Application of Regression Analysis Based on Genetic Particle Swarm Algorithm in Financial Analysis

Xiaorong Cheng
Department of Computer
North China Electric Power University
Baoding, Hebei Province, China
cheng3100@sohu.com

Lin Sun
Department of Computer
North China Electric Power University
Baoding, Hebei Province, China
5daidaizhu521@163.com

Ping Liu
Information Operation Center
Hebei Electric Power Research Institute
Shijiazhuang, Hebei Province, China
sdmislp@sina.com

***Abstract*—Slow convergence speed and premature are two key problems existing in the regression analysis techniques based on genetic algorithm. To overcome the shortcomings, this paper proposes an improved regression analysis based on the genetic particle swarm algorithm. The basic principle is that a new operator is constructed to use PSO. This algorithm has the choice of genetic algorithms and genetic features, and drawing on the searching capabilities of particle towards the optimal forward. Finance analysis evaluates the efficiency of the algorithm. The experimental results show, the improved regression analysis is steady and greatly improve the convergent speed.**

Keywords- Particle swarm algorithm, Genetic algorithm, Regression analysis, Mutation

I. INTRODUCTION

Regression analysis is a multivariate statistical analysis; regression analysis is a mathematical method that studies the dependent relationship between the variables. Sometimes between the variables exist dependence of non-deterministic, that the variables not only exist the number of close relationship but also can not be obtained accurately by one or more variable values. Based on the statistics, such variables that can be identified variation of such relations are usually referred to as correlativity; regression analysis is used to solve such correlativity. Regression analysis include the linear regression based on the basic principle of a least squares, multiple linear regression, nonlinear regression, logarithmic regression, Poisson regression. Regression analysis in all fields has a wide range of applications. Now it has gradually inosculated many other disciplines approaches in the application. The function of regression analysis continues to improve in the integration of computer technology, mathematical modeling, and artificial intelligence.

Wen-Sheng Liu et al proposed genetic algorithm-based regression analysis, which was high intelligence and was also very effective, but slower execution. For this reason, I draw on particle swarm optimization algorithm based on genetic algorithm that has been proposed. For filling up the deficiencies of regression analysis based on genetic algorithm, I improve on this algorithm and make use for regression analysis.

II. PARTICLE SWARM ALGORITHM INDUCTION

The vector of the particle i is $X_i=(x_{i1}\cdots\cdots\cdots x_{iD})$, $i=1, 2\cdots\cdots..N$, and the position of the particle i is X_i in the D-dimensional search space. The flight speed of the particle i is $V_i=(v_{i1}\cdots\cdots\cdots v_{iD})$. So far searching the optimal position of the particle i $P_i=(p_{i1}\ldots\ldots\ldots p_{iD})$, and searching for the optimal location of the entire population is $P_g=(p_{g1}\ldots\ldots\ldots p_{gD})$. Particle action formula is shown below:

$$v_{id}(t+1) = wv_{id}(t)+c_1r_1(p_{id}(t)-x_{id}(t))+c_2r_2(p_{gd}(t)- x_{id}(t)) \quad (1)$$

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

t is set the evolution generation, w is the inertia weight for controlling the impact between the speed of the previous and velocity on the current. When w is large, the previous rate was affected greatly and the global search ability is strong; When w is small, the previous speed is small, and the local search ability is strong. So the algorithm jumps out of local minimum by adjusting the size of w. Generally ,use weight $w = 0.8$ is better. c_1 and c_2 are the learning factor, and are the random numbers, usually ranging between 0 and 2. r_1 and r_2 are also the random numbers in [0,1]. Each dimension of the particle velocity is limited to one between the maximum speed V_{max}. If $V_i>V_{max}$, then $V_i=V_{max}$; if $V_i<-V_{max}$ then $V_i=-V_{max}$. Termination condition is based on the maximum number of iterations taken on specific issues ,or the particle swarm search the best position to meet the target minimum threshold adaptation.

After the introduction of improved shrinkage factor λ, the particle swarm algorithm is shown below:

$$v_{id} = \lambda\ (wv_{id}+c_1r_1(p_{id}- x_{id})+c_2r_2(p_{gd}- x_{id})) \quad (3)$$

$$\lambda=\frac{2}{\left|2-\varphi-\sqrt{\varphi^2-4*\varphi}\right|},\varphi=c_1+c_2\ ,\varphi>4 \quad (4)$$

The role of λ is similar to the parameter V_{max} to control and bound the particles' flight speed. The results show that,

λ is more effective than the V_{max} to control the vibration of the particles velocity. Eberhart and Yuhui Shi detailed analysis and comparison of the two parameters on algorithm performance, they consider the shrinkage factor is more effective to control the particles' flight speed, while enhance the efficiency of local search ability.

III. REALIZATION

A. Modified GPSO(MGPSO)

PSO with a similar genetic algorithm is an optimization algorithm based on iterative. System is initialized to a set of random solution, an iterative search for the optimal value. PSO does not use the crossover and mutation like the genetic algorithm, the particles follows the best particle in the solution space to search. Compared with the genetic algorithm, PSO's advantage is to maintain the historical status and particle current state. Between individual particles it is to help rather than competition, making it the maintenance of population diversity can be well. Traditional genetic algorithm variate a part of a group of individual randomly, with the history and current state of independent, and does not fully use the experience of other individual optimization. For the problem, this paper introduces particle swarm algorithm to the genetic algorithm, using PSO to reconstruction operator, and full use of particles and other individual record information. The new algorithm improve on the general GPSO.It doesn't adopt population intersected strategy that has been proposed, avoiding to islanded these subpopulations. It has the choice of genetic algorithms and genetic features, and drawing on the searching capabilities of particle towards the optimal forward. View from the global search GPSO uses genetic algorithms, and from the local perspective GPSO uses particle swarm optimization. Here the hybrid algorithm of the most important steps:

1) Individual coding

Encoding is the first step in the algorithm and is also the very important step. Individual coding is the way of structuring individual chromosomes vector, It's the conversion method that a feasible solution to a problem from its solution space is transformed into the search space which the genetic algorithm can handle. The genetic algorithm have a variety of encoding, which encoding have their specific adaptation problems. Common coding methods are: fixed-length binary encoding, dynamic coding, real coding, ordered string of code, structure, coding, etc. In this paper, I use the real number encoding, that is, each individual gene in place to take a certain range (eg from 1 to n) of real numbers. Although the real-coded inside the computer has to convert binary, but it avoids the operator operate to base on the internal point, to facilitate the introduction of other genetic operators.

2) Crossover

According to the chromosomes vector the corresponding real particle swarm string Q are randomly generated when the algorithm is initialized.X_m^i is a single particle, where $i = 1 \ldots n$, m is the evolution generation and M is set the largest evolution generation. And for any i $(1 \leqslant i \leqslant n)$, both $X_m^i \leqslant Xmj(i \neq j)$, Q is order. The hybrid algorithm uses the arithmetic crossover, and two individuals X_m^i and X_m^j $(i \neq j)$ are crossed by the arithmetic crossover:

$$X_{m+1}^i = \alpha X_m^j + (1-\alpha) X_m^i \tag{5}$$

$$X_{m+1}^j = \alpha X_m^i + (1-\alpha) X_m^j \tag{6}$$

α is a parameter, where is the random number determined by the evolution generation. This is a non-uniform arithmetic crossover. Once every evolutionary, this algorithm randomly generated a random number.

3) Mutation

Particle swarm optimization evolution equation is used to refactoring mutation operator. The individuals determine their mutation rate of the direction and magnitude according to theirselves optimal solution, the optimal solution of the sub-populations and individual evolution speed so far. The individual in the evolution make the evolutionary history be as a mantra. Realization:

X_m^i replace X_{id} in the particle swarm algorithm. X^i_{max} for the optimal solution f^i_{max} of the particle i in m generation replace P_{id} in the particle swarm algorithm. X^j_{max} for the optimal solution F^j_{max} of the population (j for the particle group in the position of population)replace P_{gd} in the particle swarm algorithm. $\overline{X}^i_{max}$ is the cumulative difference arithmetic mean of X^i_{max} to replace V_{id}. The cumulative difference is obtained by the following formula:

$$\overline{X}^i_{\max m+1} = \frac{\sum_{k=2}^{m} (\overline{X}^i_{\max k} - \overline{X}^i_{\max k-1})}{m} = \overline{X}^i_{\max m-1} + \frac{X^i_{\max m} - X^i_{\max m-1}}{m} \tag{7}$$

After the introduction of variation factor ,the particle updating formula of particle swarm optimization algorithm is shown below:

$$\begin{aligned} \overline{x}^i_{\max,m+1} &= \lambda(\overline{x}^i_{\max,m} + c_1 r_1 (\overline{x}^i_{\max} - x^i_m) + c_2 r_2 (\overline{x}^j_{\max} - x^i_m)) \\ x^i_{m+1} &= x^i_m + \overline{x}^i_{\max,m+1} \end{aligned} \tag{8}$$

B. Regression analysis based on MGPSO

Regression analysis use the algorithm to compute the coefficients of the function. Take example for multiple linear regression, $Y=\omega_0+\omega_1 X_1+\omega_2 X_2^2+\omega_3 X_3^3 \ldots\ldots +\xi$, the algorithm is used to confirm the coefficients ω_i $(i=0 \cdots n)$. Each individual X_m^i is formed by a group of the chromosomes vector that is composed of coefficients $\{\omega_0^{im}, \omega_1^{im} \ldots\ldots \omega_n^{im}\}$. Individual in the evolutionary process go through crossover and mutation, crossover operator is arithmetic crossover fashion on the use of type (5), (6), on all the chromosomes of individual X_m^i in the implementation of crossover, that is the overall crossover. The results of individual X_m^i *and* X_m^j $(i \neq j)$ after the overall crossover:

$$X^i_{m+1}\left\{\omega_0^{im},\omega_1^{im}.......\omega_n^{im}\right\}=\alpha\omega_K^{jm}+(1-\alpha)\omega_K^{im} \quad (9)$$

$$X^j_{m+1}\left\{\omega_0^{j(m+1)},\omega_1^{j(m+1)}......\omega_n^{j(m+1)}\right\}=\alpha\omega_K^{im}+(1-\alpha)\omega_K^{jm} \quad (10)$$

Mutation operator use the formula (8) of the particle swarm update formula, then

$$\begin{aligned}&\vec{X}_{\max,m+1}\{\omega_0^{im},\omega_1^{im}.....\omega_n^{im}\}=\lambda(\vec{X}_{\max,m}+c_1r_1(\vec{X}^i_{\max}-\omega^i_k)+c_2r_2(\vec{X}^j_{\max}-\omega^j_k))\\&\vec{X}^i_{m+1}\{\omega_0^{i(m+1)},\omega_1^{i(m+1)}.....\omega_n^{i(m+1)}\}=\vec{X}^i_m\{\omega_0^{im},\omega_1^{im}.....\omega_n^{im}\}+\vec{X}_{\max,m+1}\end{aligned} \quad (11)$$

In the course of the genetic, using the appropriate selection strategy, the individual in the individual winners were hybridized and the inferior individuals were variated. Usually the selection strategies include the fitness proportional method and the ranking selection method. This article will combine the two, with the proportional ranking method. Specific method: the individuals are to set in descending order according to the size of fitness values. According to the number of population ,the interval was set *[0, number * (number +1)]*.The individuals of good sort were distributed the interval segments. If the location of the individual *i* is location, then its interval segment is *(2*(location+1)/number*(number+1))*.When the size of the population is *number*, the system randomly generate the random numbers. The size of the random numbers is also *number*. There are *number* selected individuals in the interval segment, these selected individuals reproduce and produce offspring.

Flow chart for the regression analysis of the proposed algorithm is shown in Fig.1:

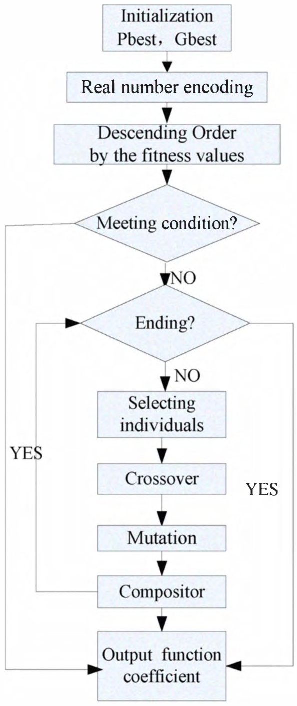

Figure 1. Algorithm flow chart

IV. CASE

A. *Stability Test One*

Testing the stability of the regression model, first, the system generate the three sets of randomly test data.

Step 1: Randomly generate x value

Step 2: accordingly $y=(1+0.0005r)x^2r$ randomly generate y value in the interval *[-1,1]*

Step 3: Back to Step 1

Step 4: Randomly generate the noise x, y values that do not meet $y=x^2$

The second set of data is y values in accordance with $y=(1+0.0005r)x^2$.The third set of data is y values in accordance with $y=(1+0.002r)x^2y$. Set the size of population is 50, the size of sample is 200, the search interval is *[-100,100]*, results as shown in the table 1.Can be seen from the examples, regression analysis based on genetic particle swarm optimization has a certain stability, and it can get a good regression equation.

TABLE I. THE REGRESSION EQUATION

Data	**R**	**The regression equation**
The first set of data	0.99	$Y=0.0015-0.0072x+1.0189x^2$
The second set of data	0.99	$Y=0.0075-0.0453x+1.0538x^2$
The third set of data	0.99	$Y=0.0015-0.0072x+1.0189x^2$

B. Stability Test Two

The regression analysis use 6 consecutive years financial data from one company, as shown in Fig.2:

A	B	C	D
Item	2000	2001	2002
Current assets	293510.18	312933.45	324095.85
Long-term assets	166581.54	355853.37	452077.65
Current liabilities	138620.1	196312.5	191713.5
Long-term liabilities	2350	2250	2250
capitalization	48260.47	87753.34	114479.65
capital accumulation fund	104201.48	162403.13	195029.43
surplus accumulation fund	30998.95	53286.41	71976.07
undistributed profit	135660.72	166781.44	200724.84
shareholder's equity	318861.15	469374.97	530705.55
equity	460091.72	668786.82	776173.5
Main business income	395364.01	474209.53	570651.74

Figure 2. Financial Data

These financial data was analysis by the SPSS tool. The result is that current assets, long-term assets and liabilities are linear correlation with the main business income, and these data is adapted to the regression analysis, in Fig.3.

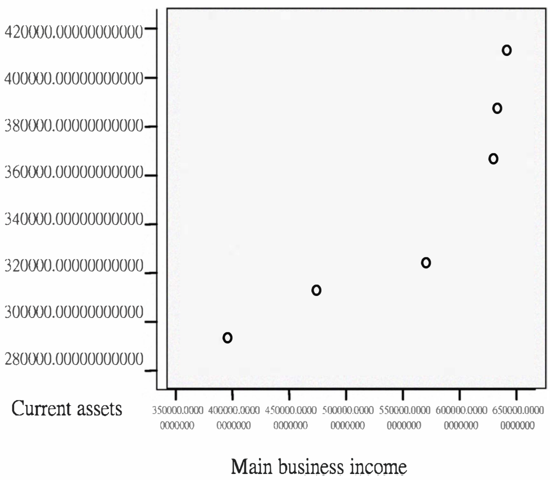

Figure 3. Linear Correlation Figure

Using the SPSS linear regression analysis tool, the results shown in Fig 4.The regression equation by the regression analysis-based genetic algorithm is:

Main business income=394510.2+0.390*Current assets+0.831*Long-term assets-1.327* Current liabilities

Multiple correlation coefficient is 0.990, Regression sum of squares is 7E+013, Residual sum of squares is 9E+009。The regression equation by the regression analysis-based genetic particle swarm optimization algorithm is:

Main business income=400361.2+0.265*Current assets+0.801*Long-term assets-1.499* Current liabilities

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.993a	.987	.979	**********

a. Predictors: (Constant),

ANOVAb

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	7E+010	3	2.174E+010	125.888	.000a
	Residual	9E+008	5	172682909.4		
	Total	7E+010	8			

Coefficientsa

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	320665.5	87741.162		3.655	.015
	Current assets	.381	.193	.204	1.975	.105
	Long-term assets	.736	.146	1.153	5.049	.004
	Current liabilities	-1.135	.684	-.356	-1.658	.158

a. Dependent Variable:

Figure 4. Result Show

Multiple correlation coefficient is 0.995, Regression sum of squares is 7E+012, Residual sum of squares is 9E+005. Comparison of the three can be seen that multiple correlation coefficients of our method is closer to 1 than the linear regression and regression analysis-based genetic algorithm , so this algorithm is better than that two. At the same time, this algorithm run longer than this linear regression algorithm, but better stability, quicker convergence speed than the regression analysis based on genetic algorithm. Thus, this method of regression analysis has good stability and convergence speed, improved these defects of regression analysis based on genetic algorithm.

REFERENCES

[1] Yu Wan-Xia, Zhang Wei-Cun, Zheng Hong-Xing. Genetic and particle swarm algorithm－based optimization solution for high-dimension complex functions [J]. 2007, 43(36) : 31-33.

[2] Peng Xiao-Bo, Gui Wei-Hua, Huang Zhi-Wu, Hu Zhi-Kun, Li Yong-Gang. GAPSO : Effective Genetic Particle Swarm Algorithm and Its Application[J]. 2008,20(18) : 5025-5031.

[3] Liu Song-Bing, Li Zhi-Yong, Wang Yong et a1. Novel evolutionary Particle Swarm Optimization for Traveling Salesman Problem[J]. 2008, 44(28) : 62-64.

[4] Wang Zhi-Gang . A modified particle swarm optimization[J]. 2009,25(4) : 464-466.

[5] Fevrier Valdez, Patricia Melin and Oscar. Castillo.Evolutionary Method Combining Particle Swarm Optimization and Genetic Algorithms using Fuzzy Logic for Decision Making. FUZZ-IEEE. 2009.

[6] Ahmed Yousuf Saber, Toshihisa Funabashi,Tomonobu Senjyu and so on. Thermal Unit Commitment Strategy with Solar and Wind Energy Systems Using Genetic Algorithm Operated Particle Swarm Optimizati . IEEE International Conference on Power and Energy. 2008.

[7] Said E. El-Khamy, Mohamed A, Aboul-Dahab and Mohamed M. Attia. A Hybrid of Particle Swarm Optimization and Genetic Algorithm for Multicarrier Cognitive Radio. 26th national radio science conference. 2009.