

Self-adaptive Particle Swarm Optimization Algorithm with Mutation Operation based on K-means

Xue-mei Wang, Yi-zhuo Guo, Gui-jun Liu
 Department of Computer Science and Technology,
 Cheng-dong College of Northeast Agricultural University
 Harbin, 150025, China
 master2@163.com

Abstract—Adaptive Particle Swarm Optimization algorithm with mutation operation based on K-means is proposed in this paper, this algorithm Combined the local searching optimization ability of K-means with the global searching optimization ability of Particle Swarm Optimization, the algorithm self-adaptively adjusted inertia weight according to fitness variance of population. Mutation operation was peocessed for the poor performative particle in population. The results showed that the algorithm had solved the poblems of slow convergence speed of traditional Particle Swarm Optimization algorithm and easy falling into the local optimum of K-Means, and more effectively improved clustering quality.

Keyword: *k-means cluster algorithm ; Particle Swarm Optimization; mutation;*

I. INTRODUCTION

Particle Swarm Optimization originates from the investigation into behavior of bird swarm prey mentioned by Eberhart and Kennedy in 1995[1].The algorithm which is global optimization algorithm based on swarm intelligence has strong ability of global search optimum.It was widely used in the fields such as clustering problem,Genetic Algorithm,Neural networks,Image Processing because of the characteristics such as simple structure,needs fewer adjustment parameters and so on.But there are some defects in evolution process of PSO which exist the premature convergence,the poor local search ability,slower convergence rate in the later stage of evolution and higher iterations.In recent years,much work has been done to these disadvantages by different authors at home and abroad then puts forward all kinds of improved PSO algorithm by increasing the scale of population[2-3], but these improved PSO different degree to reduce convergence speed.

Mentioned by J.B.MacQueen, the K-means clustering algorithm was a typical partition algorithm[1].It was widely used in scientific research and industrial application because of its advantages such as simplification,rapid convergence and suitable for processing large data sets and so on.In initialization of clustering center, classical K-means clustering algorithm just randomly assigned initial point, and usually found a local optimum clustering results.So,the poor stability affected the accuracy of classification.In order to overcome the defluts of k-means, optimized cluster algorithm combined Simulated Annealing Algorithm with cluster algorithm in literature 4,but Simulated Annealing

Algorithm must adjust many parameters,human factors may cause the difference of the clustering results, better parameter were collected by a large number of numerical simulation,then prevent the algorithm promotion.

According to the defects,adaptive Particle Swarm Optimization algorithm with mutation operation based on K-means (KMPSO)is proposed in this paper. The randomness which produced the next generation of particle swarm has great in this algorithm. The algorithm come fastly local search that accelate declined speed of inertia weight. According to judgement about fitness variance of population whether fell into the local optimum, mutation operation was done to the particle part of population,then get rid of local searching trap and increase the ability of global search.The realization which get rid of constraint fell into the local optimum in later stage and retain the prosperity of fastly search speed in earlier stage.

II. RELEVANT WORK

A. The classical k-means clustering algorithm

1) Initialization

K objects which is randomly selected serve as center of initial k clustering collection ;

2) Distribution

Searching the nearest clustering center of each sample, and distribute the sample to the corresponding clustering ;

3) Modifying clustering center

Calculate the mean value of every new divided clustering center.

4) Calculation of deviation and judgment

Repeat (2)(3) until there is not changing.

B. Particle Swarm Optimization Algorithm

The bird in PSO based on the poputation and fitness is called particle which correspond solution of optimization problem in search space[5]. The speed of each particle decide the direction and distance of fly. A group of particles randomly initialized begin to move follow the current optimal particle until find the optimal solution in the whole solution space.

In every iteration, the particle update the speed and position through the tracking two extremums.One p^{best} is

the best station of particle itself passed, another g^{best} is the best station in global of particle swarm passed.

$$v_{t+1} = w \cdot v_t + c_1 \cdot r_1(p^{best} - x_t) + c_2 \cdot r_2(g^{best} - x_t) \quad (1)$$

$$x_{t+1} = x_t + v_{t+1} \quad (2)$$

Among them, w is inertia weight factor; v_t is current speed of particle; v_{t+1} is new speed of particle; x_t is current station; x_{t+1} is new station which particle generated; c_1, c_2 (nonnegative number), which are the acceleration coefficient respectively, adjust the largest stride length of flying direction to the best individual particle and the best global particle. r_1, r_2 are the random number between [0, 1].

The influence previous speed to the latter speed was controlled by inertia weight factor w [6]. Larger w can enhance the global search ability of PSO, and smaller w can strengthen the local search ability. At present, w was setted which linear decreasing strategy was adopted by most of the algorithm, then adjust the search ability limited, The search process can not adapt to the practical problems of complicated situation. It can better meet the complicated practical environment according to the iterations to adjust inertia weight value.

$$w = w_{min} + (w_{max} - w_{min}) \exp\left(-\frac{1}{1 + \frac{t}{t_{max}}}\right) \quad (3)$$

The algorithm quickly enter the local search through the acceleration the declined speed of inertia weights of in early iterative algorithm. The algorithm have rapid convergence compared with PSO, algorithm make into local search as soon as possible and gain a better solution efficiency.

C. The decision of premature phenomenon of and mutation operation

There is a smaller difference between the individuals with the population continuously evolved, particle appear "gather" phenomenon. The size of the fitness is determined by Individual position, how to judge whether the particle premature gathered. Initially, the fitness of each particle of the particle group ascending sort, then take out m particle. Accord formula (4) calculate fitness variance σ^2 which was evidence as judgment of convergence.

$$\sigma^2 = \frac{\sum_{i=1}^m (f_i - f_{avg})^2}{m} \quad (4)$$

Among them, f_i is fitness of particle i , f_{avg} is average fitness of m particles.

$$f_{avg} = \frac{\sum_{i=1}^m f_i}{m} \quad (5)$$

The algorithm entered the later stage was thought that Particle swarm tends to convergence when fitness variance

σ^2 less than a given threshold θ (tending to zero, general), fell into easily the local optimum, appear the phenomenon of premature convergence. K-Means operation was happen between m particles. The fitness of particles which was setted individual optimal was calculated after mutation operation. Particle swarm jump out of local optimal solution and flight to direction of the global optimal solution after global optimum and location was updated.

D. Description of algorithm based on KMPSO Improving

Step1 Initiation of population swarm. The total number of particles whose position and speed were initialized randomly set $n.k$ ($k < n$) particles were choosed randomly and regarded as center of clustering. The initial speed of center was 0.

Step2 The fitness of particle was calculated, initiation station which are the best station of particle itself passed $pbest$ and the best station $gbest$ in global of particle swarm passed.

Step3. According to formula (3) adjust inertia weight, according to formula (1),(2) adjust the speed and location of particle.

Step4 A new clustering center was calculated for each particle in accordance with the corresponding clustering division, and then updated the value of fitness of particle and individual optimization and global optimization and location.

Step5 σ^2 of former m particles which were selected that sort from small to large in accordance with the fitness of particle calculated according to formula (4), (5). If σ^2 was less than the given threshold, take the next step, otherwise, turn to step8, and then these speed of particle were not limited in the next iteration.

Step6 K-Means operation was happen to m particles which participated in calculation of σ^2 , updated the speed of particle and individual optimization and global optimization and location.

Step7 The mutation of particle was happen after optimization of K-means. The distance to centers of k clustering were calculated from each samples.

The samples which had the nearest distance to the clustering center were divided into the same cluster again. The fitness of particle after mutation were setted individual optimization and updated the global optimization and location.

Step8 The algorithm terminate if fitness of population met the maximum evolution algebra of convergence conditions. Otherwise, algorithm returned to step3.

III. EXPERIMENTAL ANALYSIS AND RESULTS

In order to test the effectiveness of clustering algorithm based on mutation of particle which was been proposed in this paper. Three data sets were used to respectively test the k-means clustering algorithm, the PSO

algorithm, the improved algorithm which was been proposed in this article.

Table I showed number of data sample, attribute number of data sample and number of category in each data set.

TABLE I. THE PERFORMANCE CLASSIFICATION COMPARISON OF THE EVERY ALGORITHM IN IRIS DATA SET

	number of data samples	number of data attribute	number of category
data setI	12	4	2
data setII IRIS	150	3	4
data setIII	2400	7	2

The parameters was setted in process of particle swarm optimization as follows, Numer of selected particles are 40, Inertia weight $w_{min}=0.4$, $w_{max}=0.9$, the maximum of iterations is 1000, Number of population of particle is 5, Acceleration coefficient $c_1 = c_2 = 1$, threshold of fitness variance $\theta=0.1$, Numbers of particles m in calculation of fitness variance was repectively 4,7,10.

TABLE II. THE RESULTS OF K-MEANS ALGORITHM IN THREE DATA SETS

k-mesns	data set1	data set2 (IRIS)	data set 3
Iterations achieved optimal solution	2.6	7.1	10.8
Average time achieved optimal solution	0.000693	0.018835	0.511108
Maximum of optimal solution	12.833	145.279	20114.1
Average value of optimal solution	9.700	98.478	15698.2
Minimum of optimal solution	5.333	78.94	11594.1
Value of optimal solution	5.333	78.94	11594.1
Times achieved optimal solution	4	7	5

TABLE III. THE RESULTS OF TRADITIONAL PSO ALGORITHM IN THREE DATA SETS

k-mesns	data set1	data set2 (IRIS)	data set 3
Iterations achieved optimal solution	24.1	161.3	455.1
Average time achieved optimal solution	0.013357	0.978232	23.23743
Maximum of optimal solution	12.833	78.94	206181
Average value of optimal solution	6.317	78.94	205238.6
Minimum of optimal solution	5.333	78.94	204038
Value of optimal solution	5.333	78.94	204038
Times achieved optimal solution	8	10	0

The results of table II, table III ,table IV showed that K-mean algorithm had the fast clustering operation speed, but not get the optimal solution every time due to the sensitive selection of initial clustering center. Through

traditional PSO algorithm, small scale data could reach the optimal solution, but had the slower convergence rate. For the bigger scale data, the optimal solution was difficult to be obtained through traditional PSO..

TABLE IV. THE RESULTS OF IMPROVED ALGORITHM IN THREE DATA SETS

k-mesns	data set1	data set2 (IRIS)	data set 3
Iterations achieved optimal solution	3	8.3	7.5
Average time achieved optimal solution	0.004770	0.239138	1.711048
Maximum of optimal solution	5.333	78.94	11594.1
Average value of optimal solution	5.333	78.94	11594.1
Minimum of optimal solution	5.333	78.94	11594.1
Value of optimal solution	10	10	10
Times achieved optimal solution			

Especially, the results which showed in data 3 no optimal solution was obtained through traditional PSO. The improved algorithm in this paper had not only reached the optimal solution for three groups of data, but also had the faster convergence speed, and effectively overcomed the disadvantage of slow convergence speed of traditional Particle Swarm Optimization algorithm and easily falling into the local optimum of K-Means.

IV. CONCLUSION

According to respective advantages of K-means and PSO, self-adaptive Particle Swarm Optimization algorithm with mutation operation, which was based on K-means, was proposed. The algorithm organically combined two algorithms, gave full play to the performances of respective algorithm.

The mutation operation of algorithm was introduced in right time to strengthen ability of global search optimum. From the results of experiment, we can gain conclusions as follows: Introduce of mutation operation made algorithm jump out of local extremum. The combination of K-means with PSO algorithm enhanced efficiency of global searching optimization, saved computation-consuming time, effectively improved the stability of algorithm.

V. REFERENCES

- [1] Qing- Kennedy J, Eberhart R. "Particle swarm optimization"[C]Proc of IEEE International Conference on Neural Networks (ICNN) ,Australia,1995:1942-1948.
- [2] Hu Yan-wei, Qin Zheng, Zhang Zhong-zhi. "Intrusion detection algorithm based on simulated annealing and K-mean clustering"[J].Computer Science,2010,6(6) :122-124.
- [3] LIU Junfang, GAO Yuelin. "Quantum particle swarm optimization algorithm with adaptive mutation".Computer Engineering and Applications, 2011, 47 (3) : 41-43.
- [4] TAO Xinmin, XU Peng, ZHANG Dongxue, HAO Siyuan

- “Particle swarm optimization clustering algorithm with mutation based on K-means” Applied Science and Technology,2011,12:25-28
- [5] PAN Junliang, SHI Yuexiang, LI Pingting. New particle swarm optimization clustering method. Computer Engineering and Applications,2012, 48 (8) :179-181.
- [6] YANG Chun-hua, GU Li-shan, GUI Wei-hua“ Particle Swarm Optimization Algorithm with Adaptive Mutation” Computer Engineering. 2008, Aug:188-190