The Application of Macro-Economic Prediction Based on Improved Gene Expression Programming

YAN Jingfeng
College of Computer Science and Technology
Xuchang University, Xuchang, Henan, 461000
jeffery8224@126.com

LI Guoqing
College of Computer Science and Technology
Xuchang University, Xuchang, Henan, 461000
lgq0419@sohu.com

Abstract-**An Improved Gene Expression Programming(IGEP) is proposed in this paper. It has some new features: 1) introducing a new individual coding; 2) introducing a new way of creating constants; 3) introducing a hybrid self-adaptive crossover-mutation operator, which can enhance the search ability and exploit the optimum offspring. To validate the performance of IGEP, this paper applies IGEP into the solution of the macro-economic predictions. The experimental results demonstrate that Improved GEP can automatically find better Optimization Model, based on which prediction will be generated much more exactly.**

Keywords-component; Gene Expression programming; genetic algorithms, macro-economy；prediction

I. INTRODUCTION

Gene expression programming (GEP[1]) is an new evolutionary algorithm presented by the Portuguese scientist C. Ferreira in 2001. GEP and genetic algorithms (GA)[2] and genetic programming (GP)[3], have the same Mechanics of survival of the fittest . The fundamental difference between the three algorithms lies in the nature of the individuals: in GAs the individuals are linear strings of fixed length (chromosomes); in GP the individuals are non-linear entities of different sizes and shapes expressed as parse trees; and in GEP the individuals are encoded as linear strings of fixed length (the genome or chromosomes) which are afterwards expressed as non-linear entities of different sizes and shapes (simple diagram representations or expression trees).

Above all, GEP absorbs the merits of both GA and GP and meanwhile eliminates their demerits. GA facilitates genetic operation with its fixed-length linear serial coding, but it can only deal with some simplematters. GP, with its individual non-linear tree graph, could be used for handling complicated problems, however, it is not suitable for genetic operations and difficult to understand due to its very complexity. In GEP the individuals are encoded as linear strings of fixed length which is later conversed to expression trees while calculating. It is able to deal with many matters such as prediction, classification, symbol regression and image processing through the automatic creation of programing.

Gene Expression Programming is simple, fast, robust and easy to use; however, Original Gene Expression programming (GEP) may give rise to a slow convergence speed in the last of evolution.

II. RELATED RESEARCH WORK

Portuguese scientist C. Ferreira encodes the individual genes in GEP with head and tail[4]. The head is made up of elements from operator sets and terminal sets; and the tail consists of element from termianal sets. A problem arises when initialization individuals are gennerated randomly. The first element generated belongs to terminal sets and the corresponding chorosome expression is a constant or a variable which is of no practical meaning.

GEP can make unlimited search in problem solving space and generate effective program structures because of the mapping between its linear serial and expression tree[5][6]. However, after the creating of program structure, the system established can still be instable if improper parameter is selected[7][8].

Reference[9][10] adopts an evolutionary hill climbing algorithm based on Multi-parent Crossover. It adopts a comparing strategy capable of maintaining the diversity of population to judge the quality of individuals, However, this algorithm only applies to simple problems and requires repeated evaluation of functions many times, which can not be widely used in practical engineering predictions.

In view of this, some improvements are made on Original Gene Expression programming in this paper, and a new individual coding is proposed, which can reduce the no practical meaning expressing and a new creating of constant is proposed, which establishs random constant and makes adaptive changes continuously according to fitness parameter during the process of program running; Meanwhile a novel hybrid self-adaptive crossover-mutation operator is also introduced to enhance the global search ability of the algorithm and the search of non-convex areas, and maintains the diversity of population.

III. GEP OF SOLVING FUNCTION MODEL

GEP[5] solves function model problems as follows:

Step 1. Produce an initial population with NP individuals, generation k=1; each individual is a expression tree

Step 2. Calculate the fitness of each individual in the population

Step 3. Find out from the population the best individual X_{best} and the worst individual X_{worst} .

978-0-7695-3645-3/09 $25.00 © 2009 IEEE
DOI 10.1109/CINC.2009.31

Step 4. Stop if evolutionary generation $k = MAXGEN$ or $|f(X_{best}) - f(X_{worst})| < \varepsilon$; or else, continue.

Step 5. For each individual i, $i \in \{1,2,\cdots,NP\}$, random select 5 individuals r_1, r_2, r_3, r_4, r_5, individuals form the population, $r_1 \neq r_2 \neq r_3 \neq r_4 \neq r_5 \neq i$.

Step 6. Use crossover-mutation operator to recombine each individual:

Mutation:
$$X_i^{'} = X_{r_1} + F \times (X_{r_2} + X_{r_3}) \tag{1}$$

F > 0 is a user-defined scaling factor, X_{r_1} is mutation vector. new individual may not always use $X_i^{'}$ directly; it may produce new individual Y_i, by crossover operator as showed in formula(2) with the original parent-individual.

Crossover :
$$Y_i(j) = \begin{cases} X_i^{'}(j), & if\,(R(j) \le CR) or (j = t) \\ X_i(j), & if\,(R(j) > CR) and (j \neq t) \end{cases} \tag{2}$$

$t \in \{1,2,\cdots,n\}$ is a random integer, $R(j) \in [0,1]$, j is a random number, CR is user-defined crossover probability.

CR = 1 $Y_i = X_i^{'}$; $CR = 0$ $Y_i = X_i$

Step 7. choose parent X_i and child Y_i based on tournament selection strategy, better individuals descend into next generation population. $k = k + 1$, goto Step 3.

IV. IMPROVED GEP

In order to enhance the convergence speed and the ability of searching function model and keep the diversity of population, improvement on GEP is proposed as follows.

1) an new coding**:** individual genes are encoded as three parts of head, body and tail. The head consists of elements from operator sets; the body consists of elements from both operator sets and terminal sets; and the tail consists of elements from terminal sets.

h(tail)=h(body+head)*(n-1)+1 (3)

h(tail) indicates the length of the node, h(body+head) indicates accumulated length of head and body; n refers to the maximum number in operator symbol. for example, ＋ ， － ， *, / is monadic operation; sin,ln q, is binary operation, and IF is three-demension operation. In this way, it will be guranteed that the chorosome generated at random will produce a correct expression; examples are shown as follows;

fig1 shows linear string of individual coding and fig2 shows expression tree of individual coding.

head			body			tail								
0	1	2	3	4	5	6	7	8	9	0	1	2	3	4
*	sin	+	+	a	-	ln	a	a	b	b	a	a	b	a

Fig1. linear string of individual coding

The respective arithmetic expression is sin(lnb+a)*(a+(a-b))

2) an new way of creating constant: brings in the symbol "?" to represent temporary random constant in terminate sets and meanwhile uses an array to save random constant; and in the process of decoding, replaces the symbol "?" with the elements in the array in turn, the random constant's array X[10]={0.43, 0.05, 0.576, 0.001, -2.33, 1.178, -0.25, 2.93, 1.33, 0.98}.

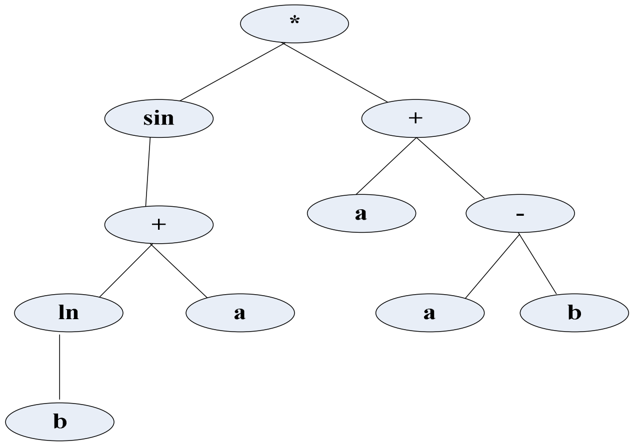

fig2. expression tree of individual coding

3) Heuristic self-adaptive crossover-mutation operator. Two mutation operators are designed as follows formula (4), (5)

$$P_1 = \sum_{i=1}^{M} a_i * X_i \tag{4}$$

$$P_2 = X_{r_1} + F * (X_{r_2} - X_{r_3}) + (1 - F)(X_{r_4} - X_{r_5}) \tag{5}$$

a is a random real number, $a_i \in [-0.5, 1.5]$, $\sum_{i=1}^{M} a_i = 1$, M stands for the number of parent individual selected. Formula 8 is mentioned in reference[10]; random select a_i from interval [-0.5, 1.5] may enhance the search ability of non-convex domain.

V. EXPERIMENT AND RESULTS

The experimental data of this paper comes from Rob Hydman's web site[11] where data specially used for time series forecasting are sorted out, and from which this paper selected some data about macro-economic prediction, that is, data of telephone expenditure from 1915 to 1996 in US New York, San Francisco. Specific information about the data is displayed in the website. As is shown, 984 groups of data in all are included in this data sets. This paper takes the first 8 groups of data as training sample data and the last 30 groups of data as predicting data.

A. Individual parameters

. Table 1 Common Parameters setsting for GEP experiment

Parameter	value	Parameter probability	value
Sizes of population	100	Inversion transformation	0.1
Length of head	8	IS transformation	0.1
Length of body	4	RIS transformation	0.1
Length of tail	13	Gene transformation	0.1
Number of genes	10	One-point crossover	0.3
Time delay	1	Two-point crossover	0.3
Numbers of prediction	5	Mutation probability	0.4

B. *Experiment Results*

The experiment chooses function sets F = {+, -, *, /}, endpoint sets T = {x0, x1, x2, x3, x4, x5, x6, x7, x8, x9} , and uses random constant . Running the experiment 10 times, a preferable model as follows is generated:

$$F = x0/(((x8+x2)*x3)/x8)-x4)+x1/(x3*x3)+x7+x7/(x2*x2)+$$
$$(((x6/x3)-(x6-x9))/(x1*x2))*x5+((x0*(x9-x7))-1.203*x8)/x3$$
$$+0.56*x0/(x1+((x4+x9)-x3)/x9+(x1-x8)/x9 \tag{6}$$

$$Q = \sqrt{\sum_{i=1}^{n} (y - y')^2} \tag{7}$$

The fitting error Q=0.987490 while the predicting error is 0.359066. Experimental result is displayed in Fig3.

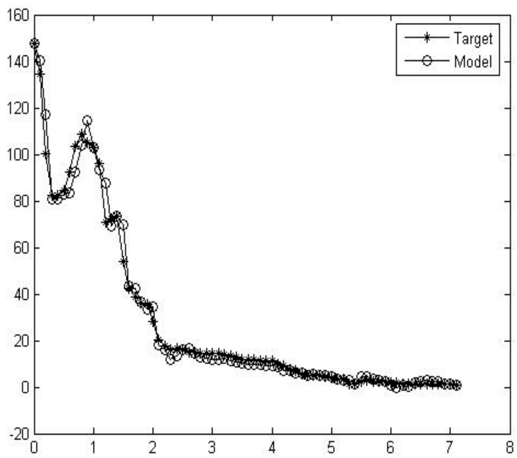

Fig3. Predicting curve of training data sets

C. *Analysis of the result*

As is shown in Fig3, the using of GEP could promote the fitting of training data. Models established through improved GEP can better reflect real situation due to its smaller fitting error. Furthermore, by fitting previous data, the improved GEP can make better prediction on future situation; the value of predicting error is 0.359066.

VI. CONCLUSION

GEP is an adaptive, new evolutionary algorithm. choosing proper fitness function, this paper comes up with a method to design individual coding including three parts (head, body, tail) and adopts a new method of creating constant to improve GEP's searching efficiency, and uses improved GEP to solve macro-economic predicting problems. The experimental result of this paper shows that improved GEP can automatically find better models and make more precise predication based on the model created above. As is shown, improved GEP has promising prospects in the field of rules exploration and knowledge discovery in large data warehouse.

REFERENCES

[1] Ferreira, C. Gene Expression programming: A New Adaptive Algorithm for Solving Problems[J]. *Complex Systems*, 2001, 13(2): 87-129.

[2] Mitch M. An introduction to genetic algorithms[M]. MIT Press, 1996.

[3] Koza J R. Genetic programming: On the programming of computers by means of natural selection[M]. Cambridge, MA: MIT Press, 1992.

[4] Ferreira, C. Gene Expression programming [M]. Portugal, Angra do Heroismo, 2002.

[5] Tang Changjie，Zhang Tianqing.etc. Knowledge discovery based on Gene Expression Programming-History, achievements and futuredirections [J].Computer Application，2001，24(10)，7-10.

[6] Ferreira, C. Mutation, Transposition, and Recombination: An Analysis of the Evolutionary Dynamics[J]. *The 4th International Workshop on Frontiers in Evolutionary Algorithms*, 2002, 614-617.

[7] Ferreira, C. Gene Expression programming in Problem Solving[J]. *Soft Computing and Industry – Recent Applications*, 2002,635-654.

[8] Ferreira, C. Combinatorial Optimization by Gene Expression programming: Inversion Revisited[J]. Proceedings of the Argentine Symposium on Artificial Intelligence, 2002, 160-174.

[9] Duan Lei，Tang Changjie. etc. An Anti-Noise Method for Function Mining Based on GEP [J]. Computer Research and Development，2004，41(10)，1684-1689.

[10] Huang Xiaodong，Li Zhi, etc. Function Mining Based on GEP [J]. Software，2004，15，97-109.

[11] http://www-personal.buseco.monash.edu.au/~hyndman