Spatial Clustering with Obstacles Constraints by Dynamic Piecewise-Mapped and Nonlinear Inertia Weights PSO

Xueping Zhang[1], Haohua Du[2], and Jiayao Wang[1,3]

[1] School of Information Science and Engineering, Henan University of Technology, Zhengzhou 450052, China
[2] School of Computer Science and Engineering, Beihang University,Beijing 100191, China
[3] School of Surveying and Mapping, PLA Information Engineering University, Zhengzhou 450052, China
zhang_xpcn@yahoo.com.cn

Abstract. Spatial clustering with constraints has been a new topic in spatial data mining. A novel Spatial Clustering with Obstacles Constraints (SCOC) by dynamic piecewise-mapped and nonlinear inertia weights particle swarm optimization is proposed in this paper. The experiments show that the algorithm can not only give attention to higher local constringency speed and stronger global optimum search, but also get down to the obstacles constraints and practicalities of spatial clustering; and it performs better than PSO K-Medoids SCOC in terms of quantization error and has higher constringency speed than Genetic K-Medoids SCOC.

Keywords: Spatial Clustering with Obstacles Constraints, Particle Swarm Optimization, Dynamic Piecewise Linear Chaotic Map, Dynamic Nonlinear Inertia Weights.

1 Introduction

Spatial clustering with constraints has been a new topic in spatial data mining. Spatial clustering with constraints has two kinds of forms [1]. One kind is Spatial Clustering with Obstacles Constraints (SCOC), such as bridge, river, and highway etc. whose impact on the result should be considered in the clustering process of large spatial data. Ignoring the constraints leads to incorrect interpretation of the correlation among data points. The other kind is spatial clustering with handling operational constraints, it consider some operation limiting conditions in the clustering process. In this paper, we mainly discuss SCOC.

Since K.H.Tung put forward a clustering question COE (Clustering with Obstacles Entities) [2] in 2001, a new studying direction in the field of clustering research have been opened up. To the best of our knowledge, only four clustering algorithms for clustering spatial data with obstacles constraints have been proposed very recently: COD-CLARANS [2] based on the Partitioning approach of CLARANS,

M.J. Zaki et al. (Eds.): PAKDD 2010, Part I, LNAI 6118, pp. 254–261, 2010.

AUTOCLUST+ [3] based on the Graph partitioning method of AUTOCLUST, DBCluC [4-5] based on the Density-based algorithm, and GKSCOC [6] based on Genetic algorithms (GAs) and Partitioning approach of K-Medoids. Although these algorithms can deal with some obstacles in the clustering process, many questions exist in them [6].

PKSCOC based on Particle Swarm Optimization (PSO) and K-Medoids is presented [7] by us. However, the performance of simple PSO depends on its parameters, it often getting into local optimum and fails to converge to global optimum. A lot of improved methods were presented by many scholars, e.g. the paper [8-9] presented the Quantum PSO algorithm, and the paper [10-12] presented the Chaotic PSO algorithm. Recently, Dynamic Piecewise-mapped and Nonlinear Inertia Weights PSO (PNPSO) is proposed in [13]. Experiments and comparisons demonstrated that PNPSO outperformed several other well-known improved PSO algorithms on many famous benchmark problems in all cases.

This article developed a novel spatial clustering with obstacles constraints by PNPSO to cluster spatial data with obstacles constraints, which called PNPKSCOC. The contrastive experiments show that PNPKSCOC is better than PKSCOC in terms of quantization error and has higher constringency speed than GKSCOC.

The remainder of the paper is organized as follows. Section 2 introduces PNPSO algorithm. Section 3 presents PNPKSCOC. The performances of PNPKSCOC are showed in Section 4, and Section 5 concludes the paper.

2 Dynamic Piecewise-Mapped and Nonlinear Inertia Weights PSO

2.1 Classical PSO

PSO is a population-based optimization method first proposed by Kennedy and Eberhart. The mathematic description of PSO is as the following. Suppose the dimension of the searching space is D, the number of the particles is n. Vector $X_i = (x_{i1}, x_{i2}, \dots, x_{iD})$ represents the position of the i^{th} particle and $pbest_i = (p_{i1}, p_{i2}, \dots, p_{iD})$ is its best position searched by now, and the whole particle swarm's best position is represented as $gbest = (g_1, g_2, \dots, g_D)$ .Vector $V_i = (v_{i1}, v_{i2}, \dots, v_{iD})$ is the position change rate of the i^{th} particle. Each particle updates its position according to the following formulas:

$$v_{id}(t+1) = wv_{id}(t) + c_1 rand()[p_{id}(t) - x_{id}(t)] + c_2 rand()[g_d(t) - x_{id}(t)] \tag{1}$$

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1)\ ,\ 1 \le i \le n,\ 1 \le d \le D \tag{2}$$

where c_1 and c_2 are positive constant parameters, $Rand()$ is a random function with the range [0, 1], and w is the inertia weight. Eq.1 is used to calculate the particle's new velocity, the particle flies toward a new position according to Eq.2.

2.2 Coordinate PSO with Dynamic Piecewise Linear Chaotic Map and Dynamic Nonlinear Inertia Weights

2.2.1 Dynamic Piecewise Linear Chaotic Map

The well-known Piecewise linear chaotic map is defined as follows [10]:

$$x_{t+1} = \begin{cases} x_t / p_c, & x_t \in (0, p_c) \\ (1-x_t)/(1-p_c), & x_t \in (p_c, 1) \end{cases} \tag{3}$$

where p_c is the control parameter and X is a variable. Although Eq.3 is deterministic, it exhibits chaotic dynamics in (0, 1) when $p_c \in (0, 0.5) \cup (0.5, 1)$. The newly introduced dynamic Piecewise linear chaotic map [14] is incorporated into the PSO inertia weight which is described in equations (4) and (5).

$$\alpha = \alpha_{\max} - (\alpha_{\max} - \alpha_{\min})\left(\frac{iter}{iter_{\max}}\right) \tag{4}$$

$$w = \alpha + (1-\alpha)Pmap \tag{5}$$

where α is the dynamic chaotic inertia weight adjustment factor, $\alpha_{\max}$ and $\alpha_{\min}$ represent the maximum and minimum values of α respectively, $iter$ is the current iteration number, $iter_{\max}$ is the maximum iteration number, and $Pmap$ is the result of Piecewise linear chaotic map.

2.2.2 Dynamic Nonlinear Equations

To achieve trade-off between exploration and exploitation, two types of dynamic nonlinear inertia weights are introduced [13]. In this paper, the first type is proposed in equations (6) and (7):

$$dnl = dnl_{\min} + (dnl_{\max} - dnl_{\min})\left(\frac{iter}{iter_{\max}}\right) \tag{6}$$

$$w = w_{\min} + (w_{\max} - w_{\min})\left(\frac{iter_{\max} - iter}{iter_{\max}}\right)^{dnl} \tag{7}$$

where dnl represents the dynamic nonlinear factor, w represents the inertia weight, $w_{\max}$ and $w_{\min}$ represent the maximum and minimum value of w respectively, $dnl_{\max}$ and $dnl_{\min}$ represent the maximum and minimum value of dnl respectively, $iter$ represents the current iteration number, and $iter_{\max}$ represents the maximum iteration number.

2.2.3 Parallel Inertia Weight Adjustment

To avoid the premature convergence problem and to achieve the balance between global exploration and local exploitation, dynamic Piecewise linear chaotic map and dynamic nonlinear equations are used in parallel to dynamically adjust PSO inertia weight w, which is described as follows[13]:

Initialize all the parameters.
repeat
Evaluate the fitness values of all the particles
if $f_i \succ f_{avg}$
Equations (4), (5), (1) and (2) are employed
Elseif $f_i \le f_{avg}$
Equations (6), (7), (1) and (2) are employed
endif
until (a termination criterion is met)

where f_i is the fitness value of particle i and f_{avg} is the average fitness value of the whole population.

3 Spatial Clustering with Obstacles Constraints by PNPSO

3.1 Motivating Concepts

To derive a more efficient algorithm for SCOC, the following definitions are first introduced.

Definition 1 (Visibility graph). Given a set of m obstacle, $O = (o_1, o_2, \ldots, o_m)$,the visibility graph is a graph $VG = (V, E)$ such that each vertex of the obstacles has a corresponding node in V , and two nodes v_1 and v_2 in V are joined by an edge in E if and only if the corresponding vertices they represent are visible to each other.

To generate VG , we use VPIA (VGRAPH Point Incorporation Algorithm) as presented in [14].

Definition 2 (Obstructed distance). Given point p and point q , the obstructed distance $d_o(p,q)$ is defined as the length of the shortest Euclidean path between two points p and q without cutting through any obstacles.

We can use Dijkstra Algorithm to compute obstructed distance. The simulation result is in Fig.1 and the red solid line represents the obstructed distance we got.

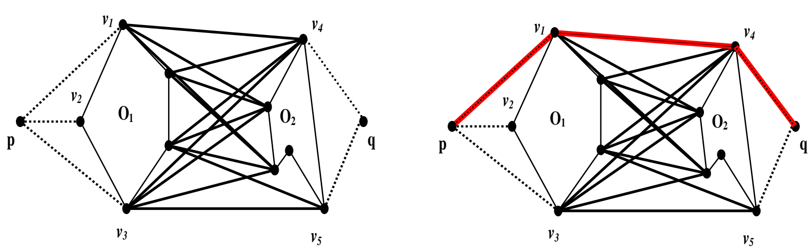


Fig. 1. Visibility graph and Obstructed distance

3.2 Spatial Clustering with Obstacles Constraints by Improved K-Medoids

K-Medoids algorithm is adopted for SCOC to avoid cluster center falling on the obstacle. Square-error function is adopted to estimate the clustering quality, and its definition can be defined as:

$$E = \sum_{j=1}^{N_c} \sum_{p \in C_j} (d(p, m_j))^2 \tag{8}$$

where N_c is the number of cluster C_j, m_j is the cluster centre of cluster C_j, $d(p,q)$ is the direct Euclidean distance between the two points p and q.

To handle obstacle constraints, accordingly, criterion function for estimating the quality of spatial clustering with obstacles constraints can be revised as:

$$E_o = \sum_{j=1}^{N_c} \sum_{p \in C_j} (d_o(p, m_j))^2 \tag{9}$$

where $d_o(p,q)$ is the obstructed distance between point p and point q.

The method of IKSCOC is adopted as follows [4].

1. Select N_c objects to be cluster centers at random;
2. Assign remaining objects to nearest cluster center;
3. Calculate E_o according to Eq.9;
4. While (E_o changed) do {Let current $E = E_o$;
5. Select a not centering point to replace the cluster center m_j randomly;
6. Assign objects to the nearest center;
7. Calculate E according to Eq.8;
8. If $E >$ current E , go to 5;
9. Calculate E_o ;
10. If $E_o <$ current E , form new cluster centers}.

While IKSCOC still inherits two shortcomings, one is selecting initial value randomly may cause different results of the spatial clustering and even have no solution, the other is that it only gives attention to local constringency and is sensitive to an outlier.

3.3 Spatial Clustering with Obstacles Constraints Based on PNPSO and Improved K-Medoids

In the context of clustering, a single particle represents the N_c cluster centroid. That is, each particle X_i is constructed as follows:

$$X_i = (m_{i1}, \ldots, m_{ij}, \ldots, m_{iN_c}) \tag{10}$$

where m_{ij} refers to the j^{th} cluster centroid of the i^{th} particle in cluster C_{ij}. Here, the objective function is defined as follows:

$$f(x_i) = 1/J_i \tag{11}$$

$$J_i = \sum_{j=1}^{N_c} \sum_{p \in C_{ij}} d_o(p, m_j) \tag{12}$$

The PNPKSCOC is developed as follows.

1. Execute the IKSCOC algorithm to initialize one particle to contain N_c selected cluster centroids;
2. Initialize the other particles of the swarm to contain N_c selected cluster centroids at random;
3. For $t=1$ to t_{max} do {
4. For $i = 1$ to no_of_particles do {
5. For each object p do {
6. Calculate $d_o(p,m_{ij})$;
7. Assign object p to cluster C_{ij} such that $d_o(p,m_{ij}) = \min \forall c = 1,...,N_c \{d_o(p,m_{ic})\}$;
8. Evaluate fitness of particle according to Eq.11;
9. if $f_i \succ f_{avg}$ Update particles using equations (4), (5), (1) and (2);
10. Elseif $f_i \le f_{avg}$ Update particles using equations (6), (7), (1) and (2) ;
11. Update $Pbest$;
12. Update $Pgbest$;
13. If $\|v\| \le \varepsilon$, terminate }
14. Select two other particles j and k ($i \ne j \ne k$) randomly;
15. Optimize new individuals using IKSCOC}
16. Output.

where t_{max} is the maximum number of iteration for PNPSO. STEP 16 is to improve the local constringency speed of PNPSO.

4 Results and Discussion

We have made experiments separately by K-Medoids, IKSCOC, GKSCOC, PKSCOC, and PNPKSCOC. $n = 50$, $c_1 = c_2 = 2$, $t_{max} = 100$. Fig.2 shows the results on real Dataset. Fig.2 (a) shows the original data with river obstacles. Fig.2 (b) shows the results of 4 clusters found by K-Medoids without considering obstacles constraints. Fig.2(c) shows 4 clusters found by IKSCOC. Fig.2(d) shows 4 clusters found by GKSCOC. Fig.2 (e) shows 4 clusters found by PNPKSCOC. Obviously, the results of the clustering illustrated in Fig.2(c), Fig.2 (d) , and Fig.2(e) have better practicalities than that in Fig.2 (b), and the ones in Fig.2 (e) and Fig.2 (d) are both superior to the one in Fig.2(c). So, it can be drawn that PNPKSCOC is effective and has better practicalities.

Fig.3 is the value of J showed in every experiment on Dataset1 by IKSCOC, PKSCOC, and PNPKSCOC respectively. It is showed that IKSCOC is sensitive to initial value and it constringes in different extremely local optimum points by starting at different initial value while PNPKSCOC constringes nearly in the same optimum points at each time, and PNPKSCOC is better than PKSCOC.

Fig.4 is the constringency speed in one experiment on Dataset1. It is showed that PNPKSCOC constringes in about 12 generations while GKSCOC constringes in nearly 25 generations. So, it can be drawn that PNPKSCOC is effective and has higher constringency speed than GKSCOC.

Therefore, we can draw the conclusion that PNPKSCOC has stronger global constringent ability than PKSCOC and has higher convergence speed than GKSCOC.

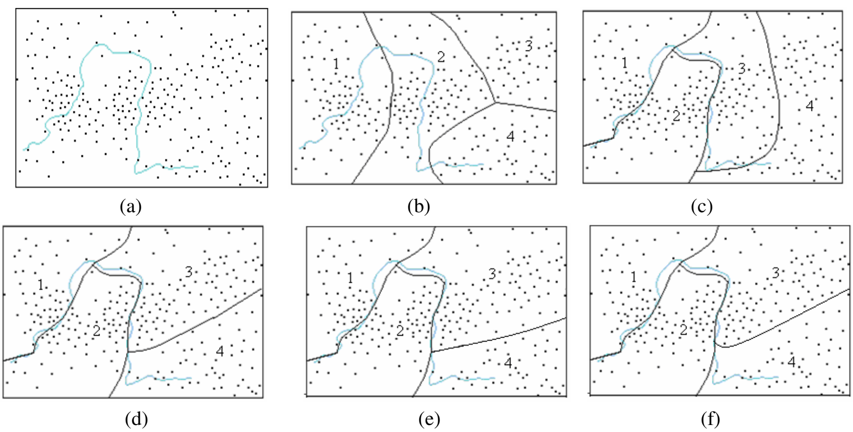
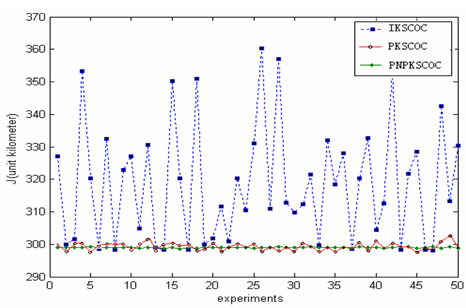


Fig. 2. Clustering Dataset



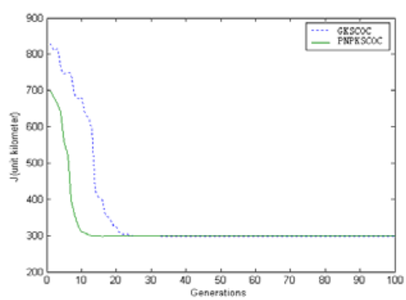


Fig. 3. PNPKSCOC vs. IKSCOC, PKSCOC

Fig. 4. PNPKSCOC vs. GKSCOC

5 Conclusions

In this paper, we developed a novel spatial clustering with obstacles constraints by dynamic piecewise-mapped and nonlinear inertia weights particle swarm optimization to cluster spatial data with obstacles constraints. The proposed method is also compared with some other algorithms to demonstrate its efficiency and the experimental results are satisfied.

Acknowledgments. This work is partially supported by Program for New Century Excellent Talents in University (NCET-08-0660),the Supporting Plan of Science and Technology Innovation Talent of Colleges in Henna Province (Number: 2008HASTIT012),and the Opening Research Fund of Key Laboratory of Spatial Data Mining & Information Sharing of Ministry of Education (Number:200807).

References

1. Tung, A.K.H., Han, J., Lakshmanan, L.V.S., Ng, R.T.: Constraint-Based Clustering in Large Databases. In: Van den Bussche, J., Vianu, V. (eds.) ICDT 2001. LNCS, vol. 1973, pp. 405–419. Springer, Heidelberg (2000)
2. Tung, A.K.H., Hou, J., Han, J.: Spatial Clustering in the Presence of Obstacles. In: Proceedings of International Conference on Data Engineering (ICDE 2001), Heidelberg, Germany, pp. 359–367 (2001)
3. Estivill-Castro, V., Lee, I.J.: AUTOCLUST+: Automatic Clustering of Point-Data Sets in the Presence of Obstacles. In: Proceedings of the International Workshop on Temporal, Spatial and Spatial-Temporal Data Mining, Lyon, France, pp. 133–146 (2000)
4. Zaïane, O.R., Lee, C.H.: Clustering Spatial Data When Facing Physical Constraints. In: Proceedings of the IEEE International Conference on Data Mining (ICDM 2002), Maebashi City, Japan, pp. 737–740 (2002)
5. Wang, X., Hamilton, H.J.: Gen and SynGeoDataGen Data Generators for Obstacle Facilitator Constrained Clustering (2004) http://Ftp.cs.uregina.ca/Research/Techreports/2004-08.pdf
6. Zhang, X., Wang, J., Wu, F., Fan, Z., Li, X.: A Novel Spatial Clustering with Obstacles Constraints Based on Genetic Algorithms and K-Medoids. In: Proceedings of the Sixth International Conference on Intelligent Systems Design and Applications (ISDA 2006), Jinan Shandong, China, pp. 605–610 (2006)
7. Zhang, X., Wang, J.: A Novel Spatial Clustering with Obstacles Constraints Based on Particle Swarm Optimization and K-Medoids. In: Zhou, Z.-H., Li, H., Yang, Q. (eds.) PAKDD 2007. LNCS (LNAI), vol. 4426, pp. 1105–1113. Springer, Heidelberg (2007)
8. Sun, J., Feng, B., Xu, W.: Particle Swarm Optimization with particles having Quantum Behavior. In: Proceedings of Congress on Evolutionary Computation, Portland (OR, USA), pp. 325–331 (2004)
9. Mikki, S.M., Kishk, A.A.: Quantum Particle Swarm Optimization for Electromagnetics. IEEE transactions on antennas and propagation 54(10), 2764–2775 (2006)
10. Baranovsky, A., Daems, D.: Design of one-dimensional chaotic maps with prescribed statistical properties. International Journal of Bifurcation and Chaos 5(6), 1585–1598 (1995)
11. Liu, B., Wang, L., Jin, Y.-H., Tang, F., Huang, D.-X.: Improved particle swarm optimization combined with chaos. Chaos, Solitons and Fractals 25, 1261–1271 (2005)
12. Xiang, T., Liao, X.F., Wong, K.W.: An improved particle swarm optimization algorithm combined with piecewise linear chaotic map. Appl. math. and Compute., 1637–1645 (2007)
13. Liu, H., Su, R., Gao, Y.: Coordinate Particle Swarm Optimization with Dynamic Piecewise-mapped and Nonlinear Inertia Weights. In: Proceedings of the 2009 International Conference on Artificial Intelligence and Computational Intelligence (AICI 2009), Shanghai, China, pp. 124–128 (2009)
14. Tokuta, A.: Extending the VGRAPH Algorithm for Robot Path Planning, http://wscg.zcu.cz/wscg98/papers98/Tokuta_98.pdf