

The time complexity analysis of a class of gene expression programming

Xin Du · Youcong Ni · Datong Xie · Xin Yao ·
Peng Ye · Ruliang Xiao

Published online: 13 December 2014
© Springer-Verlag Berlin Heidelberg 2014

Abstract This paper studies the time complexity of gene expression programming based on maintaining elitist (ME-GEP). Using the theory of Markov chain and the technique of artificial fitness level, the properties of transition matrices of ME-GEP are analyzed. Based on the properties, the upper and lower bounds of the average time complexity of ME-GEP are obtained. Furthermore, the upper bound is estimated, which is determined by the parameters of ME-GEP algorithm. And the theoretical results acquired in this paper are used to analyze ME-GEP for solving function modeling and clustering problem. At last, a set of experiments are performed on these problems to illustrate the effectiveness of theoretical results. The results show that the upper bound of expected first hitting time can be used to direct the algorithm design of ME-GEP.

Keywords GEP · ME-GEP · Average time complexity · Artificial fitness level · Markov chain

Communicated by U. Fiore.

X. Du (✉) · Y. Ni · R. Xiao
Faculty of Software, Fujian Normal University, Fujian 350108, China
e-mail: xindu79@126.com

X. Du · X. Yao
The Centre of Excellence for Research in Computational Intelligence and Applications(CERCIA), School of Computer Science,
The University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

P. Ye
College of Mathematics and Computer, Wuhan Textile University,
Wuhan 430073, China

D. Xie
Department of Information Management Engineering,
Fujian Commercial College, Fujian 350012, China

1 Introduction

Gene expression programming (GEP) (Ferreira 2001) is a genotype/phenotype evolutionary algorithm. As a descendant of genetic algorithms (GAs) and genetic programming (GP), GEP combines their advantageous features and eliminates their main handicaps. The essential feature among GEP, GA and GP is that GEP separates genotype (linear structure) from phenotype (non-linear structure) of chromosomes. GA and GP only have one entity, such as linear strings of fixed length (chromosomes) or nonlinear entities of different sizes and shapes (parse trees).

From the history of life on the earth (Smith and Szathmry 1995), this kind of system such as GA and GP is condemned to have one of two limitations (Ferreira 2001): if they are easy to be manipulated genetically, they lose in functional complexity (the case of GA); if they exhibit a certain amount of functional complexity, they are extremely difficult to reproduce with modification (the case of GP). In GEP, the individuals are encoded as linear strings of fixed length (chromosomes) which are afterwards expressed as nonlinear entities of different sizes and shapes (the phenotype). It not only is manipulated genetically in an easy way but also exhibits a certain amount of functional complexity. Thus, GEP overcomes the limitation of representation of GA and makes up the defect of complex genetic operators of GP. Those excellent properties of GEP have attracted wide attention of scholars around the world and have aroused their research interests. In recent years, GEP has rapidly become a powerful tool of automatic programming. And it has been widely applied in symbolic regression (Peng et al. 2005; Xin et al. 2006), time series prediction (Lopes et al. 2004), data mining (Chi et al. 2003; Bin et al. 2005; Karakasis and Stafylopatis 2008) and many other fields (Shi et al. 2010; Canakci et al. 2009; Teodorescu 2004; Teodorescu and Sherwood 2008).

Up to now, theoretical results of EA on computational time complexity are comparatively abundant. The representative works are as follows: for specific problems (e.g. Rudolph 1996; Garnier et al. 1999) and simple single-individual EA, such as the (1 + 1) EA (e.g. Droste et al. 2002; Wegener and Witt 2005). It is to be noted that He and Yao have done a series of works about the time complexity for several kinds of EAs and different problems (He and Yao 2002, 2003, 2004; Lehre and Yao 2012). Further, some more profound results about the computation time of EAs by using the drift analysis and other tools such as Dynkin's formula in stochastic process theory were given Ding and Yu (2005). A survey of the results obtained in the past decade along the time complexity of EAs was presented Oliveto et al. (2007). After that, a simplified drift theorem was introduced Happ et al. (2008). A further simplification of the drift theorem which is particularly suited for the analysis of EAs was presented Oliveto and Witt (2008). By exploiting the relationship between the convergence rates and the expected first hitting times, the upper bounds of the first hitting times of evolutionary algorithms based on He and Kang works (He and Kang 1999) were given Yu and Zhou (2008). A general idea for analyzing population-based EAs on unimodal problems using Markov chain theory and Chernoff bounds was given Chen et al. (2009). However, how to make the upper or lower bounds of the expected first hitting times associate with the parameters of EAs is still unresolved.

The individual of GEP has linear genotype and non-linear phenotype. Simultaneously, GEP not only contains all the genetic operators of traditional EA but also introduces some new genetic operators. All above characters of GEP bring some challenges to the time complexity analysis. Now, in the case of GEP, most studies of GEP focus on the realizations, improvements and applications of the algorithm. There have been few studies involved in theoretical analysis of GEP, especially the time complexity. To the best of our knowledge, the convergence of GEP under the smallest residual sum of squares was studied Yuan et al. (2004, 2006). In view of above results, we Du et al. (2009, 2010) studied the convergence and convergence rate of ME-GEP. We not only proved that ME-GEP algorithm converged to the global optimal solution in probability but also obtained the upper bound of the convergence rates of GEP. The results were that the upper bound was dependent on the second largest eigenvalue of the transition matrix of the Markov chain by the method of Jordan Blocks of matrix and the second largest eigenvalue of the transition matrix was determined by the parameters of ME-GEP algorithm.

This paper concerns theoretical analysis of the average time complexity of ME-GEP on optimization problems with a unique global optimum. A general and easy-to-apply approach to estimate the average time complexity for ME-GEP algorithm is presented. By using the proper-

ties of Markov chain and artificial fitness level technique, the upper and lower bounds of the average time complexity of ME-GEP algorithm are obtained. More importantly, we get the relation between the upper bound and the parameters of ME-GEP algorithm. These results obtained in this paper are the general theoretical results. Moreover, it has been shown that the analytical methods and techniques adopted in this paper have broad suitability and useful reference for the researchers in the theory of GEP. The rest of this paper is organized as follows: Sect. 2 introduces GEP and ME-GEP algorithm. In Sect. 3, the characteristic of nonnegative matrix and artificial fitness level technique are introduced. In addition, starting with the Markov chain model associated with ME-GEP algorithm, some fundamental properties of transition matrix of ME-GEP algorithm are proved. Section 4 gives the upper and lower bound of average time complexity of ME-GEP algorithm and discusses the relations between the upper bound and the parameters of ME-GEP. As an application of the theoretical results acquired in this paper, in Sect. 5, the upper bounds of average time complexity of ME-GEP for function modeling and clustering problem are also explored. And the results in Sect. 4 are verified. Finally, some short comments on the results obtained in this paper and future work are summarized.

2 Background

2.1 GEP algorithm

In the following, we will first introduce the genotype, phenotype and genetic operators in GEP.

The genotype is composed of head and tail. Let x be a genotype and set $x = \{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_l\}$ where

$$\hat{x}_i \in \begin{cases} T \vee F, & 1 \leq i \leq h \\ T, & h < i \leq l \end{cases}.$$

Here, l represents the length of genotype. F represents function set, which is composed of operators, the primary functions or user-defined functions. T represents terminal set, which is composed of constants and variables. Both F and T are finite. The elements in head are from F or T , and the elements in tail are from T . Variables h and tal denote the length of head and tail, respectively. For any given h , tal is evaluated by formula $tal = h \times (n_{\max} - 1) + 1$, $l = h + tal$, where $n_{\max} = \max(\text{paraNum}(f)_{f \in F})$, $\text{paraNum}(f)$ is the number of parameters of function.

For a given genotype, the phenotype is inferred as follows by Karva Language (Ferreira 2001), which reads genotype from the start character depending on the number of arguments of each element (functions may have a different number of arguments, whereas terminals have an arity of zero). And the representation (algebraic expression) is

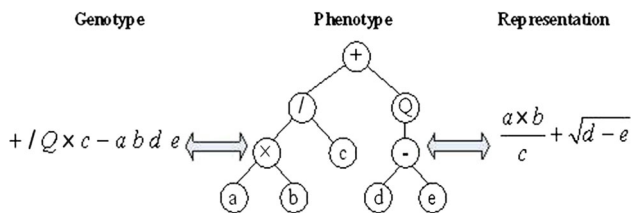


Fig. 1 Transformation relation between genotype, phenotype and representation

inferred from the phenotype according to the method of inorder traversal. The transformation relation between genotype, phenotype and representation is shown as Fig. 1.

There are three kinds of genetic operators in basic GEP (Ferreira 2001), including mutation, transposition and recombination. Each operation occurs with an active probability. Next, we show these operators adopted in GEP.

Mutation is allowed to occur anywhere in the chromosome and each bit flips with a probability independently at time t . In the head of the gene, a function can be flipped into another function or terminal. In the tail, a terminal can only be flipped into another terminal.

In GEP, there are two recombination operators: one-point and two-point. Similar to crossover operators in GA, the two operations act the same way to one-point and two-point crossover operators. In all cases of recombination, both offspring are conserved into successive evolution.

IS transposition Any sequence in the chromosome can become an IS element. The chosen sequence copies itself to any position in the head of a chromosome except the start position. A sequence with the same length is excluded from the end of the head to maintain the structural organization of the gene. For example, there is a parent chromosome whose head length is 10 and tail length is 11.

Q** + abbb* a a **bba** aaabbab

Suppose that the sequence “bba” becomes an IS element, which is highlighted, and its target position is 7, as the result of IS transposition, the last three elements “b*a” of the head are deleted. The offspring is as follows:

Q** + ab **bba** b abbaaaabbab

2.2 ME-GEP algorithm

Although the advantages of a system like GEP are clear from nature, the most important should be emphasized. First, the chromosomes (genotype) are simple entities: linear, compact, relatively small, easy to be manipulated genetically (mutate, recombine, transpose, etc.). Second, Karva language was created to read and express the information encoded in the chromosome (genotype). It always guarantees that a genotype can be converted into a valid phenotype, which

allows the implementation of a very powerful set of unconstrained genetic operators acting on genetic space. Clearly, it is that the unconstrained genetic operators highly improving the search capability, also making GEP algorithm difficult to converge. ME-GEP is a kind of GEP algorithm with elitist strategy. Thus, ME-GEP not only has efficient search capability but also ensures the convergence proved by our previous work Du et al. (2010).

ME-GEP algorithm with population size $n(n \geq 1)$ can be described as follows:

Step 1 Initialization: generate randomly an initial subpopulation of n individuals, denote it by $X(0) = (x_1(0), \dots, x_n(0))$, where $x_i(0) \in H, i = 1, \dots, n$ and let $t \leftarrow 0$. The optimal individual in generation 0, denoted by $x_0(0)$, is then defined as the individual among $x_1(0), \dots, x_n(0)$ with the biggest fitness. The population in generation 0 is defined as $\xi(0) = (x_0(0), x_1(0), \dots, x_n(0))$, where $x_i(0) \in H, i = 0, 1, \dots, n$.

Step 2 Recombination: generate a new (intermediate) subpopulation by recombination operation based on subpopulation $X(t)$ with probability $p_c(0 < p_c < 1)$ and denote it as $X_C(t)$.

Step 3 Transposition: generate a new (intermediate) subpopulation by transposition operation based on subpopulation $X_C(t)$ with probability $p_t(0 < p_t < 1)$ and denote it as $X_T(t)$.

Step 4 Mutation: generate a new (intermediate) subpopulation by mutation operation based on subpopulation $X_T(t)$ with probability $p_m(0 < p_m < 1)$ and denote it as $X_{Mu}(t)$.

Step 5 Select n individuals from subpopulation $X_{Mu}(t)$ and $X(t)$ according to certain select strategy, where the probability for each individual to be selected is >0 and then obtain the next subpopulation $X(t + 1)$.

Step 6 Find optimal individual $x_0(t + 1)$ from subpopulation $X(t + 1)$ and optimal individual $x_0(t)$.

Step 7 Let $t \leftarrow t + 1$; then the next population $\xi(t + 1) = (x_0(t + 1), x_1(t + 1), \dots, x_n(t + 1))$.

Step 8 If $f(x_0(t + 1)) = f_{\max}$, then stop; otherwise go to step2.

Remark In this paper, we always assume that all the genetic operations are independent of each other and fitness function is a single-valued function. The individual space is denoted by H and $|H| = N$

3 Markov chain model of ME-GEP

In reference Du et al. (2010), we have proved that both $\{X(t); t \geq 0\}$ and $\{\xi(t); t \geq 0\}$ are Homogeneous Markov chains with finite states space H^n and $V = H^{n+1}$, respectively; this is because subpopulation $X(t)$ only depends on subpopulation $X(t - 1)$, and optimal individual $x_0(t)$ depends on both subpopulation $X(t)$ and optimal individual $x_0(t - 1)$.

Hence, population $\xi(t)$ only depends on population $\xi(t - 1)$; further, all genetic operators adopted in different generation are fixed.

Since the limit behaviors of Markov chain are determined by its corresponding transition matrices, some definitions and symbols about matrix are stated as below.

3.1 The characteristic of nonnegative matrix

Definition 1 Rudolph (1994) A square matrix $A = [a_{ij}]_{m \times m}$ is said to be

- (1) nonnegative ($A \geq 0$), if $a_{ij} \geq 0$ for all $i, j \in \{1, 2, \dots, m\}$,
- (2) positive ($A > 0$), if $a_{ij} > 0$ for all $i, j \in \{1, 2, \dots, m\}$. A nonnegative matrix $A = [a_{ij}]_{m \times m}$ is said to be
- (3) primitive if there exists a positive integer k such that A^k is positive,
- (4) stochastic, if $\sum_{j=1}^m a_{ij} = 1$ for all $i \in \{1, 2, \dots, m\}$,
- (5) column-allowable, if it has at least one positive entry in each column.

Clearly, the product of stochastic matrices is again a stochastic matrix and that every positive matrix is also primitive.

Apart from providing a smaller Markov chain transition matrix, artificial fitness levels lead to simplified calculations. Next, the artificial fitness level technique will be introduced.

3.2 Artificial fitness levels

Artificial fitness levels (Oliveto et al. 2007) is a kind of partition technique of the search space based on the fitness function. The whole search space S as a set of $|S| = n$ different states is divided into $m < n$ states A_1, \dots, A_m , such that for all points $a \in A_i$ and $b \in A_j$ it happens that $f(a) < f(b)$ if $i < j$. Thus, A_m contains only optimal search points, in such a way the Markov chain can be constructed considering only m different states rather than n .

3.3 The transition matrices of ME-GEP

Note that the evolutionary process of ME-GEP can be illustrated by the following formula:

$$\begin{aligned} \xi(t) &= (x_0(t), X(t)) \xrightarrow{C} (x_0(t), X_C(t)) \xrightarrow{T} (x_0(t), X_T(t)) \\ &\xrightarrow{M} (x_0(t), X_{Mu}(t)) \xrightarrow{S} (x_0(t), X(t+1)) \\ &\xrightarrow{U} (x_0(t+1), X(t+1)) = \xi(t+1), \end{aligned} \tag{1}$$

where C, T, M, S and U represent recombination, transposition, mutation, select and update operation, respectively.

Let P^X and P^ξ be the transition matrices associated with homogeneous Markov chains $\{X(t); t \geq 0\}$ and $\{\xi(t); t \geq$

$0\}$. Then the properties and relations between P^X and P^ξ will be analyzed in the following. First, some conclusions of P^X can be stated as below.

Lemma 1 For any two chromosomes x and y , the mutation probability from x to y is

$$P \left\{ x \xrightarrow{M} y \right\} = (1 - p_m)^{h+tal-k_1-k_2} p_m^{k_1+k_2} \times \left(\frac{1}{|F| + |T|} \right)^{k_1} \left(\frac{1}{|T|} \right)^{k_2}, \tag{2}$$

where k_1 and k_2 are the numbers of different bits in gene head and tail of x and y separately.

Proof For $\forall x, y \in H$, suppose that there are k_1 bits in head and k_2 bits in tail of individual x to be mutated. Each bit in head has $|F| + |T|$ conditions and each bit in tail has $|T|$ conditions; hence the probability that individual x becomes individual y after mutation can be aggregated to

$$\begin{aligned} P \left\{ x \xrightarrow{M} y \right\} &= (1 - p_m)^{h+tal-k_1-k_2} \left(\frac{p_m}{|F| + |T|} \right)^{k_1} \left(\frac{p_m}{|T|} \right)^{k_2} \\ &= (1 - p_m)^{h+tal-k_1-k_2} p_m^{k_1+k_2} \left(\frac{1}{|F| + |T|} \right)^{k_1} \left(\frac{1}{|T|} \right)^{k_2} \end{aligned}$$

□

Lemma 2 Through IS transposition, population in last generation could generate any chromosome with the probability of

$$p_T = 2p_t / ((h - 2) \cdot l \cdot (l + 1))$$

Proof Suppose individual $X = x_1, x_2, \dots, x_l$ is selected as IS transpose with probability p_t . Let the beginning bit of IS string be $f_j, j = 1, 2, \dots, l$. There would be $l - j + 1$ possibilities of the IS string, and $h - 2$ possibilities of its insertion bit, except the first bit and the bit. Thus IS transposition with the beginning bit f_j could generate $(h - 2) \cdot (l - j + 1)$ kinds of chromosomes. When IS transposition occurs on X , there would be

$$\sum_{j=1}^l (h - 2) \cdot (l - j + 1) = (h - 2) \cdot l \cdot (l + 1) / 2$$

kinds of chromosomes. So any chromosome could be generated by IS transposition with the probability of $2p_t / ((h - 2) \cdot l \cdot (l + 1))$. □

Lemma 3 P^X is primitive, which can be decomposed in a natural way into a product of stochastic matrices $P^X = C \cdot T \cdot M \cdot S$, where $C = [c_{ij}]_{N^n \times N^n}$, $T = [t_{ij}]_{N^n \times N^n}$, $M = [m_{ij}]_{N^n \times N^n}$ and $S = [s_{ij}]_{N^n \times N^n}$ are the intermediate transition matrices induced by recombination, transposition, mutation and selection, respectively.

Proof Crossover, mutation and selection operators may be regarded as random total functions whose domain and range are H^n , i.e., each state of H^n is mapped probabilistically to another state since $0 < p_c, p_m < 1$. Therefore, C, M and S are stochastic matrices.

According to Lemma 1, we have

$$P \left\{ a \xrightarrow{M} b \right\} = (1 - p_m)^{h+tal-k_1-k_2} p_m^{k_1+k_2} \left(\frac{1}{|F| + |T|} \right)^{k_1} \left(\frac{1}{|T|} \right)^{k_2} \geq (1 - p_m)^{h+tal} p_m^{h+tal} \left(\frac{1}{|F| + |T|} \right)^h \left(\frac{1}{|T|} \right)^{tal},$$

where $k_1 \leq h, k_2 \leq tal$.

Denote the minimal mutation probability by $p_{\min m} = \min_{a,b \in H} p \left\{ a \xrightarrow{M} b \right\}$; then

$$p_{\min m} \geq (1 - p_m)^{h+tal} p_m^{h+tal} \left(\frac{1}{|F| + |T|} \right)^h \left(\frac{1}{|T|} \right)^{tal} > 0.$$

For any $i, j \in H^n$, we have $m_{ij} \geq p_{\min m}^n > 0$. Thus M is positive.

According to selection operation, the probability of each individual to be selected is greater than zero. Then for $\forall i \in H^n$, we have $s_{ii} > 0$. Thus S is column-allowable.

Let $A = [a_{ij}]_{N^n \times N^n} = T \cdot M, Q = [q_{ij}]_{N^n \times N^n} = C \cdot A, P^X = [p_{ij}^X]_{N^n \times N^n} = Q \cdot S$. Since C and T are stochastic matrices with at least one positive entry in each row, and $M > 0$, then, by matrix multiplication $a_{ij} = \sum_{k=1}^{N^n} t_{ik} m_{kj} > 0$ for all $i, j \in \{1, 2, \dots, N^n\}$, i.e. $A > 0$. Similarly, $q_{ij} = \sum_{k=1}^{N^n} c_{ik} a_{kj} > 0$, and hence $p_{ij}^X = \sum_{k=1}^{N^n} q_{ik} s_{kj} > 0$ for all $i, j \in \{1, 2, \dots, N^n\}$, because S is column-allowable. Thus P^X is positive. Since every positive matrix is primitive. \square

In order to analyze the properties of transition matrix P^ξ , we will use artificial fitness level technique to partition the state space $V = H^{n+1}$. First, set $H = \{s_1, s_2, \dots, s_N\}$ with $f(s_1) > f(s_2) > \dots > f(s_N)$; thus H is a totally ordered set, and the elements in H can be ordered by $s_1 > s_2 > \dots > s_N$. Let

$$V_i = \{(s_i, x_1, \dots, x_n); x_j \in H, j = 1, \dots, n\}, \forall i = 1, \dots, N. \tag{3}$$

It is obvious that $V_1 \cup \dots \cup V_N = H^{n+1}$ and V_1 is the optimal state set. Then the following facts can be reached immediately:

- (1) By step 2 to 5 of ME-GEP, we know that the optimal individual is not affected by the genetic operators, and the extended transition matrices on H^{n+1} for recombination C' , transposition T' , mutation M' and selection S' can be written as block diagonal matrices.

$$C' = \begin{pmatrix} C & & & \\ & C & & \\ & & \ddots & \\ & & & C \end{pmatrix} \quad T' = \begin{pmatrix} T & & & \\ & T & & \\ & & \ddots & \\ & & & T \end{pmatrix}$$

$$M' = \begin{pmatrix} M & & & \\ & M & & \\ & & \ddots & \\ & & & M \end{pmatrix} \quad S' = \begin{pmatrix} S & & & \\ & S & & \\ & & \ddots & \\ & & & S \end{pmatrix}$$

$$C'T'M'S' = \begin{pmatrix} CTMS & & & \\ & CTMS & & \\ & & \ddots & \\ & & & CTMS \end{pmatrix} = \begin{pmatrix} P^X & & & \\ & P^X & & \\ & & \ddots & \\ & & & P^X \end{pmatrix}$$

- (2) Update operation of step 6 is only to update the optimal individual. And transition matrix U can be written by matrices $U_{rk} (1 \leq r, k \leq N)$ as follows:

$$U = \begin{pmatrix} U_{11} & U_{12} & \dots & U_{1N} \\ U_{21} & U_{22} & \dots & U_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ U_{N1} & U_{N2} & \dots & U_{NN} \end{pmatrix},$$

where $U_{ij} (1 \leq i, j \leq N)$ are $N \times N$ matrices.

- (3) Transition matrix $P^\xi = (C'T'M'S') \cdot U$.

Some useful properties of P^ξ can be obtained by analyzing matrix U . Let $n_k = |\{(x_1, \dots, x_n) : \max_{1 \leq j \leq n} f(x_j) = f(s_k)\}|$, that is, n_k is the number of those subpopulations with best individual s_k . Then we have

Lemma 4 For each subspace $V_i = \{(s_i, x_1, \dots, x_n) : x_j \in H, j = 1, \dots, n\}, \forall i = 1, \dots, N$ in H^{n+1} , we have $\sum_{k=1}^N n_k = N^n$ and $n_k = (N - k + 1)^n - (N - k)^n$, where $k = 1, \dots, N$.

Proof According to the definition of V_i , we can know immediately that the size of each V_i is equal to N^n ; thus it is clear that

$$\sum_{k=1}^N n_k = N^n.$$

In subspaces V_i , if subpopulation (x_1, \dots, x_n) satisfies $\max_{1 \leq i \leq n} f(x_i) < f(s_k)$, then subpopulation (x_1, \dots, x_n) can-

not contain individuals s_1, s_2, \dots, s_{k-1} . Thus, there are $(N - k + 1)^n$ states which do not contain individuals s_1, s_2, \dots, s_{k-1} . Likewise, there exist $(N - k)^n$ states that do not contain individuals s_1, s_2, \dots, s_k . So we have

$$n_k = (N - k + 1)^n - (N - k)^n.$$

□

Lemma 5 *U Matrix has the following properties:*

- (1) *In each row of matrix U, there is exactly one entry equaling to 1, all others are 0, and U_{11} is a unit matrix.*
- (2) $U_{rk} = 0 (r < k, 1 \leq r < N)$.
- (3) $U_{rk} (2 \leq r \leq N, 1 \leq k \leq r - 1)$ are diagonal matrices such that n_k elements are 1 in primary diagonal line and the remaining elements are 0.
- (4) $U_{rr} (2 \leq r \leq N)$ are diagonal matrices such that $\sum_{i=r}^N n_i$ elements are 1 in primary diagonal line and the remaining elements are 0.

Proof (1) The upgrade operation is represented by an upgrade matrix U . Let $b(i)$ be the best individual of the subpopulation (x_1, \dots, x_n) at any state i and write $b(i) = \arg \max\{f(x_i): i = 1, \dots, n\}$. Then $u_{ij} = 1$ if $f(x_0) < f(b(i))$ with $j \triangleq (b(i), x_1, \dots, x_n) \in V, i = (x_0, x_1, \dots, x_n) \in V$; otherwise $u_{ij} = 0 (j \neq i)$. That is, there is exactly one entry equal to 1, and all others are 0 in each row of matrix U . Notice that subpopulation (x_1, \dots, x_n) is unchanged by upgrade operation. In other words, a state either gets upgraded or remains unaltered. So there is only one in primary diagonal line for each sub-matrix $U_{ij} (1 \leq i, j \leq N)$. In particular, the optimal individual of subspace V_1 is the global optimal solution. That is $f(x^*) \geq f(b(i))$. Then we have $u_{ii} = 1, u_{ij} = 0 (j \neq i)$ according to (1). So U_{11} is a unit matrix.

- (2) The state of population remains unaltered when $r < k$, that is $s_r > s_k$. Then we have $u_{ij} = 0 (j \neq i)$. Thus $U_{rk} = 0 (r < k, 1 \leq r < N)$.
- (3) There is $s_k > s_r$ when $1 \leq k \leq r - 1$. The total state number with the best individual s_k in V_k is n_k according to Lemma 4. In addition, for any diagonal matrix $U_{rk} (2 \leq r \leq N, 2 \leq k \leq r)$, there are n_k elements whose value equal to 1, while others equal to 0 according to (1).
- (4) There is $s_k = s_r$ when $2 \leq k = r \leq N$. The total state number with the best individual s_k in V_r is n_k . The best individual can be s_r, s_{r+1}, \dots, s_N . Then $U_{rr} (2 \leq r \leq N)$ are diagonal matrices such that $n_r + n_{r+1} + \dots + n_N$ elements are 1 in primary diagonal line and the remaining elements are 0. □

Based on the above, the transition matrix P^ξ becomes

$$P^\xi = \begin{pmatrix} P^X & & & \\ & P^X & & \\ & & \ddots & \\ & & & P^X \end{pmatrix} \begin{pmatrix} U_{11} & & & \\ U_{21} & U_{22} & & \\ \vdots & \vdots & \ddots & \\ U_{N1} & U_{N2} & \cdots & U_{NN} \end{pmatrix} = \begin{pmatrix} P^X U_{11} & & & \\ P^X U_{21} & P^X U_{22} & & \\ \vdots & \vdots & \ddots & \\ P^X U_{N1} & P^X U_{N2} & \cdots & P^X U_{NN} \end{pmatrix} = \begin{pmatrix} Y & 0 \\ R & T \end{pmatrix} \tag{4}$$

where

$$R = \begin{pmatrix} P^X U_{21} \\ P^X U_{31} \\ \vdots \\ P^X U_{N1} \end{pmatrix}, T = \begin{pmatrix} P^X U_{22} & & \\ \vdots & \ddots & \\ P^X U_{N2} & \cdots & P^X U_{NN} \end{pmatrix}, R \neq 0, T \neq 0, Y = P^X U_{11} = P^X > 0.$$

According to the properties of absorbing Markov chain (Iosifescu 1980), we have

$$P^\xi = \begin{matrix} & V_1 & V \setminus V_1 \\ V_1 & \begin{pmatrix} Y & 0 \\ R & T \end{pmatrix} \\ V \setminus V_1 & \end{matrix} \tag{5}$$

where Y concerns the transition from one absorbing state to the other, the region 0 consists entirely of 0's, the matrix R concerns the transition from transient state to absorbing one, and the matrix T concerns the transition from transient to transient states. The matrix Y corresponds to the fact that once the Markov chain reaches an absorbing state, it will never leave that absorbing state.

Let $P^\xi = (p_{ij}^\xi; i, j \in N^{n+1})$, $P^X = (p_{ij}^X; i, j \in N^n)$ where p_{ij}^ξ, p_{ij}^X are the transition probability from state i to state j . According to Lemma 5, we know that matrices $U_{rk} (2 \leq r \leq N, 1 \leq k \leq r - 1)$ and matrices $U_{rr} (2 \leq r \leq N)$ can be written as follows:

$$U_{rk} = \begin{pmatrix} \overbrace{0}^{n_1} & & & & & & \\ & \ddots & & & & & \\ & & 0 & & & & \\ & & & \overbrace{1}^{n_k} & & & \\ & & & & \ddots & & \\ & & & & & 1 & \\ & & & & & & 0 \\ & & & & & & & \ddots \\ & & & & & & & & 0 \end{pmatrix}$$

expected first hitting time (expected FHT) is the average time that ME-GEP requires to find the optimal solution, which implies the average computational time complexity of ME-GEP.

The first hitting time on V_1 can be defined by

$$\tau(V_1) = \min \{k; k \geq 0, \xi(k) \in V_1\} \tag{9}$$

Let E denote the expectation operators. The first hitting time $\tau(V_1)$ is a random variable, and what we are interested is its expectation value $E[\tau(V_1)]$.

Hence, we can obtain the following conclusion of EFHT of ME-GEP:

Theorem 1 Let $\beta = \min\{p_{ij}^X; i, j \in N^n\}$; then the EFHT of ME-GEP is as follows:

$$P\{\xi(0) \notin V_1\} \frac{1}{1 - (N^n - n_1)\beta} \leq E[\tau(V_1)] < \frac{1}{(N^n - (N - 1)^n)\beta}$$

Proof Note that $P\{\tau(V_1) \geq 0\} = 1, P\{\tau(V_1) \geq 1\} = P\{\xi(0) \notin V_1\}$.

By Markov property of $\{\xi(t); t \geq 0\}$, for any $l \geq 2$, we have

$$\begin{aligned} P\{\tau(V_1) \geq l\} &= P\{\xi(0) \notin V_1, \xi(1) \notin V_1, \dots, \xi(l-1) \notin V_1\} \\ &= P\{\xi(0) \notin V_1\} \prod_{k=0}^{l-2} P\{\xi(k+1) \notin V_1 | \xi(k) \notin V_1\} \\ &= P\{\xi(0) \notin V_1\} \prod_{k=0}^{l-2} \frac{P\{\xi(k+1) \notin V_1, \xi(k) \notin V_1\}}{P\{\xi(k) \notin V_1\}} \\ &= P\{\xi(0) \notin V_1\} \prod_{k=0}^{l-2} \frac{\sum_{i,j \notin V_1} P\{\xi(k+1) = i | \xi(k) = j\} P\{\xi(k) = i\}}{\sum_{s \notin V_1} P\{\xi(k) = s\}} \\ &= P\{\xi(0) \notin V_1\} \prod_{k=0}^{l-2} \frac{\sum_{i \notin V_1} [P\{\xi(k) = i\} \cdot \sum_{j \notin V_1} p_{ij}^{\xi}]}{\sum_{s \notin V_1} P\{\xi(k) = s\}}. \end{aligned}$$

Then for any $i \notin V_1$, according to Eqs. (6)–(8), we have

$$(N^n - n_1)\beta \leq \sum_{j \notin V_1} p_{ij}^{\xi} \leq 1 - n_1\beta.$$

So that for any $l \geq 2$, we have

$$\begin{aligned} P\{\tau(V_1) \geq l\} &\geq P\{\xi(0) \notin V_1\} \prod_{k=0}^{l-2} (N^n - n_1)\beta \\ &\geq P\{\xi(0) \notin V_1\} \prod_{k=0}^{l-2} (N^n - n_1)\beta \\ &\geq P\{\xi(0) \notin V_1\} \{(N^n - n_1)\beta\}^{l-1} \\ P\{\tau(V_1) \geq l\} &\leq P\{\xi(0) \notin V_1\} \prod_{k=0}^{l-2} (1 - n_1\beta) \end{aligned}$$

$$\begin{aligned} &\leq P\{\xi(0) \notin V_1\} \prod_{k=0}^{l-2} (1 - (N^n - (N-1)^n)\beta) \\ &\leq P\{\xi(0) \notin V_1\} \{1 - (N^n - (N-1)^n)\beta\}^{l-1} \end{aligned}$$

Note that $0 < n_1\beta < 1, 0 < (N^n - n_1)\beta < 1$. Hence

$$\begin{aligned} E[\tau(V_1)] &= \sum_{l \geq 1} P\{\tau(V_1) \geq l\} = P\{\xi(0) \notin V_1\} \\ &+ \sum_{l \geq 2} P\{\tau(V_1) \geq l\} \end{aligned}$$

So the upper bound for $E[\tau(V_1)]$ is

$$\begin{aligned} E[\tau(V_1)] &= \sum_{l \geq 1} P\{\tau(V_1) \geq l\} \\ &\leq P\{\xi(0) \notin V_1\} + P\{\xi(0) \notin V_1\} \sum_{l=2}^{+\infty} \\ &\quad \{1 - (N^n - (N - 1)^n)\beta\}^{l-1} \\ &\leq P\{\xi(0) \notin V_1\} \left\{ 1 + \sum_{l=2}^{+\infty} \right. \\ &\quad \left. \times \{1 - (N^n - (N - 1)^n)\beta\}^{l-1} \right\} \end{aligned}$$

The lower bound for $E[\tau(V_1)]$ is

$$\begin{aligned} E[\tau(V_1)] &= \sum_{l \geq 1} P\{\tau(V_1) \geq l\} \\ &\geq P\{\xi(0) \notin V_1\} \\ &+ P\{\xi(0) \notin V_1\} \sum_{l=2}^{+\infty} \{(N^n - n_1)\beta\}^{l-1} \\ &\geq P\{\xi(0) \notin V_1\} \{1 + \sum_{l=2}^{+\infty} \{(N^n - n_1)\beta\}^{l-1}\} \\ &\geq P\{\xi(0) \notin V_1\} \frac{1}{1 - (N^n - n_1)\beta} \end{aligned}$$

That is,

$$\begin{aligned} P\{\xi(0) \notin V_1\} \frac{1}{1 - (N^n - n_1)\beta} &\leq E[\tau(V_1)] \\ &\leq P\{\xi(0) \notin V_1\} \frac{1}{(N^n - (N - 1)^n)\beta} \\ P\{\xi(0) \notin V_1\} \frac{1}{1 - (N^n - n_1)\beta} &\leq E[\tau(V_1)] < \frac{1}{(N^n - (N - 1)^n)\beta} \end{aligned}$$

□

In order to make the EFHT associated with parameters of ME-GEP directly, Corollary 1 is given.

Corollary 1 Let $p_{\min t} = \min_{\forall a, b \in H} p \{a \xrightarrow{T} b\}$, $p_{\min s} = \min_{\forall a \in H} p \{a \xrightarrow{S} a\}$, $p_{\min m} = \min_{\forall a, b \in H} p \{a \xrightarrow{M} b\}$; then we have

$$E[\tau(V_1)] < \frac{1}{(N^n - (N - 1)^n) \cdot P_{\min t}^n \cdot P_{\min m}^n \cdot P_{\min s}^n}$$

Proof For any $i, \delta, \eta, k, j \in H^n$, we have

$$\sum_{i, \delta \in H^n} P_{i\delta}^C = 1, P_{\delta\eta}^T \geq P_{\min t}^n, P_{\eta k}^M \geq P_{\min m}^n, P_{j j}^S \geq P_{\min s}^n.$$

Then we have

$$\begin{aligned} p_{ij}^X &= P(X(t + 1) = j | X(t) = i) \\ &= \sum_{\eta, \delta, k \in H^n} P_{i\delta}^C \cdot P_{\delta\eta}^T \cdot P_{\eta k}^M \cdot P_{kj}^S \\ &\geq \sum_{k=j} \sum_{\delta, \eta \in H^n} P_{i\delta}^C \cdot P_{\delta\eta}^T \cdot P_{\eta k}^M \cdot P_{kj}^S \\ &\geq P_{j j}^S \cdot P_{\min t}^n \cdot P_{\min m}^n \cdot \sum_{\delta \in H^n} P_{i\delta}^C \\ &\geq P_{\min s}^n \cdot P_{\min t}^n \cdot P_{\min m}^n \end{aligned}$$

So we get

$$\beta \geq P_{\min t}^n \cdot P_{\min m}^n \cdot P_{\min s}^n$$

By Theorem 1, we have

$$E[\tau(V_1)] < \frac{1}{(N^n - (N - 1)^n) \cdot P_{\min t}^n \cdot P_{\min m}^n \cdot P_{\min s}^n}$$

□

Theorem 1 estimates the upper and lower bounds of EFHT of ME-GEP algorithm. Further, Corollary 1 makes the upper bound associate with parameters of ME-GEP algorithm directly. Thus, Theorem 1 and Corollary 1 can be used to direct algorithm design of ME-GEP.

5 Case study

In order to justify our theoretical results, the EFHT of ME-GEP for solving function modeling and clustering problems are analyzed below.

To guarantee the validity of results, each execution of each algorithm on each parameter set was independently repeated 50 times.

5.1 Function modeling problem

5.1.1 The description of function modeling problem

Function modeling problem can be described as follows: given a group of datum $(x_1^i, x_2^i, \dots, x_m^i, y^i), i = 1, 2, \dots, k$,

find a function $\varphi(x_1, x_2, \dots, x_m)$ in a certain function set $\Phi(\bar{x}), \bar{x} = (x_1, \dots, x_m)$, which satisfies

$$\begin{aligned} & \text{Min}_{g(\bar{x}) \in \Phi(\bar{x})} \sum_{i=1}^k \left(g(x_1^i, x_2^i, \dots, x_m^i) - y^i \right)^2 \\ &= \sum_{i=1}^k \left(\varphi(x_1^i, x_2^i, \dots, x_m^i) - y^i \right)^2, \end{aligned}$$

where $g(\bar{x})$ is an arbitrary function in set $\Phi(\bar{x})$.

Next, the function modeling about the amount of gas emitted from coalfaces (Xin et al. 2006) is used as an example to be described as follows. The amount of gas emitted from coalfaces in coal mine is the main unsafe factor during the production of coal mine. However, this amount is hard to predict because it is affected by many factors and has complicated nonlinear relations with these factors. In Table 1, the sampling cases are given. There are six main factors as input variables: the depth of coalface x_1 , the height of coalface x_2 , the amount of coalfaces x_3 , the layer interval between working coalface and neighboring coalface x_4 , average daily advance of working face x_5 and average daily output x_6 . The amount of gas emitted from coalfaces y is the output variable. By ME-GEP, the function can be found to model accurately the nonlinear relations between input variables and y .

5.1.2 The EFHT of ME-GEP for solving the amount of gas emitted from coalfaces

The parameters of ME-GEP algorithm are designed as follows: Let $n = 80, h = 15$ then $tal = 16, l = 31, F = \{+, -, *, /, \text{sqrt}, \text{exp}, \text{sin}, \text{cos}, \text{ln}\}, T = \{R, x\}$ where R is a set of random constants, and we assume that R has 20 elements. The genetic operators adopted are one-point recombination, IS transposition, selection and mutation operators. Especially, mutation operators include uniform mutation that flips independently each bit with probability p_m and one-point mutation that flips one bit with probability p_m randomly. Each individual is selected with same probability in selection strategy. Next, the EFHT of ME-GEP with two different mutation operators is analyzed below.

The fitness (Ferreira 2001) of ME-GEP is evaluated by Eq. (10).

$$f_i = \max \left(M - \sum_{j=1}^k \left(\left| \frac{C_{ij} - y_j}{y_j} \right| \cdot \text{factor} \right), 0 \right). \quad (10)$$

where k is the number of sampling, and M is the range of selection. For this example, let $M = 1,000, \text{factor} = M/k \approx 56, C_{ij}$ be the value returned by the individual i for sample case j , and y_j be the target value for sample case j .

Table 1 Sample cases of gas emitted from coalfaces

x_1	x_2	x_3	x_4	x_5	x_6	y
408	2.0	1.92	20	4.42	1,825	3.34
411	2.0	2.15	22	4.16	1,527	2.97
420	1.8	2.14	19	4.13	1,751	3.56
432	2.3	2.58	17	4.67	2,078	3.62
456	2.2	2.40	20	4.51	2,104	4.17
516	2.8	3.22	12	3.45	2,242	4.60
527	2.5	2.80	11	3.28	1,979	4.92
531	2.9	3.35	13	3.68	2,288	4.78
550	2.9	3.61	14	4.02	2,325	5.23
563	3.0	3.68	12	3.53	2,410	5.56
590	5.9	4.21	18	2.85	3,139	7.24
604	6.2	4.03	16	2.64	3,354	7.80
607	6.1	4.34	17	2.77	3,087	7.68
634	6.5	4.80	15	2.92	3,620	8.51
640	6.3	4.67	15	2.75	3,412	7.95
450	2.2	2.43	16	4.32	1,996	4.06
544	2.7	3.16	13	3.81	2,207	4.92
629	6.4	4.62	19	2.80	3,456	8.04

Applying Theorem 1 and Corollary 1, we have

$$E[\tau(V_1)] < \frac{1}{(N^n - (N - 1)^n) \cdot p_{\min t}^n \cdot p_{\min m}^n \cdot p_{\min s}^n}$$

Let $p_{\min m1}$ and $p_{\min m2}$ be the minimum probabilities when an individual x becomes individual y after uniform mutation and one-point mutation, respectively. Next, we will estimate N , $p_{\min t}$, $p_{\min m1}$, $p_{\min m2}$ and $p_{\min s}$.

(1) The size of individual space N .

Denote the tree height of phenotype by variable ht . Because the individual length is 31, we have $ht \leq 5$. It follows that

$$N = |H| \leq 9^{(2^0+2^1+2^2+2^3)} \times (|R| + 6)^{2^4} = 9^{15} \times 26^{16}$$

(2) $p_{\min m1}$

According to Lemma 1, we have

$$P\{x \xrightarrow{M} y\} = (1 - p_m)^l \left(\frac{p_m}{1 - p_m}\right)^{k_1+k_2} \times \left(\frac{1}{|F| + |T|}\right)^{k_1} \left(\frac{1}{|T|}\right)^{k_2}$$

If $1 - p_m > p_m$, that is $p_m < \frac{1}{2}$, it follows that

$$P\{x \xrightarrow{M} y\} > (1 - p_m)^l \left(\frac{p_m}{1 - p_m}\right)^l \left(\frac{1}{|T| + |F|}\right)^l \left(\frac{1}{|T|}\right)^l = \left(\frac{p_m}{(|T| + |F|) \cdot (|T|)}\right)^l = \left(\frac{p_m}{910}\right)^{31}$$

It is easy to find that the minimum probability that any individual x mutates to another individual y is $\left(\frac{p_m}{910}\right)^{31}$ where $p_m \in (0, 0.5)$. Thus, we have $p_{\min m1} \geq \left(\frac{p_m}{910}\right)^{31}$.

(3) $p_{\min m2}$

There are two situations for one-point mutation. One is that one bit flips in head. The other is that one bit flips in tail.

For the first situation according to Lemma 1, we have

$$P\{x \xrightarrow{M} y\} = (1 - p_m)^{l-1} \frac{p_m}{|F| + |T|} = (1 - p_m)^l \frac{p_m}{1 - p_m} \frac{1}{|F| + |T|}$$

If $1 - p_m > p_m$, that is $p_m < \frac{1}{2}$, it follows that

$$P\{x \xrightarrow{M} y\} > (1 - p_m)^l \left(\frac{p_m}{1 - p_m}\right)^l \left(\frac{1}{|T| + |F|}\right) = \frac{p_m^l}{|T| + |F|} = \frac{p_m^{31}}{35}$$

For the second situation according to Lemma 1, we have

$$P\{x \xrightarrow{M} y\} = (1 - p_m)^{l-1} \frac{p_m}{|T|} = (1 - p_m)^l \frac{p_m}{1 - p_m} \frac{1}{|T|}$$

If $1 - p_m > p_m$, that is $p_m < \frac{1}{2}$, it follows that

$$P \left\{ x \xrightarrow{M} y \right\} > (1 - p_m)^l \left(\frac{p_m}{1 - p_m} \right)^l \left(\frac{1}{|T|} \right) = \frac{p_m^l}{|T|} = \frac{p_m^{31}}{26}$$

From above analysis, the minimum probability that any individual x mutates to another individual y is $\frac{p_m^{31}}{35}$ where $p_m \in (0, 0.5)$. Thus, we have $p_{\min m2} \geq \frac{p_m^{31}}{35}$.

(4) $p_{\min t}$

According to Lemma 2, for $\forall x, y \in H$, we have

$$P \left\{ x \xrightarrow{T} y \right\} = \frac{2}{(h - 2) \cdot l \cdot (l + 1)} \cdot p_t > \frac{2}{(l - 2) \cdot l \cdot (l + 1)} \cdot p_t > \frac{2}{l^2(l + 1)} \cdot p_t$$

Thus, we have

$$p_{\min t} = \frac{2}{l^2(l + 1)} \cdot p_t = \frac{1}{15376} \cdot p_t$$

(5) $p_{\min s}$

According to selection operator, obviously, we have

$$p_{\min s} = \frac{1}{2n}$$

According to Corollary 1, the EFHT of ME-GEP with uniform mutation is as follows:

$$E[\tau(V_1)] < \frac{1}{(N^n - (N - 1)^n) \cdot p_{\min t}^n \cdot p_{\min m2}^n \cdot p_{\min s}^n} < \frac{1}{2^{80} \cdot 80^{80} \cdot 910^{2480} \cdot 15376^{80}} < \frac{1}{((9^{15} \times 26^{16})^{80} - (9^{15} \times 26^{16} - 1)^{80}) \cdot p_m^{2480} \cdot p_t^{80}} \tag{11}$$

and the EFHT of ME-GEP with one-point mutation is as follows:

$$E[\tau(V_1)] < \frac{1}{(N^n - (N - 1)^n) \cdot p_{\min t}^n \cdot p_{\min m2}^n \cdot p_{\min s}^n} < \frac{1}{2^{80} \cdot 80^{80} \cdot 35^{80} \cdot 15376^{80}} < \frac{1}{((9^{15} \times 26^{16})^{80} - (9^{15} \times 26^{16} - 1)^{80}) \cdot p_m^{2480} \cdot p_t^{80}} \tag{12}$$

5.1.3 Numerical experiment

From above formulas (11) and (12), the EFHT of ME-GEP can be decreased by increasing mutation probability and transposition probability.

The following experiments to observe the actual average first hitting time of ME-GEP, especially considering different mutation operators, are done:

- (1) When IS transposition probability is 0.1, uniform and one-point mutation probabilities are changed;
- (2) When uniform and one-point mutate probabilities are 0.04, the IS transposition probability is changed.

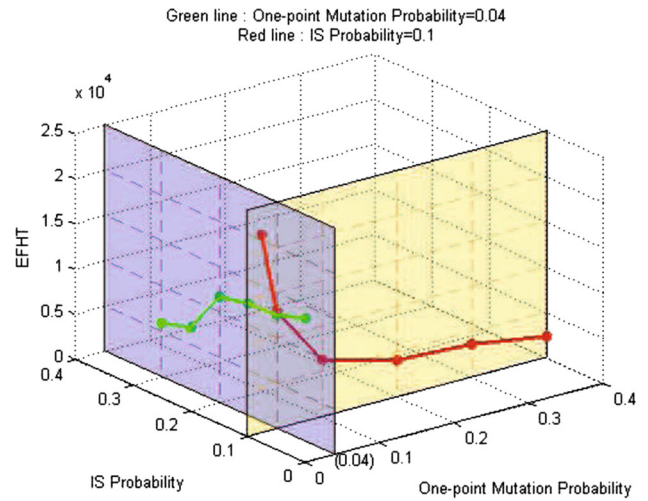


Fig. 2 The EFHT of ME-GEP for solving gas emitted from coalfaces modeling with different one-point mutation and IS transposition probabilities

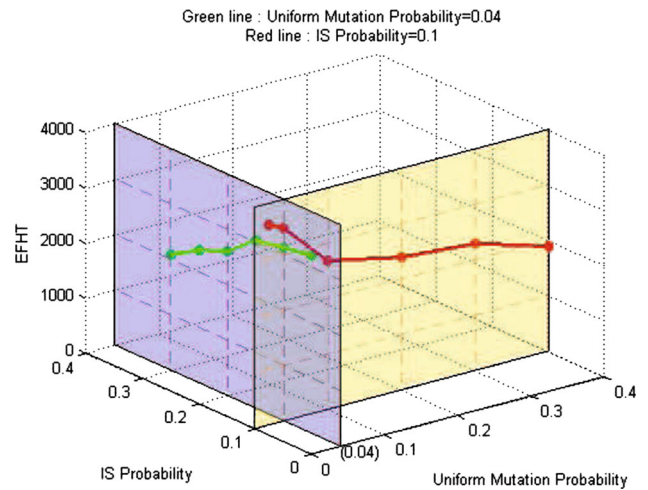


Fig. 3 The EFHT of ME-GEP for solving gas emitted from coalfaces modeling with different uniform mutation and IS transposition probabilities

The results obtained are shown in Figs. 2 and 3. The EFHT of ME-GEP decreased with the increasing of one-point and uniform mutation probabilities when IS transposition probability was 0.1. And the EFHT of ME-GEP decreased with the increasing of IS transposition probabilities when one-point and uniform mutation probabilities were 0.04. In this example, the experimental results conformed to formulas (11) and (12).

5.2 Clustering problem

5.2.1 The description of clustering problem

Clustering problem can be described as follows: given a set of observations (P_1, \dots, P_m) , where each observation is a

d -dimensional real vector, partition the m observations into $q(\leq m)$ sets $C = \{C_1, \dots, C_q\}$, which satisfies

Minimize $\sum_{i=1}^q \sum_{P_j \in C_i} \|P_j - \mu_i\|^2$, where μ_i is the mean of points in C_j .

In ME-GEP for solving clustering problems, two special clustering operators, namely ‘ \cup ’ and ‘ \cap ’, are introduced as binary operators used to define aggregation and segmentation of clusters respectively.

Let O_1 and O_2 be the centers of cluster C_1 and C_2 .

- (1) The aggregation operator ‘ \cap ’ is defined as $O_1 \cap O_2 = \text{centroid}(O_1, O_2)$, where the return of function centroid is the center vector of O_1 and O_2 . And C_1 and C_2 are aggregated into a cluster C with the center $O = \text{centroid}(O_1, O_2)$.
- (2) The segmentation operator ‘ \cup ’ is defined as $O_1 \cup O_2 = \{O_1, O_2\}$. C_1 and C_2 are two distinct clusters, and they still keep the segmentation, whose centers of the clusters are O_1 and O_2 , respectively.

The chromosome is also composed of head and tail. The function set is $F = \{\cup, \cap\}$, and the terminals set is $T = \{x_1, x_2, \dots, x_i\}$, where $i \in (1, n)$. Here, the symbol x_i represents the center of a cluster in the data set.

Based on above definitions, the clusters are easily defined in the chromosome of ME-GEP framework.

The main procedure of ME-GEP for solving clustering problem is described as follows. The detailed content can be got in reference (Chen et al. 2007):

- Step 1 Initialize the population;
- Step 2 Select and aggregate cluster centers by coding and decoding expression trees;
- Step 3 Calculate the Euclidean distances between each data object and cluster centers and tag the corresponding data object;
- Step 4 Update the cluster centers and compute the fitness of each individual;
- Step 5 Preserve the best individual to the next generation;
- Step 6 Select individuals with same probability;
- Step 7 Mutate and recombination with a certain probability;
- Step 8 Generate a new population, and if the optimal solution has been found, then turn to step 9; otherwise, turn to step 2;
- Step 9 Select the best individual;
- Step 10 Auto-clustering algorithm of best individual;
- Step 11 Output the result of clusters.

The EFHT of ME-GEP (Chen et al. 2007) for solving two actual clustering examples of Iris and Yeast (benchmark databases) (Zhou et al. 2003) with different data dimensions and sizes is analyzed as follows:

To solve clustering problems in examples, the parameters of ME-GEP are designed as Table 2. The genetic opera-

Table 2 Parameters set of ME-GEP for clustering two benchmark databases

Database	Size	Data dimensions	The length of gene head (h)	Popsiz (n)
Iris	150	4	4	20
Yeast	1,484	8	15	30

tors adopted include one-point recombination, selection with equal probability, uniform and one-point mutation operators.

The fitness of ME-GEP algorithm for solving the clustering problem is evaluated by Eq. (13).

$$f = \frac{1}{\sum_{i=1}^q \sum_{P_j \in C_i} \|P_j - \mu_i\|^2}, \quad \mu_i = \frac{1}{n_i} \sum_{P_j \in C_i} P_j, \quad (13)$$

where q is the number of clusters, C_i is the i th cluster, P_j is the observation point of cluster C_i and μ_i is the center of i th cluster. The bigger the individual fitness, the more compact and independent the cluster, namely the corresponding individual is better.

5.2.2 The EFHT of ME-GEP for sloving clustering Iris database

The parameters of ME-GEP for clustering Iris database are as follows: Let $h = 4$; then $tal = 5, l = 9$. $T = \{x_1, x_2, x_3, x_4, x_5\}$.

Applying Theorem 1 and Corollary 1, we have

$$E[\tau(V_1)] < \frac{1}{(N^n - (N - 1)^n) \cdot p_{\min m}^n \cdot p_{\min s}^n}$$

Let $p_{\min m1}$ and $p_{\min m2}$ be the minimum probabilities when an individual x changes to individual y after uniform mutation and one-point mutation, respectively. Next, we will evaluate and estimate $N, p_{\min m1}, p_{\min m2}$ and $p_{\min s}$.

(1) The size of individual space N .

Denote the tree height of phenotype by variable ht . Because the individual length is 9, we have $ht \leq 4$. It follows that

$$N = |H| \leq 2^{(2^0+2^1+2^2)} \times 5^{2^3} = 2^7 \times 5^8$$

(2) $p_{\min m1}$

According to Lemma 1, we have

$$P \left\{ x \xrightarrow{M} y \right\} = (1 - p_m)^l \left(\frac{p_m}{1 - p_m} \right)^{k_1+k_2} \left(\frac{1}{|F|} \right)^{k_1} \left(\frac{1}{|T|} \right)^{k_2}$$

If $1 - p_m > p_m$, that is $p_m < \frac{1}{2}$, it follows that

$$\begin{aligned} P \left\{ x \xrightarrow{M} y \right\} &> (1 - p_m)^l \left(\frac{p_m}{1 - p_m} \right)^l \left(\frac{1}{|F|} \right)^l \left(\frac{1}{|T|} \right)^l \\ &= \left(\frac{p_m}{(|F|) \cdot (|T|)} \right)^l = \left(\frac{p_m}{10} \right)^9 \end{aligned}$$

From the above analysis, the minimum probability that any individual x mutates to another individual y is $(\frac{p_m}{10})^9$ where $p_m \in (0, 0.5)$. Thus, we have $p_{\min m1} \geq (\frac{p_m}{10})^9$.

(3) $p_{\min m2}$

There are two situations for one-point mutation. One is that one bit flips in head. The other is that one bit flips in tail.

For the first situation according to Lemma 1, we have

$$P\{x \xrightarrow{M} y\} = (1 - p_m)^{l-1} \frac{p_m}{|F|} = (1 - p_m)^l \frac{p_m}{1 - p_m} \frac{1}{|F|}$$

If $1 - p_m > p_m$, that is $p_m < \frac{1}{2}$, it follows that

$$\begin{aligned} P\{x \xrightarrow{M} y\} &> (1 - p_m)^l \left(\frac{p_m}{1 - p_m}\right)^l \left(\frac{1}{|F|}\right) \\ &= \frac{p_m^l}{|F|^l} = \frac{p_m^9}{2} \end{aligned}$$

For the second situation according to Lemma 1, we have

$$P\{x \xrightarrow{M} y\} = (1 - p_m)^{l-1} \frac{p_m}{|T|} = (1 - p_m)^l \frac{p_m}{1 - p_m} \frac{1}{|T|}$$

If $1 - p_m > p_m$, that is $p_m < \frac{1}{2}$, it follows that

$$\begin{aligned} P\{x \xrightarrow{M} y\} &> (1 - p_m)^l \left(\frac{p_m}{1 - p_m}\right)^l \left(\frac{1}{|T|}\right) \\ &= \frac{p_m^l}{|T|^l} = \frac{p_m^9}{5} \end{aligned}$$

From above analysis, the minimum probability that any individual x mutates to another individual y is $\frac{p_m^9}{5}$, where $p_m \in (0, 0.5)$. Thus, we have $p_{\min m2} \geq \frac{p_m^9}{5}$.

(4) $p_{\min s}$

According to selection operator, obviously, we have $p_{\min s} = \frac{1}{2n}$.

According to Corollary 1, the EFHT of ME-GEP with uniform mutation is as follows:

$$\begin{aligned} E[\tau(V_1)] &< \frac{1}{(N^n - (N - 1)^n) \cdot p_{\min m1}^n \cdot p_{\min s}^n} \\ &< \frac{10^{180} \cdot 2^{20} \cdot 20^{20}}{((2^7 \times 5^8)^{20} - (2^7 \times 5^8 - 1)^{20}) \cdot p_m^{180}} \end{aligned} \quad (14)$$

And the EFHT of ME-GEP with one-point mutation is as follows:

$$\begin{aligned} E[\tau(V_1)] &< \frac{1}{(N^n - (N - 1)^n) \cdot p_{\min m2}^n \cdot p_{\min s}^n} \\ &< \frac{5^{20} \cdot 2^{20} \cdot 20^{20}}{((2^7 \times 5^8)^{20} - (2^7 \times 5^8 - 1)^{20}) \cdot p_m^{180}} \end{aligned} \quad (15)$$

5.2.3 The EFHT of ME-GEP for clustering yeast database

The parameters of ME-GEP for clustering yeast database are as follows: Let $h = 15$; then $tal = 16, l = 31$. $T = \{x_1, x_2, \dots, x_{16}\}$.

Applying Theorem 1 and Corollary 1, we have

$$E[\tau(V_1)] < \frac{1}{(N^n - (N - 1)^n) \cdot p_{\min m}^n \cdot p_{\min s}^n}$$

Let $p_{\min m1}$ and $p_{\min m2}$ be the minimum probabilities when an individual x mutates to individual y after uniform mutation and one-point mutation, respectively. Next, we will evaluate and estimate $N, p_{\min m1}, p_{\min m2}$ and $p_{\min s}$.

(1) The size of individual space N .

Denote the tree height of phenotype by variable ht . Because the individual length is 31, we have $ht \leq 5$. It follows that

$$N = |H| \leq 2^{(2^0+2^1+2^2+2^3)} \times 16^{2^4} = 2^{15} \times 16^{16}$$

(2) $p_{\min m1}$

According to Lemma 1, we have

$$P\{x \xrightarrow{M} y\} = (1 - p_m)^l \left(\frac{p_m}{1 - p_m}\right)^{k_1+k_2} \left(\frac{1}{|F|}\right)^{k_1} \left(\frac{1}{|T|}\right)^{k_2}$$

If $1 - p_m > p_m$, that is $p_m < \frac{1}{2}$, it follows that

$$\begin{aligned} P\{x \xrightarrow{M} y\} &> (1 - p_m)^l \left(\frac{p_m}{1 - p_m}\right)^l \left(\frac{1}{|F|}\right)^l \left(\frac{1}{|T|}\right)^l \\ &= \left(\frac{p_m}{(|F|) \cdot (|T|)}\right)^l = \left(\frac{p_m}{32}\right)^{31} \end{aligned}$$

It is easy to find that the minimum probability that any individual x mutates to another individual y is $(\frac{p_m}{32})^{31}$ where $p_m \in (0, 0.5)$. Thus, we have $p_{\min m1} \geq (\frac{p_m}{32})^{31}$.

(3) $p_{\min m2}$

There are two situations when an individual does one-point mutation. First is that one bit flips in head. Second is that one bit flips in tail.

For the first situation according to Lemma 1, we have

$$\begin{aligned} P\{x \xrightarrow{M} y\} &= (1 - p_m)^{l-1} \frac{p_m}{|F|} \\ &= (1 - p_m)^l \frac{p_m}{1 - p_m} \frac{1}{|F|} \end{aligned}$$

If $1 - p_m > p_m$, that is $p_m < \frac{1}{2}$, it follows that

$$\begin{aligned} P\{x \xrightarrow{M} y\} &> (1 - p_m)^l \left(\frac{p_m}{1 - p_m}\right)^l \left(\frac{1}{|F|}\right) \\ &= \frac{p_m^l}{|F|^l} = \frac{p_m^{31}}{2} \end{aligned}$$

For the second situation according to Lemma 1, we have

$$P \left\{ x \xrightarrow{M} y \right\} = (1 - p_m)^{l-1} \frac{p_m}{|T|} = (1 - p_m)^l \frac{p_m}{1 - p_m} \frac{1}{|T|}$$

If $1 - p_m > p_m$, that is $p_m < \frac{1}{2}$, it follows that

$$\begin{aligned} P \left\{ x \xrightarrow{M} y \right\} &> (1 - p_m)^l \left(\frac{p_m}{1 - p_m} \right)^l \left(\frac{1}{|T|} \right) \\ &= \frac{p_m^l}{|T|} = \frac{p_m^{31}}{16} \end{aligned}$$

From above analysis, the minimum probability that any individual x mutates to another individual y is $\frac{p_m^{31}}{16}$ where $p_m \in (0, 0.5)$. Thus, we have $p_{\min m2} \geq \frac{p_m^{31}}{16}$.

(4) $p_{\min s}$

According selection operator, obviously, we have $p_{\min s} = \frac{1}{2n}$.

According to Corollary 1, the EFHT of ME-GEP with uniform mutation is as follows:

$$\begin{aligned} E[\tau(V_1)] &< \frac{1}{(N^n - (N - 1)^n) \cdot p_{\min m1}^n \cdot p_{\min s}^n} \\ &< \frac{1}{32^{930} \cdot 2^{30} \cdot 30^{30}} \\ &< \frac{1}{((2^{15} \times 16^{16})^{30} - (2^{15} \times 16^{16} - 1)^{30}) \cdot p_m^{930}} \end{aligned} \tag{16}$$

And the EFHT of ME-GEP with one-point mutation is as follows:

$$\begin{aligned} E[\tau(V_1)] &< \frac{1}{(N^n - (N - 1)^n) \cdot p_{\min m2}^n \cdot p_{\min s}^n} \\ &< \frac{1}{16^{30} \cdot 2^{30} \cdot 30^{30}} \\ &< \frac{1}{((2^{15} \times 16^{16})^{30} - (2^{15} \times 16^{16} - 1)^{30}) \cdot p_m^{930}} \end{aligned} \tag{17}$$

5.2.4 Numerical experiments

From the above formulas (14)–(17), the upper bound of the EFHT of ME-GEP for solving clustering Iris and Yeast databases can be decreased by increasing mutate probability.

In order to verify formulas (14)–(17), the following experiments to observe the actual average first hitting times of ME-GEP, especially considering different mutation operators: (1) the uniform mutation probability is changed; (2) the one-point mutation probability is changed.

The results obtained are shown in Figs. 4 and 5. The EFHT of ME-GEP decreased with the increasing of mutate probabilities. In the above mentioned examples, the experimental results conformed to formulas (14)–(17).

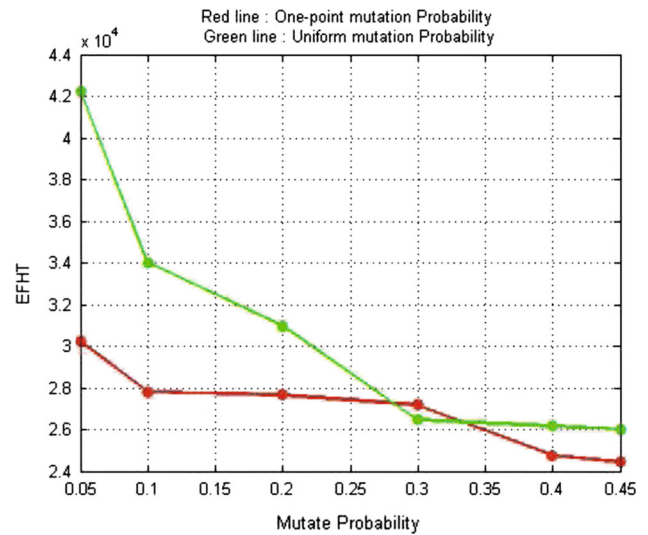


Fig. 4 The EFHT of ME-GEP algorithm for clustering Iris database with different one-point and uniform mutation probabilities

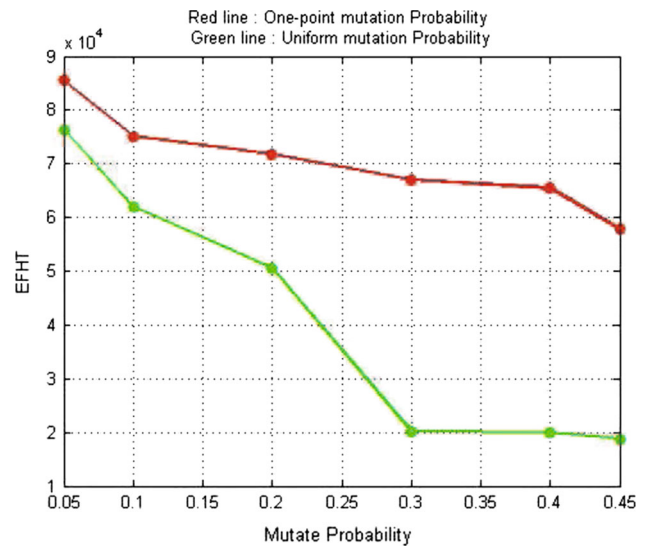


Fig. 5 The EFHT of ME-GEP algorithm for clustering Yeast database with different one-point and uniform mutation probabilities

Conclusion and future work

This paper studies the computational time complexity of ME-GEP. By Markov chain theory and artificial fitness levels technique, the upper and lower bounds of expected first hitting time of ME-GEP are obtained. Furthermore, the relations between the upper bound and the parameters of the ME-GEP are given. Moreover, the relations between the expected first hitting time and the parameters of the algorithm are verified by actual examples for function modeling and clustering problems. Our results obtained in this paper are the general theoretical results, which can be used to guide the design and implementation of GEP. What is more, it has shown that the

analytical methods and techniques adopted in this paper have broad suitability and useful reference for the researchers in relevant research area of GEP theory.

Crossover operator has an important role in GEP. The influence of crossover operator on EFHT of GEP will be discussed in our future work. We will also research on the approach to judging whether a problem belongs to ME-GEP-easy class or ME-GEP-hard class.

Acknowledgments This work was supported by the National Natural Science Foundation of China (No. 61305079, 61305086, 61203306), the project of preeminent youth fund of Fujian province (JA12471), outstanding young teacher training fund of Fujian Normal University (No. fjsdjk2012083), the open fund of State Key Laboratory of Software Engineering (No. SKLSE2012-09-28). Finally, we thank Prof. Jinghu Yu for discussing some issues about this paper.

References

- Bin JX, Jie TC et al (2005) Mining frequent function set based on gene expression programming. *Chin J Comput* 128(8):1247–1254
- Canakci H, Baykasoglu A, Gullu H (2009) Quantitative structure-activity relationships studies of ccr5 inhibitors and toxicity of aromatic compounds using gene expression programming. *Eur J Med Chem* 18(8):1031–1041
- Chen Yu, Tang Changjie et al (2007) Clustering without prior knowledge based on gene expression programming. *Third Int Conf Nat Comput* 3:451–455
- Chen TS, He J et al (2009) A new approach for analyzing average time complexity of population-based evolutionary algorithms on unimodal problems. *IEEE Trans Syst Man Cybern B Cybern* 39(5):1092–1106
- Chi Z, Xiao WM, Thomas MT, Peter CN (2003) Evolving accurate and compact classification rules with gene expression programming. *IEEE Trans Evol Comput* 7(6):519–531
- Ding LX, Yu JH (2005) Some theoretical results about the computation time of evolutionary algorithms. In: *Proceedings of the 7th annual conference on genetic and evolutionary computation*. ACM, Washington, DC, USA, pp 1409–1415
- Droste S, Jansen T, Wegener I (2002) On the analysis of the $(1 + 1)$ evolutionary algorithms. *Theor Comput Sci* 276(1–2):51–81
- Du X, Ding LX (2010) About the convergence rates of a class of gene expression programming. *Sci China Inform Sci* 53(4):715–728
- Du X, Ding LX et al (2009) Convergence analysis of gene expression programming based on maintaining elitist. In: *Proceedings of the first ACM/SIGEVO Summit on Genetic and Evolutionary Computation*, vol 6, pp 823–826
- Ferreira C (2001) Gene expression programming: a new adaptive algorithm for solving problems. *Complex Syst* 12(2):87–129
- Garnier J, Kallel L, Schoenauer M (1999) Rigorous hitting times for binary mutations. *Evol Comput* 7(2):173–203
- Happ E, Daniel J, Christian K, Frank N (2008) Rigorous analyses of fitness-proportional selection for optimizing linear functions. In: *Proceedings of the 10th annual conference on genetic and evolutionary computation*. ACM, Atlanta, Georgia, USA, pp 953–960
- He J, Kang LS (1999) On the convergence rates of genetic algorithms. *Theor Comput Sci* 229(1–2):23–39
- He J, Yao X (2002) From an individual to a population: an analysis of the first hitting time of population-based evolutionary algorithms. *IEEE Trans Evol Comput* 6(5):495–511
- He J, Yao X (2003) Towards an analytic framework for analyzing the computation time of evolutionary algorithms. *Artif Intell* 145(1–2):59–97
- He J, Yao X (2004) A study of drift analysis of estimating computation time of evolutionary algorithms. *Nat Comput* 3(1):21–35
- Iosifescu M (1980) *Finite Markov processes and their application*. Wiley, Chichester
- Karakasis V, Stafylopatis A (2008) Efficient evolution of accurate classification rules using a combination of gene expression programming and clonal selection. *IEEE Tran Evol Comput* 12(6):662–678
- Lehre PK, Yao X (2012) On the impact of mutation-selection balance on the runtime of evolutionary algorithms. *IEEE Trans Evol Comput* 16(2):225–241
- Lopes HS, Weinert WR (2004) A gene expression programming system for time series modeling. In: *Proceedings of XXV Iberian Latin American Congress on Computational Methods in Engineering*, vol 11. CILAMCE, Recife, pp 10–12
- Oliveto PS, Witt C, (2008) Simplified drift analysis for proving lower bounds in evolutionary computation. In: *Proceeding of the 10th International Conference on Parallel Problem Solving from Nature*, pp 82–91
- Oliveto PS, He J, Yao X (2007) Computational complexity analysis of evolutionary algorithms for combinatorial optimization: a decade of results. *J Autom Comput* 4(3):281–293
- Peng J, Jie TC et al (2005) M-GEP: a new evolution algorithm based on multi-layer chromosomes gene expression programming. *Chin J Comput* 9(28):1459–1466
- Rudolph G (1994) Convergence analysis of canonical genetic algorithm. *IEEE Trans Neural Netw* 5:96–101
- Rudolph G (1996) How mutation and selection solve long path problems in polynomial expected time. *Evol Comput* 4(2):194–205
- Shi W, Zhang X, Shen Q (2010) Quantitative structure-activity relationships studies of ccr5 inhibitors and toxicity of aromatic compounds using gene expression programming. *IEEE Trans Evol Comput* 45(1):49–54
- Smith MJ, Szathmry E (1995) *The major transitions in evolution*. Oxford University Press, Oxford
- Teodorescu L (2004) Gene expression programming approach to event selection in high energy physics. *IEEE Trans Nuclear Sci* 53(6):2221–2227
- Teodorescu L, Sherwood D (2008) High energy physics event selection with gene expression programming. *Comput Phys Commun* 178(6):409–419
- Wegener I, Witt C (2005) On the analysis of a simple evolutionary algorithm on quadratic pseudo-boolean function. *J Discrete Algorithm* 3(1):61–78
- Xin D, Qiao LY, Tong XD, Shan KL (2006) A new algorithm of automatic programming: GEDGEP. In: *The 6th International Conference on Simulated Evolution and Learning*, vol 10. Springer, Hefei, pp 292–301
- Yu Y, Zhou ZH (2008) A new approach to estimating the expected first hitting time of evolutionary algorithms. *Artif Intell* 172(15):1809–1832
- Yuan CA et al (2004) Function mining based on gene expression programming-convergence analysis and remnant-guided evolution algorithm. *J Sichun Univ* 36(6):100–105
- Yuan CA et al (2006) Convergence of genetic regression in data mining based on gene expression programming and optimized solution. *Int J Comput Appl* 28(4):359–366
- Zhou Chi, Xiao W, Tirpak TM, Nelsom PC (2003) Evolving accurate and compact classification rules with gene expression programming. *IEEE Trans Evol Comput* 7(6):519–531