Proceedings of the Fifth International Conference on Computing Methodologies and Communication (ICCMC 2021)
IEEE Xplore Part Number: CFP21K25-ART

2021 5th International Conference on Computing Methodologies and Communication (ICCMC) | 978-1-6654-0360-3/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICCMC51019.2021.9418026

Fuzzy clustering using Hybrid CSO-PSO search based on Community mobility during COVID 19 lockdown

Parvathavarthini S
Asst. Professor (Sr. Grade), School of Computer Technology and Applications,
Kongu Engineering College, Perundurai,
Erode, India
varthinis@gmail.com

Naveenkumar R V
Final year student, School of Computer Technology and Applications,
Kongu Engineering College, Perundurai,
Erode, India

Sanjay Chowdry B, Siva Prakash K
Final year students, School of Computer Technology and Applications,
Kongu Engineering College, Perundurai,
Erode, India

***Abstract*— Recently COVID-19 virus had made a potential threat to humanity. The effect of mobility habits might be considerably high in the spread of disease. This work analyzes the mobility patterns in selected districts of Tamilnadu and applies Fuzzy C-Means clustering based hybrid variant of cuckoo search combined with best features from particle swarm optimisation. Google Community Mobility reports are taken for the experimental purpose and 26 districts in Tamilnadu are considered. The decision regarding the restricted movement of people can be made from the results. The correlation between human mobility during and after lockdown to the disease spread is analysed. The results are evaluated using internal indices like Silhouette and DB index. The regions are classified into high, medium and low risk regions with relevance to the human mobility.**

Keywords—COVID-19; Mobility; Fuzzy C-Means, Particle Swarm Optimization, Cuckoo Search Optimization, Modified Levy flight, Gaussian Distribution, Hybrid CSO-PSO optimization;

I. INTRODUCTION

Covid 19 outbreak has its severe impact on health, livelihood, environment, psychology, education and transportation of the people living all over the world. Several critical decisions have to be made to ensure the safety of public during this crisis. In order to avoid the spread of virus, several restrictions were posted on the public mobility. Mobility in several forms were reduced by providing vacation to schools, offices started work from home procedures and quarantining the regions with more number of positive cases. Lockdown was announced during the month of March 2020. This work aims to cluster the regions in Tamilnadu based on the mobility rate during the period of February to August 2020.

About 47% of people in the world were affected by Covid19 virus. The main reason for COVID spread is due to mobility causes. Mobility occurs when people are gathering at various places and this has been the reason why the disease has been spreading from one to another. This work analyzes the COVID affected areas and categorizes them into High, Medium and Low risk areas based on mobility rate.

Clustering [1] is one of the unsupervised learning techniques where objects are grouped into various categories based on their features. Partitional clustering divides data into k clusters based on the distance measure. Fuzzy C-Means (FCM) [2] is a popular technique that allocates objects to multiple clusters each with a belongingness given by the membership value. The object belongs to the cluster that possesses a higher membership. The drawback with FCM is that it picks up the initial centroids in a random fashion. This leads to local optimal results and causes delay in convergence. To avoid this, optimization algorithms can be utilized so as to reach a global optimal value for the objective function. To utilize the strengths of both Particle Swarm Optimization (PSO) and Cuckoo Search Optimization (CSO), a hybrid PSO-CSO algorithm is used to optimize the initial centroids.

Mobility data is taken from the Google Community Mobility Reports (CMR) [3]. This data is available for all the countries in the world. The paper is organized as follows: Section II focuses on literature review, Section III narrates the proposed methodology, Section IV discusses the experimental results and Section V presents the conclusion and future enhancements.

II. LITERATURE REVIEW

Several recent works are found in the literature as COVID-19 pandemic has been the core area of research in the year 2020. A forecasting model is designed by [4] for predicting COVID-19 cases in Arizona with the help of Google CMR. A Partial Differential Equation (PDE) based model for clustering county level regions based on historical data is formulated and fourth order Runge Kuttam method is used to solve the PDE. The activities of human across the boundaries of various counties are considered and the prediction accuracy is above 94%.

Ref [5] simulated the effect of reduced mobility over the spread of COVID-19 in Shenzhen city of China by collecting mobile phone data from the service provider named China Unicom. People from the age group of 15 to 65 are considered and their movements are tracked. People from various stages of COVID such as suspected, exposed, infected and recovered stages are modeled using differential equation and a simulation of various ranges of mobility restrictions are applied so that further decisions regarding mobility may be taken.

The effect of mobility routines over the spread of the pandemic in Italy is analyzed by [6]. Data is collected from daily reports given by health and transport ministries of Italy. The past 21 days travel history of infected persons is tracked. Multiple linear regression model is developed by considering geographical area, environmental, mobility and healthcare variables. The results confirmed that the increase in mobility contributes to the severe spread of disease.

The effectivenss of quarantining during the pandemic is analyzed by [7]. The author identified how much population moved out of Wuhan and developed a risk model for deciding the spread ratio. Levenberg–Marquardt algorithm is used for the analysis and the results can be used to make decisions regarding the mobility. [8] assessed the impact of lockdown in Italy by constructing a proximity network to estimate the radius of gyration thus extracting the weekly and daily movements of people. The results are compared with those of other data sources like Google CMR.

The analysis of spread pattern of COVID-19 infection in Iran [9] has been done using clustering and geographical information systems. The data is scaled and a dendrogram is formed to find the infection spread and Tehran and Qom are found to be the most vulnerable sources of disease in Iran [9]. The correlation between air pollution and COVID-19 mortality rates in London is analysed in [10]. Due to the various restrictions in mobility, the air pollution decreased to a greater extent. This analysis helps the government to take appropriate measures in restricting the mobility.

An epidemiological model is designed by [11] to find out the number of people affected at any point of time. The actual infected number could be found by training a neural network model. The impact of community mobility in India is assessed by [12]. The data is analysed using excel sheet and the averge increase or decrease in mobility in the sates and union territories of India are colored by mapping the spatio-temporal data.

The impact of mobility in terms of own vehicles, walking or public vehicles in England is identified [13] and reduction of mobility rate during lockdown has serious effect on disease spread. This study helps the government to take effective measures and decide about the strategies to tackle the situation. [14] studied the trends in mobility before and during lockdown in China that were adapted to control the spread and severity of Covid19. On an average of 6.5 days delay of COVID-19 is observed because of reduced mobility. The modified cuckoo search was introduced in [15] to optimize the selection of relay in a joint beam forming system. The

Still, this analysis is done for most of the foreign countries like Iran, China, UK and so on. It would be fruitful to analyse this for every district of Tamilnadu.

III. PROPOSED METHODOLOGY

A. Dataset description

The dataset obtained from Google CMR is a time-series data. The places where people gather for essential purposes are categorized into six groups such as Retail and Recreation, Grocery and Pharmacy, Parks, Transit Stations, Workplaces and Residential areas. The period from January 3,2020 to February 6, 2020 is taken to calculate the median value for the corresponding day of the week during this five weeks timeline. Under each gathering places, the increase or decrease in mobility in denoted with a positive and negative value respectively.

B. Hybrid CSO-PSO based Fuzzy Clustering

This work aims at clustering selective districts of Tamilnadu based on their mobility values. The sequence of steps involved in this process is given in the block diagram of Figure 1.

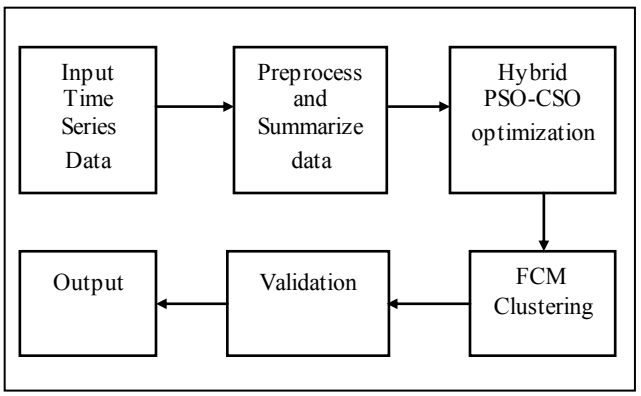

Fig. 1. Steps in Proposed Methodology

1) Preprocessing

Preprocessing is cleaning of data thus converting it into a suitable form for running the algorithm. Since, this data is a time series data, the data has to be summarized or consolidated so that the data can be converted into pivot table. The pivot table takes the mean values for all 6 categories of places in selected districts of Tamilnadu and this summary can be taken as input for the algorithm to execute. Out of the 38 districts in

Tamilnadu, the data is not available for some districts. So, a total of 26 districts are taken for the purpose of evaluation.

2) Hybrid PSO-CSO optimization

Optimization is essential to find the best fit value for some objective function. In case of clustering, it facilitates global optimal results that yield the minimum value for the fitness function. CSO and PSO are the two most famous optimization algorithms but each has its own drawback. CSO finds the global optimal results but convergence is little bit slower whereas PSO is converging fast towards the local optimum. To maintain a better balance between local and global search, a combination of these two algorithms is needed.

Cuckoo search simulates the breeding behaviour of the bird and PSO simulates the food search of a bird swarm. The cuckoo searches for a host nest to lay eggs via Levy flight distribution. If the host bird identifies the cuckoo eggs, they will be evicted or the bird may abandon this nest and build a new one. Nests with quality eggs are retained for next iteration.

CSO is applied to search the best centroid which is simulated as the nest. Other cuckoos also update their position related to the best nest which is simulated as updating the position of the swarm towards food source in the PSO. Due to this hybridization which exploits the power of both the algorithms, the convergence occurs at a faster rate. In this work, Levy flight is modified on Gaussian distribution and updation of speed is done using the formula from PSO.

The new solution x (t +1) for the worst nest (Yang, Xin-She, and Suash Deb, 2014) is performed by

$$x_i^{(t+1)} = x_i^t + \alpha \oplus Levy() + \sigma \tag{1}$$

where α denotes the step size chosen based on the problem of interest. The product $\oplus$ means entrywise multiplications and $\sigma = \sigma_0 e^{-ct}$ where σ_0 and c are constants and t denotes the current iteration.

The velocity is updated using the following equation

$$Velo(k+1) = wt.Velo(k) + (c1.rand1).(pbest(k) - Xpos(k)) + (c2.rand2).(gbest(k) - Xpos(k)) \tag{2}$$

where *c1* and *c2* are constants, *wt* indicates inertia weight, *rand1* and *rand2* are the random values whose range is from 0 to 1.

3) Pseudocode for Hybrid CSO-PSO

begin
Generate n *host nests as the initial population* $\mathbf{x}_i$ (I = 1, 2, ..., n)
while (t <*Maxiteration) or (stop criterion)*
Get a cuckoo randomly by Levy flight modified on Gaussian distribution
Compute its fitness F_i
Choose a nest j randomly
if ($F_i > F_j$),
replace j *by the new solution;*
end
Abandon the worst nests and build new solutions
Update solutions using modified Levy flight
Update particle velocity using PSO
Keep the best solutions
Rank the solutions and find the current best
end while
end

4) FCM clustering

Clustering tends to find unknown patterns from a dataset based on its features. When class labels are not known, this kind of unsupervised methods can help draw conclusions from the dataset. Fuzzy clustering has been a prevailing technique that allocates data points to more than one cluster. It preserves uncertainty in data by mentioning the membership and non-membership values. The data has to be converted into fuzzy representation and then the algorithm partitions the given data into n clusters. The number of clusters should be specified by the user in advance.

Distance measure is the first step that where Euclidean distance between the centres and the data objects is found. The membership matrix is found by

$$U_{ij} = \frac{1}{\sum_{r=1}^{C}\left(\frac{dis(d_j', v_i)}{dis(d_j', v_r)}\right)^{\frac{2}{m-1}}}, 1 \le i \le C, 1 \le j \le n, m = 2 \tag{3}$$

The fitness function of the FCM method is given as follows

$$J_m(x,y) = \sum_{i=1}^{c}\sum_{j=1}^{n} U_{ij}{}^m \left\| X'_j - C_i \right\|, \quad 1 \le m \le \infty \tag{4}$$

To iteratively, update the centroids, the following formula is applied

$$v_j = \left(\sum_{i=1}^{n}(\mu_{ij})^m x_i\right) / \left(\sum_{i=1}^{n}(\mu_{ij})^m\right), \forall \ j = 1, 2, \ldots, c \tag{5}$$

During every iteration, the process of centroid updation and membership calculation are repeated. The stopping criterion is reached when the the objective function is saturated, or the centroid value remains unchanged in the consecutive iterations. Based on the highest membership value allotted to a cluster, the conversion from fuzzy to crisp values has to happen. This process is known as defuzzification process and this will serve as the index of the cluster for that object.

These steps produce three clusters that are representing high, medium and low mobility in various districts.

IV. EXPERIMENTAL RESULTS

Each nest or cuckoo represents a potential solution to the problem. The initial cluster values are taken as the nests. The parameters set for the hybrid algorithm is given in Table 1.

Table 1 Parameters for Hybrid CSO-PSO

Parameter	Value
No. of nests	20
No. of eggs	1 egg per nest
The probability of a nest being abandoned	0.25
Fuzziness parameter m	2
Population	20
Max Iterations	200

Since the data does not contain class labels, the clustering quality can only be evaluated using internal indices. The DB Index and Silhouette index are used to validate the results.

The Davis-Bouldin index (DB) [16] computes the ratio of within cluster and between cluster distances. A lower value indicates better outcome of the clustering algorithm.

The formula for DB index can be given as

$$\frac{1}{k}\sum_{i=1}^{k} max_j \frac{s(C_i)+s(C_j)}{d_c(C_i,C_j)} \tag{6}$$

where k is the number of clusters. The DB Index value obtained is 0.0830.

The formula for silhouette coefficient [17] is given by

$$SC = \frac{1}{N}\sum_{i=1}^{N} s(x) \tag{7}$$

where the term s(x) is defined as

$$s(x) = \frac{b(x)-a(x)}{\max\{a(x),b(x)\}} \tag{8}$$

where a(x) denotes compactness and b(x) indicates separation and $s(x)$ lies in the range [-1, +1]. The value -1 indicates that the solution is bad, 0 indicates that it is indifferent and 1indicates a good and appreciable value. Silhouette measure reached the value of 0.9361. Since there are no class labels available, these two internal measures are capable of assuring the quality of clustering results.

The average mobility in all the 26 districts is computed and the Figure 2 shows the corresponding plot. Figure 3 shows the district level map of Tamilnadu in which the clusters are marked using different colors.

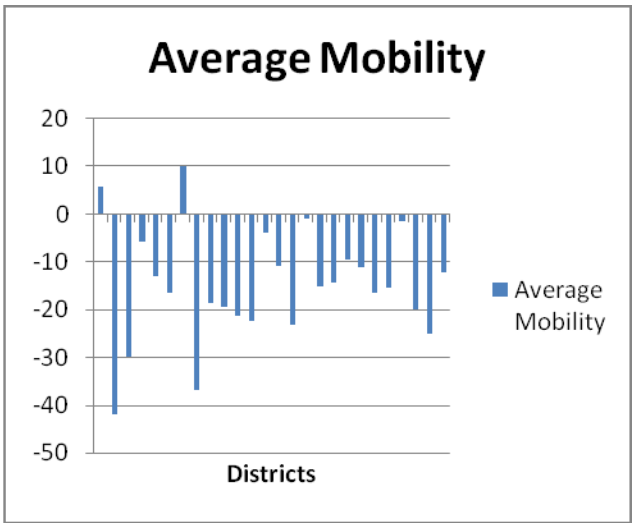

Fig. 2. Average mobility in Districts

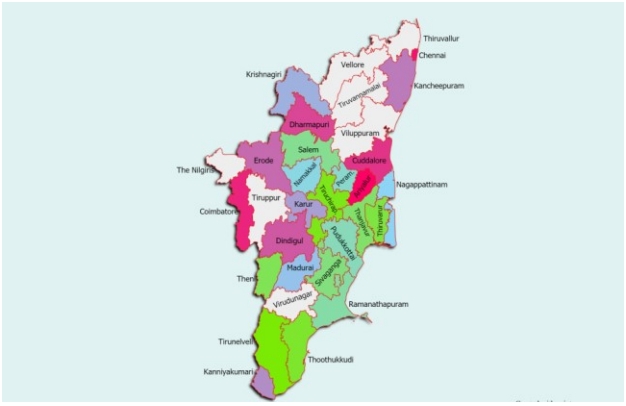

Fig. 3. Clustering Result

The map illustrates various districts of Tamilnadu that are colored based on the mobility. Selected districts are considered for evaluation on the basis of data available in Google CMR. The pink, blue and green colors indicate the high, medium and low range of mobility respectively. The white colored regions are those which are not considered for evaluation. The obtained results assist government to make decision regarding the transition of goods and movement of people.

Several possibilities to avoid the spread of disease including the quarantine of people who travel across the boundaries can be assessed. Based on daily affected cases, the districts can be separated into various zones. Within the zone travel can be encouraged and E-pass may be availed to travel to other zones. Essential commodities can be supplied directly to home and the gathering of people can be monitored through drones. Other public health strategies like limited opening of groceries, mobile hospitals can also be applied.

Erode and Ariyalur districts showed the high rate of mobility including all 6 categories. The total percentage of mobility in all the 6 categories namely Retail and Recreation, Grocery and Pharmacy, Parks, Transit Stations, Workplaces

and Residential areas throughout the period from February 2020 to August 2020 is depicted as a pie chart in Figure 4.

Fig. 4. Category wise Mobility

V. CONCLUSION AND FUTURE ENHANCEMENTS

The mobility to residential places has considerably increased due to the announcement of lockdown and people moved to their hometown or stayed at home. Immediately after the announcement, there was a hike in mobility in grocery and pharmacy as people purchased the essential commodities and medicines for emergency situation. Suggestions and measures to be followed regarding lockdown can be made as a result of this work. In future, this can be used for prediction of decrease in the spread of COVID-19 due to the measures taken. In future, the work can be extended using intuitionistic fuzzy c-means algorithm. The correlation between daily reported cases and mobility can be analysed.

References

[1] AK Jain, MN Murty, PJ Flynn, "Data clustering: a review", ACM Comput. Surv. vol. 31, pp. 264-323, Sep 1999.

[2] JC Bezdek, R Ehrlich, W Full, " FCM: The fuzzy c-means clustering algorithm", Comput and Geosci, vol.1., pp. 191-203, Jan 1984.

[3] https://www.google.com/covid19/mobility/

[4] H Wang, N Yamamoto, "Using a partial differential equation with Google Mobility data to predict COVID-19 in Arizona", Math. Biosci. Eng, vol. 17, Jan 2020.

[5] Y Zhou, R Xu,D Hu, Y Yue, Q Li, J Xia, "Effects of human mobility restrictions on the spread of COVID-19 in Shenzhen, China: a modelling study using mobile phone data", The Lancet Digital Health, vol. 2.pp.417-424, Aug 2020

[6] A Cartenì, L Di Francesco,M Martino, " How mobility habits influenced the spread of the COVID-19 pandemic: Results from the Italian case study", Science of the Total Environment, vol. 741, Nov 2020.

[7] JS Jia, X Lu,Y Yuan, G Xu,J Jia, NA Christakis, "Population flow drives spatio-temporal distribution of COVID-19 in China", Nature. Vol. 582, pp. :389-394, Jul 2020.

[8] E Pepe, P Bajardi, L Gauvin, F Privitera, B Lake, C Cattuto,M Tizzoni, "COVID-19 outbreak response, a dataset to assess mobility changes in Italy following national lockdown", Scientific data. Vol.7, pp. 1-7, Jul 2020.

[9] M Azarafza, M Azarafza, H Akgun, "Clustering method for spread pattern analysis of corona-virus (COVID-19) infection in Iran", medRxiv, Jan 2020

[10] M Sasidharan, A Singh, ME Torbaghan, AK Parlikad, " A vulnerability-based approach to human-mobility reduction for countering COVID-19 transmission in London while considering local air quality" Sci. Total Environ, vol. 741, Nov 2020.

[11] R Dandekar, G Barbastathis, "Quantifying the effect of quarantine control in Covid-19 infectious spread using machine learning", medRxiv, Jan 2020

[12] J Saha, B Barman, P Chouhan, "Lockdown for COVID-19 and its impact on community mobility in India: An analysis of the COVID-19 Community Mobility Reports", Child Youth Serv. Rev. vol. 116, 105160, Sep 2020

[13] GM Hadjidemetriou,M Sasidharan, G Kouvialis, AK Parlikad, "The impact of government measures and human mobility trend on COVID-19 related deaths in the UK". Transportation research interdisciplinary perspectives, Vol. 6, Jul 2020.

[14] MU Kraemer, CH Yang, B Gutierrez, CH Wu, B Klein, DM Pigott, L Du Plessis, NR Faria, R Li, WP Hanage,JS Brownstein, " The effect of human mobility and control measures on the COVID-19 epidemic in China", Science. vol. 368, pp. 493-497, May 2020.

[15] A Kuruppath, S Thivyakat, "Power-and time-optimized MMSE-based joint beam-forming with relay selection for future generation MIMO networks using Modified Cuckoo-Search Optimization algorithm", *Sādhanā*, vol. *45*, pp.1-14, Dec 2020.

[16] DL Davies,DW Bouldin, "A cluster separation measure", IEEE Trans. Pattern Anal. Mach. Intell. vol. 2, pp. 224-227, April 1979.

[17] PJ Rousseeuw , "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis", *J. Comput. Appl. Math.*, vol. *20*, pp.53-65, 1987.