

Gravitational Optimization Using Fuzzy C-Means Clustering Algorithm

Nayana.R¹, K.R.Kundhavai²

P.G. Student, Department of Computer Science, New Horizon College of Engineering, Bangalore, Karnataka, India¹

Assistant Professor, Department of Computer Science, New Horizon College of Engineering, Bangalore,
Karnataka, India²

ABSTRACT: Nowadays the growth of spatial data encountered in different application domains and their analysis has become more important. Clustering is one of the important tasks in spatial data mining. Spatial clustering aims group similar spatial objects into group or clusters so that objects within a cluster have high similarity in comparison to one another but are dissimilar to objects in other clusters. In this paper, we propose a new approach for data clustering. Initially the proposed approach exploits spatial fuzzy c-means for clustering the vertex into homogeneous regions, Fuzzy c-Means Clustering performs clustering by iteratively searching for a set of fuzzy clusters and the associated cluster centres that represent the structure of the data as best as possible. In order to improve the performance of fuzzy c-means to cope with segmentation problems, we employ gravitational search algorithm which is inspired by Newton's rule of gravity. The improved GSA (namely Fuzzy-GSA) is employed to construct a novel data mining algorithm for classification rule discovery from reference data sets. Gravitational search algorithm is incorporated into fuzzy c-means to take advantage of its ability to find optimum cluster centres which minimizes the fitness function of fuzzy c-means.

KEYWORDS: Clustering, Fuzzy c-means, Gravitational Search.

I. INTRODUCTION

Data mining is the process used to turn raw data into useful information by discovering interesting and potentially useful patterns from databases. Spatial data are the data related to objects that occupy space. Spatial clustering aims group similar spatial objects into group or clusters so that objects within a cluster have high similarity in comparison to one another but are dissimilar to objects in other clusters [1]. Moreover, Geographical information systems are becoming rich deposits of spatial data in many applications such as traffic planning etc. The GISs provide the user with the possibility of querying a territory for extracting areas that exhibit certain properties. This explosively growing spatial data creates the necessity of knowledge discovery from spatial data, which leads to a promising emerging field, called spatial data mining or knowledge discovery in spatial databases [2]. Regionalization has been an important and challenging problem for a large spectrum of research and application domains, Regionalization is essentially a special form of classification where spatial units are grouped together, based on a set of defined criteria and a set of contiguity or adjacency constraints [3]. Spatial clustering can be applicable for solving many problems. An important application area for the spatial clustering algorithm is social and economic geography. In the scope a classical methodical problem of social geography, "regionalization" can be considered. In this paper, we propose a new approach for data clustering making use of fuzzy c-means algorithm and gravitational search optimization. Fuzzy c-Means Clustering performs clustering by iteratively searching for a set of fuzzy clusters and the associated cluster centres that represent the structure of the data as best as possible. In order to improve the performance of fuzzy c-means to cope with segmentation problems, we employ gravitational search algorithm which is inspired by Newton's rule of gravity. Therefore, we employ gravitational search algorithm in order to improve the performance of fuzzy c-means.

II. RELATED WORK

In recent years, a new intelligence algorithm has been introduced based on the gravity forces between the masses. This method is called Gravitational Search Algorithm (GSA). The concept of intelligently controlling the search process of

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

gravitational search algorithm (GSA) is introduced to develop a novel data mining technique. The GSA could be considered as an isolated system of masses. It is like a small artificial world of masses obeying the Newtonian laws of gravitation and motion. The position of the mass corresponds to a solution of the problem, and its gravitational and inertial masses are determined using a fitness function. Thus, agents are considered as objects and their performance is measured by their masses. All these objects attract each other by the gravity force, and this force causes a movement of all objects towards the objects with heavier masses. Hence, masses cooperate using a direct form of communication, through gravitational force. It has been shown that the GSA is able to find the optimum solution for many benchmarks [4]. GSA has some important parameters (e.g. swarm size and gravitational coefficient) which play major roles in its search characteristics, including premature convergence, convergence rate, local capturing, exploitation, exploration, etc. We know that the value of swarm size considerably affects the powerfulness and effectiveness of GSA.

A large value of swarm size reduces the convergence rate and increases the powerfulness of algorithm; whereas a small value of swarm size causes a local minimum capturing and reduces the performance of GSA. On the other hand, the search process of GSA (like other swarm intelligence techniques) is very complicated. Thus, it is hard if it is not possible, to model mathematically the search process of GSA. Thus, linearly decreasing the effective objects and an exponential schedule for decreasing the gravitational coefficient is not an optimal selection for tackling different complex engineering problems. On the other hand, there are some linguistic descriptions and understandings of the search process of GSA. These understandings and linguistic descriptions make a fuzzy system a good candidate for controlling intelligently the parameters of GSA. This idea leads to designing more powerful and efficient gravitational search algorithm. Another topic which is related to this paper is data mining. Data mining is defined as the process of model abstraction from the data sets and searching for valid, nontrivial patterns, and symptoms within the abstracted model [5]. DM covers a wide range of knowledge discovery methods from the databases including classification, clustering, dependence modelling, etc. An important branch of DM includes rule discovery techniques for data classification. In this part, a rule based classifier is designed for assigning each pattern (object, record, or instance) in the feature space to one distinct class. Many approaches, methods and goals have been tried out for designing rule based classifiers. Biology inspired algorithms such as Genetic Algorithms and swarm intelligence based approaches like particle swarm optimization and ant colony optimization have been successfully used ([6], [7], [8], and [9]).

III. PROBLEM DEFINITION

The process of discovering useful patterns from spatial datasets is more difficult than extracting the corresponding patterns from traditional numeric and categorical data due to the complexity in spatial database. Because of the huge amounts of spatial data that may be obtained from satellite images, medical equipment, video cameras, etc., it is costly and often unrealistic for users to examine spatial data in detail. A spatial database is a database that is optimized to store and query data that is related to objects in space, including points, lines and polygons. Regionalization is one of the important tasks in spatial data mining. Regionalization is a process of dividing regions into small areas. Regionalization is a classification procedure applied to spatial objects with an area representation, which group them into a homogenous contiguous region. Regionalization is one of the biggest problems faced by spatial data mining while representing economic and social geography. To solve regionalization issues, we propose a new hybrid approach for data clustering making use of fuzzy c-means clustering algorithm and gravitational search optimization.

IV. ALGORITHM AND MATHEMATICAL MODEL

A. Fuzzy c-means clustering:

Fuzzy c-Means Clustering performs clustering by iteratively searching for a set of fuzzy clusters and the associated cluster centres that represent the structure of the data as best as possible. Fuzzy clustering by contrast allows data points to belong to more than one group. The resulting partition is therefore a fuzzy partition. Each cluster is associated with a membership function that expresses the degree to which individual data points belong to the cluster. The algorithm relies on the user to specify the number of clusters and data size present in the set of data to be clustered. Given a

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

number of clusters c , FCMC partitions the data $X = \{x_1, x_2, \dots, x_n\}$ into c fuzzy clusters by minimising the within group sum of squared error objective function as follows:

ALGORITHM

1. Randomly select cluster centre c
2. Initialize $U = [u_{ij}]_{matrix}$
The u_{ij} is calculated using

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{(m-1)}}$$
3. At k -step calculates the centres vectors

$$C^{(k)} = [C_j] \text{ with } U^{(k)}$$

$$C_j = \frac{\sum_{i=1}^N (\mu_{ij}^m \cdot x_i)}{\sum_{i=1}^N \mu_{ij}^m}$$
4. Update $U^{(k)}, U^{(k+1)}$
5. If, $\|U^{(k+1)} - U^{(k)}\| < \epsilon$ then stop or return to step 2

B. Gravitational search algorithm:

The gravitational search algorithm is the one to solve the optimization problems based on the Law of gravity. Gravitational search algorithm is applied on large numbers of problems because it requires only two parameters and having ability to find near global optimum solution and provides better results as compare to other nature inspired algorithms. In GSA, each particle has associated with four specifications: particle position, its inertial mass, active gravitational mass and passive gravitational mass. The gravitational force between two particles is directly proportional to the product of their masses and inversely proportional to the square of the distance between them. The particles are considered as objects and their performance has evaluated with their masses. The position of particles provides the solution of problem while fitness function is used to calculate the gravitational and inertial masses. A measure μ is added with fitness measure in order to gain high clustering accuracy. After this, update the global best and worst fitness among them. Further calculate the mass and acceleration for each candidate solution. For each candidate solution update the velocity and position. Finally, best candidate solution is obtained as output at $t + 1$ iteration.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

The flow diagram of the GSA algorithm is shown in the figure below:

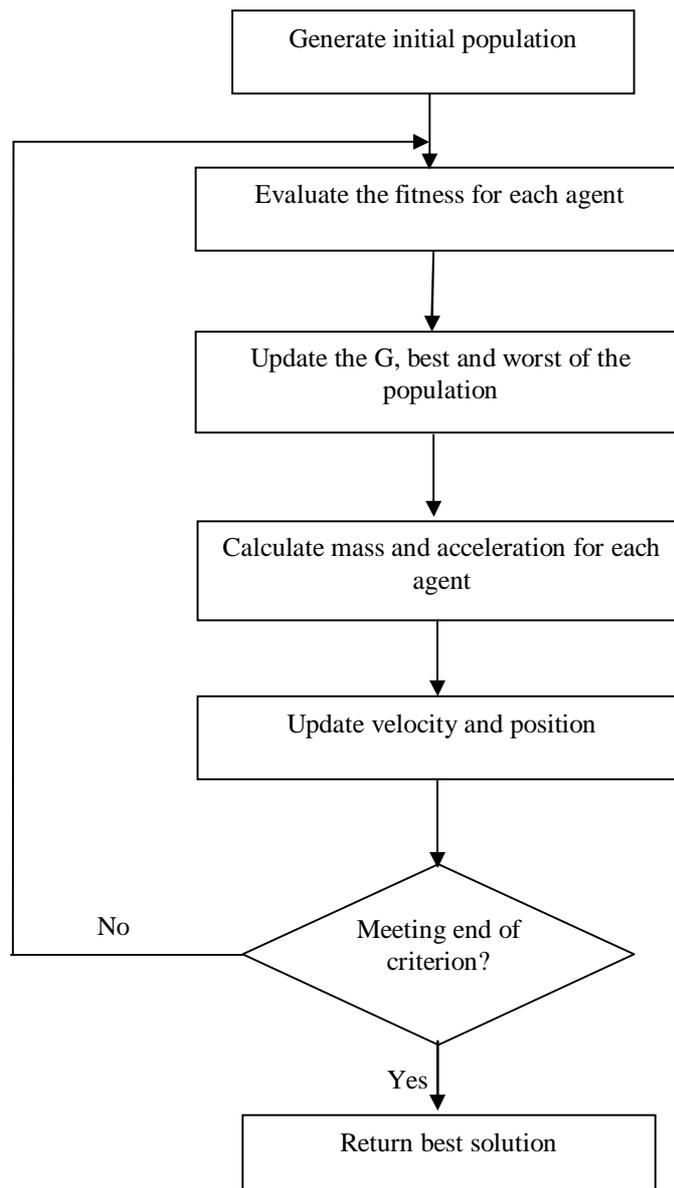


Figure.1 Proposed Block Architecture

The process of gravitational search algorithm starts with the identification the search space followed by initialization the swarm and masses. After finding masses of each object, fitness evaluation of objects is computed. Update Global best and worst fitness i.e., $G(t)$, $best(t)$, $worst(t)$. Find the total force in different directions to calculate of acceleration and velocity, after finding velocity and acceleration update objects' position. Repeat steps until the stop criteria is reached.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

V. IMPLEMENTATION AND RESULTS

We have implemented the proposed method using Java Programming Language in NetBeans IDE in a system having 6 GB RAM and 2.6 GHz Intel processor. The output shows the randomized and finalized centroids along with the optimized values and members included in clusters.

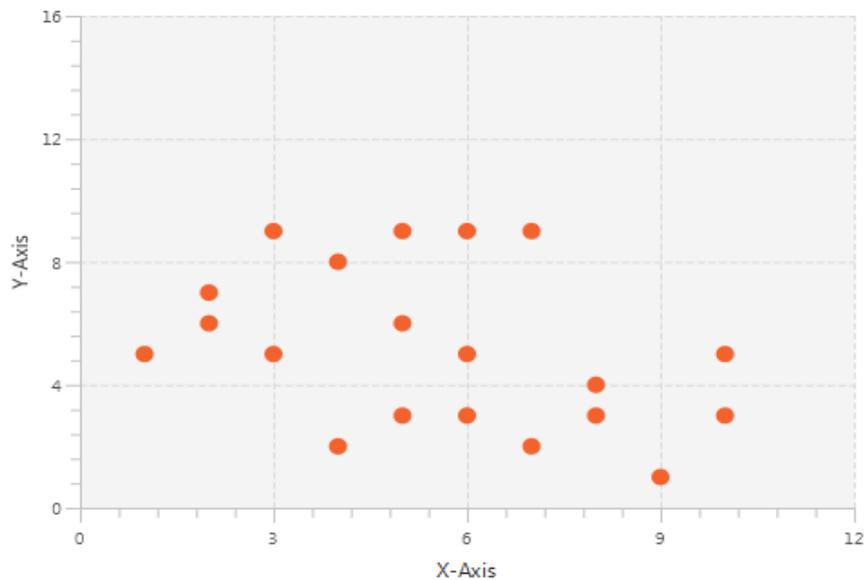


Figure.2 The graph shows the position of the data points given as input

The graph plotted Figure.2 shows the position of the data points given as input, graph plotted in Figure.3 shows the centroids with the objective function value for clusters size and data points which are similar are included in their respective clusters as members.

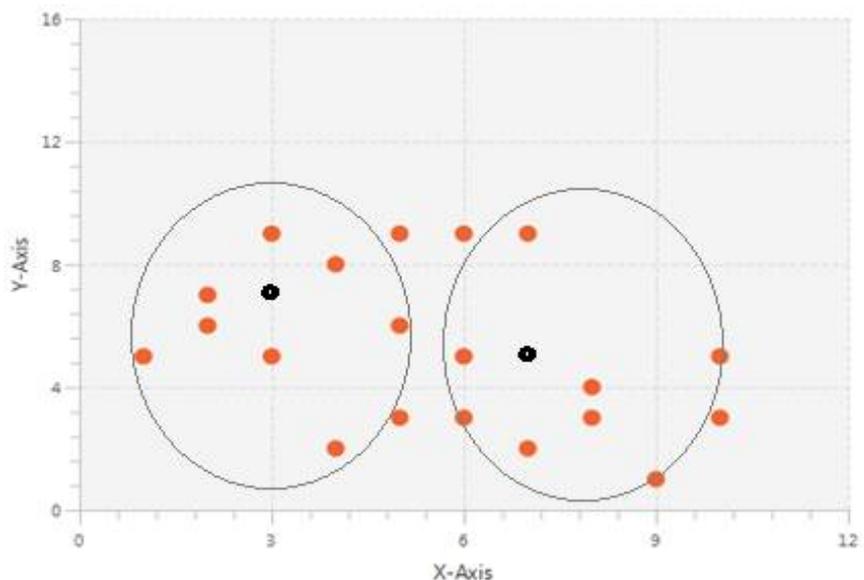


Figure.3 Fuzzy c-means and Gravitational search algorithm

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

VI. CONCLUSION

In this paper the fuzzy c-means clustering Algorithm and gravitational search optimization is used for optimal clustering. Fuzzy c-Means Clustering performs clustering by searching for a set of fuzzy clusters and the associated cluster centres using fuzzy algorithm. Gravitational search algorithm which is inspired by Newton's rule of gravity has ability to find optimum cluster centres which minimizes the fitness function of fuzzy c-means. Thereby, increasing the performance and efficiency of clustering process.

REFERENCES

- [1] PVS Srinivas, Susanta K Satpathy., Lokesh K Sharma., and Ajaya K Akasapu (2011), "Regionalisation as Spatial Data Mining Problem: A Comparative Study", Proc. International Journal of Computer Trends and Technology, Vol.18 No.5 pp. 577-589
- [2] M. Ester, H. P. Kriegel, J. Sander and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise", In: Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining (KDD'96), Portland, AAAI Press, pp. 291-316, 1996.
- [3] E. Rashedi, H. Nezamabadi-pour, and S. Saryazdi, GSA: "A Gravitational Search Algorithm" Information Sciences, 179(13) (2009), 2232-2248.
- [4] R. Sarkar, H. abbas and C. Newton, "Introducing data mining and knowledge discovery, Heuristic and Optimization for Knowledge Discovery", Idea Group Publishing, (2008), 1-23.
- [5] T. Sousa, A Silva and A. Neves, "Particle swarm based data mining algorithms for classification tasks", Parallel Computing, 30(1)(2004), 767-783.
- [6] R. S. Parpinelli, H. S. Lopes and A. A. Freitas, "Data mining with an ant colony optimization algorithm", IEEE Transactions on Evolutionary Computing, 6(4) (2002), 321-332.
- [7] A. Freitas, A survey of evolutionary algorithms for data mining and knowledge discovery, In A. Ghosh, S. Tsutsui, eds., Advances in Evolutionary Computation, Springer-Verlag, 2001.
- [8] B. Liu, H. A. Abbass and B. McKay Classification rule discovery with ant colony optimization, In Proceeding of the IEEE/WIC International Conference on Intelligent Agent Technology, Beijing, China, (2003), 83-88.
- [9] Peng Gao, John A. Kupfer, Diansheng Guo and Ting L. Lei, "Identifying functionally connected habitat compartments with a novel regionalization technique", Landscape Ecol, Springer, Vol.28, pp.1949-1959, 2013.
- [10] Yugalkumar and G. Sahoo, "A Review on Gravitational Search Algorithm and its applications to Data Clustering and Classification", IJ Intelligent Systems and Applications, Vol.06, pp.79-93, 2014.
- [11] Guo D, "Flow mapping and multivariate visualization of large spatial interaction data", IEEE Trans Vis Comput Graph, vol. 15, no.6, pp.1041-1048, 2009.
- [12] N. Sumathi, R. Geetha, "spatial data mining - techniques trends and its applications" Journal of Computer Applications, Vol - 1, No.4, Oct- Dec 2008
- [13] L.K. Sharma, S. Scheider, W. Kloesgen and O. P. Vyas, "Efficient clustering technique for regionalisation of a spatial database", International Journal Business Intelligence and Data Mining. 3 No. 1 pp. 66-81, 2008