International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 02 Issue: 04 | July-2015 www.irjet.net p-ISSN: 2395-0072

# A HYBRID FIREFLY BASED APPROACH FOR DATA CLUSTERING

Gunjan Dashora[1], Payal Awwal[2]

[1] *Student, Computer Science, Govt. Women Engineering College,Ajmer, Rajasthan, India*
[2] *Assistant Professor,Computer Science, Govt.Women Engineering College, Ajmer ,Rajasthan,India*

---

Abstract - *In data mining, clustering or cluster analysis is a common technique. Clusters is a set of objects which are assigned into a group. The objects in a single cluster are similar to each other but they are different from the objects in other clusters. A collective behavior of social systems like insects such as ants, fish schooling ,honey bees and birds are called as swarm intelligence(SI).In this paper, a swarm intelligence based technique for data clustering is proposed using firefly and nelder mead search method .K-means algorithm tends to converge faster than firefly algorithm but usually trapped in a local optimal area. A new way of integrating firefly with nelder mead search method proposed in this paper.*

*Key Words: Firefly algorithm, nelder mead simplex search, swarm intelligence, K-Means.*

## 1. INTRODUCTION

The process of grouping of data into number of clusters are known as data clustering. The aim of data clustering is to keep the objects in a single cluster are similar to each other but they are different from objects in other clusters.[1]. K-means is the most popular and widely used method for clustering.

In k-means , we use Euclidean distance for good clustering results.[2].However k-means contains several drawbacks like trapped into local optima and local minima and it is also sensitive to initial cluster centers.[3][4].

A new prototype Swarm Intelligence(SI) is being used in research settings to improve the management and control of large numbers of collaborating entities like computer and sensor networks, communication, satellite constellations and many more. It is a collective behavior of insects.[5] Firefly algorithm is used to overcome the problem of local optima in k-means algorithm. The firefly algorithm works on the flashing behaviour of firefly, but the convergence rate of Firefly algorithm is slower than those of local search technique( Nelder Mead simplex search). To deal with the the slow convergence of firefly ,we combine firefly with nelder- mead simplex search.

## 2. K-MEANS ALGORITHM

Semi structured or unstructured datasets are classify with the help of k-means clustering. K-means clustering is simple and it has the ability to handle voluminous datasets. Therefore, this is the one of the most common and effective method to classify data.

The parameter used in k-mean clustering is the number of clusters and the initial set of centroids. The distance of each item in the dataset is calculated with each of the centroids of the respective cluster. The item is then assigned to the cluster with which the distance of the item is least. The centroid of the cluster to which the item was assigned is recalculated.

The standard k- means algorithm is as follows-

Initial positions of K cluster centers are determined randomly. Following phases are repeated:

a)For each data vector : the vector is allocated to a cluster which its Euclidean distance from its center is less than the other cluster centers. The distance to cluster center is calculated by Eq. (1):

$$Dis(X_p, Z_j) = \sqrt{\sum_{i=1}^{D}(X_{pi} - Z_{ji})^2} \quad (1)$$

In Eq (1), $X_p$ is $p^{th}$ data vector, $Z_j$ is $j^{th}$ cluster center and D is the dimension of data and cluster center.

b) Cluster center are updated by Eq (2):

$$Z_j = \frac{1}{n_j}[\sum_{\forall Xp \in Cj} Xp] \quad (2)$$

In Eq.(2), $n_j$ is the number of data vectors corresponding to $j^{th}$ cluster and $C_j$ is a subset of the total data vectors which constitute $j^{th}$ cluster and are in it.

Phases (a) and (b) are repeated until stop criterion is satisfied.

## 3. FIREFLY ALGORITHM

Short and rhythmic flashes for communication and attracting the potential hunt are used by most of the fireflies. Yang introduce this firefly algorithm in 2008[6].

Firefly works on three rhapsodize rules-

1)All fireflies are unisex, so that one firefly will be attracted to other fireflies regardless of their sex.

2)Attractiveness is proportional to their brightness. Thus, for any two flashing fireflies, the less brighter one will move towards the brighter one. The attractiveness is proportional to the brightness and they both decrease as their distance increases. If there is no brighter one than a particular firefly, it will move randomly.

 

3) The brightness of a firefly is determined by the landscape of the objective function. The original firefly algorithm pseudo code is described below-

```
Firefly algorithm
Initialize algorithm parameters:
MaxGen: the maximum number of generations
Objective function of f(x), where x=(x1,……,xd)T
Generate initial population of fireflies or xi( i=1,2…n)
Define light intensity of li at xi via f(xi)
While (t< MaxGen)
        For i=1 to n (all n fireflies);
        For j=1 to n (all n fireflies)
If (lj>li), move firefly i towards j; end if
Evaluate new solutions and update light intensity;
   End for j;
   End for i;
Rank the fireflies and find the current best;
End while;
Post process results and visualization;
End procedure;
```

Fig-1 : the pseudo code of firefly algorithm

Variation of light intensity and the formulation of the attractiveness are the two important issues in the firefly algorithm. Brightness is associated with the objective function of the optimization problem and attractiveness of a firefly is determined by the brightness.
The attractiveness of a firefly can be determined by

$$\beta(r) = \beta_o e^{-\gamma r2} \quad (1)$$

where, $\beta_o$ is the attractiveness at r=0, and $\gamma$ is the light absorption coefficient at the source.
The distance between two fireflies i and j at $x_i$ and $x_j$ are determined by Cartesian distance-

$$r_{ij}=||x_i-x_j|| \sqrt{\sum_{k=1}^{d}(Xi,k - Xj,k)} \quad (2)$$

where $x_i$ and $x_j$ are the spatial coordinate of the fireflies I and j, respectively.
The movement of a firefly i, which is attracted to another more attractive firefly j is determined by

$$X_i= x_i+\beta_o e^{-\gamma r2_{ij}} (x_j-x_i)+\alpha(rand-1/2) \quad (3)$$

Where the second term is the attraction while the third term is randomization including randomization parameter $\alpha$ and the random number generator rand which is uniformly distributed in interval [0,1]

## 4. NELDER-MEAD SIMPLEX SEARCH

Nelder and Mead search is a derivative free line search method proposed by Nelder and Mead in(1865).It is used to solve the problems of nonlinear least squares, non linear simultaneous equation and other types of function minimization[7].
The procedure of Nelder Mead simplex search is as follows-

First evaluate the function values of initial simplex which is polyhedron in the factor space of N input variables at (N+1) vertices. Then the vertex with has highest function value is replaced by a newly reflected and better point for minimization case. This new reflected and better point is located in the negative gradient direction. Basically NM search method uses four basic operation: reflection, expansion, contradiction and shrinkage. Using this operation the simplex can improve itself and come closer and closer to a local optimum point successfully.

## 5.PROPOSED WORK

### 5.1 Hybrid NM-Firefly

For solving N dimensional problem, the population size of this hybrid NM-Firefly approach is set at 5N+1.The initial 5N+1 fireflies are randomly generated and sorted by fitness and the top N+1 fireflies are then fed into the simplex search method to improve the $(N+1)^{th}$ firefly. The other 4N fireflies are adjusted by the firefly method by taking into account the position of the N+1 best fireflies. This step of adjusting the 4N fireflies involves selection of the global best fireflies, selection of the neighborhood best firefly and finally movement update. The global best firefly of the population is determined according to the sorted fitness value. The neighborhood best fireflies are selected by first evenly dividing the 4N fireflies into N neighborhood and designating the fireflies with better fitness value in each neighborhood as the neighborhood best fireflies. By eq.(3), a movement update for each of the 4N fireflies is then carried out. The 5N+1 fireflies are sorted again in preparation for reporting the entire run. The whole process terminates when certain convergence criteria are met. The pseudo code of Hybrid NM- Firefly is described below-

```
1)Initialization
Generate a population of size 5N+1.
2)Evaluation and Ranking
Evaluate the fitness of each firefly. Rank them on the basis of fitness.
3)Simplex Method
Apply NM operator to the top N+1 fireflies and replace the (N+1)th firefly with the movement.
4) Firefly Method
Apply firefly operator for updating the remaining 4N fireflies.
Selection: From the population select the global best firefly and neighborhood best firefly.
Movement: Apply movement of firefly to the 2N fireflies with worst fitness according to eq (3)
5) If the termination conditions are not met, go back to equation 2.
```

Fig-2: the pseudo code of NM firefly

 

## 6. CONCLUSIONS

In this paper, a new hybrid algorithm based on NM search method and firefly is proposed to cluster data. Since the convergence rate of firefly algorithm is slow. To deal with the slow convergence rate of firefly, we proposed to combine Nelder-Mead simplex search method with firefly. The rationale behind it being that such a hybrid approach will enjoy the merits of both firefly and Nelder-Mead simplex search method.

## REFERENCES

[1] Jiawei Han, Micheline Kamber, "Data Mining: Concepts and Technique".

[2] K. Jain, M. N. Murty, P. J. Flynn, "Data Clustering: A Review", ACM Computing Surveys, Vol. 31, No. 3, September 1999.

[3] Xiong, H., J. Wu and J. Chen, 2009. K-Means clustering versus validation measures: A data distribution perspective. IEEE Trans. Syst.,Man,Cybernet.PartB,39:318331.http://www.ncbi.nlm.nih.gov/pubmed/19095536.R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[4] M.N.Joshi. Parallel K-Means Algorithm Distributed MemoryMultiprocessors[J].Computer, 2003, 9:3-15.

[5] Mark Fleischer," Foundations of Swarm Intelligence: From Principles to Practice", Swarming: Netwrk Enabled C4ISR 2003 By Mark Fleischer.

[6] X. S. Yang, "Nature-Inspired etaheuristic Algorithms".Luniver Press, 2008.

[7] Nelder, J.A. and Mead, R..(1965). A simplex method for function minimization. Computer Journal, 7, 308-313.

 