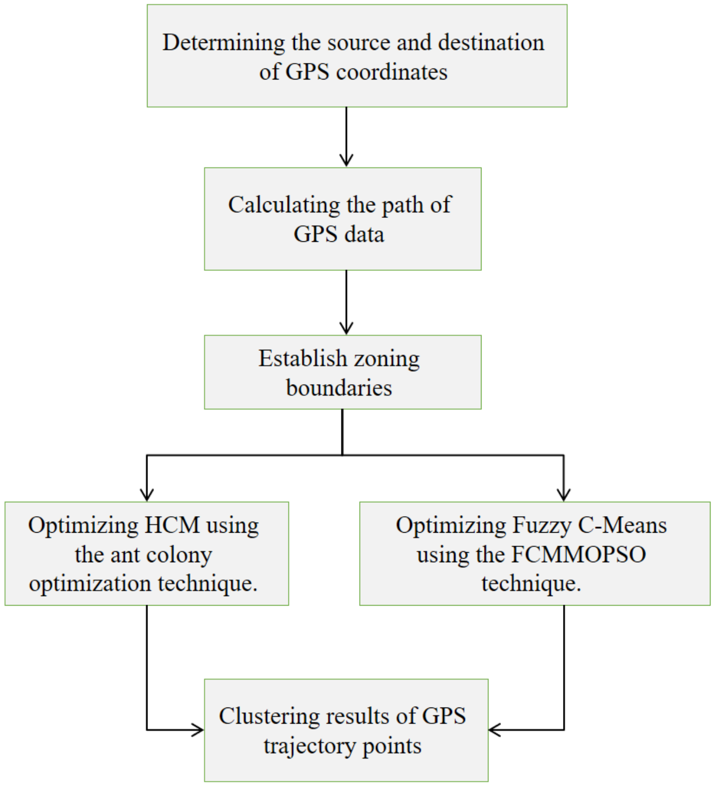

Fig. 1. General structure of the proposed method.

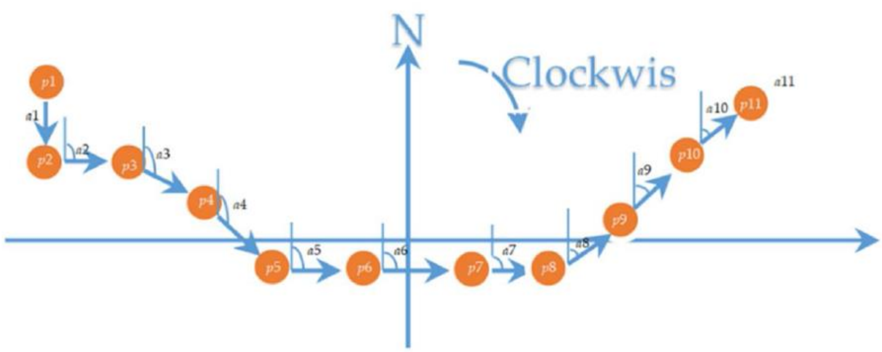

Fig. 2. A GPS moving point example with an angle-based trajectory.

C. Similarity Criteria

Among the newly-introduced criteria for comparing the similarity of linear sections is the cosine similarity criterion. Similarity results of linear regions form the basis for both trajectory clustering and sub-trajectory computing. The proposed method expresses the trajectory of Lj as a function of angle changes. The similarity criterion is also produced by relationship (1 [27].

$$sim(S_{j'}, S_j) = \begin{cases} e^{\forall l} \times \frac{e^{\partial g} - e^{-\partial g}}{e^{\partial g} + e^{-\partial g}} & S_{j'} \neq S_j \\ 1 & otger \end{cases} \quad (1)$$

The results pertaining to the clustering which involves the regionalization of the trajectories are comparable. Fig. 3 illustrates an instance of the clustering of garlic lines.

D. Establish Zoning Lines

The process of dividing the two-part line areas is detailed here. Every possible angle for the specified GPS locations is computed and saved in normal mode when the connecting line between two consecutive points forms a right angle. The angle between the lines and the vertical line is considered. First, we find the tenth neighboring point's shortest distance from the selected point. The sums of these points can vary. This leads to the identification of a region that matches the three nearest existing places, as illustrated in Fig. 4. If, among these three locations, the largest angle measures more than 180 degrees, the guiding angle is computed counterclockwise. In cases when the maximum angle falls short of 180 degrees, the guiding angle is calculated in the opposite direction. The values associated with the angle are changed according to relations (2) and (3) to create a standard that is both normal and harmonious [28, 29, 30].

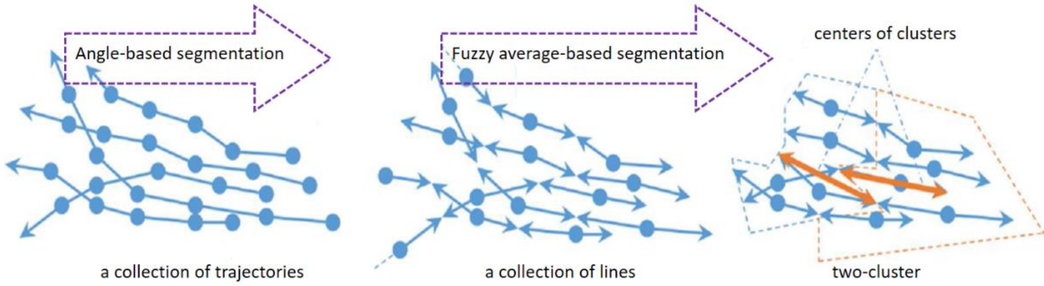

Fig. 3. Clustering of trajectories.

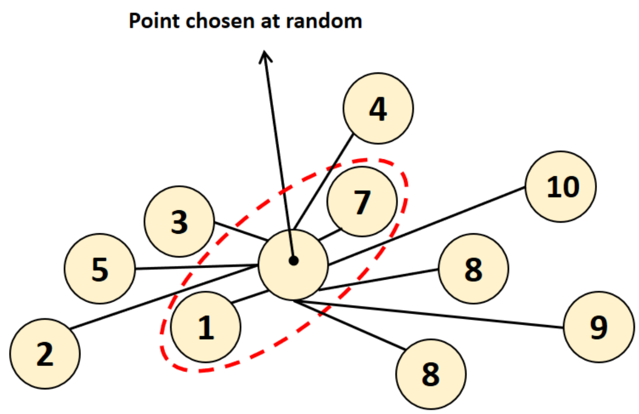

Fig. 4. The process of choosing three adjacent locations.

$$Nita = 2\gamma - (Nita_2 - Nita_1) \quad (2)$$

$$Nita = Nita_2 - Nita_1 \quad (3)$$

usage of angle-based segmentation necessitates limiting the intersection angle of the lines, denoted as $\gamma_t = (f = 1,2,\ldots,e)$ The theorem of cosines, as stated in equation (4), can be used to determine the angle of the intersecting lines when three GPS points are represented as (S^-, S, S^+) and P is taken as the vertex. (a) Point P is initially selected at random. (b) The vertex is selected at this position, and two nearby points are identified by calculating the lowest distance. (c) If the values of If γ <T, where T is the threshold value, are extracted from the data set and saved in memory, then step 4 is carried out; otherwise, we return to step 1. (d) In order to build linear regions, GPS points are traveled until each point is checked. Using angle-based segmentation and cosine-based limiting, linear areas are segmented in this part. Depending on the angle between them, a collection of journey lines tied to human or point device movement based on GPS points is expressed as specific lines. A greater amount of data is available for grouping by the trend lines.

E. HCM Algorithm-based Traffic Line Clustering

In order to tackle this problem, researchers are investigating meta-heuristic algorithms. The HCM method's implementation and assessment do not effectively optimize the output of the objective function algorithm, resulting in the clustering centers being situated in local optima. Using meta-heuristic approaches improves the HCM algorithm. The defining feature of optimization and random search algorithms that rely on collective intelligence is the collective behavior and self-organization of individuals inside the community. This work involves the grouping of trajectory data using a combination of classical and meta-heuristic clustering algorithms.

F. Using the Particle Swarm Technique to Optimize HCM Clustering

The clustering problem is reframed as an optimization problem in the proposed strategy. In equation (4), C is a specific clustering of the data set f of the optimization function, and the optimization problem is specified as the set of all possible optimal clusterings (D*= {D1, D2,..., Dk}). C* is the best clustering that an iterative algorithm can provide [31].

$$E(D^*) = \min E(D); \quad (4)$$

The particle swarm optimization approach is employed to get an optimal segmentation. The HCM algorithm is a clustering technique that relies on an objective function. The objective function, denoted by equation (5), is defined for the data set Y = {Y1, Y2, ..., Yn} of dimension s.

$$I_n = \sum_{i=1}^{p} \sum_{j=1}^{r} \mu_{lb}^{n} \, \|P_i - d_r\|^2 \quad (5)$$

When determining the distance and degree of similarity of data with the center of the cluster, Let S represent the number of linear segments extracted in the preceding phases. D represents the number of clusters. m specifies the fuzzy degree of overlap of the clusters. The value of µid, which is the degree of the membership function Dj for cluster k, is determined using equation (6) [32].

$$\mu_{id} = \frac{1}{\sum_{d'=1}^{d} \left(\frac{\|P_j - E_d\|}{\|P_j - E_{d'}\|} \right)^{\frac{2}{n-1}}} \quad (6)$$

The suggested methodology utilizes HCM clustering, employing the particle swarm optimization technique. In particle swarm optimization, the fitness function is replaced by the objective function in HCM clustering. The sequence of actions performed by the algorithm is as follows:

- Each particle in the particle swarm optimization process is assigned a membership function ujk, which is determined based on equation (6).
- Equation (7) [33] reevaluates the centroids of the categories determined using HCM clustering.
- The fitness function of the present particle is defined by Equation (5).
- The user's text is a bullet point. The ultimate outcome is the global optimum, which represents the most optimal solution when taking into account all the particles. This clustering represents the highest level of optimization and is considered the global optimum.

$$W_i(d+1) = \frac{\sum_{i=1}^{m} y_{ij}^{n}(d) Y_i}{\sum_{i=1}^{m} y_{ij}^{n}(d)} \quad (7)$$

G. Optimization of HCM Clustering by Ant Ga Algorithm

As mentioned earlier, HCM clustering aims to determine the appropriate values for cluster centers and membership function values in order to achieve optimal clustering and minimize the desired objective function. Currently, there exist two separate optimization concerns. The data is subsequently divided into d clusters. The values of the pheromone matrix p are responsible for performing this task. Following each step, the pheromone matrix p undergoes modification, and the updated values for ujk are computed using equation (6) and the cluster centers. The pheromone levels are updated using Equation (8) [31].

$$s_{id} = s_{id} \times (1-\partial) + (y_{id}/(I - I_{min} + \epsilon)^{\partial} \quad (8)$$

for cluster d's route i, the pheromone concentration is denoted as s_{id}. While j is the number of samples, d is the number of clusters. The variables ϵ, which precludes division by zero, ∂, which controls the pace of convergence, and ∂, which controls the amount of pheromone evaporation, are all parameters in the system. The Algorithm 1 uses an ant colony method to show the structure of HCM clustering optimization. Included in this algorithm are the input data set Y, the fuzzy power m, the number of clusters c, and the maximum number of steps needed to run the ant colony process.

Algorithm 1: Optimization of HCM clustering by GA algorithm

dataset%
Initialise Y, d, n;
Initialise Jmin = inf, **Y** = 0
Initialise GA parameters - smax, ∂, _, β;
Initialise the pheromone matrix, **S**, using Eq. (5);
for t = 1 to smax **do**
  repeat
    for k = 1 to m do
      With probability $s_{jd}/\sum_{i=1}^{s} s_{jd}$
      set j^{th} cluster membership value = 1,
      for i ≠ j, set the membership value= 0;
    end for
  until there is at least one point per cluster;
  Using Eq. (3), compute centroids **W**;
  Using Eq. (2), compute the new fuzzy membership matrix, **Y**;
  Using Eq. (3), compute the new centroids **W**;
  Using Eq. (1), calculate objective function I;
  if $I < I_{min}$ then
    $I_{min} = I$;
  end if
  Using Eq. (4), update pheromone matrix **s**;
end for

- Data Collection:

A large dataset consisting of GPS trajectories of cabs in operation in a big Chinese city is used in the study. All sorts of spatial and temporal dimensions are present in the dataset, which records cab travels over a long period of time. For accurate data representation, the trajectories are sampled frequently. Each trajectory point is annotated with details like timestamp, latitude, and longitude. These details are used for further research.

- Preprocessing:

The trajectory data is thoroughly preprocessed to remove noise, outliers, and missing values before grouping. In order to make the following clustering more resilient, we exclude outliers found via spatial and temporal analysis. To preserve the flow of time, values that are missing are filled in by extrapolating from nearby trajectory points. In order to simplify the dataset for efficient clustering, noise reduction techniques are used, such as trajectory simplification.

- Evaluation Metrics:

The suggested algorithms are evaluated using a suite of thorough metrics. The silhouette score, the Davies-Bouldin index, and the internal cluster cohesion are all examples of cluster validity indices. The agreement between the algorithm-generated clusters and ground truth clusters is also measured using external validation measures, such as the adjusted Rand index and the Fowlkes-Mallows index.

- Experimental Setup:

A computer platform with industry-standard hardware is used to conduct the experiments. Languages like Python are used to implement the algorithms, with libraries like scikit-learn being utilized for optimization and HCM clustering. In order to examine how well an algorithm performs when applied to new scenarios, the researchers used a stratified sampling method that split the dataset in half.

IV. Evaluation and Simulation

Here we discuss how to put the proposed idea into practice and evaluate it. The proposed methodology includes an optimization strategy for grouping motion trajectories. This was accomplished by optimizing the HCM clustering method using the evolutionary techniques of particle swarm and ant colony optimization. Throughout the implementation process, the MATLAB environment was utilized. The clustering procedure on the real data set containing the routes of the passenger-transporting taxis was improved in two separate circumstances using two optimization methods based on this. The primary clustering algorithm was used in both cases. The itineraries incorporate the regularly acquired GPS location data of the taxis. The results related to the evaluation criteria are recorded and inputted into the Excel software at every stage of execution. The desired charts are extracted by Excel program.

A. Datasets

On routes that convey passengers, the suggested strategy has been tried. Data is stored using GPS coordinates that are sampled every two minutes. This is completely accurate and applies to a lot of cities in China. The data utilized pertains to the whereabouts of Beijing taxis. On March 22, 2017, during the hours of 7:50 and 7:59, the data pertains to the movements of thirty thousand taxis. We acquired the information about the origin and destination of taxis and ran additional analyses on these sites, as mentioned in reference [34]. In Fig. 5, we can see the 69514 data point pairings that include origin and destination points. With these two coordinates, you may draw out patterns of trajectories. Fig. 6 shows the results of applying the angle-based segmentation method to these points and trajectories; the resulting data set contains 13584 linear segments. Fig. 5 shows the spatial representation of the distribution of pairs of starting and ending points within a latitude and longitude range of [0.19×0.4]. Passenger traffic in a certain location of Beijing is represented by these figures. The data in question might provide a treasure trove of information useful for traffic management and control. When constructing new streets or tearing down or fixing old ones, these numbers can also be highly useful. We also run tests of the proposed methodology on these data to see how well it works.

B. Clustering Evaluation Criteria

Clustering algorithms can be evaluated using a variety of criteria; however, trajectory clustering requires its own set of criteria due to its distinct characteristics. Considered with other evaluation criteria such as the DB-index [14], Dunn's index [25], and XB index [26] for assessing the clustering of geographical patterns, the MBP-index [15] emerges as the most accurate and relevant. Use this metric to compare the proposed approach to other clustering techniques. The value of MBP for clustering with K classes is represented by Equation (9).

$$MBS(D) = \left(\frac{1}{D} \times \frac{F_1}{F_D} \times K_D\right)^2 \tag{9}$$

where K_D and F_D are calculated from relations (10) and (11) respectively.

$$F_D = \sum_{d=1}^{D} \sum_{j=1}^{S} Hausd(C_j, s_d) \tag{10}$$

$$K_D = max_{i,j=1}^{D}\{k(s_i, s_j)\} \tag{11}$$

Where $k(s_i, s_j)$ is the distance between two cluster centres and $Hausd(C_j, s_d)$ is the distance between the sub-trajectory of C_j and the cluster centre c_k. The segmentation of the data location is not crucial, and pre-specifying it is unneeded, since the MBP evaluation criterion is considered an unsupervised evaluation criterion. That is to say, the results of the clustering routes are unrelated to the road map or the actual roads.

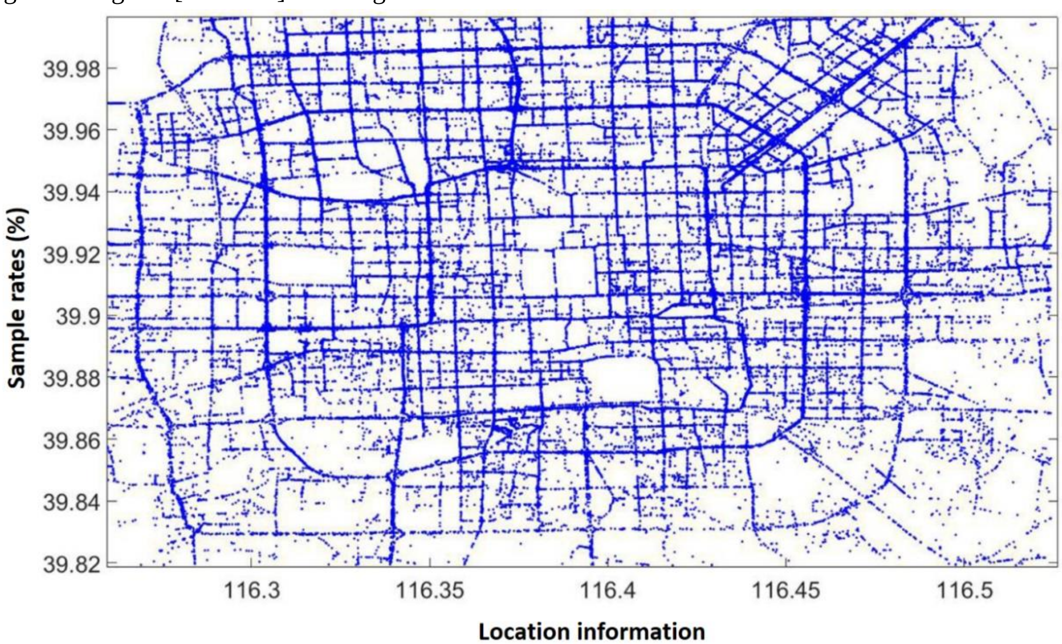

Fig. 5. Points of origin and destinations for taxis and their locations.

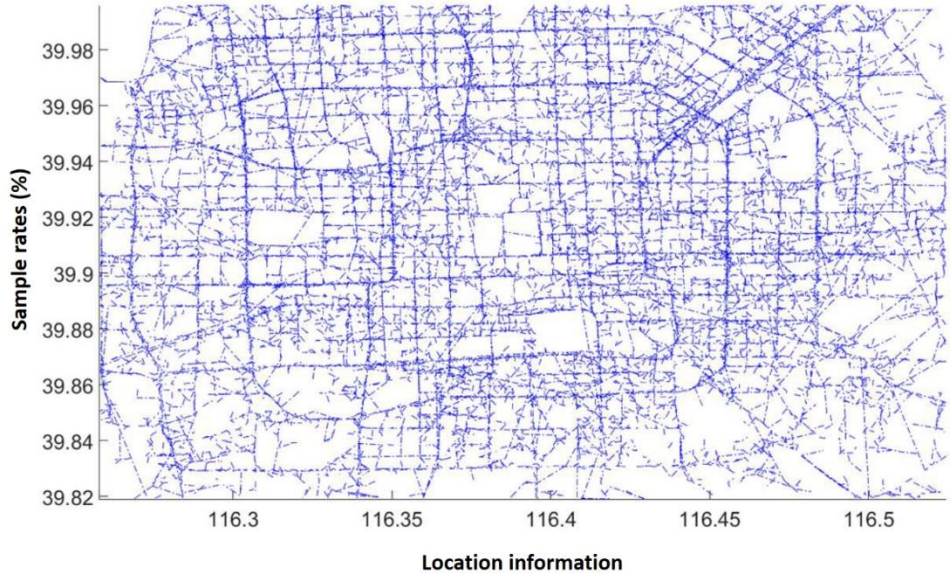

Fig. 6. The use of angles for movement line segmentation.

C. Methods for Comparing and Assessing the Proposed Approach

Before introducing an algorithm, it is essential to choose the correct method to evaluate and compare it. All of the method's pros and downsides should be taken into account in fair evaluations. To implement a novel approach to grouping motion routes based on optimisation of the HCM clustering algorithm, the ant colony algorithm and two-particle swarm optimisation methods were used separately in the proposed method. Consequently, the performance of each part is examined and evaluated independently. But these two suggested algorithms are tested against HCM, KMeans, and basic HCM algorithms to see how well they perform [35, 36]. An innovative approach to classifying motion trajectories, the HCM algorithm has been refined from its foundation in the cubic regression model and the Lagrange equations. In what follows, you will see the results of these comparisons according to several criteria.

V. Evaluation of MBP Value for Different K's

When evaluating the clustering process, one of the most important parameters is the number of clusters, denoted as K. Experiments have been carried out for K values ranging from 10 to 80. What this means is that the number of clusters, as defined by the proposed algorithm and other methods, is determined before the clustering process even begins. One important part of clustering techniques is setting up the cluster centres [37]. When AGNES or HCM performs a clustering method with K clusters, the centres of these clusters are originally chosen at random, and this selection process affects the clustering results. Therefore, in order to deliver a more precise assessment of the proposed method and other results, each algorithm is executed in a 20-step process. Table I displays the results of the minimum, maximum, and average MBP values for different values of K for all methods.

TABLE I. Clustering Outcomes Compared for Varying K-Values

	AGNES	HCM	FCML	HCMMOPSO	HCMGA
			K=10		
Max	0.070	0.079	0.089	0.0770	0.076
Mean	0.057	0.079	0.086	0.0774	0.075
Min	0.043	0.078	0.084	0.0760	0.074
			K=20		
Max	0.043	0.050	0.057	0.0670	0.059
Mean	0.037	0.050	0.057	0.0665	0.058
Min	0.030	0.050	0.056	0.0660	0.058
			K=40		
Max	0.032	0.036	0.040	0.041	0.043
Mean	0.025	0.034	0.039	0.041	0.042
Min	0.019	0.033	0.038	0.041	0.052
			K=80		
Max	0.025	0.031	0.034	0.036	0.039
Mean	0.022	0.031	0.033	0.035	0.039
Min	0.017	0.030	0.032	0.034	0.038

Fig. 7 to 10 display the overarching findings from comparing the suggested method to alternative ways for various K. By comparing the outcomes produced by the particle swarm optimisation and ant colony optimisation approaches, we can observe that the suggested method outperforms HCM, albeit to a lesser extent, for small k, i.e., when K = 10. The results show that HCMMOPSO outperforms HCMGA, the other recommended approach. However, when K is larger, the two scenarios of the suggested method outperform HCM and the two fundamental methods. Getting caught in the local minimum becomes more of an issue and the optimisation problem becomes significantly more hard as the number of clusters increases. However, the suggested approach combines the clustering algorithm with optimisation methods in an effort to achieve optimal clustering. Furthermore, when K grows, the HCMGA scenario outperforms HCMMOPSO. We will further study the ant colony optimisation approach, which outperforms the particle swarm with respect to issue complexity but has slower convergence.

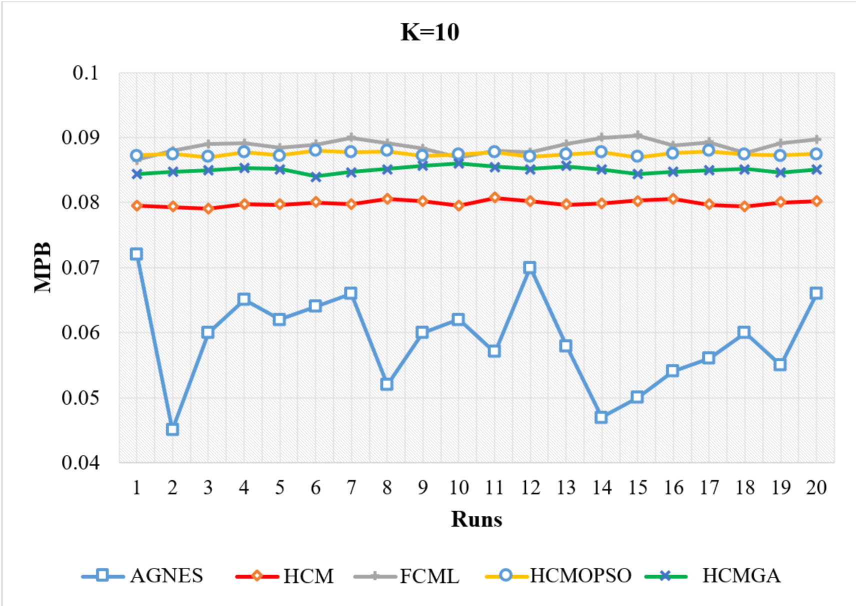

Fig. 7. Evaluation of MBP values during 20 iterations with K=10.

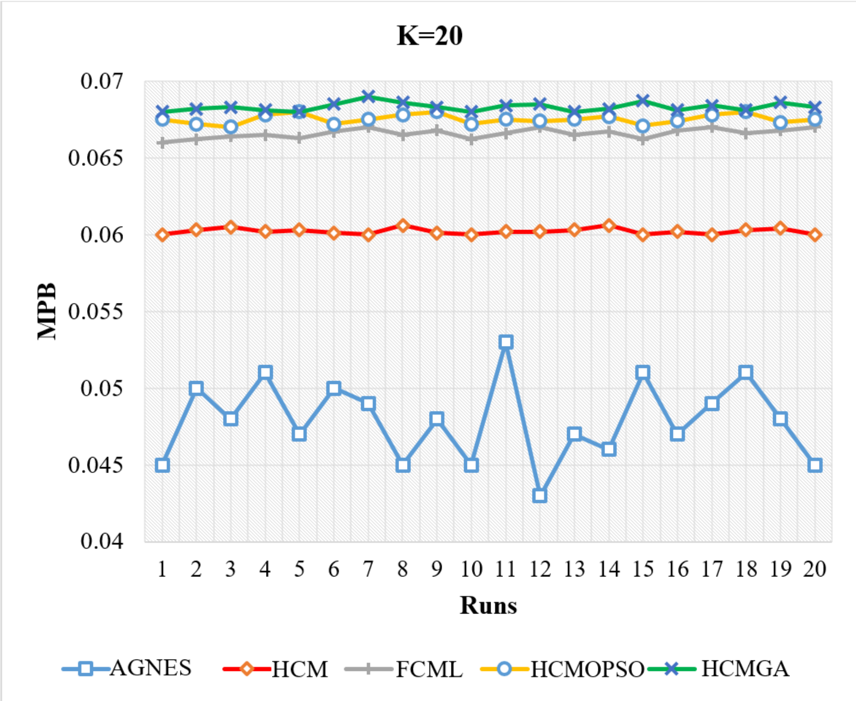

Fig. 8. Evaluation of MBP values during 20 iterations with K=20.

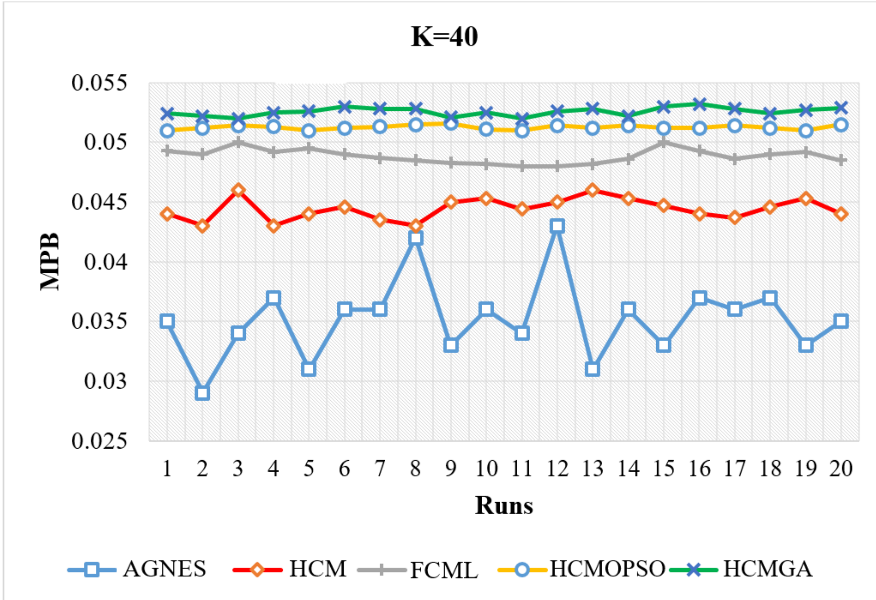

Fig. 9. Evaluation of MBP values during 20 iterations with K=40.

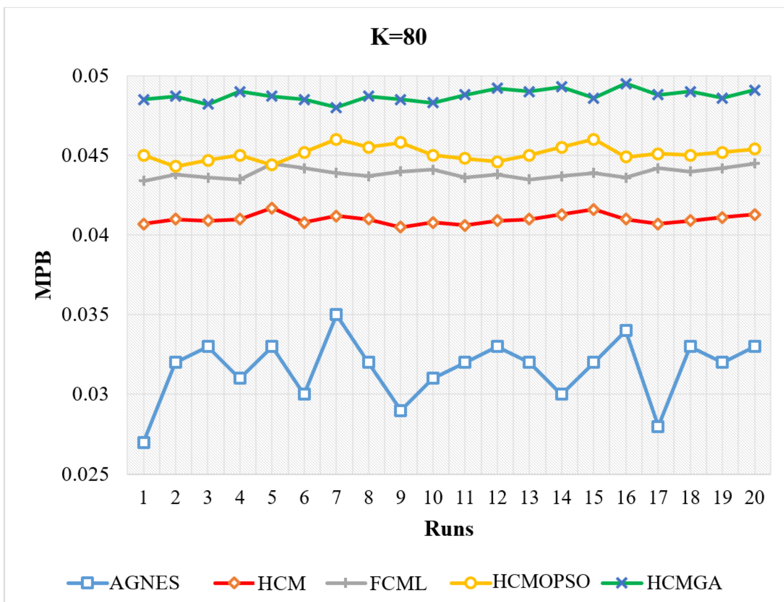

Fig. 10. Evaluation of MBP values during 20 iterations with K=80.

A. *Clustering Procedure Termination Criteria*

Clustering techniques iteratively update the cluster centres at each iteration to obtain the best clusters. But the most important thing is to figure out when this process should conclude. There is more than one way to finish the job. For example, it is possible to perform the same number of clustering steps if the initial number of steps is fixed. This strategy, however, appears to be somewhat illogical. On the other hand, you may consider the evaluation function and how the algorithm's efficiency is improving at each stage as an alternative approach. This value must remain over a specified threshold for the stages to be executed; otherwise, they will not be executed. The proposed method use equation (12) to halt the clustering procedure.

$$\frac{I(h+1)-I(h)}{I(h)} < \varepsilon \tag{12}$$

In the proposed approach, the threshold value is assumed to be 0.002, and I(h) represents the value of the evaluation function at step h. The MBP standards are used to compute the I(h) value. There will be no end to the clustering processes until relation (12) is created. Nevertheless, several algorithms can achieve this with different iterations. As the algorithm gets better at meeting these requirements, its convergence speed goes up. Convergence speed alone will not be enough to achieve better performance, though. Since this rapid convergence may also lead to the local optimum, it demands consideration. As a result, optimal performance and rapid convergence are compatible. In order to compare the convergence rates of the two situations and different techniques, we considered the execution stage of the algorithms as 100 steps and used equation (13) to get the normalised value of I(h). This criterion is used to compare the rate of convergence; after normalisation, its value will range from zero to one.

$$\frac{I-I_{min}}{I_{max}-I_{min}} \tag{13}$$

Where are the maximum and minimum values of the evaluation function, I_{max} and I_{min}, respectively. Fig. 11 to 14 show a comparison of the results for different cluster densities with respect to the rate of convergence. For the case where there are precisely 10 clusters, this comparison is displayed in Fig. 14. The AGNES approach converges rapidly, as seen, but the best result it produces is inacceptably low for our issue. Steps 5 and 6 mark the end of the threshold technique, which, when applied to the question of whether the job is done, produces a locally optimal value. As a basic clustering algorithm, the HCM approach converges at the slowest rate. The HCM method, however, differs in its performance from the two alternatives. The HCM algorithm does not converge as quickly as HCMMOPSO. Almost as fast as HCMGA's convergence speed is HCM. So, it's safe to say that the particle swarm method achieves the fastest convergence when applied to clustering optimisation. With PSO's convergence speed, this conclusion was close. Nevertheless, the HCMGA method outperforms the optimal value for the clustering process.

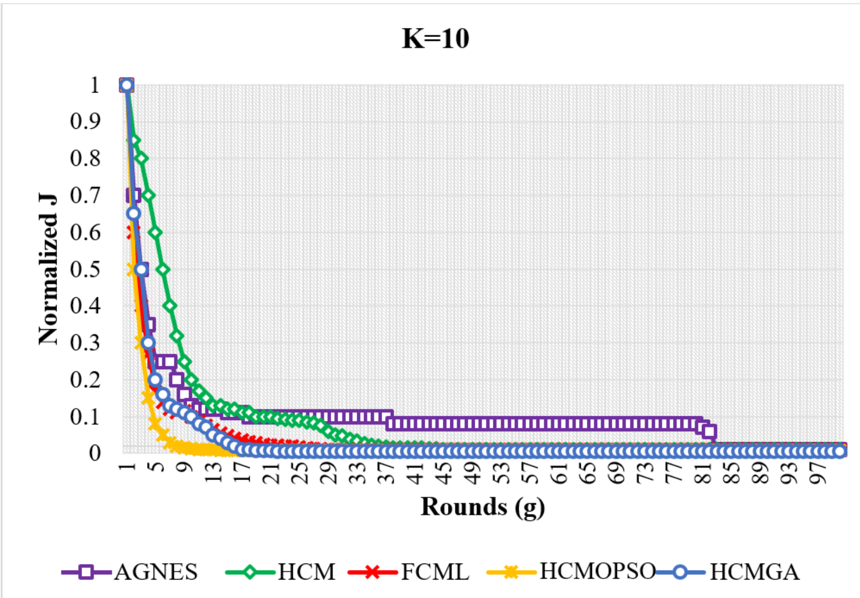

Fig. 11. Speed of convergence comparison for K=10.

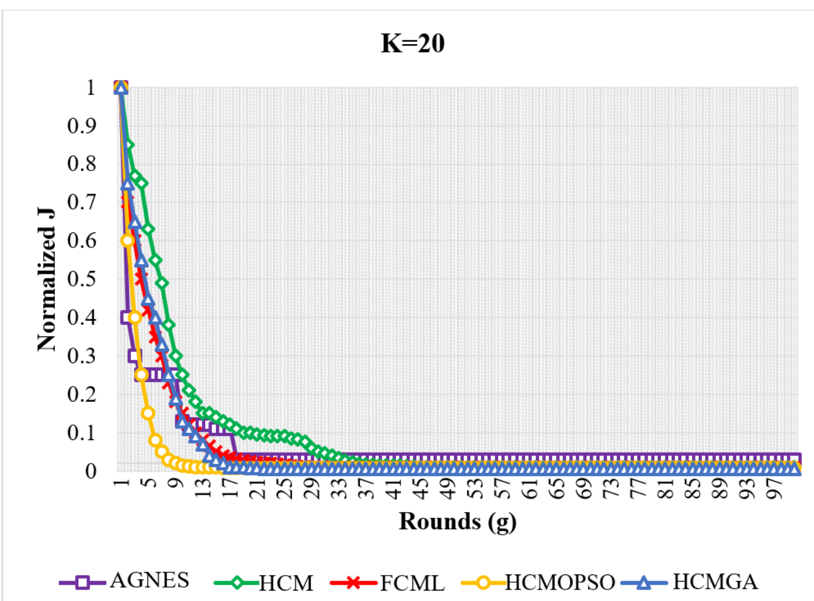

Fig. 12. Speed of convergence comparison for K=20.

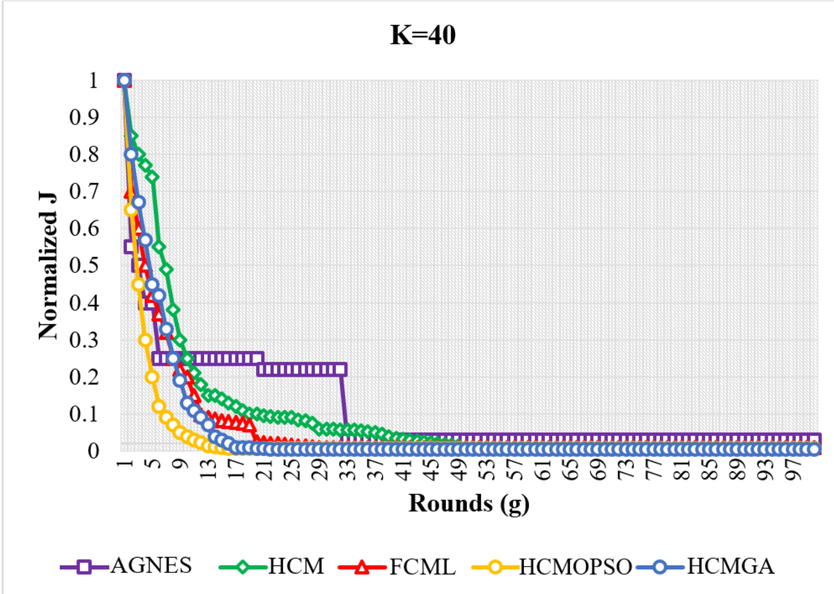

Fig. 13. Speed of convergence comparison for K=40.

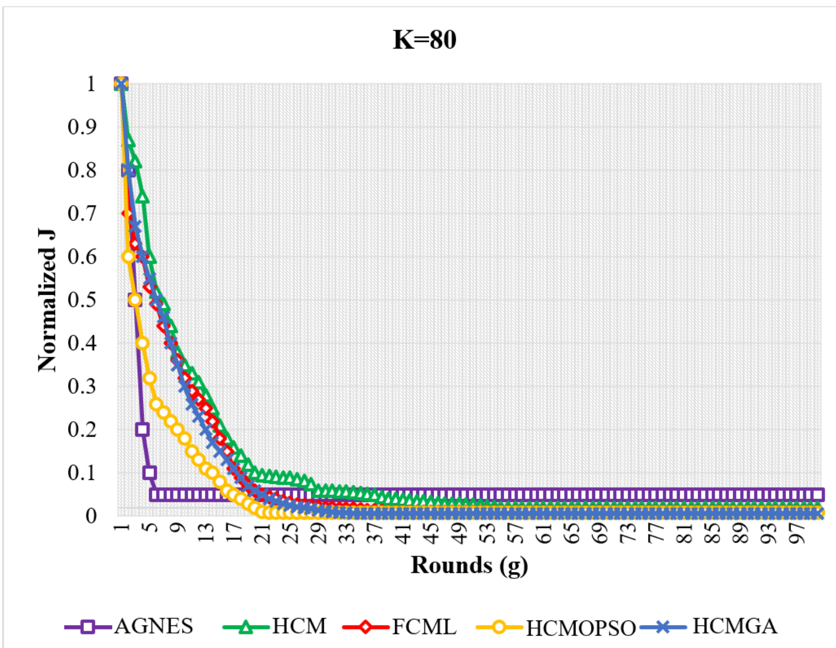

Fig. 14. Speed of convergence comparison for K=80.

The convergence process for 20 clusters is compared in Fig. 12, while Fig. 13 and 14 display the similar process for clusters 40 and 80, respectively. The tests also demonstrate that HCMGA converges at the fastest rate, with the convergence speeds of HCM and HCM approaches being nearly identical.

Use of a data set pertaining to the movement lines of taxis in one of China's cities, which includes the current GPS locations of taxis, allowed for the evaluation of the suggested method. Various numbers of clusters have been clustered in order to provide a more precise assessment of the suggested strategy. In situations with a small number of clusters, the findings demonstrate that the suggested method performs nearly as well as the HCM method. However, the suggested method's performance improves as the number of clusters increases. Optimisation using ant colonies achieves better clustering accuracy, while optimisation using particle swarms achieves faster convergence, according to the data.

Using HCM clustering with Ant Colony Optimisation (HCMGA) and HCM clustering with Particle Swarm Optimisation (HCMMOPSO), this study found that trajectory clustering became much more accurate and efficient. A clear and concise presentation of the results highlights the important contributions of the algorithms that were suggested.

- Clustering Accuracy Comparison:

In the first part of our investigation, we compare HCMMOPSO and HCMGA to other HCM clustering methods to see how well they cluster data. These metrics—internal cluster cohesiveness, Davies-Bouldin index, and silhouette score are presented in Table II for every algorithm. In terms of detecting significant trajectory clusters, HCMMOPSO and HCMGA both routinely beat baseline approaches.

TABLE II. CLUSTERING ACCURACY METRICS

Metric	HCMMOPSO	HCMGA	Baseline 1	Baseline 2
Silhouette Score	0.71	0.80	0.61	0.53
Davies-Bouldin Index	0.40	0.31	0.54	0.60
Internal Cohesion (avg)	0.91	0.92	0.83	0.80

Higher silhouette scores and lower Davies-Bouldin indices show that the clusters are better defined and well-separated, indicating a significant improvement in clustering quality. HCMMOPSO and HCMGA demonstrate exceptional internal cohesiveness, which is a direct result of the algorithms' ability to produce dense clusters.

- Statistical Significance:

With p-values lower than the 0.05 threshold, statistical analyses, such as ANOVA and t-tests, validate the importance of the noticed improvements. Statistics confirm that HCMMOPSO and HCMGA are better than baseline approaches, which gives the suggested trajectory clustering method more credibility.

- Comparison to Existing Studies:

We compare our results to those of related studies so that you can put them in context. Importantly, as compared to studies that used conventional clustering techniques, our silhouette scores are far higher. The optimisation procedures that were introduced allow for more accurate trajectory segmentation, which in turn improves the performance of HCMMOPSO and HCMGA.

- Discussion:

Results showing an increase in clustering accuracy show that HCMMOPSO and HCMGA are useful in the real world, especially for optimising taxi dispatch systems. The overall efficiency of transport services is improved by effectively identifying and differentiating between distinct trajectory patterns.

Finally, HCMGA and HCMMOPSO are great examples of how trajectory analysis has progressed thanks to the integration of HCM clustering and optimisation approaches. This study's findings add to the expanding literature on trajectory clustering and point the way towards further investigations on transportation system optimisation. Our approach has the potential to improve the efficiency of trajectory-based applications, and the shown improvements highlight its practical value.

VI. CONCLUSION

An approach based on the HCM clustering algorithm for route clustering was proposed as a means to extract city passenger flow patterns. Using a sequence of GPS markers that the user specifies, this method plots a route that is accessible to both pedestrians and drivers. The data shown here pertain to a journey that starts in one place and finishes in another. At regular intervals (often less than two minutes) during this process, the location of the moving object is captured using sampling. This is how the information about the moving cab's many points is recorded. The suggested method specifies the route information as a pair consisting of the geographical dimensions of the spot and the angle of the line that connects it to the next place. Before extracting the sublines, the data undergo pre-processing as part of the clustering process. An initial step in improving clustering accuracy and speed is to partition the sublines according to the angle between them. After that, HCM is used to cluster the data. Cluster centres were found in the local optimum, and the HCM algorithm failed to optimise the objective function, according to the evaluation and implementation of the algorithm. In order to fix this, we talked about meta-heuristic algorithm research and how to apply it to improve the HCM algorithm.

In two different cases, the HCM clustering method was fine-tuned using ant colony and particle swarm methods. Using a real data set containing cab movement data in Beijing, China, and the MATLAB environment, the suggested method was implemented. To test and compare the suggested method, a number of experiments were conducted. By applying the MBP criterion to the clustering accuracy performance, we found that the two scenarios outperformed other methodologies as the number of clusters increased. For a large number of clusters, the evaluation results show that the proposed method outperforms existing approaches in terms of clustering accuracy. The HCMGA algorithm improved clustering accuracy for MBP values by 0.61% compared to prior approaches, while the HCMMOPSO methodology improved it by 0.64%. The findings of the implementation also show that the particle swarm method converges faster and the ant colony approach is the most accurate.

A specific dataset concerning cab traffic in a Chinese city, namely Beijing, is used to test the proposed method. Data collected from GPS coordinates every two minutes form the basis of the evaluation. As an additional metric for cluster evaluation, the MBP index is utilised in the assessment.

For future studies, the proposed method might be expanded and generalised in other ways. For example, this method can be employed to classify various motion trajectories. Alternately, use more evolutionary optimisation methods to improve grouping. You may also test how well the method works by changing the pre-processing and sub-line segmentation steps.

REFERENCES

[1] H.-L. Ling, J.-S. Wu, Y. Zhou, and W.-S. Zheng, "How many clusters? A robust PSO-based local density model," Neurocomputing, vol. 207, pp. 264–275, 2016.

[2] Y. Wu, "A survey on population-based meta-heuristic algorithms for motion planning of aircraft," Swarm Evol Comput, vol. 62, p. 100844, 2021.

[3] M. S. Choudhry and R. Kapoor, "Performance analysis of fuzzy C-means clustering methods for MRI image segmentation," Procedia Comput Sci, vol. 89, pp. 749–758, 2016.

[4] Yumeng Cao, Ning Xu, Huanqing Wang, Xudong Zhao, A. Ahamed. Neural Networks-Based Adaptive Tracking Control for Full-State Constrained Switched Nonlinear Systems With Periodic Disturbances and Actuator Saturation, International Journal of Systems Science, 54(14): 2689-2704, 2023.

[5] Y. R. Alekya Rani and E. Sreenivasa Reddy, "An optimal communication in WSN enabled by HCM clustering and improved meta-heuristic model," International Journal of Pervasive Computing and Communications, vol. 19, no. 2, pp. 233–254, 2023.

[6] Pedditi, R. B., & Debasis, K. (2024). MACR: A Novel Meta-Heuristic Approach to Optimize Clustering and Routing in IoT-based WSN. International Journal of Intelligent Systems and Applications in Engineering, 12(1), 346-359.

[7] Tengda Wang, Liang Zhang, Ning Xu, and K. H. Alharbi. Adaptive critic learning for approximate optimal event-triggered tracking control of nonlinear systems with prescribed performances, International Journal of Control, https://doi.org/10.1080/00207179.2023.2250880, 2023.

[8] Wang, G., Wu, J., & Trik, M. (2023). A Novel Approach to Reduce Video Traffic Based on Understanding User Demand and D2D Communication in 5G Networks. IETE Journal of Research, 1-17.

[9] J. Sun, Y. Zhang, and M. Trik, "PBPHS: a profile-based predictive handover strategy for 5G networks," Cybern Syst, pp. 1–22, 2022.

[10] M. Trik, H. Akhavan, A. M. Bidgoli, A. M. N. G. Molk, H. Vashani, and S. P. Mozaffari, "A new adaptive selection strategy for reducing latency in networks on chip," Integration, vol. 89, pp. 9–24, 2023.

[11] Ding, X., Yao, R., & Khezri, E. (2023). An efficient algorithm for optimal route node sensing in smart tourism Urban traffic based on priority constraints. Wireless Networks, 1-18.

[12] Yue, S., Niu, B., Wang, H., Zhang, L. and Ahmad, A.M. (2023), "Hierarchical sliding mode-based adaptive fuzzy control for uncertain switched under-actuated nonlinear systems with input saturation and dead-zone", Robotic Intelligence and Automation, Vol. 43 No. 5, pp. 523-536. https://doi.org/10.1108/RIA-04-2023-0056.

[13] Kadir, D. H. (2021). Statistical evaluation of main extraction parameters in twenty plant extracts for obtaining their optimum total phenolic content and its relation to antioxidant and antibacterial activities. Food Science & Nutrition, 9(7), 3491-3499.

[14] Xiao, L., Cao, Y., Gai, Y., Khezri, E., Liu, J., & Yang, M. (2023). Recognizing sports activities from video frames using deformable convolution and adaptive multiscale features. Journal of Cloud Computing, 12(1), 1-20.

[15] Omer, A. W., Blbas, H. T., & Kadir, D. H. (2021). A Comparison between Brown's and Holt's Double Exponential Smoothing for Forecasting Applied Generation Electrical Energies in Kurdistan Region.

[16] M. Braik, M. H. Ryalat, and H. Al-Zoubi, "A novel meta-heuristic algorithm for solving numerical optimization problems: Ali Baba and the forty thieves," Neural Comput Appl, vol. 34, no. 1, pp. 409–455, 2022.

[17] M. Trik, A. M. N. G. Molk, F. Ghasemi, and P. Pouryeganeh, "A hybrid selection strategy based on traffic analysis for improving performance in networks on chip," J Sens, vol. 2022, 2022.

[18] Saleh, D. M., Kadir, D. H., & Jamil, D. I. (2023). A Comparison between Some Penalized Methods for Estimating Parameters: Simulation Study. QALAAI ZANIST JOURNAL, 8(1), 1122-1134.

[19] J. Riedy, "Updating pagerank for streaming graphs," in 2016 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), IEEE, 2016, pp. 877–884.

[20] Shen, Y. (2024). Robotic trajectory tracking control system based on fuzzy neural network. Measurement: Sensors, 101006.

[21] Sajadi, S. M., Kadir, D. H., Balaky, S. M., & Perot, E. M. (2021). An Eco-friendly nanocatalyst for removal of some poisonous environmental pollutions and statistically evaluation of its performance. Surfaces and Interfaces, 23, 100908.

[22] W. Qiao and Z. Yang, "An improved dolphin swarm algorithm based on Kernel Fuzzy C-means in the application of solving the optimal problems of large-scale function," Ieee Access, vol. 8, pp. 2073–2089, 2019.

[23] Blbas, H., & Kadir, D. H. (2019). An application of factor analysis to identify the most effective reasons that university students hate to read books. International Journal of Innovation, Creativity and Change, 6(2), 251-265.

[24] Chen Cao, Jianhua Wang, Devin Kwok, Zilong Zhang, Feifei Cui, Da Zhao, Mulin Jun Li, Quan Zou. webTWAS: a resource for disease candidate susceptibility genes identified by transcriptome-wide association study. Nucleic Acids Research.2022, 50(D1): D1123-D1130.

[25] Fakhri, P. S., Asghari, O., Sarspy, S., Marand, M. B., Moshaver, P., & Trik, M. (2023). A fuzzy decision-making system for video tracking with multiple objects in non-stationary conditions. Heliyon.

[26] Y. Zheng, "Trajectory data mining: an overview," ACM Transactions on Intelligent Systems and Technology (TIST), vol. 6, no. 3, pp. 1–41, 2015.

[27] Kadir, D. (2018). Bayesian inference of autoregressive models (Doctoral dissertation, University of Sheffield).

[28] Haoyu Zhang, Quan Zou, Ying Ju, Chenggang Song, Dong Chen. Distance-based Support Vector Machine to Predict DNA N6-methyladine Modification. Current Bioinformatics. 2022, 17(5): 473-482.

[29] Wang, G., Wu, J., & Trik, M. (2023). A Novel Approach to Reduce Video Traffic Based on Understanding User Demand and D2D Communication in 5G Networks. IETE Journal of Research, 1-17.

[30] Khezri, E., Yahya, R. O., Hassanzadeh, H., Mohaidat, M., Ahmadi, S., & Trik, M. (2024). DLJSF: Data-Locality Aware Job Scheduling IoT tasks in fog-cloud computing environments. Results in Engineering, 101780.

[31] Heng Zhao, Huanqing Wang, Ning Xu, Xudong Zhao, Sanaa Sharaf, Fuzzy approximation-based optimal consensus control for nonlinear multiagent systems via adaptive dynamic programming, Neurocomputing, 533: 126529, 2023.

[32] Trik, M., Boukani, B., Ansari, B., Emtiyaz, S., Azar, S. R., & Mohammadi, F. (2014). An Overview of through-silicon via–based three dimensional integrated circuits (3D IC) to placement to optimize timing. Research Journal of Recent Sciences.

[33] Khezri, E., Zeinali, E., & Sargolzaey, H. (2023). SGHRP: Secure Greedy Highway Routing Protocol with authentication and increased privacy in vehicular ad hoc networks. Plos one, 18(4), e0282031.

[34] Khosravi, M., Trik, M., & Ansari, A. (2024). Diagnosis and classification of disturbances in the power distribution network by phasor measurement unit based on fuzzy intelligent system. The Journal of Engineering, 2024(1), e12322.

[35] Pedroche, D. S., Herrero, J. G., & López, J. M. M. (2024). Context learning from a ship trajectory cluster for anomaly detection. Neurocomputing, 563, 126920.

[36] Liu, L., Wang, X., Wang, X., Xie, J., Liu, H., Li, J., ... & Yang, X. (2024). Path Planning and Tracking Control of Tracked Agricultural Machinery Based on Improved A* and Fuzzy Control. Electronics, 13(1), 188.

[37] Kapnopoulos, A., Kazakidis, C., & Alexandridis, A. (2024). Quadrotor trajectory tracking based on backstepping control and radial basis function neural networks. Results in Control and Optimization, 14, 100335.