http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (14): 3030-3033, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.3030.3033

Application Research of Improved Particle Swarm Algorithm in Online Trading Customer Classification

Li Xinwu
School of International Trade and Economics, Jiangxi University of Finance and Economics, Nanchang, Jiangxi, China,330013

Abstract: Classifying customer correctly and effectively according to customers' characteristics and behaviors plays a key role for network enterprises to realize the full values of modern customer relationship management. Aiming at the shortages of the existing particle swarm and K-means algorithm, this paper improves two algorithms through integrating them together and presents a new customer classification algorithm. First 21 customer classification indicators is designed based on consumer characteristics and behaviors analysis; Second, the study analyzes the shortages of particle swarm optimization algorithm, improves its optimal performance, redesigns a new particle swarm optimization algorithm for classifying online trade customer; Finally the experimental results verify that the new algorithm can improve effectiveness and validity of customer classification when used for classifying network trading customers practically.

Key words: Particle swarm optimization algorithm, K-means algorithm, online customer classification, customer relationship management

INTRODUCTION

Customer Relationship Management (CRM) is one of the core problems of modern enterprises in which three basic problems needs to solve are how to get customers, how to keep customers and how to maximize customer value and the core of analyzing the three problems CRM needs to solve is to classify customers. "Getting customers" and "Retaining customers" need to ascertain which customers are attainable which customers need to be kept which customers are kept for a long term and which customers are kept for a short term, therefore, customer classification is needed. It is the same case with "Maximizing Customer Value". Due to different values of different customers, "Maximum Customer Value" of different customers should be distinguished. Thus, customer classification is becoming a more and more popular research hotspot, also a research difficulty, becoming one of the urgent problems of CRM (Zhou, 2012).

The widely-used methods of enterprises for customer classification at present are mainly qualitative method and quantitative method. Quantitative classification method is to apply quantitative analysis technology to conduct customer classification on the basis of some specific customer variables (credit level of customers, purchasing power of customers, characteristics of demand of customers, etc.). K-means clustering method is wildly used for its powerful data process ability but the shortages such as insufficient classification accuracy or poor "robustness", limits the application of K-means in customer classification (Jiang and Zhang, 2011).

Particle swarm optimization, PSO, is a new method used in customer classification because it can obtain global optimal solution and get high classification accuracy when it combined with K-means algorithm. But PSO algorithm has slow convergence rate and is easy to fall into local extreme point, so the paper tries to present some improvements to overcome the limitations of the algorithm and takes advantage of powerful classification ability of the algorithm to classify online trading customer.

CUSTOMER CLASSIFICATION INDICATORS SELECTION

The selection of reasonable classification variables is the basis of correct and effective customer classification. In view of the nature of trading and own characteristics of online trading, this study adopts customer characteristics type variable and customer behaviors type variable in the specific selection of customer classification variables. ?Based on analyzing and summarizing existing literatures, the customer characteristics type variables designed in this paper include: Customer No., Post Code, Date of Birth, Sex, Educational Background, Occupation, Monthly Income, Time of First Website Browsing and Marital Status □ Based on analyzing and summarizing existing literatures, the customer behaviors type variables designed in this study include Monthly Frequency of Website Login, Monthly Website Staying Time, Monthly Times of Purchasing, Monthly Amount of Purchasing, Type of Consumer Products Purchased, Times of Service Feedback, Service Satisfaction, Customer Profitability, Customer Profit, Repeat Purchases, Recommended Number of Customers, Purchasing Growth Rate (Jiang and Zhang, 2011).

CUSTOMER CLASSIFICATION ALGORITHM DESIGN

Particle swarm optimization algorithm: PSO is an evolutionary computation technology based on swarm intelligence, mainly used for obtaining global optimal solution, the basic concept of which is that the potential solution of each optimization problem is a bird in search space, called "particle"; all the particles have an adaptive value determined by optimization function and each particle has a speed vector determining their flying direction and distance; then particles follow the search of current optimal particle in solution space. PSO is initialized as a group of random particles and then finds optimal solution through iteration. In every iteration, particle updates itself through tracking two extreme values, among which one is the best solution particle itself finds till the current moment, called individual best value; the other is the best solution the entire group find till the current moment which is global best value (Ding, 2013).

Suppose that there are m particles forming a group in D-dimensional target search space, among which the ith particle is expressed as a d-dimensional vector, $x_i = (x_{i1}, x_{i2},.....x_{id})$ i.e. the location of the ith particle in D-dimensional search space being the location of each particle is a potential solution. To substitute x_i into a target function can get its adaptive value, measuring the strength and weakness of $\vec{x}_i$ according to the size of adaptive value. The flying speed of the ith particle is also a d-dimensional vector, recorded as:

$$\vec{v} = (v_{i1}, x_{i2}, ...v_{id})$$

recording the optimal location searched till now of the ith particle as:

$$\vec{p}_{id} = (p_{i1}, p_{i2}, ...p_{id})$$

and the optimal location searched till now of the entire particle swarm as:

$$\vec{p}_{gd} = (p_{g1}, p_{g2}, ...p_{gd})$$

Traditional PSO carries out the operation on particles according to equation 1 and 2 (Bradley and Managasarian, 2000):

$$v_{id}^{t+1} = \varpi v_{id}^{t} + c_1 r_1 (p_{id} - x_{id}^{t}) + c_2 r_2 (p_{gd} - x_{id}^{t}) \quad (1)$$

$$x_{id}^{t+1} = x_{id}^{t} + \alpha v_{id} \quad (2)$$

Among which i = 1, 2, 3...m, d = 1, 2, 3... D, learning factors c_1 and c_2 are nonnegative constants r_1 and r_2 are random numbers, obeying to the uniform distribution of [0, 1]. In the formula, ω is nonnegative number, called momentum coefficient, controlling the influence of previous speed on current speed; the greater ω is, the bigger the influence of previous speed is, the stronger the global searching ability is; the smaller ω is, the less the influence of previous speed is, the stronger the global searching ability is, the local minimum able to be jumped out through adjusting the size of ω. α is called constraints factor, aiming to control the weight of speed. Condition for terminating iteration is, according to specific problems, generally selected as maximum iterations or optimal location particle swarm searched up till now meeting desired minimum adaptive threshold.

Optimal performance improvement of PSO: Two shortcomings of basic particle swarm optimization: One is that PSO is slow in convergence rate; the other is that PSO is easy to fall into local extreme point, causing that the algorithm cannot obtain global optimal solution (Managasarian, 2009).

PSO based on growth stage of particle mainly improves it on the speed update formula of particle. The core concept of algorithm improvement is that in the early phase of iteration of particle swarm algorithm while particle is falling into local extreme point, adopt mutation mechanism to make it jump out of local extreme value and dynamically adjust the proportion of "individual cognition part (pbst)" and "social cognition part (gbst)" of particle in different evolution phases, thus improving optimal performance of algorithm. As the setting of inertia weight ϖ influences the balance between global searching ability and local searching ability of particle it plays a critical role in the success of algorithm. Algorithm adopting larger inertial weight in the early phase can make particle swarm have strong global searching ability; adopting smaller inertial weight in the later phase can improve the local searching ability of particle swarm. Therefore, linear adjustment of inertial weight is adopted in this thesis.

Through the learning of parameters of PSO and large quantity of simulation experiment, parameters ϖ, c_1 and c_2 are expressed with mathematical expression as shown in equartion (3),equartion (4) and equartion (5):

$$\varpi = \varpi_{max}-(\varpi_{max}-\varpi_{min})t/T \tag{3}$$

$$c_1 = \begin{cases} 1.4962 & 0 \prec t \prec 2T/3 \\ 1.4962 - t/T & 2T/3 \le t \le T \end{cases} \tag{4}$$

$$c_2 = \begin{cases} t/T & 0 \prec t \prec T/3 \\ 0.4926 & T/3 \le t \prec 2T/3 \\ 1.4926 + t/T & 2T/3 \le t \le T \end{cases} \tag{5}$$

In the above equations, t is the current iterations of particles and T is the maximum iterations of particle swarm. The calculation steps of PSO1 are as follows. Inputs are particle swarm scale N and particle dimension D, inertial weight parameters of particle swarm, ϖ_{max} and target function fitness(x); outputs are optimal location of particle swarm gbst; optimal solution fitness (gbst):

- Initialize particle location x and particle speedv
- t = 1
- Calculate the parameters of iterative formula through equation (3), equation (4) and equation (5)
- If gbst fails to be updated or is subtle in change in recent iterations and particle hasn't mutated in recent iterations, carry out mutation on the particle
- Update particle speed and location
- According to fitness function fitness(x)
- If the current location of particle is better than the optimal location experienced, pbst = current location of particle
- Return to the second step until t = T

Application of PSO and K_means in customer classification: As there are two inherent defects of K-means clustering algorithm: One is that the random selecting of initial value may cause different clustering results, even no solution; the other is that the algorithm is the one based on gradient descent, thus constantly falling into local optimal solution inevitably. These two defects restrain its application scope. Thus, this thesis will make use of strong parallel computation capability of particle swarm, combing with K-means clustering algorithm, so as to conquer the defects of K-means algorithm.

In particle swarm algorithm, the location of each particle is comprised of K clustering centers, particles having speed and adaptive values besides location. As the dimension of sample vector is D, the location and speed of particle are K*D dimensional vector. Besides, particle has an adaptive value (Tao and Liu, 2012).

In PSO, fitness function is used to judge the strength and weakness of particle location in the evolution progress of swarm. Clustering results depend on the results of target function:

$$J = \sum_{i=1}^{k}\sum_{p\in c_i} d(p - Z_i)$$

The smaller the J is, the better the clustering result is. Hence, fitness function of particle adopted in this thesis is shown in equation (6):

$$\text{fitness}(Z_i)\frac{1}{1+J} = \frac{1}{!\sum_{i=1}^{k}\sum_{p\in c_i} d(p - Z_i)} \tag{6}$$

In formula 11, J means intra-class distance sum; algorithm search the maximum value of fitness function fitness (Z_i) through the iteration of particle swarm.

The steps of K-means clustering algorithm based on PSO are as follows. Inputs are Data Set, category number K and particle swarm algorithm parameter; output is optimal location coding gbst:

- One-off dispatching each sample as certain category
- Calculate each clustering center Z_k
- Initialize particle swarm with clustering center Z_k
- Carry out clustering division on each sample according to Nearest Neighbor Principle
- If $x \le x_{min}$, then $x = x_{min\ (1-0.01\times rand())}$
- As for each particle, calculate new clustering center, update fitness value of particle, replace original coded value (Shen and Wang, 2011)

EXPERIMENT CONFIRMATION

The instance data of the experiment conduct empirical research on the customer data of the B2C transaction of certain enterprise website of the recent three years (totaling data of 41351 customers, 21 attributes in the data table are listed in the third part of the paper including customer characteristics type variables and customer behaviors type variables), making statistics on attribute values like annual transaction frequency, total amount, product cost, etc. of certain customer according to customer transaction records in information base, forming an information table (among which the decision attribute set D is null) . Establish training sample set according to domain (prior) knowledge. Times of purchasing and total amount of purchasing of each customer are two major

Table 1: Classification performance comparison of different algorithms

Algorithm	Algorithm in this paper	Ordinary k-means algorithm	Ordinary PSO algorithm
Accuracy rate	99.7%	88.47%	94%
E Value	104.33	159.81	119.96

factors of customer classification (this is the prior knowledge of domain), so select 400 pieces of typical data among all the customers to form training sample set. And divide them into five types as Gold Customers, Silver Customers, Copper Customers, General Customers and Negligible Customers according to ABC management theory.

The study uses the customer classification algorithm above-mentioned and the customer classification results can be expressed like these. In the autonomous learning of algorithm of this study, such five factors as the educational background, income, occupation, times of purchasing and total amount of purchasing of customers have a relatively great influence on customer classification. Through the classification results it can be seen that Gold Customers take up 6.71% of the total number of customers while the profit takes up 53.19% of the total profit. These customers play a significant role in the existence and development of enterprises. However, the negligible customers account for 16.89%, who not only do not bring profit to enterprises but also make enterprise lose money. These customers should be either further cultivated or eliminated according to the actual situation.

In the specific algorithm realization, this paper simultaneously realizes ordinary K_means algorithm (Jiang and Zhang, 2011)) and customer classification algorithm based on ordinary PSO algorithm (Shen and Wang, 2011). The performance comparison of these three algorithms can be expressed in Table 1. In Table 1 E means square errors The smaller the E value is, the smaller the possibility of wrong classification is. Thus it can be seen that the square error and E value of the algorithm in this paper during the classification are far more less than ordinary K-means algorithm and ordinary PSO algorithm. Therefore it shows that the improvement algorithm clustering algorithm in this study turns out to be a success, with reasonable classification results.

CONCLUSION

Customer relations management of online trading is still developing. But to correctly and effectively classify online trading customers is the critical issue for reforming network marketing mode, improving customer management and service level and enhancing competitiveness of network enterprises. On account of the shortcomings of the typical K_means and PSO classification algorithm in data mining, this paper puts forward several improvement measures and applies them into the classification of online trading customers. Simulation results indicate that the improved online trading customer classification has higher accuracy rate on customer classification and more reasonable classification results.

ACKNOWLEDGEMENTS

This work is supportedbyscientific research project of the education department of Jiangxi Province (GJJ13300) and52nd Chinese Postdoctoral Fund underthegrantNo. 2012M521284.

REFERENCES

Bradley, P.S. and L. Managasarian, 2000. K-plane Clustering. J. Global Optimiz., 16: 23-32.

Ding, R.T., 2013. The resource scheduling strategy for occupation ideal network games based on improved particle swarm algorithm. J. Digi. Cont. Tech. Applied, 2: 117-123.

Jiang, L.G. and L.S. Zhang, 2011. Review of customer classification methods based on the customer value. J. Zhej Univ., 8: 1209-1215.

Managasarian, L.K., 2009. A modified particle swarm optimization. Trans. Comput. Math, 12: 1174-1183.

Shen, Y.X. and G.Y. Wang, 2011. Correlative particle swarm optimization model. J. Soft, 4: 695-708.

Tao, X.M. and F.R. Liu, 2012. Multi-scale cooperative mutation particle swarm optimization. J. Soft, 7: 1806-1805.

Zhou, H., 2012. Study of classifying customers method in CRM. Comput. Eng. Des., 3: 659-661.