International Journal of Advanced Computational Engineering and Networking, ISSN: 2320-2106, Volume-4, Issue-10, Oct.-2016

CLUSTERING BASED ON BLACK HOLE PHENOMENON

[1]KUO-LUNG WU, [2]AVUS HOU, [3]CHUN-MIN HUNG, [4]MIIN-JYE WEN

[1,3]Department of Information Management, Kun Shan University, Tainan, Taiwan, ROC
[2]Department of Marketing and Distribution Management, Oriental Institute of Technology, Taipei, Taiwan, R.O.C.
[4]Department of Statistics, National Cheng Kung University, Tainan, Taiwan, ROC
E-mail: [1]klwu@mail.ksu.edu.tw

Abstract- Big data analysis is a popular topic in recent years and clustering is one tool can be used to discover the data knowledge. Many clustering optimization problems are solved by a Nature-inspired heuristic algorithm. In this paper, we will propose a black hole clustering algorithm (BHC) by using the space gravitational phenomenon. Based on the concept of universal gravitation, we define a function to measure the gravitation of a star. Since the gravitational power of the black hole is so strong, the location of the black will have a maximum gravitation estimation function. At each step of BHC, the stars (data objects) will then attracted to those invisible black holes and finally be swallowed.

Keywords- clustering; black hole; universal gravitation.

I. INTRODUCTION

Many clustering optimization problems are solved by a Nature-inspired heuristic algorithm [1] such as genetic algorithms (GA) [2], simulated annealing (SA) [3], ant colony optimization (ACO) [4], particle swarm optimization (PSO) [5], firefly algorithm (FA) [6], bat algorithm (BA) [7] and so on. Nature-inspired heuristic algorithms have now been applied in many fields such as computer science, data mining, computer vision, forecasting, medicine and biology, scheduling, economy and engineering [8-18].

Black hole concept was first proposed by John Michell and Pierre Laplace which formulated the Newton's law of a star becoming invisible. A black hole in space is formed when a star of massive size collapses and the gravitational power of the black hole are too high that even the light cannot escape from it. The black hole has so strong gravity because matter has been squeezed into a tiny space and it will swallow anything that crosses to its boundary. Nothing can get away from the sphere-shaped boundary of a black hole known as the event horizon. The radius of the event horizon is termed as the Schwarzschild radius that can be calculated by the following equation:

$$R = \frac{2GM}{C^2} \qquad (1)$$

where G is the gravitational constant, M is the mass of the black hole, and C is the speed of light. The radius is proportional to the black hole mass. If anything moves close to a black hole's Schwarzschild radius it will be absorbed and then permanently disappear. Since the black hole is invisible, the existence can be discerned by its effect over the objects surrounding it [19,20].

Using black hole phenomenon in optimization problem can be found in [21,22]. In [21], the method introduces a new mechanism into particle swarm optimization (PSO). This method is an extension of the PSO and a new generated particle called the black hole attracts other particles under certain conditions. The objects of this method are to accelerate the convergence speed of the PSO and also to prevent the local optima problem. The algorithm only little connects to the phenomenon of black hole. In [22], a k-means objective function is used to identify the randomly generated candidates and a star gets too close to the black hole will be swallowed and gone forever. Since no clear description about the detail processes, the readers do not know how to use this black hole algorithm. Moreover, these two methods do not mention about a very important phenomenon of black the invisible and also the black holes cannot be randomly generated.

In this paper, we will propose a black hole clustering algorithm (BHC) by using the space gravitational phenomenon. Based on the concept of universal gravitation, we define a function to measure the gravitation of a star and the black holes will lie on a local-maximum gravitation location. At each step of BHC, the stars (data objects) will then attracted to those black holes and finally be swallowed.

II. BLACK HOLE CLUSTERING ALGORITHM

2.1. The Gravitation Estimation Function

In Newton's law of universal gravitation, every point mass attracts every single other point mass by a force pointing along the line intersecting both points. The force is proportional to the product of the two masses and inversely proportional to the square of the distance between them. Suppose s_i and s_j denote two stars, m_i and m_j are the mass of s_i and s_j. The gravitation of s_i and s_j can be defined by:

$$F(s_i, s_j) = G\frac{m_1 m_2}{R_{ij}^{\ 2}} \qquad (2)$$

where F is the force between the masses, G is the gravitational constant and R_{ij} is the distance between the centers of the masses. The gravitation of s_i and s_j is proportional to R_{ij} and we rewrite (2) as

$$F(s_i, s_j) = f(R_{ij}) = f(R(s_i, s_j))\,. \quad (3)$$

We then define the gravitation of a location x in space as

$$F(x) = \sum_i F(s_i, x) = \sum_i f(R(s_i, x))\,. \quad (4)$$

The following is a simple example of gravitation estimation function. Figure 1(a) illustrates the distribution of the stars in space (the data set) and the estimated gravitation distribution is shown in Fig. 1(b).

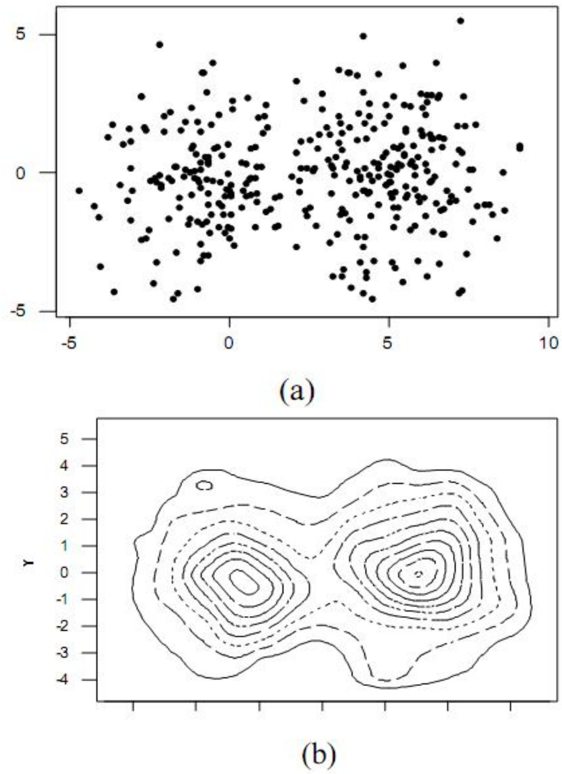

Fig.1. (a) The distribution of the stars in space (the data set). (b) The estimated gravitation distribution

2.2. The proposed black hole clustering algorithm

Since the gravitational power of the black hole is so strong, the location of the black hole will have a maximum gravitation estimation function and we can find it by

$$\begin{aligned} \frac{\partial}{\partial x} F(x) &= \frac{\partial}{\partial x} \sum_i f(R(s_i, x)) \\ &= \sum_i \frac{\partial}{\partial x} f(R(s_i, x)) = 0 \end{aligned} \quad (5)$$

If we suppose $R(s_i, x)$ is a function of the inner product with $(x - s_i)^T (x - s_i)$, Eq. (5) can be written as

$$\begin{aligned} &\sum_i \frac{\partial}{\partial x} f(R(s_i, x)) \\ &= \sum_i \frac{\partial}{\partial x} f(R((x - s_i)^T (x - s_i)))\,. \quad (6) \\ &= \sum_i f'(R(x - s_i))(x - s_i) = 0 \end{aligned}$$

We then have

$$x = \frac{\sum_i f'(R(s_i, x)) s_i}{\sum_i f'(R(s_i, x))}\,. \quad (7)$$

Although both right and left hand side have the variable x, we can use a Fix-Point iterative method to solve it. If we set the stars (all data points) as the initial values for the Fix-Point iteration, the iterative process will pull the stars into the black holes.

III. NUMERICAL EXAMPLES

We now illustrate some numerical examples. Figure 2 shows a three clusters data set and the estimated gravitation distribution. Figure 3 shows the process of the proposed black hole clustering algorithm (BHC). In each iterative step, the stars are pulled into the invisible black holes and finally be swallowed.

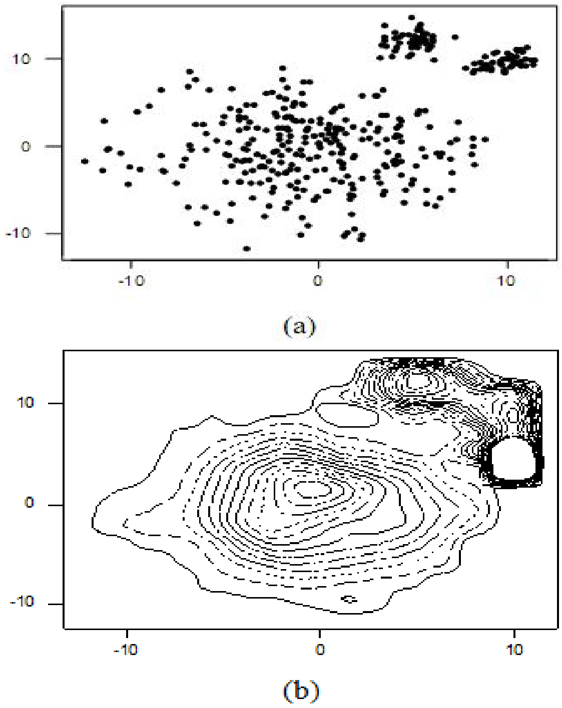

Fig.2. (a) The distribution of the stars in space (the data set). (b) The estimated gravitation distribution

Here we give another example. Figure 4(a) is an artificial data set and the estimated gravitation distribution is shown in Fig. 4(b). All data points (stars) are assigned as the initial values in the proposed BHC and the locations of all stars after 5 iterations are shown in Fig. 4(c). In each iterative step, the stars are pulled into the invisible black holes. Finally, all data points will converge to the black holes.

CONCLUSIONS

According to Black Hole phenomenon, we proposed a new Nature-inspired algorithm. We first use Newton's law of universal gravitation to defined the gravitation of a location x in space as Equation (4) and then the location of the black hole will have a maximum gravitation estimation function. In order to find the black holes, we use a Fix-Point iterative method (Eq. (7)) and all data points (stars) are assigned as the initial values. This hence the iterative process (black hole clustering) will pull all stars into the invisible black holes. In general, the gravitation estimation function should to a monotone decreasing function of $R(s_i, x)$ and we adopt the Gaussian function in our examples.

REFERENCES

[1]. X.S. Yang, Nature-Inspired Metaheuristic Algorithms, Luniver Press, 2008.

[2]. R.L. Haupt, S.E. Haupt, Practical Genetic Algorithms, second ed., John Wiley & Sons, 2004.

[3]. D.S. Johnson, C.R. Aragon, L.A. McGeoch, C. Schevon, Optimization by simulated annealing. An experimental evaluation. Part I. Graph partitioning, Operations Research 37 (1989) 65–892.

[4]. M. Dorigo, C. Blum, Ant colony optimization theory: a survey, Theoretical Computer Science 344 (2005) 243–278.

[5]. J. Kennedy, R. Eberhart, Particle swarm optimization, in: Proceedings of IEEE International Conference on Neural Networks, vol. 1944, 1995, pp. 1942–1948.

[6]. M. Bramer, R. Ellis, M. Petridis, X.S. Yang, Firefly algorithm, Lévy flights and global optimization, Research and Development in Intelligent Systems, vol. XXVI, Springer, London, 2010, pp. 209–218.

[7]. J. Gonzalez, D. Pelta, C. Cruz, G. Terrazas, N. Krasnogor, X.S. Yang, A new metaheuristic bat-inspired algorithm, in: Nature Inspired Cooperative Strategies for Optimization (NICSO 2010), Springer, Berlin/Heidelberg, 2011, pp. 65–74.

[8]. J. Christmas, E. Keedwell, T.M. Frayling, J.R.B. Perry, Ant colony optimization to identify genetic variant association with type 2 diabetes, Information Sciences 181 (2011) 1609–1622.

[9]. J.F. Connolly, E. Granger, R. Sabourin, An adaptive classification system for video-based face recognition, Information Sciences 192 (2012) 50–70.

[10]. E. Cuevas, D. Oliva, D. Zaldivar, M. Perez-Cisneros, H. Sossa, Circle detection using electro-magnetism optimization, Information Sciences 1 82 (2012) 40–55.

[11]. M. El-Abd, Performance assessment of foraging algorithms vs. evolutionary algorithms, Information Sciences 182 (2012) 243–263.

[12]. S. Ghosh, S. Das, S. Roy, S.K. Minhazul Islam, P.N. Suganthan, A differential covariance matrix adaptation evolutionary algorithm for real parameter optimization, Information Sciences 182 (2012) 199–219.

[13]. V.J. Manoj, E. Elias, Artificial bee colony algorithm for the design of multiplier-less nonuniform filter bank transmultiplexer, Information Sciences 192 (2012) 193–203.

[14]. D. Picard, A. Revel, M. Cord, An application of swarm intelligence to distributed image retrieval, Information Sciences 192 (2012) 71–81.

[15]. S. Rana, S. Jasola, R. Kumar, A review on particle swarm optimization algorithms and their applications to data clustering, Artificial Intelligence Review 35 (2011) 211–222.

[16]. J. Wang, H. Peng, P. Shi, An optimal image watermarking approach based on a multi-objective genetic algorithm, Information Sciences 181 (2011) 5501–5514.

[17]. W.C. Yeh, Novel swarm optimization for mining classification rules on thyroid gland data, Information Sciences 197 (2012) 65–76.

[18]. Y. Zhang, D.-W. Gong, Z. Ding, A bare-bones multi-objective particle swarm optimization algorithm for environmental/economic dispatch, Information Sciences 192 (2012) 213–227.

[19]. L. Kaper, E. Heuvel, P. Woudt, R. Giacconi, Black hole research past and future, in: Black Holes in Binaries and Galactic Nuclei: Diagnostics, Demography and Formation, Springer, Berlin/Heidelberg, 2001, pp. 3–15

[20]. C. Pickover, Black Holes: A Traveler's Guide, John Wiley & Sons, 1998.

[21]. L. Zhang, Q. Cao, A novel ant-based clustering algorithm using the kernel method, Information Sciences 181 (2010) 4658–4672.

★★★