

Dominik Ślęzak
Guoyin Wang
Marcin Szczuka
Ivo Düntsch
Yiyu Yao (Eds.)

LNAI 3641

Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing

10th International Conference, RSFDGrC 2005
Regina, Canada, August/September 2005
Proceedings, Part I

1
Part I

 Springer

Acknowledgements

This work was financed by Chinese Nation Nature Science Foundation (60374029), and Visiting Scholar Foundation of Shanxi Province, P. R. China (2004-18).

References

1. Burnet, F. M.: The Clonal Selection Theory of Acquired Immunity. Cambridge University Press (1958)
2. Jerne, N. K.: The immune system. *Scientific American*, Vol.229(1) (1973) 52–60
3. Castro, L. N. D., Zuben, F. J. V.: An evolutionary immune system for data clustering. *Proceedings of Sixth Brazilian Symposium on Neural Network* (2000) 84–89
4. Sun, C.Y., Xie, K.M., Cheng, M.Q.: Mind-Evolution-Based Machine Learning Framework and New Development. *Journal of Taiyuan University of Technology*, Vol.30(5) (1999) 453–457
5. Xie, K.M., Du, Y.G., Sun, C.Y.: Application of the mind-evolution-based machine learning in mixture-ratio calculation of raw materials cement. *Proceedings of the World Congress on Intelligent Control and Automation (WCICA)*, Vol.1 (2000) 132–134
6. Wang, C.L., Xie, K.M., Sun, C.Y.: A Study of convergence of mind evolution based machine learning. *Journal of computer research and development*, Vol.37(7) (2000) 838–842
7. Castro, L. N. D., Zuben, F. J. V.: The Clonal Selection Algorithm with Engineering Applications. Proc. of GECCO'00, Workshop on Artificial Immune Systems and Their Applications, Nevada (2000) 36–37
8. Zhou, G.Y.: Principles of Immunology. Shanghai Press of Science and Technology Literature (2000)
9. Jiao, L.C.: Intelligent Signal and Image Process. Xidian University Intelligent Information Process Institute (2003) 282–283

A Study on the Global Convergence Time Complexity of Estimation of Distribution Algorithms

R. Rastegar and M.R. Meybodi

Computer Engineering Department, Amirkabir University, Tehran, Iran
 {rrastegar, meybodi}@ce.aut.ac.ir

Abstract. The Estimation of Distribution Algorithm is a new class of population based search methods in that a probabilistic model of individuals are estimated based on the high quality individuals and used to generate the new individuals. In this paper we compute 1) some upper bounds on the number of iterations required for global convergence of EDA 2) the exact number of iterations needed for EDA to converge to global optima.

1 Introduction

Genetic Algorithms (GAs) are a class of optimization algorithm motivated from the theory of natural selection and genetic recombination. It tries to find better solution by selection and recombination of promising solution. It works well in wide varieties of problem domains. The poor behaviors of genetic algorithms in some problems, in which the designed operators of crossover and mutation do not guarantee that the building block hypothesis is preserved, have led to the development of other type of algorithms. The search for techniques to preserve building blocks has led to the emergence of new class of algorithm called Probabilistic Model Building Genetic Algorithm (PMBGA) also known as Estimation of Distribution Algorithm (EDA). The principle concept in this new technique is to prevent disruption of partial solutions contained in an individual by giving them high probability of being presented in the child individual. It can be achieved by building a probabilistic model to represent correlation between variables in individual and build model to generate next population.

The EDAs are classified into three classes based on the interdependencies between variables in individuals [9]. Instances of EDAs algorithm include Population-based Incremental Learning (PBIL) [1], Univariate Marginal Distribution Algorithm (UMDA) [10], Learning Automata-based Estimation of Distribution Algorithm (LAEDA) [14], Compact Genetic Algorithm (cGA) [7] for no dependencies model, Mutual Information Maximization for Input Clustering (MIMIC) [3], Combining Optimizer with Mutual Information Trees (COMIT) [2] for bivariate dependencies model, and Factorized Distribution Algorithm (FDA) [11], Bayesian Evolutionary Algorithm (BOA) [13] for multiple dependencies model, to name a few.

Some researchers have studied the working mechanism of EDAs. Mühlenbein [10], González et al [4][5], Höhfeld and Rudolph [6] have studied the behavior of UMDA and PBIL. Mühlenbein and Mahnig [12] discussed the convergence of FDA for

separable additively decomposable functions. In [15], Zhang and Mühlenbein proved that EDAs with infinite population size globally converge. Despite the fact that working mechanisms of EDAs has been studied, the time complexity and the speed of convergence of EDAs algorithm are not known. In this paper we propose some results on the number of iterations needed for EDAs to converge globally when population size is infinite. Our approach is proposed in two sections. At first some upper bounds on the number of iterations required for global convergence of EDA are calculated and then in the second section the exact number of iterations needed for EDA to converge to global optima is calculated.

The rest of paper is organized as follows. Section 2 briefly presents the EDA algorithm and its modeling when EDA uses an infinite population size. Section 3 and 4 demonstrate some theorems about time complexity of EDAs. Conclusion is given in final section.

2 Estimation of Distribution Algorithm with Infinite Population Size

Given a search space D and a positive and continuous function $f(x): D \rightarrow \mathbb{R}^{>0}$, find $\max\{f(x); x \in D\}$. (1)

Let D^* be a set of all points at which function f reaches its maximum value f_{\max} . The steps of the EDA algorithm for solving such an optimization problem are described below.

- 1-Initialization: generate an initial population, $\xi(0)$, of N individuals.
- 2-Selection: choose S_e ($S_e < N$) individuals from population $\xi(n)$ (i.e. population in iteration n) to form the parent population $\xi^S(n)$ using a selection schema such as truncation, tournament or proportional selection schema.
- 3-Updating: perform updating operations on individuals of parent population at iteration n , $\xi^S(n)$, and generate new individuals to form the new population at time n , $\xi(n+1)$, e.g. $\xi(n+1) = \xi^S(n)$.
- 4- If $E\{f(X)|\xi(n+1)\} = f_{\max}$, then stop else go to step 2.

Condition of step 4 of the above algorithm is met when every individual in the population is an optimal solution, that is, the EDA has globally converged.

Let the underlying probability distribution functions for the individuals of $\xi(n)$ and $\xi^S(n)$ be $P(X|\xi(n))$ and $P(X|\xi^S(n))$, respectively. By the famous Glivenko-Cantelli theorem [17], the empirical probability density functions induced by individuals in $\xi(n)$ and $\xi^S(n)$ will converge to $P(X|\xi(n))$ and $P(X|\xi^S(n))$ respectively, as the sizes of $\xi(n)$ and $\xi^S(n)$ tend to infinity. Therefore $P(X|\xi(n))$ and $P(X|\xi^S(n))$ can be thought of as the population and the parent population at iteration n in EDA with infinite population [15]. Below we describe the selection schemes used in this paper.

The Truncation Selection Schema: Truncation selection ranks all the individuals in population $\xi(n)$ according to their fitness and selects the best ones as the set of parents $\xi^S(n)$. In truncation selection with threshold $0 < \mu < 1$ only the $100\mu\%$ of best individuals

are selected to become the parents for the next generation. When the population size is infinite, it can be modeled as [15]:

$$P(X=x|\xi^S(n)) = \begin{cases} \frac{P(X=x|\xi(n))}{\mu} & f(x) \geq \beta(n) \\ 0 & \text{otherwise} \end{cases} . \quad (2)$$

Where:

$$\mu = \int_{f(x) \geq \beta(n)} P(X=x|\xi(n)) .$$

The Two-Tournament Selection Schema: In the two-tournament selection model, 2 individuals are chosen from the current population $\xi(n)$ and the best individual is selected to be a parent. This selection must be repeated S_e times to generate the set of parents $\xi^S(n)$. When the population size is infinite then this schema can be modeled as [15][16].

$$P(X=x|\xi^S(n)) = 2P(X=x|\xi(n)) \int_{f(y) \leq f(x)} P(X=y|\xi(n)) . \quad (3)$$

Remark 1. In [15] Zhang and Mühlenbein have proved that EDA with truncation, or two-tournament selection schema, when using infinite population size, will globally converge.

Let $d(\xi)$ be the ratio of the number of individuals in population ξ , that do not belong to D^* , to the size of ξ . The sequence $\{d(\xi(n)); n=0,1,2,\dots\}$ generated by EDA is a random sequence in general. If $d(\xi)$ is 0, then all individuals of population ξ , are members of D^* . If the population size tends to infinity, then according to Glivenko-Cantelli theorem [17] $d(\xi(n))$ can be computed as follows,

$$d(\xi(n)) = \sum_{x \in D - D^*} P(X=x|\xi(n)) = 1 - \sum_{x \in D^*} P(X=x|\xi(n)) . \quad (4)$$

Define the global convergence stopping time of EDA as $\tau = \min\{n | E\{f(X)|\xi(n)\} = f_{\max}\}$ where for every $t \leq \tau$ $E\{f(X)|\xi(t)\} = f_{\max}$. According to the definition of it, τ is the first time that EDA globally converges. τ can be infinite or finite. In the same manner, we define τ' as $\min\{n | d(\xi(n)) = 0\}$. In the following we state two lemmas to show the relationship between τ' and τ .

Lemma 1. The global convergence stopping time of EDA, τ , is equal to $\tau' = \min\{n | d(\xi(n)) = 0\}$.

Proof. We prove this lemma by contradiction. First assume that $\tau < \tau'$, by the definition of τ' we have $d(\xi(\tau)) > 0$ i.e. there exists at least one $y \in \xi(\tau)$ that doesn't belong to D^* (i.e. $f(y) < f_{\max}$) and $P(X=y|\xi(\tau)) = b > 0$. Thus we have

$$E\{f(X)|\xi(\tau)\} = \sum_{x \in D} f(x) P(X=x|\xi(\tau)) \leq bf(y) + f_{\max}(1-b) . \quad (5)$$

Using (5) and the fact that $E\{f(X)|\xi(\tau)\} = f_{\max}$ we can conclude that $f_{\max} \leq f(y)$ and hence a contradiction. Second, assume that $\tau > \tau'$. Using definitions of τ and τ' , and (4) we have

$$d(\xi(\tau')) = \sum_{x \in D - D^*} P(X = x | \xi(\tau')) = 0 . \quad (6)$$

Using (6) and $0 \leq P(X = x | \xi(\tau')) \leq 1$, we have,

$$P(X = x | \xi(\tau')) = 0 \quad \text{if } x \in D - D^* . \quad (7)$$

By definition of $E\{f(X) | \xi(\tau')\}$ we can write

$$\begin{aligned} E\{f(X) | \xi(\tau')\} &= \sum_{x \in D} f(x) P(X = x | \xi(\tau')) \\ &= \sum_{x \in D^*} f(x) P(X = x | \xi(\tau')) + \sum_{x \in D - D^*} f(x) P(X = x | \xi(\tau')) . \end{aligned} \quad (8)$$

Using (7) and (8) we have

$$E\{f(X) | \xi(\tau')\} = f_{\max}(1) + \sum_{x \in D - D^*} 0 \times f(x) = f_{\max} .$$

Which contradicts the assumption that $\tau' > \tau$ and hence the proof. Q.E.D.

Lemma 2. If $\tau' = \min\{n | d(\xi(n)) = 0\}$, then for every $\tau' \leq \tau$, $d(\xi(\tau)) = 0$.

Proof. Proof is done by contradiction. Assume that $d(\xi(\tau)) \neq 0$, then there exists at least one $y \in \xi(\tau)$ that doesn't belong to D^* (i.e., $f(y) < f_{\max}$) and $P(X = y | \xi(\tau)) = b > 0$. So we have,

$$E\{f(X) | \xi(\tau)\} = \sum_{x \in D} f(x) P(X = x | \xi(\tau)) \leq bf(y) + f_{\max}(1 - b) . \quad (9)$$

Using (9) and $E\{f(X) | \xi(\tau)\} = f_{\max}$ (By lemma 1), we have $f_{\max} \leq f(y)$ and hence a contradiction. Q.E.D.

Lemma 1 indicates that $\tau = \tau'$ and Lemma 2 and Remark 1 state that τ' is the stopping time of $\{d(\xi(n)) ; n=0,1,2,\dots\}$. That is the stopping time of $\{E\{f(X) | \xi(n)\} ; n=0,1,2,\dots\}$ is the same as the stopping time of $\{d(\xi(n)) ; n=0,1,2,\dots\}$ and for this reason in the rest of the paper we study the time complexity of $d(\xi(n))$ rather than the time complexity of $\{E\{f(X) | \xi(n)\} ; n=0,1,2,\dots\}$.

Using above notations and lemmas the EDA algorithm can be described as follows.

- 1- **Initialization:** $P(X = x | \xi(0)) > 0$ for all x (That is $P(X = x | \xi(0)) = p$ for all x where $0 < p < 1$).
- 2- **Selection:** generate $P(X | \xi^S(n))$ from $P(X | \xi(n))$ according to a selection schema.
- 3- **Updating:** $P(X | \xi(n+1))$ is set to $P(X | \xi^S(n))$.
- 4- if $d(\xi(n+1)) = 0$ then stop; otherwise go to step 2.

3 Upper Bounds on Time Complexity of Global Convergence

The Results for upper bounds on time complexity of global convergence of EDA reported in this paper can be summarized by the following two theorems.

Theorem 1. If an EDA with infinite population size and truncation selection schema is used for optimizing function f , then the termination condition is met at most after $(\mu d(\xi(0)) / ((I - \mu)(I - d(\xi(0)))) + I$ iterations.

It is obvious that $0 < \mu < 1$ is an important parameter for the stopping time of EDA when EDA use truncation selection schema. Lower values for μ will impose a lower upper bound on the stopping time and higher values for μ will exert a higher upper bound on the stopping time of EDA.

Theorem 2. If an EDA using infinite population size and 2-tournament selection schema is considered for optimizing f , at most after $(d(\xi(0)) / (I - d(\xi(0)))) + I$ iterations the termination condition is met.

Before we give the proofs of the above two theorems we state one useful lemma.

Lemma 3. If $d(\xi) \leq h_0$, $h_0 > 0$, for any given population ξ and $\{E/d(\xi(n)) - d(\xi(n+1)) | d(\xi(n)) > 0\} \geq (1/h_1)$ then starting from any initial population $\xi(0)$ with $d(\xi(0)) > 0$,

$$E\{\tau | d(\xi(0)) > 0\} \leq h_0 h_1 . \quad (10)$$

Proof¹. Since $\{E/d(\xi(n)) - d(\xi(n+1)) + d(\xi(n)) > 0\} \geq (1/h_1)$ we have $\{d(\xi(n)) ; n=0,1,2,\dots\}$ as a super-martingale. Since $h_0 \geq d(\xi(n)) \geq 0$, it ultimately converges, that is

$$\lim_{n \rightarrow \infty} E\{d(\xi(n)) | d(\xi(0)) > 0\} = 0 .$$

From the definition of stopping time τ , we have $d(\xi(\tau)) = 0$. Therefore,

$$E\{d(\xi(\tau)) | d(\xi(0)) > 0\} = 0 .$$

For $n \geq 1$, we have

$$\begin{aligned} E\{d(\xi(n)) | d(\xi(0)) > 0\} &= \\ E\{E\{d(\xi(n-1)) + d(\xi(n)) - d(\xi(n-1)) | \xi(n-1)\} | d(\xi(0)) > 0\} & \end{aligned}$$

Since $E/d(\xi(n)) - d(\xi(n+1)) + d(\xi(n)) > 0 \geq (1/h_1)$, for $n-1 < \tau$, we have

$$E\{d(\xi(n-1)) + d(\xi(n)) - d(\xi(n-1)) | \xi(n-1)\} \leq d(\xi(n-1)) - \frac{1}{h_1} . \quad (11)$$

From (11) we can write

$$E\{d(\xi(n)) | d(\xi(0)) > 0\} \leq E\{d(\xi(n-1)) - \frac{1}{h_1} | d(\xi(0)) > 0\} . \quad (12)$$

Using (12) and by induction on n , we can get

$$\begin{aligned} E\{d(\xi(n)) | d(\xi(0)) > 0\} &\leq E\{d(\xi(0)) - \frac{n}{h_1} | d(\xi(0)) > 0\} \text{ and} \\ 0 &= E\{d(\xi(\tau)) | d(\xi(0)) > 0\} \leq E\{d(\xi(0))\} - \frac{1}{h_1} E\{\tau | d(\xi(0)) > 0\} \end{aligned} \quad (13)$$

From (13) and $d(\xi) \leq h_0$, we have

$$E\{\tau | d(\xi(0)) > 0\} \leq E\{d(\xi(0))\} h_1 \leq h_0 h_1 ,$$

and hence the proof. Q.E.D.

¹ The idea of the proof is borrowed from [8].

Now we are ready to prove theorems 1 and 2. To do this we first prove that conditions of lemma 3 stand and then using lemma 3 we conclude the theorems.

Proof of theorem 1: We first show that conditions of lemma 3 hold and then use lemma 3 to conclude the theorem.

Using the definition of $d(\xi(n))$ and steps 2 and 3 of EDA algorithm we can write

$$\begin{aligned} E\{d(\xi(n)) - d(\xi(n+1)) \mid d(\xi(n)) > 0\} &= E\left\{\sum_{x \in D^*} P(X=x \mid \xi(n+1)) - \right. \\ &\quad \left. \sum_{x \in D^*} P(X=x \mid \xi(n)) \mid d(\xi(n)) > 0\right\} = \\ E\left\{\sum_{x \in D^*} P(X=x \mid \xi^S(n)) - \sum_{x \in D^*} P(X=x \mid \xi(n)) \mid d(\xi(n)) > 0\right\} \end{aligned} \quad (14)$$

Using (2) and the fact that for all $x \in D^*$ we have $f(x) = f_{\max} \geq \beta(n)$, (14) can be rewritten as

$$\begin{aligned} E\left\{\sum_{x \in D^*} \frac{P(X=x \mid \xi(n))}{\mu} - \sum_{x \in D^*} P(X=x \mid \xi(n)) \mid d(\xi(n)) > 0\right\} = \\ E\left\{\left(\frac{1}{\mu} - 1\right) \sum_{x \in D^*} P(X=x \mid \xi(n)) \mid d(\xi(n)) > 0\right\} = \left(\frac{1}{\mu} - 1\right)(1 - d(\xi(n))) \end{aligned} \quad (15)$$

Using (15) and induction on n we have

$$d(\xi(n)) \leq d(\xi(0)) = h_0. \quad (16)$$

From (16) we have,

$$\begin{aligned} E\{d(\xi(n)) - d(\xi(n+1)) \mid d(\xi(n)) > 0\} &\geq \left(\frac{1}{\mu} - 1\right)(1 - d(\xi(0))) = \\ &= \frac{1}{\mu} = \frac{1}{h_1} \\ &= \frac{1}{(1-\mu)(1-d(\xi(0)))} \end{aligned} \quad (17)$$

From (16) and (17) we conclude that conditions of Lemma 3 are satisfied and therefore we can write,

$$E\{\tau \mid d(\xi(0)) > 0\} \leq h_0 h_1 = \frac{\mu d(\xi(0))}{(1-\mu)(1-d(\xi(0)))}.$$

Hence the proof. Q.E.D.

Proof theorem 2: Using the definition of $d(\xi(n))$ and steps 2 and 3 of EDA algorithm, we can write,

$$\begin{aligned} E\{d(\xi(n)) - d(\xi(n+1)) \mid \xi(n)\} &= E\left\{\sum_{x \in D^*} P(X=x \mid \xi(n+1)) - \right. \\ &\quad \left. \sum_{x \in D^*} P(X=x \mid \xi(n)) \mid \xi(n)\right\} = \\ E\left\{\sum_{x \in D^*} P(X=x \mid \xi^S(n)) - \sum_{x \in D^*} P(X=x \mid \xi(n)) \mid \xi(n)\right\} \end{aligned} \quad (18)$$

Using (3) and the fact that for all $y \in D$ we have $f_{\max} \geq f(y)$, we can rewrite (18) as

$$\begin{aligned} E\left\{\sum_{x \in D^*} \left\{2P(X=x \mid \xi(n)) \int_{f_{\max} = f(x) \geq f(y)} P(X=y \mid \xi(n))\right\} - \sum_{x \in D^*} P(X=x \mid \xi(n)) \mid \xi(n)\right\} \\ E\left\{\sum_{x \in D^*} P(X=x \mid \xi(n)) \mid \xi(n)\right\} = (1 - d(\xi(n))) \end{aligned} \quad (19)$$

Using (19) and induction on n we can write

$$d(\xi(n)) \leq d(\xi(0)) = h_0. \quad (20)$$

Using the (18), (19) and (20) we have

$$E\{d(\xi(n)) - d(\xi(n+1)) \mid \xi(n)\} \geq (1 - d(\xi(0))) = \frac{1}{h_1}.$$

Since conditions of Lemma 3 are satisfied and we have,

$$E\{\tau \mid d(\xi(0)) > 0\} \leq h_0 h_1 = \frac{d(\xi(0))}{1 - d(\xi(0))},$$

Hence the theorem. Q.E.D.

4 Computation of Global Convergence Stopping Time

In this section, some strong results about the convergence of EDA are derived. As stated before $\{d(\xi(n)); n=0,1,2,\dots\}$ is a random sequence in general and when population size tends to infinity this sequence becomes a deterministic sequence. In other words by knowing $d(\xi(n-1))$ we can compute the exact value of $d(\xi(n))$. We can use these properties to derive some strong results about the convergence of EDA.

Definition 1. (Convergence Rate). Let $\{a_n; n=0,1,2,\dots\}$ be a sequence that converges to a^* . If we have

$$\lim_{n \rightarrow \infty} \frac{|a_{n+1} - a^*|}{|a_n - a^*|} = \beta$$

then $\{a_n; n=0,1,2,\dots\}$ converges to a^* with convergence rate β .

The results for the exact number of iterations needed for EDA to converge to global optima reported in this paper can be summarized by the following two theorems

Theorem 3. If we use an EDA with infinite population size and truncation selection method having threshold μ then a) After $I + (\log(I \cdot d(\xi(0))) / \log \mu)$ iterations the condition of termination is met. b) $\{d(\xi(n)); n=0,1,2,\dots\}$ converges to 0 with convergence rate $1/\mu$.

Theorem 4. If we use an EDA with infinite population size and 2-tournament selection method, then a) After $I + (\log(I \cdot d(\xi(0))) / \log 0.5)$ iterations the condition of termination is met. b) $\{d(\xi(n)); n=0,1,2,\dots\}$ converges to 0 with convergence rate 2.

Before we give the proofs of theorems 3 and 4, we state two lemmas for the computation of $d(\xi(n))$.

Lemma 4. For EDA algorithm with infinite population size and truncation selection method $d(\xi(n))$ can be computed as follows,

$$d(\xi(n)) = 1 - (1 - d(\xi(0))) \left(\frac{1}{\mu}\right)^n,$$

where $0 < \mu < 1$ is the selection threshold.

Proof. By the definition $d(\xi(n))$ and (2) we have,

$$\begin{aligned} d(\xi(n+1)) &= 1 - \sum_{x \in D^*} P(X=x | \xi(n+1)) = 1 - \sum_{x \in D^*} P(X=x | \xi^S(n)) \\ &= 1 - \frac{1}{\mu} \sum_{x \in D^*} P(X=x | \xi(n)) = 1 - \frac{1}{\mu} (1 - d(\xi(n))) \end{aligned} \quad (21)$$

From (21) we have

$$d(\xi(n+1)) - \frac{1}{\mu} d(\xi(n)) = (1 - \frac{1}{\mu}). \quad (22)$$

Characteristic equation of (22) is,

$$r^2 - \left(\frac{1}{\mu} + 1\right)r + \frac{1}{\mu} = 0. \quad (23)$$

By solving (23) we have

$$d(\xi(n)) = 1 - (1 - d(\xi(0))) \left(\frac{1}{\mu}\right)^n.$$

Q.E.D.

Lemma 5. For EDA algorithm with infinite population size and tournament selection method we have

$$d(\xi(n)) = 1 - (1 - d(\xi(0))) 2^n.$$

Proof. By the definition of $d(\xi(n))$ and (3) we have,

$$d(\xi(n+1)) - 2d(\xi(n)) + 1 = 0. \quad (24)$$

The characteristic equation of (24) is,

$$r^2 - 3r + 2 = 0. \quad (25)$$

By solving (24) we have

$$d(\xi(n)) = 1 - (1 - d(\xi(0))) 2^n.$$

Q.E.D.

Now we use Lemmas 4 and 5 and prove theorems 3 and 4.

Proof of theorem 3. (a) By definition of stopping time and Lemma 4 we have

$$d(\xi(\tau)) = 1 - (1 - d(\xi(0))) \left(\frac{1}{\mu}\right)^\tau = 0. \quad (26)$$

Using (26) we conclude,

$$\tau = \log(1 - d(\xi(0))) / \log \mu. \quad (27)$$

By (27) after $1 + (\log(1 - d(\xi(0))) / \log \mu)$ iterations the condition of termination is met.
(b) By lemma 4, we have

$$\lim_{n \rightarrow \infty} \frac{|d(\xi(n+1)) - 0|}{|d(\xi(n)) - 0|} = \frac{1}{\mu}.$$

From definition 1, $\{d(\xi(n)); n=0,1,2,\dots\}$ converges to 0 with convergence rate $1/\mu$.
Q.E.D.

Proof of theorem 4: (a) By definition of stopping time and lemma 5 we have

$$d(\xi(\tau)) = 1 - (1 - d(\xi(0))) 2^\tau = 0. \quad (28)$$

Using (28) we have,

$$\tau = \log(1 - d(\xi(0))) / \log(0.5). \quad (29)$$

By (29) after $1 + (\log(1 - d(\xi(0))) / \log 0.5)$ iterations the condition of termination is met.
(b) By lemma 5, we conclude that

$$\lim_{n \rightarrow \infty} \frac{|d(\xi(n+1)) - 0|}{|d(\xi(n)) - 0|} = 2.$$

According to definition 1, $\{d(\xi(n)); n=0,1,2,\dots\}$ converges to 0 with convergence rate 2. Q.E.D.

5 Conclusion

This paper presented some new results for global convergence computation time for EDA algorithms. The following quantities were computed: 1) some upper bounds on the number of iterations required for global convergence of EDA 2) the exact number of iterations needed for EDA to converge to global optima.

References

- Baluja, S.: Population Based Incremental Learning: A Method for Integrating Genetic Search Based Function Optimization and Competitive Learning. Technical Report, Carnegie Mellon University (1994)
- Baluja, S., Davis, S.: Fast Probabilistic Modeling for Combinatorial Optimization. In: The 15th National Conference on Artificial Intelligence, Madison, Wisconsin, AAAI Press, (1998) 469-476
- Bonet, D., Isbell, J.S., Viola, P.: MIMIC: Finding Optima by Estimating Probability Densities. Advances in Neural Information Processing Systems, Vol. 9, MIT Press, Cambridge (1997) 424-431
- González, C., Lozano, J.A., Larrañaga, P.: The Convergence Behavior of the PBIL Algorithm: A Preliminary Approach. In: The 5th International Conference on Artificial Neural Networks and Genetic Algorithms (2001)
- González, C., Lozano, J.A., Larrañaga, P.: Analyzing the PBIL Algorithm by Means of Discrete Dynamical Systems. Complex Systems, Vol. 12, (2000) 465-479

6. Höhfeld, M., Rudolph, G.: Towards A Theory of Population Based Incremental Learning. In: The 4th IEEE Conferences on Evolutionary Computation, Indianapolis, IEEE Press, (1997) 1-5
7. Harik, G.R., Lobo, F.G., Goldberg, D.E.: The Compact Genetic Algorithm. *IEEE Transactions on Evolutionary Computation*, Vol. 3, No. 4 (1999) 287-297
8. He, J., Yao, X.: Drift Analysis and Average Time Complexity of Evolutionary Computation. *Artificial Intelligence*, Vol. 127, (2001) 57-85
9. Larrañaga, P., Lozano, J.A.: Estimation of Distribution Algorithms. A New tools for Evolutionary Computation. Kluwer Academic Publishers (2001)
10. Mühlenbein, H.: The Equation for Response to Selection and Its Use for Prediction. *Evolutionary Computation*, Vol. 5, (1998) 303-346
11. Mühlenbein, H., Mahnig, T., Rodriguez, A.O.: Schemata, Distributions and Graphical Models in Evolutionary Optimization. *Journal of Heuristics*, Vol. 5, (1999) 215-247
12. Mühlenbein, H., Mahnig, T.: Evolutionary Computation and Wright's Equation. *Theoretical Computer Science* (in press)
13. Pelikan, M., Goldberg D.E., Cantz Paz, E.: BOA: the Bayesian Optimization Algorithm. In: The Genetic and Evolutionary Computation Conference, Morgan Kaufmann Publishers, (1999) 525-532
14. Rastegar, R., Meybodi, M.R.: LAEDA, A New Evolutionary Algorithm using Learning Automata. In: The 9th Annual International Computer Society of Iran Computer Conference CSICCC2004, Tehran, Iran, (2004) 456-464
15. Zhang, Q., Mühlenbein, H.: On the Convergence of a Class of Estimation of Distribution Algorithms. *IEEE Transactions on Evolutionary Computation*, Vol. 8, No. 2 (2004)
16. Bickle, T., Thiele, L.: A Mathematical Analysis of Tournament selection. In: The 6th International Conference of Genetic Algorithms, Morgan Kaufmann Publishers, San Francisco, CA (1995) 9-16
17. Devroye, L., Gyorfi, L., Lugosi, G.: A Probabilistic Theory of Pattern Recognition. Springer Verlag, Berlin (1996)

Finding Minimal Rough Set Reducts with Particle Swarm Optimization

Xiangyang Wang, Jie Yang, Ningsong Peng, and Xiaolong Teng

Inst. of Image Processing & Pattern Recognition, Jiaotong University,
Shanghai, 200030, P.R. China
wangxiangyang@sjtu.edu.cn

Abstract. We propose a new algorithm to find minimal rough set reducts by using Particle Swarm Optimization (PSO). Like Genetic Algorithm, PSO is also a type of evolutionary algorithm. But compared with GA, PSO does not need complex operators as crossover and mutation that GA does, it requires only primitive and simple mathematical operators, and is computationally inexpensive in terms of both memory and times. The experiments on some UCI data compare our algorithm with GA-based, and other deterministic rough set reduction algorithms. The results show that PSO is efficient to minimal rough set reduction.

1 Introduction

Rough set reduction method has been used for feature selection [1,2]. It is worth to find minimal reducts, with which we can generate more general decision rules and better classification quality of new samples. However, the problem of finding a minimal reduct is NP-hard [3]. So some heuristic or approximation algorithms have to be considered.

X.Hu gives a rough set reduction algorithm using positive region based attribute significance as heuristics [4]. G. Y. Wang develops a conditional information entropy reduction algorithm [5]. These hill-climbing methods, however, do not guarantee to find a global optimal or minimal reducts. Using the attribute significance discriminate between candidates may lead the search down a non-minimal path.

Wroblewski and Bazan et al. [6,8] use genetic algorithms to find minimal reducts. One of their methods uses classical genetic algorithm with individuals represented by bit strings to short reducts finding. This method, sometimes fails to find the global optimum. Another is the hybrid algorithm that exploits advantages of both genetic and heuristic algorithms. Genetic algorithm is used to produce proper attributes order, and heuristic algorithm finds the minimal reducts. Although this method is much better, it increases the computation complexity [6]. One of the most widely used hybrid algorithm is the order-based genetic algorithm [7]. Slezak and Wroblewski [10] extend it to search approximate entropy reducts.

We propose a new algorithm to find minimal rough set reducts with Particle Swarm Optimization (PSO). PSO is a new evolutionary computation technique