Vol. 7(26), Oct. 2017, PP. 3685-3689

Contents list available at IJMEC

International Journal of Mechatronics, Electrical and Computer Technology (IJMEC)

Journal Homepage: www.aeuso.org

PISSN: 2411-6173 - EISSN: 2305-0543

The Clustering of Data Streams using Particle Swarm Optimization

Seydeh Somayeh Salehi Komamradkhi [1,], Saleh Shakeri[2] and Hamid Tavkolaei[3]*

[1]*Computer engineering graduate student, Ayatollah Amoli Branch, Islamic Azad University, Amol, Iran*

[2]*Department of Mathematics, Ayatollah Amoli Branch, Islamic Azad University ,amol, p.o.Box 678, Iran*

[3]*Department of Computer Engineering, Ayatollah Amoli Branch,Islamic azad University, Amol, p.o.Box 678, Iran*

Phone Number:*+98- 1144258683*

Corresponding Author's E-mail: *Seyedesumayesalehi@gmail.com*

Abstract

Large volumes of data does not help the managers in decision-making and decision alone, but can also cause confusion managers of organizations. Therefore, managing the internal and external raw data and convert the data into information and knowledge using different techniques is very important and essential. The famous technique in the field of data mining, which can be done on the database and obtain the required knowledge. Explore clusters also one of the important techniques in growing fields, is known as data mining exploration that applied various disciplines of engineering and science, such as biology, psychology, medicine, marketing, computer and mapping satellite. The proposed methods are combination of two methods FCM and FKM. This improved to aid PSO and DCT. Studies show that the method presented in the paper is more efficient outcomes in terms of density and separation of clusters by minimizing the validity index XB. The PSO method is used in the paper for the optimum solutions and continuous and closest adjacent. In addition, the discrete cosine transform is used to reduce the dimensions and reduce the problem of search and more efficient for PSO**.**

Keywords*: clustering, data streams, DCT, FCM, PSO.*

1. Introduction

Data and pattern are the important index in the world of information. Clustering is one of the best ways of working with data that have been proposed. Clustering has been one of the most ideal mechanisms to work with data world, its ability to enter the space Data and recognize their structure. The first idea was presented in the decade 1935, today emerged Improvements and huge mutations in clustering applications and various aspects (Shrabyan and Rad, 2013). Clustering is one of learning Branches without monitoring and it is an automatic process, during which the samples are divided into categories that are members of the same, that these categories are called clusters. Therefore, cluster is a collection of objects where objects are similar and also in other clusters are non-similar objects. For the Similar Can be considered various criteria. The example can be used to measure distance for clustering and objects that are closer together as a cluster, consider that this type of clustering, also called distance-based clustering.

Article History:

Received Date: Feb. 14, 2017
Accepted Date: Aug. 16, 2017
Available Online: Oct. 01, 2017

2. Particle Swarm Optimization

Method PSO is a method of general optimization which using those can answer questions that they have a point or surface in N dimensional space. In such Space, assumptions discussed and primary velocity is assigned to Particles, also Communication channels considered to be between particles. After each period particle moving in the space and the results will be based on a calculation. Over time, Particles toward the particles that have higher eligibility criteria and located in the same communication group, accelerated.

The main advantage of the method that is Optimization strategies, large numbers of particle swarm, that Causing flexible method against the problem locally optimal solution. Each particle has a position that determines what is a multi-dimensional coordinates of the particle in the search space. With the particle motion during the time changing the position of the particle, xi (t) determines the position of the particle I in time T. Also, each particle for movement in space needs to have velocity. VI (t) determines the velocity of the particle I in time T. By adding velocity to the position of each particle, the particle can be considered a new position. Updated the position of the particle in equation are given below.

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (1)$$

$$x_i(t) \sim U(x_{min}, x_{max}) \quad (2)$$

Whether position of the particle in the search space suitable position or not? And can be evaluated by a fitness function. Particles have the ability the best position remembers where had been during his life. The best individual experience of a particle or the best position visited by the particle called yi (in some of the algorithms yi called PBEST) and particles can have the best visited position also be aware of the whole group That the position is called (in some of the algorithms called GBEST). Particle velocity vector in the optimization process reflects empirical knowledge of particle and Information Society particles.

3. Discrete cosine transform

The discrete cosine transform (DCT): Limited sequence of data points as the sum of the cosine functions oscillate at different frequencies, the conversion can play an extensive role in science and engineering; the compression in loss of audio data such as (MP3) and picture (JPEG) that Small pieces can be removed with higher frequencies. Spectral methods for the numerical solution of partial differential equation are used in the range of DCT.

Since the cosine function is less substantial for approximating in the normal signal (Relative to sine functions) also using the Cosine function Rather than sinus is essential in compression. Also, relative to differential functions, cosine functions have a more specific border in conditions. Discrete cosines transform the related conversion to the Fourier transform that Very similar to the discrete Fourier transforms (DFT). With the exception that only uses real numbers.

4. Fuzzy k-means Algorithm

The most known clustering methods, fuzzy k-Means and k-Means clustering algorithms Features of two algorithms that they act only on numerical data and this is one of the limitations of the clustering.

$$1.\text{Initialize} U = [u_{ij}] \text{matrix}, U^{(0)}$$

$$2.\text{At} k\text{-step}: \text{caculate the centers vectors} C^{(k)} = [c_j] \text{with} U^{(k)}$$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m . x_i}{\sum_{i=1}^{N} u_{ij}^m}$$

$$3.\text{Update} U^{(k)}, U^{(k+1)}$$

$$uij = \frac{1}{\sum_{k=1}^{c} \left\langle \frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right\rangle^{\frac{2}{m-1}}}$$

$$4.\text{if} \left\| U^{(K-1)} - U^{(K)} \right\| < 0 \quad \text{than} \quad \text{STOP; otherwise} \quad \text{return} \quad \text{to} \quad \text{step2.}$$

Figure 1: write something.

5. Fuzzy c-means Algorithm

Fuzzy c-means Algorithm (FCM) is the most famous of fuzzy clustering algorithm that has many applications. Fuzzy c-means Algorithm will start with an initial value from W, and between the estimation of cluster centers z given in z and estimate Belong matrix Repeated in W and as long as they are equal to two consecutive values of Z or W. Mathematically, a fuzzy clustering problem can be represented as an optimization problem as follows.

That n, the number of objects studied in the data set and k is the number of clusters. Set of object any of them described with d Features. Z a set with k cluster center and W is a fuzzy belong matrix and power weight and d is the criteria certain distance between the center of the cluster and the object. Also Fuzzy c-means Algorithm only working with numerical data.

6. The flow chart of the proposed method

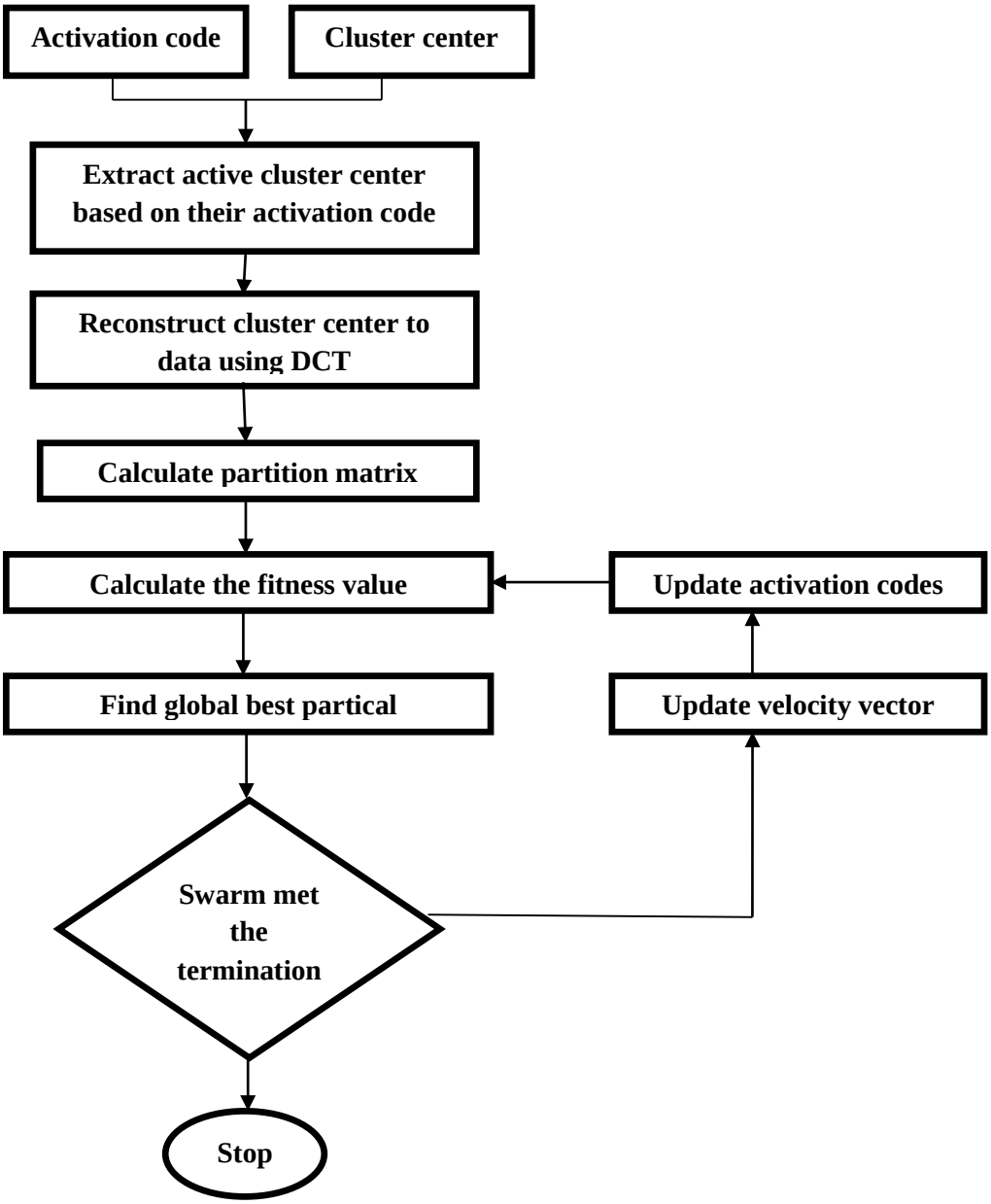

Figure 2: write something

Conclusion

The proposed method is a combination of two methods FKM and FCM that have improved with Algorithms PSO, DCT. The results show that the method presented in this research effective result in terms of density and separation clusters to minimize reliability index XB. In the paper was used for a POS method for optimal solution and continuous nearest neighbor, In addition, The Using from discrete cosine transform for reduce dimensions And reducing search problems and efficient POS. Routing problems solving and get the shortest path are many functionalities and to solve issues such as Routing vehicles vrp and tps has many applications But during the surveys found that to help Cosine transform DCT can be reduced path length and in terms of time reduced path length with DCT is faster than other algorithms And nearest neighbor can be achieved target path.

References

[1] Park, N.H.; Lee, W.S.; "*A statistical grid-based clustering over data streams*", ACM SIGMOD, Record 33(1), P.P. 32–37, 2004.

[2] Park, N.H.; Lee, W.S.; "Cell trees: an Adaptive Synopsis structure for clustering multi-dimensional on-line data streams", Data & Knowledge Engineering, 63(2), P.P.528–549, 2007.

[3] Guha, S.; Meyerson, A.; Mishra, N.; Motwani, R.; O'Callaghan, L.; *"Clustering data streams: Theory and practice"*, IEEE Trans. Knowledge Data Engineering, 15 (3), P.P. 626 515–528, 2003.

[4] Piciarelli, C., Foresti, G., & Snidara, L. (2005). Trajectory clustering and its applications for video surveillance. Proc. IEEE Conf. AVSS, Como (pp. 40–45).

[5] Vlachos, M., Gunopulos, D., & Kollios, G. (2002). Discovering similar multidimensional trajectories. 18th Int. Conf. On data engineering (pp. 673–684).

[6] Yanagisawa, Y., & Satoh, T. (2006). Clustering multidimensional trajectories based on shape and velocity. Proc. Int. Conf. on data engineering workshop (pp. 12–21).

[7] Li, X., & Hu, W. (2006). A coarse-to-fine strategy for vehicle motion trajectory clustering. Proceedings International Conference on Pattern Recognition, 1, 591–594

[8] Anjum, N., & Cavallaro, A. (2007). Unsupervised fuzzy clustering for trajectory analysis. Proceedings IEEE International Conference on Processing, 3, 213–216.

[9] Lee, J., Han, J., & Whang, K. (2007). Trajectory clustering: A partition-and- group framework. Proc. ACM SIGMOD Int. Conf. management. data (pp. 593–604).

[10] Zhang, Y., & Pi, D. (2009). A trajectory clustering algorithm based on symmetric neighborhood. Proceedings WRI World Congress on Computer Science and Information Engineering, 3, 640–645.

[11] Chen, J., Wang, R., Liu, L., & Song, J. (2011). Clustering of trajectories based on Hausdorff distance. Proc. IEEE Conf. electronics (pp. 1940–1944). Communications and Control (ICECC).

[12] O'Callaghan, L.; Mishra, N.; Meyerson, A.; Guha, S.; Motwani, R.; "*Streaming-data Algorithms for High-quality Clustering",*IEEE International Conference on Data Engineering, 2002.

[13] Aggarwal C. C.; "Data Streams Models and Algorithms", Springer Publishers, 2007.

[14] Boyd. R, Richerson. P.J, Culture and the Evolutionary Process, University of Chicago Press, Chicago, 1985.

[15] Eberhart. R, Kennedy. J, A new optimizer using particle swarm theory, in: Proceedings of the 6th International Symposium on Micro Machine and Human Science, pp. 39–43, 1995.

[16] Z. Michalewicz and D. Fogel. How to Solve It: Modem Heuristics. Springer-Verlag، Berlin، 2000.

[17] P. Van Laarhoven and E. Aarts. SimulatedAnnealing: Theoty andApplications. Kluwer Academic Publishers ،1987.

[18] Sarabian,V and Kivan Rad, M,A. Clustering and its methods. 2013. (In Persian)

[19] Salimi, N Rafih,H. Providing a combined method for clustering data using algorithm k-means, pso, Genetics. Seventh International Conference Information and Knowledge Technology ،2015. (In Persian)

[20] Sarani,M and Reazaei,H. The combination of particle swarm optimization and genetic algorithm for dynamic clustering National Conference Sustainable Development with a focus on computer engineering and computer networks, modeling and Security Systems ،2013. (In Persian).

Autor(s)

Seydeh Somayeh Salehi Komamradkhi received her MSc, degrees in Computer engineering graduate student, Ayatollah Amoli Branch, Islamic Azad University, Amol, Iran. He is a ticher. Areas of her research interests include clustering of data and distributed systems, optimization techniques, intelligent systems.

email: Seyedesumayesalehi@gmail.com
Phone Number: (+98) *01144258683*