International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 10
September 2017

Metaheuristic Algorithms Based On the Law of Universal Gravitation for Data Grouping

Mr. Syed Abdul Moeed[1] and **Dr. Niranjan Polala**[2]

1. Research Scholar, Mewar University, Rajasthan, India
2. Research Supervisor in Department of Computer Science, Faculty of Computer Studies of Mewar University, Chittorgarh, Rajasthan, India

Abstract

This document contains a review of the state of the art of the application of metaheuristic techniques based on the law of universal gravitation to the mining group task of data, the motivation was caused by the identification of the two separate trends (metaheuristic and based algorithms in the law of universal gravitation) in recent investigations related to data grouping and results outstanding that these proposals have achieved in comparison to conventional data mining techniques.

Key Words: Data Mining, Data Grouping, Law of Universal Gravitation, Metaheuristic Algorithms.

1. INTRODUCTION

The metaheuristic term was introduced in 1986 by Golver in his study on what he considered at the time innovative developments with great potential to solve optimization problems from the perspectives of the Artificial intelligence research [1]. In the years later this type of algorithms have gained great popularity and demonstrating its ability to solve a wide range of problems in various application domains, always characterized by its ability to deal with problems complexes with little or no knowledge of the space of search, relevant feature in optimization problems [2]. Parallel proposals inspired by the law of gravitation universally have solved in an outstanding manner the problem of grouping of data, the research objective of the present document was determined by a literature review if the development of metaheuristic techniques based on the law of universal

gravitation has contributed to the solution of the problem of grouping of data and can be considered as initiatives with great potential for solving this type of problem and so arouse the interest of the scientific community towards these.

2. DESCRIPTION OF THE PROBLEM

Data mining has become the favorite tool for the extraction of knowledge under the conditions that the current information society requires, thanks to its strong ability to process automatically or semi-automatically large amounts of data stored in different formats to extract new and useful knowledge hidden in the data [3], the domain of application in which they have His techniques have been used with satisfactory results. For another side grouping is one of the most data mining tasks important, refers to the process of categorization of objects of data in groups with similar objects between them and different to the objects belonging to other groups, now a days a wide variety of grouping methods with different approaches, however it is well known that these methods have different limitations. In recent research some of these limitations were addressed through the use of metaheuristic optimization techniques, proposals that have demonstrated that they bring improvements in grouping tasks in regarding quality, complexity, time and computational cost. Additionally, the recent proposals for techniques grouping based on the law of universal gravitation [4], [6], [9], [10], [11], [12], [13] have also proven to solve the problem of pooling data effectively delivering better results than conventional mining techniques data, the objective of this research is to conduct a review of state of the art and identify how the integration of techniques metaheuristics and based on gravitation have contributed to the Solution of the problem of grouping data with the purpose to serve as a foundation for future research that is relate to the topics considered.

3. METAHEURISTIC ALGORITHMS

Metaheuristic derives from the Greek verb "heuristikein" that means "find" and the prefix "meta" which means "high level". In general terms, a metaheuristic algorithm can be seen as a high-level strategy that

develops directed random searches within possible solutions for find the best solution (close to the optimal) of a problem, were originally defined as methods that orchestrate a interaction between local improvement procedures and strategies of a higher level to create a process capable of escaping the local optima and perform a robust search of a space of solutions, then procedures were included that employ strategies to overcome the trap of local optima in complex solution spaces [1]. The algorithms Metaheuristics are mainly characterized by [2]:

1) They are strategies that "guide" the search process.

2) The basic concepts of a metaheuristic algorithm are can describe at an abstract level not tied to a problem specific.

3) The most advanced algorithms use the experience of search (simulate some kind of memory) to guide the search.

4) Incorporate mechanisms to avoid falling into local optima. Metaheuristic algorithms can be classified considering different aspects, for example: based or non-based on nature, with memory or without memory, with objective function static or dynamic, however the most appropriate classification is one that considers the manipulation in each iteration of a only search space point " *trajectory* " or a " *population* " set, the term " *trajectory* " is already used that the search generates a trajectory in the space of search, in other words the search starts from a point and by exploring the neighborhood, the solution varies currently forming a trajectory, usually emerge from of improving local search methods when joining techniques that allow them to escape from local optima, algorithms based on "trajectory" incorporate criteria of termination as a maximum number of iterations, identification of a stalemate or finding a solution what Enough acceptable.

Algorithms based on " *population* " work in parallel with a set of agents (solutions) in each iteration thus allowing a natural and intrinsic way to explore the search space [2], for example, inspired algorithms in the behavior of swarms use a collection of agents (solutions) similar to a natural flock of birds or fish, where each member executes a series of operations individuals

and share their information with others, these operations are generally simple, however, their effect collective, known as swarm intelligence, produces a amazing result. Local interactions between agents provide a global result that allows the system Solve the problem without using any central driver. In In this case, the operations of the members, including the random search, positive feedback, negative feedback and multiple interactions, lead to a situation of self-organization [4]. You can recognize two common tasks in the algorithms population-based metaheuristics: *exploration* and *exploitation* . *Exploration* is the ability to probe space search and *exploitation* is the ability to find the optimal around a good solution. In the first iterations a metaheuristic search algorithm *explores* the search space to find new solutions this will avoid falling into a local optimum, reason for the importance of this task, with the passage of iterations, the *exploration* fades and it is passed to the farm, so the algorithm is refined in semi-optimal points. The essential key to have a search High performance is a proper balance between *exploration* and *exploitation* , on the one hand to quickly identify regions in the search space with high quality solutions and for the other not to waste too much time in the regions of the search space that has already been explored or not provided high quality solutions. All metaheuristic algorithms based on population use *exploration* and *exploitation* but using different approaches and operators. On the other hand the agents of a search algorithm based in population, go through three steps in each iteration to perform exploration and exploitation: self-adaptation, cooperation and competition, in the step of self-adaptation each member (agent) improves its performance, in the cooperation step, the members collaborate with each other by transferring information and finally in the competition step, the members compete to survive. The above leads us to conclude that all metaheuristic search algorithms they have a common framework. When reviewing the literature, it can be shown that, in general, metaheuristic algorithms are inspired by nature and mimic physical or biological processes, among the most

popular we found: the Swarm Optimization algorithm of Particles *(Particle Swarm Optimization - PSO)*, which simulates the behavior of a flock of birds; Genetic algorithm *(Genetic Algorithm - GA)* , inspired by the theory of evolution of Darwin; The Simulation Cooking algorithm *(Simulated Annealing - SA)* , inspired by the effects of thermodynamics; *Ant* colony algorithm *(Ant Colony Optimization - ACO)* , which simulates the behavior of an ant colony in search of food [3].

4. OPTIMIZATION IN DATA GROUPING

Grouping is a fundamental task of learning computational and data mining that takes a set of data and classifies them into different groups based on similarity calculated between them, in such a way that in the same group find the objects (data records) most similar to each other and different from objects belonging to other groups. Exists a large number of grouping algorithms, usually focused on partition or hierarchical methods, both approaches they have their own advantages and limitations in terms of number, form and overlap of groups, currently some of the new proposals focus on the use of different techniques of optimization, the participation of optimization techniques smart in grouping tasks has managed to find effective ways to improve the complexity, time and cost of the processes of data mining [4]. Mathematically, a grouping problem can be define as follows, given $O = \{O_1, ..., O_n\}$ where O is a set finite n objects (vector) in a space of elements S , the objective of a data grouping problem is to find the optimal partition of the objects $C = \{C\ 1\ , ...\ C\ D\ \}$, $O = \cup = 1$, *and* $\cap =$; *for* $i \neq j$, where *Ci* represents the *i-th* group of the partition C , in such a way that the data belonging to it group are similar while they are as different as possible to the data that belong to other groups in terms of a distance measurement function. The resulting partitions are called groups and must meet the following conditions:

a) Each group must contain at least one object; b) the different groups should not have objects in common; c) Each object must be assigned to a single group, in other words, after assigning objects to groups, the

sum of the objects of all the groups must be equal to the number of objects of the original data set [5]. However, it is possible to generate different non-optimal partitions of the data set that satisfy the conditions mentioned above, it becomes necessary Then optimize the quality of the obtained partition. As is known the basic objective of an algorithm of Optimization is to minimize or maximize a target function systematically choosing values taken from a space of search, for the problem of grouping data you should select an objective function that allows to evaluate the quality 30 of the obtained partition. The most popular function for this is the mean square error that considers the cohesion of the groups in order to evaluate the quality of a given partition:

$$f(O,C) = \sum_{i=1}^{D} \sum_{o_j \in C_i} \|O_j - Z_i\|^2 \qquad (1)$$

Where D is the number of groups and $\|Oj - Zi\|$ is the *distance Euclidean* between a data object and the center of the group i , represented by the symbol , which can be calculated from of the following equation:

$$Z_i = \frac{1}{|C_i|} \sum_{k \in C_i} O_k \qquad (2)$$

Where $||$ is the cardinality of the group, that is the number of objects that the group i possesses. In grouping problems the objective may be to find the centroid of each group by *minimizing* an objective function as the sum of the distances between each object and the center of the group to which it is assigned expressed by the *error quadratic mean* calculated by Eq. (1).

5. LAW OF UNIVERSAL GRAVITATION

The law of universal gravitation was published in 1687 by Isaac Newton in his book " *Philosophiae Naturalis Principia Mathematica* " . This law enunciates a quantitative relation of the gravitational interaction between different bodies with mass, define that the force with which two bodies of different mass are attracted only depends on the value of its masses and the square of the distance that separates them and that said force acts in such a way that it is as if the whole mass of each one of the bodies was concentrated only in its center. That is, the older

mass have the bodies and closest to each other, with greater strength will be attracted [6].

The Universal Gravitation force is calculated as:

$$F(t) = \frac{Gm_x m_y}{d(x(t), y(t))^2} \qquad (3)$$

Where,

m_x: It is the mass of the first body

m_y: It is the mass of the second object

d: is the distance between the bodies

f(t) : is the module of the force exerted between both bodies, and

Its direction is in the axis that unites both bodies.

G: It is the constant of Universal Gravitation equal to:

$$G = (6.67428 \pm 0.00067) \times 10^{-11} Nm^2 kg^{-2} \qquad (4)$$

A metaheuristic algorithm based on the law of gravitation Universal considers that objects (solutions) are attracted between yes by the force of gravity and this force causes a global movement of all objects towards objects with heavier masses, therefore the masses cooperate using a direct communication through the force of gravity, the masses heavier ones that correspond to the best solutions, move more slowly than light masses, this guarantees the *exploitation* stage in the algorithm.

6. ALGORITHMS FOR GROUP GRAVITATIONAL

In 1977 WE Wright started the line of models of grouping based on the law of universal gravitation, presenting a grouping algorithm called "Gravitational clustering", designed an algorithm to develop cluster analysis in Euclidean data, I evaluate the results through its application in several data sets and compared them with those obtained through algorithms not gravitational with successful results [7], the algorithm gravitational system proposed by Wright is a hierarchical algorithm of agglomeration, the gravitational forces are used as mechanisms to join particles until a single particle continues in the system [7].

In 1999 S. Kundu, developed a grouping method based in the notion of the existence of an attractive force gravitational between each pair of points, with the novelty of not

use a measure of " *similarity* ", groups are formed by allow each point to move slowly under the effect resulting from all the forces acting on him and through the Fusion of two points when they are very close to each other. East model was considered as a refinement of the method of nearest neighbor and the diffuse K-means method [8].

In 2003, J. Gómez, D. Dasgupta and O. Nasraoui [6], proposed a new non-supervised gravitational grouping algorithm which is named for purposes of this investigation as "NGCA",

this method automatically determines the number of classes, Each object in the database is considered as an object in space and objects move using the law ofuniversal gravitation and Newton's second law, the algorithm proposed correctly works on data with noise and is based on the research developed by Wright, improving its speed, robustness in addition to achieving a non-supervised.

In 2012, M. Sánchez and O. Castillo [9], developed a algorithm for the search of groups, also based on the law of Newton's universal gravitation "FGGC", this included the fuzzy theory to refine the output groups. In terms General incorporates granularity into grouping analysis diffuse that has as a principle to obtain the optimum granule that fully represents the knowledge of the data set. Granular computing has been gaining a lot of interest in the last years and this work continues this trend in the area.

In 2014, A. Hashempour and H. Nezamabadi-pour, proposed a new algorithm called " *Gravitational Ensemble" Clustering - GEC* "[10], The proposed method combines the results obtained by different grouping algorithms like the K-means algorithm using concepts gravity, with the aim of overcoming the weaknesses of individual algorithms and improve their performance. The results Experiments showed their versatility, robustness and ability to generate better results than algorithms used separately.

7. ALGORITHMS FOR METAHEURISTIC GRAVITATIONAL GROUPING

At present, algorithmic proposals are scarce metaheuristics based on universal gravitation, In 2009 E.

Rashedi, H. Nezamabadi-pour and S. Saryzdi proposed the Gravitational Search Algorithm (*Gravitational Search Algorithm-GSA)* the only one found in the literature review with these characteristics. In the GSA algorithm [4], each mass (agent) has four specifications: position, mass of inertia, Active gravitational mass and passive gravitational mass. The position of the mass corresponds to a solution of the problem and the inertial and gravitational masses are determined using an objective function (fitness function), in other words, each Mass represents a solution and the algorithm navigates by adjusting correctly the gravitational and inertial masses, with the over time, the masses are expected to be attracted by the heaviest, these masses represent an optimal solution in the search space. Whereas, a system with *N* agents (masses). The definition is defined position of the *i-th* agent by:

$$X_i = (x_i^1, \dots, x_i^d, \dots, x_i^n) \; para \; i = 1,2, \dots, N \quad (5)$$

Where $x_i$presents the position of the *i-th* agent in the dimension

8. CONCLUSION

The document presented a review of the investigations previous related to metaheuristic algorithms based on the law of universal gravitation applied to the task of grouping of data, it was evidenced that it is a little explored field without However, the investigations carried out so far have delivered in all cases results that exceed the results obtained by conventional mining algorithms of data like k-means and other metaheuristic algorithms as PSO (*Particle Swarm Optimization*), on the other hand that manage to overcome limitations such as the vulnerability to fall optimal localities that conventional algorithms have popular as k-medias. This leads to saying that these kinds of proposals have great potential for the development of tasks of application in different data grouping problems.

9. FUTURE WORK

The development of studies is proposed as future work comparisons between metaheuristic algorithms based on

the law of universal gravitation and the conventional algorithms of data mining applied to different grouping problems of data in order to determine in which problems the first ones solve the limitations identified for seconds.

10. REFERENCES

[1] F. Glover and G. Kochenberger, "Handbook of Metaheuristics.,» *Kluwer Academic Publishers, Norwell, MA*, 2012.

[2] E. Alba, «Parallel Metaheuiristics: A new class of algorithm,» *Wiley Interscience* ,, 2005.

[3] IH Witten and E. Frank, Data Mining: Practical MachineLearning Tools and Techniques (Second ed.), Burlington, USA: Morgan Kaufmann, 2005

[4] E. Rashedi, H. Nezamabadi-pour and S. Saryzdi, «GSA: A Gravitational Search Algorithm., » *Information Sciences*, vol.179, pp. 2232-2248, 2009.

[5] A. Shafiq and DRP Gillian, «An Evolutionary Particle Swarm Optimization Algorithm for Data Clustering, »by *IEEE Swarm Intelligence Symposium* , St. Louis MO USA, 2008.

[6] A. Hatamlou, S. Abdullah and H. Nezamabadi-pour, «A Combined Approach for Clustering Based on K-means and Gravitational Search Algorithms, » *Swarm and Evolutionary Computation*, nº, p. 47-52, 2012.

[7] J. Goméz, D. Dasgupta and O. Nasraoui, Composers, *New Algorithm for Gravitational Clustering.* [Sound recording].In Proc. of the SIAM Int. Conf. on Data Mining. 2003 [8] W. Wright, "Gravitational clustering,» *Pattern Recognition*, No. 9,pp. 151-166, 1977.

[9] S. Kundu, "Gravitational clustering: a new approach based on the spatial distribution of the points, »Computer Science Department, Louisiana State University, Baton Rouge, 1999.

[10] O. Castillo, J. Castro and A. Rodríguez, «Fuzzy granular gravitational clustering Algorithm, » *Fuzzy Information Processing Society Annual Meeting of the North American*
(NAFIPS), 2012.

[11] A. Sadeghian and H. Nezamabadi-pour, "Gravitational Ensemble Clustering, »Department of Electrical Engineering,

Shahid Bahonar University, Kerman, Iran, 2014.

[12] A. Hatamlou, S. Abdullah and Z. Othman, «Gravitational Search Algorithm with Heuristic Search for Clustering Problems, » *Data Mining and Optimization Conference On (DMO)*, 2011.

[13] M. Dowlatshahi and H. Nezamabadi-pour, «GGSA: A Grouping Gravitational Search Algorithm for Data Clustering, » *Elservier - Engineering Applications of Artificial Intelligence*, vol. 36, pp.114- 121, 2014.

[14] L. Peng, B. Yang, Y. Chen and A. Abraham, «Data gravitation based classification, »Information Sciences - Sciences Direct, vol. 179, no. 6, pp. 809-819, 2009.

[15] IH Witten and E. Frank, from Data Mining: Practical Machine Learning Tools and Techniques, Burlington, USA, Morgan Kaufmann, 2005.

[16] C. Pérez López, Data mining: techniques and tools, Madrid - Spain: Thomson, 2007.

[17] H. Barrera, J. Correa and J. & Rodríguez, «Prototype Software for the preprocessing of data "UD-Clear", »IV Symposium International Information Systems and Engineering
Software in the Knowledge Society SISOFT - Cartagena, Colombia, pp. 167-184, 2006.

[18] M. Berry and GS Linoff, Data Mining Techniques, Indianapolis, Indiana, USA: Wiley, 2004.

[19] JER Rodríguez, Fundamentals of data mining, Bogotá DC, Colombia: Francisco José de Caldas District University, 2010

[20] VJ Rayward-Smith, "Metaheuristics for Clustering in KDD," Evolutionary Computation, 2005. The 2005 IEEE Congress, Australia, 2005.

[21] Y. Kuma and G. Sahoo, "A Review on Gravitational Search Algorithm and its Applications to Data Clustering & Classification, »Modern Education and Computer Science Press,
No. 06, pp. 79-93, 2014.

[22] S. Abdullah, A. Hatamlou and Z. Othman, Composers, Gravitational Search Algorithm with Heuristic Search for Clustering Problems. [Sound recording]. Data Mining and Optimization Conference on (DMO). 2011

[23] G. Oatley and B. Ewart, "Cluster Analysis and Data Mining Applications, »Analysis, vol. 1, No. 2, pp. 147-153, 2011.

[24] J. Hernandez, J. Ramírez and C. Ferri, Introduction to Mining of data, Editorial Alhambra SA, 2004.

[25] T. Long and LW Jin, «A New Simplified Gravitational Clustering Method for Multi-prototype Learning Based on Minimum Classification Error Training Advances in Machine Vision, Image Processing, and Pattern Analysis, International Workshop on Intelligent Computing., »Computer Science, N. Zheng, X. Jiang, and X. Lan, vol. 4153, pp. 168-175, 2006.