BULGARIAN ACADEMY OF SCIENCES

CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **14**, Special Issue

Sofia • 2014 Print ISSN: 1311-9702; Online ISSN: 1314-4081
DOI: 10.2478/cait-2014-0048

A Novel Fuzzy Clustering Recommendation Algorithm Based on PSO

Zhang Hui, You Fei

School of Information Engineering, Yulin University, Yulin 719000, China
Emails: 86361566@qq.com 348553506@qq.com

Abstract: *Aiming at the problem of recommendation systems, this paper proposes a fuzzy clustering algorithm based on particle swarm optimization. This algorithm can find the best solution, using the capacity of global search in PSO algorithm with a powerful global and defining a proportion factor, which can adjust the position and reduce the search space automatically. Then using mutation particles it replaces the particles flying out the solution space by new particles during the searching process. In order to check the performance of the proposed algorithm, by testing with typical ZDT1, ZDT2, ZDT3 functions, the experimental results show that the improved method not only has a better ability to converge to the global point, but can also efficiently avoid premature convergence.*

Keywords: *Multi-objective optimization, recommender systems, cluster analysis mutation particle, Particle Swarm Algorithm, premature convergence.*

1. Introduction

In the time of information growth, with the development of the Internet of Things, the “information overload” and “Information Track” are urgent problems needed to be solved in Internet. In order to solve these difficult problems the recommender system comes into being – the recommendation system can be considered as a class of expert systems, autonomously leaning through a network. By learning the knowledge, the recommender system can predict users’ preferences, and then recommend which users may like a product, services, news, etc.

At present, the mainstream recommendation algorithm is a Collaborative Filtering (CF) [1-3] and the core idea of CF algorithm are target users with similar interests of favourite items accepted as the target users' favourite. Besides, the target users like similar items, as well as a similar user favourite. The designer of a recommendation system has to recommend to the user, the users' likes or interests. Meanwhile, it is better when the user really likes it. However, the goods manufacturers not only hope to recommended users' popular items, but also non-popular items. Based on the above considerations, this paper presents a novel PSO-based fuzzy clustering recommendation algorithm, which can efficiently improve the recommender precision. However, with the further development of Internet applications, the traditional recommendation system and its algorithm can hardly adapt to the user's scale. The concept of the rapid growth of the number of projects, recommended data and user history score, score data sparsity and user interest drift problems cause decreased recommendation quality, user satisfaction, reduction or even loss of a large number of users, which has seriously hampered the further promotion and application of the recommendation systems. Therefore, the current recommendation algorithm to solve the problems and improve the recommendation accuracy of the recommendation system theory and practice is of great significance [4]. On this background this paper will discuss the target accuracy of the recommendation system, drift issues and concepts for the sparsity recommendation system in the research work, and propose an improved recommendation algorithm with some innovations. In order to improve the accuracy of prediction, we propose a new fuzzy clustering recommendation that can generate a set of schemes that provides the decision maker with more choices. The decision maker can make use of the users' registration information to choose a personalized scheme to recommend to the user. The experiments show that this method can efficiently improve the accuracy of predictions. A general recommendation system has three important modules: user recommender module, recommended object model, recommendation algorithm. Fig. 1 shows a general model of the recommendation systems.

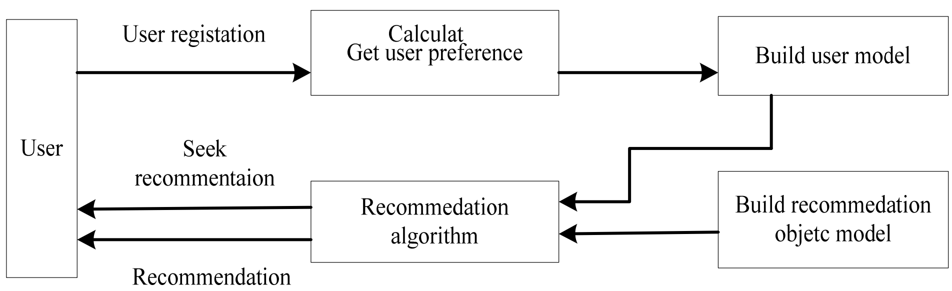

Fig. 1. Recommendation system model

2. Recommendation algorithm

2.1. Particle swarm optimization algorithm

Particle Swarm Optimization (PSO) algorithm [5] was first introduced in 1995 by Eberhart and Kennedy. It is a population-based stochastic optimization technique originally designed for continuous optimization problems. The inspiration of PSO has originated from the social behaviour, such as fish schooling and bird flocking. PSO has now become one of the most popular optimization techniques for solving continuous optimization problems.

Each individual, which is typically called a "particle" in PSO simulates a candidate solution and in order to search for the optimal solution, it updates its flying velocity and current position iteratively according to its own flying experience and the other particles' flying experience.

Let us assume that the particle swarm size is m and the particle dimension is n. Let $V_i = \{v_1, v_2, \cdots, v_n\}$ and $X_i = \{x_1, x_2, \cdots, x_n\}$ be i-th, i = 1, 2, ..., n, particle velocity vector and position vector, respectively. The updating process for each particle in the basic form can be formulated by the following equations:

(1) $$V_i \leftarrow \omega V_i + c_1 r_1 (P_i - X_i) + c_2 r_2 (G - X_i)\,,$$

(2) $$X_i \leftarrow X_i + V_i\,,$$

where $P_i = \{p_1, p_2, \cdots, p_n\}$ and $G = \{g_1, g_2, \cdots, g_n\}$ are i-th particle personal best position and the global best position of the swarm, respectively; the parameters r_1, r_2 are random numbers between 0 and 1, c_1 and c_2 are acceleration coefficients termed as cognitive and social components.

2.2. Multi-objective optimization

Multi-objective optimization problem has originated in the design of many complex systems, in modelling, planning issues. Since the 1960-ies, the multi-objective optimization problem attracted the attention of a growing number of researchers from different backgrounds [6]. Especially in recent years, the multi-objective evolutionary algorithm must do optimization of the more widely used and studied ones, resulting in a series of novel algorithms and get good application. The multi-objective optimization proposition is generally no unique global optimal solution, so this is actually a multi-objective optimization proposition of the process of seeking a Pareto set [7]. The traditional multi-objective algorithm is often converted into a single objective proposition after the use of a sophisticated single-objective optimization algorithm. The drawback is that the optimal solution can be determined only once. Now the multi-objective evolutionary strategy tends to do parallel computing that can solve a sufficient number of solutions distributed on the Pareto Front (PF) and provided to the decision-makers for the next decision. PSO as a novel evolutionary computing strategy has been more and more widely used in multi-objective optimization problems. Multi-objective optimization is described as follows:

$$\min\ y = F(X) = (f_1(x), f_2(x), ..., f_m(x)),$$
$$g_i(x) \le 0, \quad i = 1, 2, \cdots, q,$$
$$h_j(x) = 0, \quad j = 1, 2, \cdots, p,$$
$$x = (x_1, x_2, \cdots, x_n) \in X \subset R^n,$$
$$y = (y_1, y_2, \cdots, y_m) \in Y \subset R^m.$$

In the formula: $x = (x_1, x_2, \cdots, x_n) \in X \subset R^n$, denotes the decision variables; X is n-dimensional space; $y = (y_1, y_2, \cdots, y_m) \in Y \subset R^m$ is the objective function; the objective function F defines a mapping function [8] and m targets which need to be optimized; $g_i(x) \le 0$, $i = 1, 2, \cdots, q$, define q inequality constraints; $h_j(x) = 0$, $j = 1, 2, \cdots, p$, define p equality constraints.

Compared with single objective optimization [8], the complexity of multi-objective optimization has greatly increased. It needs to optimize multiple objectives, which are not comparable, and even conflicting. Improving an object may lead to reducing of another object performance. Compared to single objective, the essential difference is that the solutions are not unique, but a solution set. Pareto optimal solution set [9] in the solution space tends to form a boundary line (plane) as shown in Fig. 2.

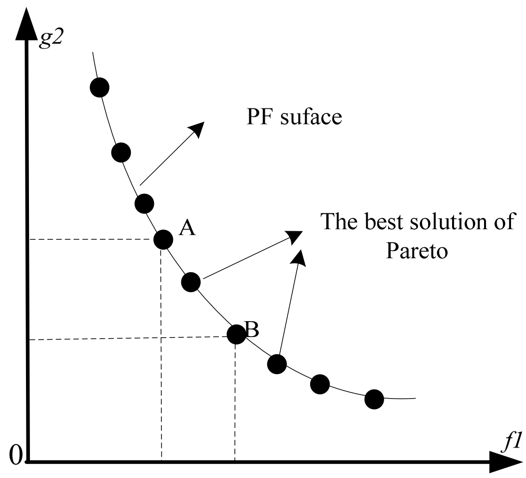

Fig. 2. Multi-objective optimization Pareto

The coordinate point of Fig. 2 demonstrates a solution, this constituent solution is called the best Pareto set. All Pareto optimal solution sets of optimal solutions, corresponding to the target vector consist of curved Pareto frontiers for PF.

3. Multi objective optimization algorithm based on PSO

3.1. Multi objective particle swarm optimization clustering algorithm

In order to improve the quality of the recommendation system, the paper proposes a multi-objective optimization PSO algorithm recommended by the relevant principles. We propose tools for optimization of the recommendation system. The fuzzy clustering algorithm [10-12] is as follows.

For evaluation of each particle in the particle swarm, define an individual fitness function

$$g(x_i) = \frac{k}{P_m(R, Z)}, \tag{3}$$

where k is constant. Define $P_m(R, Z)$ as an objective function; when P_m is smaller, $g(x_i)$ is higher. The initial cluster centers are $Z = \{z_1, z_2, \cdots, z_c\}$, z_i denotes the particles and the cluster centre coding. Each class of the n-dimensional cluster centers use real numbers, so that z is one-dimensional vector, which becomes a $c \times n$ column. Finally, the best optimal solution z can be obtained. The realization of the program algorithm is as follows:

Step 1. Define c as the number of classes, m is a fuzzy index, c_1 and c_2 are leaning factors, ω is the inertia weight and a are the iterations.

Step 2. Initialize n clustering centers and coding, forming the first generation of particles. The individual optimal solution $P_{i\,\text{best}}$ of i-th particle is initialized with its own position and fitness. The global optimal solution G_{best} of the entire population can be initialized with the first individual optimal solution.

Step 3. Calculate the fitness w_{ij} of every clustering centre, where $i = 1, 2, \cdots, n$, $j = 1, 2, \cdots, c$. Update w_{bj} to w_{b+1}:

$$w_{ij} = \begin{cases} \dfrac{1}{\sum_{k=1}^{c} \left(\dfrac{d_{ij}}{d_{ik}}\right)^{\frac{2}{m-1}}} & \text{if} \quad d_{ij} \neq 0, \\ 0 & \text{if} \quad d_{ij} = 0, \quad k = j, \\ 1 & \text{if} \quad d_{ij} = 0, \quad k \neq j. \end{cases} \tag{4}$$

Step 4. Calculate the c mean vector $\mathrm{Z}(b+1)$ according to the formula $w_{(b+1)}$,

$$\mathrm{Z}_j(b+1) = \frac{\sum_{i=1}^{n} w_{ij}^m x_i}{\sum_{i=1}^{n} w_{ij}^m} \quad \forall\ j. \tag{5}$$

Step 5. Calculate every particle fitness according to (4). If this particle is superior to the fitness of $p_{\text{best}\,i}$ [13], update the position. If this particle is superior to the fitness of G_{best}, the position and fitness of G_{best} are updated.

Step 6. Update every particle velocity and position according to (4) and (5).

When the clustering algorithm is under the condition of low constraints, the particle is likely to prematurely fall into an endless loop, which results in converging to a local optimum. Focusing on the above problem in the evolutionary process, whether PSO is in premature or global convergence, the particles of the particle swarm cause particles aggregation. In the late evolution of the population in order to enhance the local search algorithm, this section segmented evolutionary strategies. The evolutionary process is divided into two stages: the first one, using SPSO algorithm [14] updates the particle. The second phase realizes a fuzzy search to find the better populations optimal position.

3.2. Multi objective mutation particle swarm algorithm

During the process of evolution, sometimes the particles lose their abilities of exploration and will be stagnated. When some particles velocity is close to zero, the other particles will quickly go into the region near to the inactive particles position guided by p_i and p_g. Because these particles are random during the initialization and evolution process, the updating sometimes looks aimless. As a result, when p_g is trapped in a local optimum, the whole swarm does premature convergence, and the exploration performance will not be improved. Focusing on the above problem, a new improved fuzzy clustering algorithm is proposed, Fuzzy Interfering particles algorithm, abbreviated as FPSOA. Since FPSOA is a global optimization algorithm and has the chance or ability to jump out the suboptimal or local optimal solutions, it is advisable to introduce FPSOA. This strategy can readily solve the above mentioned drawbacks.

The initial $x = (x_1, x_2, ..., x_d, ..., x_D)$ is mutation probability, p_m is a mutation particle. Usually, p_m choose the small value, generally taking (0.001, 0.1). After the result of variation:

(6) $$x_d = g_{\text{best}\ d} + 0.4 \times \text{rand}\,n() \times g_{\text{best}\ d}, \quad d = 1, 2, \cdots, D,$$

where $g_{\text{best}d}$ is the first *d*-dimensional position coordinate, when the population reaches the best position, rand$n$ () obeys the distribution of random variables, where mean is 0, variance is 1.

In addition, restricting the maximum speed and the Linear Decreasing Inertia Weight does not restrict all particles searching space among PSO. In the absence of constraints, a small portion of the particles fly to the outer solution, and the solution may be illegal. So, to prevent the swarm from explosion and restrict these particles boundary is very important. In FPSOA algorithm, the position of the particle boundaries is not restricted. Moreover, the new variation particles replace the particles which are flying out of the search area. The pseudo-code is shown below:

(7) $$\text{if } x_{id} > x_d^{\max} \text{ or } x_{id} < x_d^{\min},$$

(8) $$x_{id} = x_d^{\min} + \text{rand}() \times (x_d^{\max} - x_d^{\min}),$$

end,

where $x_d^{\min}$, $\mathrm{x}_c^{\max}$ are the upper and lower bounds of the d-dimensional position, x_{id} is the i particle position in d-dimensional space. In fact, in the predation process of a bird in natural environment, there is also some minority of birds flying out of the constricted boundary and the birds outside the constricted area flow into the area to feed. To a certain extent, this phenomenon is according to the idea of a "variant" of particles. When PSO algorithm premature converges or if there is evolutionary stagnation, the variation of the particle to jump out of the current location into other areas of the solution space to search, in a subsequent search, the algorithm may produce a new individual pole value p_{best} and a global optimal g_{best}. After several iterations, the algorithm can find a better global optimal solution to avoid the standard PSO evolutionary algorithm that may appear stagnant or falling into a local optimum phenomenon. In conclusion, the algorithm steps are as follows:

Step 1. Initialize the particle position and velocity.

Step 2. Compute the fitness.

Step 3. Update every particle p_{best} and g_{best}.

Step 4. If $4\omega < Z$ the particle falls into an endless loop. Use mutation particles to substitute the accused particles, then update the position and velocity.

Step 5. Estimate the algorithm, while satisfying the convergence conditions. If met, end of the operation, otherwise go to Step 2.

4. Experiment and analysis

To evaluate the performance of MPSOA, it has been applied to classical functions and compared to SPSO, [15] MOPSO algorithm. All the experiments have been performed on Inter(R) Celeron(R)M CPU 550 machine, 3.2 GHz, 4 GB memory. The operating system is MS Windows 7 and the program compiler is Matlab 2012. The experiments were carried out with 3000 iterations for a population size of 50. All experiments were run 3000 times. $c_1 = c_2 = 2$, m = 30, $v_{\max} = 0.8$, $v_{\min} = 0.1$, ω is linearly decreased from 0.9 to 0.4 during the iterations. The following three algorithms were used to benchmark the functions ZDT1, ZDT2, ZDT6, tested separately.

As it can be seen from Tables 1, 2 and 3, the diversity indices of MPSOA are significantly superior to SPSO and show that MOPSO algorithm for [16] noninferior solution [17], is evenly distributed in the target space of Pareto optimal front.

Through the three typical function ZDT1, ZDT6, ZDT2 [18-20] from the optimization experiments it can be seen that the proposed multi-objective optimization method, based on particle swarm can approach degree , uniformity and diversity that are better than in the other two methods, which is a potential multi-objective optimization.

The comprehensive experimental results show that the clustering algorithm and FPSOA algorithm have improved the accuracy of prediction, which can improve the quality of the recommendation system.

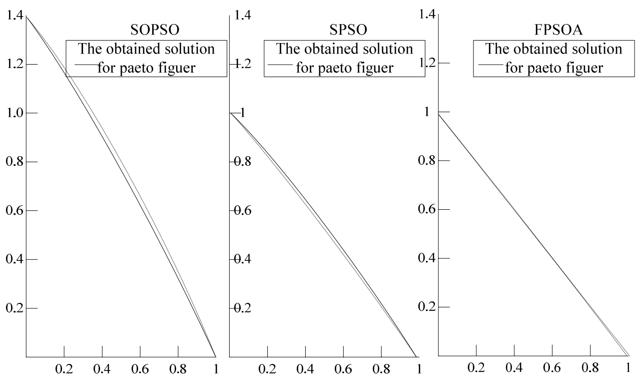

Fig. 3. Three algorithms test for ZDT1

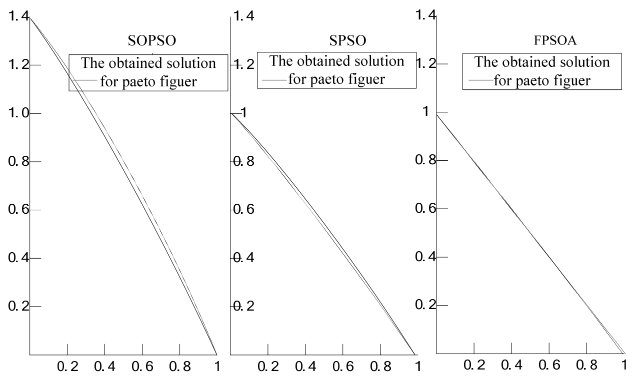

Fig. 4. Three algorithms test for ZDT6

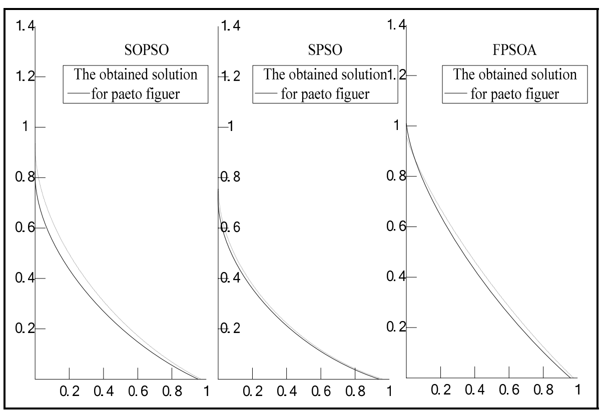

Fig. 5. Three algorithms test for ZDT2

Table1. Diversity index for ZDT1

Algorithm	Minimum	Maximum	Average	Variance
SOPSO	0.10171	0.0212	0.212	0.0012
SPSO	0.0140	0.0180	0.0160	0.0013
MPSOA	0.1002	0.0181	0.0150	0.0022

Table 2. Diversity index for ZDT6

Algorithm	Minimum	Maximum	Average	Variance
SOPSO	0.0102	0.2276	0.0911	0.1120
SPSO	0.0108	0.0827	0.0432	0.0039
MPSOA	0.1003	0.0142	0.0115	0.0015

Table 3. Diversity index for ZDT2

Algorithm	Minimum	Maximum	Average	Variance
SOPSO	0.0200	0.0320	0.0254	0.0036
SPSO	0.0140	0.02276	0.0187	0.0034
MPSOA	0.1001	0.0152	0.0108	0.0019

5. Conclusion

The recommended system for information overload provides a good solution idea. The recommendation system related research has received attention in all areas of life. This paper mainly studied Fuzzy Clustering recommendation Algorithm Based on PSO. This algorithm uses the capacity of global search in PSO algorithm. In order to overcome the particles premature falling into local populations of the extreme, mutation particles are used into an endless loop and replacing of the particles, flying out the space by new particles, in a three benchmark functions test. The experimental results show that the algorithm has a better effect and can efficiently improve the convergence accuracy and the quality of the recommendation system.

Acknowledgments: This work is partially supported by the Research and Cooperation Project for Department of Yulin city (No Gy12-12, 2012GXY3-34), Yulin University research projects in 2013 (No 13YK49). The authors would like to thank for the help.

References

1. Newman, M. E. J. The Structure of Scientific Collaboration Networks. – Proceedings of the National Academy of Sciences of the United States of America, Vol. **98**, 2001, No 2, 404-409.

2. A l b e r t, R., H. J e o n g, A. L. B a r a b a s i. Internet-Diameter of the World-Wide Web. – Nature, Vol. **401**, 1999, 6749, 130-131.
3. W a t t s, D. J., S. H. S t r o g a t z. Collective Dynamics of “Small-World” Networks. – Nature, Vol. **393**, 1998, 6684, 440-442.
4. B a r a b á s i, A. -L., R. A l b e r t. Emergence of Scaling in Random Networks. – Science, Vol. **286**, 1999, 5439, 509-512.
5. G i r v a n, M., M. E. J. N e w m a n. Community Structure in Social and Biological Networks. – Proceedings of the National Academy of Sciences of the United States of America, Vol. **99**, 2002, No 12, 7821-7826.
6. F o r t u n a t o, S. Community Detection in Graphs. – Physics Reports, Vol. **486**, 2010, No 3, 75-174.
7. P i z z u t i, C. GA-Net: A Genetic Algorithm for Community Detection in Social Networks. – Parallel Problem Solving from Nature (PPSN), Vol. **5199**, 2008, 1081-1090.
8. G o n g, M., B. F u, L. J i a o, H. D u. Memetic Algorithm for Community Detection in Networks. – Physical Review E, Vol. **84**, 2011, 56-101.
9. C a i, Q., M. G o n g, L. M a, L. J i a o. A Novel Clonal Selection Algorithm for Community Detection in Complex Networks. – Computational Intelligence, Wiley Periodical, Inc., 2014, 1-24.
10. X u, R., G. K. V e n a y a g a m o o r t h y, D. C. W u n s c h II. Modeling of Gene Regulatory Networks with Hybrid Differential Evolution and Particle Swarm Optimization. – Neural Networks, Vol. **20**, 2007, No 8, 917-927.
11. G i r v a n, M., M. E. J. N e w m a n. Community Structure in Social and Biological Networks. – Proceedings of the National Academy of Sciences of the United States of America, Vol. **99**, 2002, No 12, 7821-7826.
12. G u i m e r ò, R., M. S a l e s-Pa r d o, L. A. N. A n a r a l. Modularity from Fluctuations in Random Graphs and Complex Networks. – Physical Review E, Vol. **70**, 2004, 25-101.
13. D u c h, J., A. A r e n a s, Community Detection in Complex Networks Using Extremal Optimization. – Physical Review E, Vol. **72**, 2005, 27-104.
14. N e w m a n, M. E. J. Modularity and Community Structure in Networks. – Proceedings of the National Academy of Sciences of the United States of America, Vol. **103**, 2006, No 23, 8577-8582.
15. S h i, J., J. M a l i k. Normalized Cuts and Image Segmentation. – IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. **22**, 2000, No 8, 888-905.
16. W u, F., B. A. H u b e r m a n. Finding Communities in Linear Time: A Physics Approach. – European Physical Journal B, Vol. **38**, 2004, No 2, 331-338.
17. L a n c i c h i n e t t i, A., S. F o r t u n a t o, F. R a d i c c h i. Benchmark Graphs for Testing Community Detection Algorithms. – Physical Review E, Vol. **78**, 2008, No 4.
18. Z a c h a r y, W. W. An Information Flow Model for Confict and Fission in Small Groups. – J. Anthropol. Res., 1977, 452-473.
19. L u s s e a u, D., K. S c h n e i d e r, O. J. B o i s s e a u, P. H a a s e, E. S l o o t e n, S. M. D a w s o n. The Bottlenose Dolphin Community of Doubtful Sound Features a Large Proportion of Long-Lasting Associations. – Behavioral Ecology and Sociobiology, Vol. **54**, 2003, No 4, 396-405.
20. N e w m a n, M. E. J. Finding Community Structure in Networks Using the Eigenvectors of Matrices. – Physical Review E, Vol. **74**, 2006, 36-104.