Advances in Some Theoretical Issues of Evolutionary Computation

Zhi-Hua Zhou

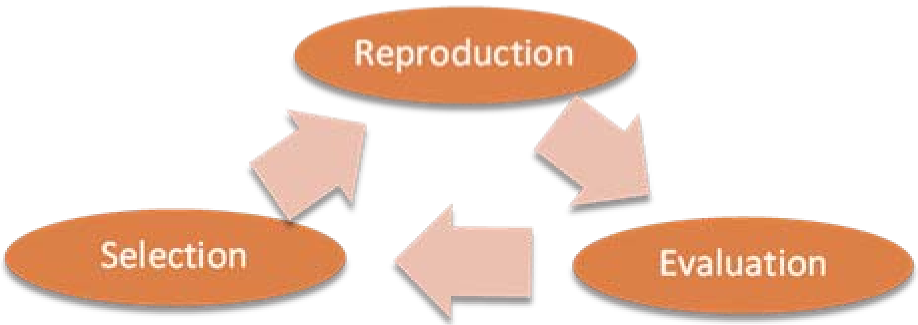
http://cs.nju.edu.cn/people/zhouzh/

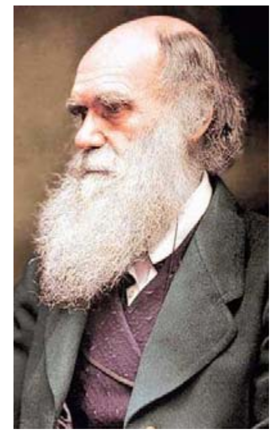
zhouzh@nju.edu.cn

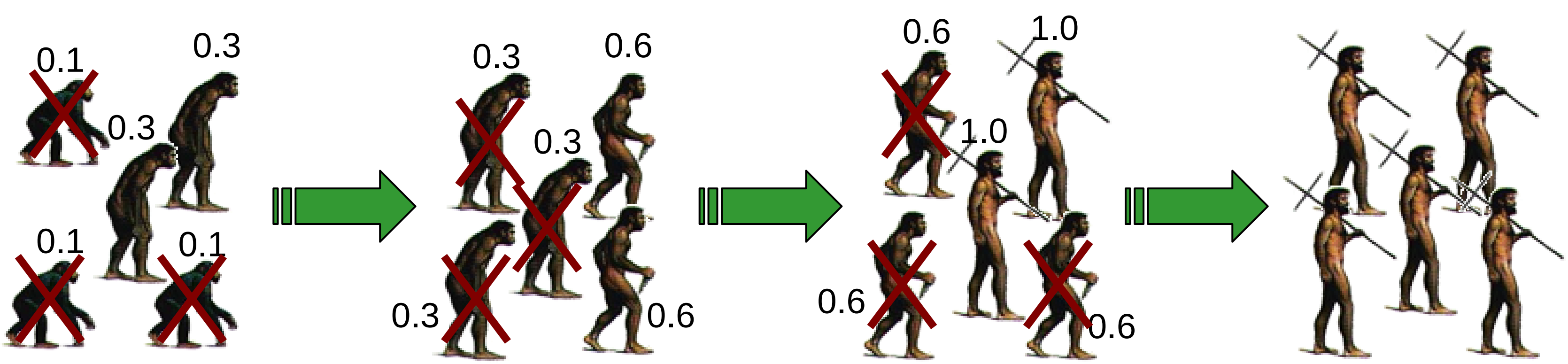
LAMDA Group,
National Laboratory for Novel Software Technology,
Nanjing University, China

Background

Evolutionary algorithms are a kind of optimization algorithms inspirited by Mother Nature: "survival of the fittest"







Background (con't)

EAs have been widely applied in real-life optimization problems

- Biology, Chemistry, Medicine, Robotics, … …, even in operating system

Home

Linux: Tuning The Kernel With A Genetic Algorithm

Posted by Jeremy on Friday, January 7, 2005 - 06:59

Jake Moilanen provided a series of four patches against the 2.6.9 Linux kernel [story] that introduce a simple genetic algorithm used for automatic tuning. The patches update the anticipatory IO scheduler [story] and the zaphod CPU scheduler [story] to both use the new in-kernel library, theoretically allowing them to automatically tune themselves for the best possible performance for any given workload. Jake says, "*using these patches, there are small gains (1-3%) in Unixbench & SpecJBB. I am hoping a scheduler guru will able to rework them to give higher gains.*"

The problem

But … Heuristics dominate

- Heuristics are very helpful. We should not overlook the usefulness of heuristics
- But if there are only heuristics, although not bad for "nature-inspired", it is not good enough for "computing"

It is the time to try to contribute to the theoretical foundation for EC

Contents of the talk

- A new approach to estimating the expected first hitting time of evolutionary algorithms

  [Yang Yu and Zhi-Hua Zhou, AAAI'06]

- The second part of the talk has not got published, which will be shared in the near future … …

A fundamental problem

A fundamental problem of computing theory

- Computational complexity

In particular, for EAs we want to know:

- What is the asymptotic behavior of an EA?
  how much time needed to solve an extremely large problem
- When an EA works?
  when its time complexity is smaller than another EA
- How to design an EA on demand?
  how to design an EA with minimal time complexity on a given task

Expected first hitting time

For evolutionary algorithms, neither worst-case analysis nor best-case analysis is suitable, since EAs are stochastic algorithms

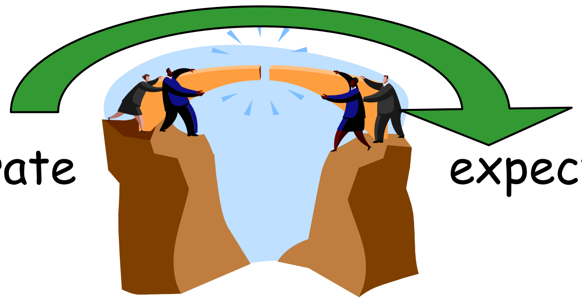
What we want to know is the average performance

First hitting time (FHT) is the time that EAs find the optimal solution for the first time

The expected FHT implies the average computational time complexity

What have we done?

- We established a bridge between the expected FHT and another important theoretical issue of EA, the convergence rate



- With this bridge, we develop a new approach to estimating the expected FHT
  - Does not need extra information, e.g. distance to the target
  - Able to analyze EAs that cannot be analyzed before, e.g. EAs with dynamic operators

Related works

Early approaches to estimating expected FHT are specific for simple EAs on simple problems

- (1+1)-EA on the long path problem [Rudolph, ECJ96]
- (1+1)-EA on separable functions [Droste et al, ECJ98]
- (1+1)-EA on linear functions [Wegener, IWGTCCS'00]
- …

Hard to be generalized

Related works (con't)

General approaches to estimating expected first hitting time:

- Drift analysis [He & Yao, AIJ01, NaturalComp04]

  a distance function is required to be manually determined, without any guidance

- Analytic approach [He & Yao, AIJ03]

  only for EAs with static operators, can hardly be applied to dynamic operators

Notations

Notations	For EAs	For Markov chains
X	populations space	state space
$X^* \subset X$	optimal populations	target subspace
$t = 0, ..., +\infty$	steps of evolution process	discrete time
$\xi_t \in X$	population at step t	state at time t
τ	first hitting time (FHT)	first hitting time (FHT)
$\mathbb{E}[\tau]$	expected FHT	expected FHT

Related works (con't)

Drift analysis [He & Yao, AIJ01, NaturalComp04]

EAs are modeled by supermartingales

- first, define a distance function $d: X \rightarrow R$ which measures the distance from current state to the target

- estimate the one-step mean drift

$$c_{\text{up}} \geq \mathbb{E}[d(\xi_t) - d(\xi_{t+1}) | \xi_t] \geq c_{\text{low}}$$

- divided by path length $d(\xi_0)$, so get the expected FHT

$$\frac{d(\xi_0)}{c_{\text{up}}} \leq \mathbb{E}[\tau | \xi_0] \leq \frac{d(\xi_0)}{c_{\text{low}}}$$

Related works (con't)

Drift analysis [He & Yao, AIJ01, NaturalComp04]

EAs are modeled by supermartingales

- first, define a distance function $d: X \rightarrow R$ which measures the distance from current state to the target
- estimate the one-step ...

$$c_{up} \geq \mathbb{E}[d(\xi_t) - d(\xi_{t+1}) | \xi_t] \geq c_{low}$$

- divided by path length ... ected FHT

$$\frac{d(\xi_0)}{c_{up}} \leq \mathbb{E}[\tau | \xi_0] \leq \frac{d(\xi_0)}{c_{low}}$$

> One need to measure the path length and step size in order to know how many steps are needed
>
> Unfortunately, there is no guide on how to design a good distance function yet

Related works (con't)

Analytical approach [He & Yao, AIJ03]

EAs are modeled by homogeneous Markov chains

- the analytic solution for the expected FHT

$$\mathbf{m} = \begin{bmatrix} \mathbb{E}[\tau|\xi_0 = 1] \\ \vdots \\ \mathbb{E}[\tau|\xi_0 = |\mathbf{T}|] \end{bmatrix} = (\mathbf{I} - \mathbf{T})^{-1}\mathbf{1}$$

- derived theorems: (**d** is a distance vector of states)

$$\mathbf{d} - \mathbf{T}\mathbf{d} \le \mathbf{1} \Rightarrow \mathbf{m} \ge \mathbf{d} \qquad \mathbf{d} - \mathbf{T}\mathbf{d} \ge \mathbf{1} \Rightarrow \mathbf{m} \le \mathbf{d}$$

Related works (con't)

Analytical approach [He & Yao, AIJ03]

EAs are modeled by homogeneous Markov chains

- the analytic solution for the expected FHT

Homogenous Markov chains can only model EAs with static operators

Popularly used dynamic operators are beyond the scope of its ability

- derived theorems ... distance vector of states)

$$\mathbf{d} - \mathbf{T}\mathbf{d} \le \mathbf{1} \Rightarrow \mathbf{m} \ge \mathbf{d} \qquad \mathbf{d} - \mathbf{T}\mathbf{d} \ge \mathbf{1} \Rightarrow \mathbf{m} \le \mathbf{d}$$

Our main result

LAMDA
Learning And Mining from DatA
http://lamda.nju.edu.cn

Theorem. Given an absorbing Markov chain $\{\xi_t\}_{t=0}^{+\infty} (\xi_t \in X)$ and a target subspace $X^* \subset X$

if two sequences $\{\alpha_t\}_{t=0}^{+\infty}$ and $\{\beta_t\}_{t=0}^{+\infty}$ satisfy

1. $\prod_{t=0}^{+\infty}(1-\alpha_t) = 0$ convergence guarantee
2. $\sum_{x \notin X^*} P(\xi_{t+1} \in X^* | \xi_t = x) P(\xi_t = x) \geq \alpha_t (1-\mu_t)$ lower bound
3. $\sum_{x \notin X^*} P(\xi_{t+1} \in X^* | \xi_t = x) P(\xi_t = x) \leq \beta_t (1-\mu_t)$ upper bound

then the chain converges to X^* and the expected FHT is bounded by

$$\mathbb{E}[\tau] \leq \alpha_0 + \sum_{t=2}^{+\infty} t\alpha_{t-1} \prod_{i=0}^{t-2}(1-\alpha_i)$$

$$\mathbb{E}[\tau] \geq \beta_0 + \sum_{t=2}^{+\infty} t\beta_{t-1} \prod_{i=0}^{t-2}(1-\beta_i)$$

Once we can bound the probability of successful transition at each step, we can bound the expected FHT

Convergence rate

Our result was obtained by establishing a bridge between the EFHT and convergence rate

What is the convergence rate?

Given a Markov chain $\{\xi_t\}_{t=0}^{+\infty}$ $(\xi_t \in X)$ and a target subspace $X^* \subset X$, $\{\xi_t\}_{t=0}^{+\infty}$ is said to converge to X^* if

$$\lim_{t \to +\infty} \mu_t = 1$$

The convergence rate is measured by $1 - \mu_t$ at step *t*

[He & Yu, JSA01]

$\mu_t = \sum_{x \in X^*} P(\xi_t = x)$ is the probability of ξ_t in X^*

On convergence rate [He & Yu, JSA01]

Lemma 1. Given an absorbing Markov chain $\{\xi_t\}_{t=0}^{+\infty} (\xi_t \in X)$ and a target subspace $X^* \subset X$

if two sequences $\{\alpha_t\}_{t=0}^{+\infty}$ and $\{\beta_t\}_{t=0}^{+\infty}$ satisfy

1. $\prod_{t=0}^{+\infty}(1-\alpha_t) = 0$ convergence guarantee
2. $\sum_{x \notin X^*} P(\xi_{t+1} \in X^* | \xi_t = x) P(\xi_t = x) \geq \alpha_t (1-\mu_t)$ lower bound
3. $\sum_{x \notin X^*} P(\xi_{t+1} \in X^* | \xi_t = x) P(\xi_t = x) \leq \beta_t (1-\mu_t)$ upper bound

then the chain converges to X* and the convergence rate is bounded by

$$(1-\mu_0)\prod_{i=0}^{t-1}(1-\alpha_i) \geq 1-\mu_t \geq (1-\mu_0)\prod_{i=0}^{t-1}(1-\beta_i)$$

Two questions

Learning And Mining from DatA
http://lamda.nju.edu.cn

- Is there any relationship between the convergence rate and the expected FHT?
- If there is, can we bound the expected FHT based on the bounds of the convergence rate?

EFHT and τ

Let's take a close look on the expected FHT

Given an absorbing Markov chain $\{\xi_t\}_{t=0}^{+\infty} (\xi_t \in X)$
and a target subspace $X^* \subset X$
Let a r.v. τ denote events: $\tau = 0$: $\xi_0 \in X^*$

$\tau = 1$: $\xi_1 \in X^* \wedge \xi_i \notin X^* (\forall i = 0)$

$\tau = 2$: $\xi_2 \in X^* \wedge \xi_i \notin X^* (\forall i = 0,1)$

...

Expected FHT

$$\mathbb{E}[\tau]$$

Convergence rate and τ

Let's take a close look on the convergence rate

Given an absorbing Markov chain $\{\xi_t\}_{t=0}^{+\infty} (\xi_t \in X)$

and a target subspace $X^* \subset X$

Let a r.v. τ denote events: $\tau = 0: \xi_0 \in X^*$

$\tau = 1: \xi_1 \in X^* \wedge \xi_i \notin X^* (\forall i = 0)$

$\tau = 2: \xi_2 \in X^* \wedge \xi_i \notin X^* (\forall i = 0,1)$

...

Convergence Rate

Define [He&Yu01]

$$1 - \mu_t$$
$$= 1 - \Sigma_{x \in X^*} P(\xi_t = x)$$

while

$$\mu_{t+1} - \mu_t$$
$$= \Sigma_{x \in X^*} P(\xi_{t+1} = x) - \Sigma_{x \in X^*} P(\xi_t = x)$$
$$= P(\tau = t + 1)$$

We find …

Learning And Mining from DatA
http://lamda.nju.edu.cn

Given an absorbing Markov chain $\{\xi_t\}_{t=0}^{+\infty} (\xi_t \in X)$
and a target subspace $X^* \subset X$

Let a r.v. τ denote events: $\tau = 0: \xi_0 \in X^*$

$\tau = 1: \xi_1 \in X^* \wedge \xi_i \notin X^* (\forall i = 0)$

$\tau = 2: \xi_2 \in X^* \wedge \xi_i \notin X^* (\forall i = 0,1)$

...

* Expected FHT
  expectation of τ

$\longleftarrow \tau \longrightarrow$

* Convergence Rate
  $1 -$ probability distribution of τ

A lemma

How about the relationship between the expectation and the probability distribution of random variables?

Lemma 2. Suppose U and V are two discrete r.v.s on nonnegative integer, let $D_u(\cdot)$ and $D_v(\cdot)$ be their distribution functions respectively. If the distributions satisfy

$$\forall t = 0,1,\ldots: D_u(t) \geq D_v(t)$$

then the expectations of the random variables satisfy

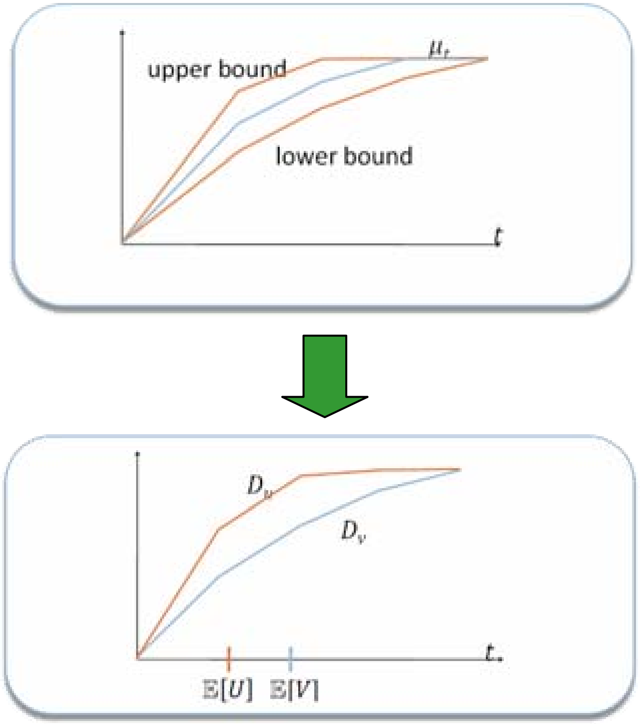
$$\mathbb{E}[U] \leq \mathbb{E}[V]$$

when the distribution of one r.v. is always higher than the other one, expectation of the former is smaller

To get our result

So, we got an idea of bounding the expected FHT:

1. The convergence rate is bounded at every step
2. The convergence rate is an inverse of the probability distribution
3. Bounds of the probability distribution comes to bounds of the expectation
4. Obtain the bounds of the expected FHT



How to use?

Two sequences are to be determined to apply our theorem, they are …

Theorem. Given an absorbing Markov chain $\{\xi_t\}_{t=0}^{+\infty}(\xi_t \in X)$ and a target subspace $X^* \subset X$

if two sequences $\{\alpha_t\}_{t=0}^{+\infty}$ and $\{\beta_t\}_{t=0}^{+\infty}$ satisfy

1. $\prod_{t=0}^{+\infty}(1-\alpha_t)=0$
2. $\sum_{x\notin X^*} P(\xi_{t+1} \in X^* | \xi_t = x) P(\xi_t = x) \geq \alpha_t(1-\mu_t)$
3. $\sum_{x\notin X^*} P(\xi_{t+1} \in X^* | \xi_t = x) P(\xi_t = x) \leq \beta_t(1-\mu_t)$

then the chain converges to X^* and the expected FHT is bounded by

$$\mathbb{E}[\tau] \leq \alpha_0 + \sum_{t=2}^{+\infty} t\alpha_{t-1} \prod_{i=0}^{t-2}(1-\alpha_i)$$

$$\mathbb{E}[\tau] \geq \beta_0 + \sum_{t=2}^{+\infty} t\beta_{t-1} \prod_{i=0}^{t-2}(1-\beta_i)$$

How to use?

Two sequences are to be determined to apply our theorem

$$\beta_t \geq \frac{\sum_{x \notin X^*} P(\xi_{t+1} \in X^* | \xi_t = x) P(\xi_t = x)}{(1-\mu_t)} \geq \alpha_t$$

$$\beta_t \geq \sum_{x \notin X^*} P(\xi_{t+1} \in X^* | \xi_t = x) \frac{P(\xi_t = x)}{\sum_{x \notin X^*} P(\xi_t = x)} \geq \alpha_t$$

probability of success

normalized distribution

Theorem

and a tar

if two seq

1. $\prod_{t=0}^{+\infty}($

2. $\sum_{x \notin X^*} P(\xi_{t+1} \in X^* | \xi_t = x) P(\xi_t = x) \geq \alpha_t (1-\mu_t)$

3. $\sum_{x \notin X^*} P(\xi_{t+1} \in X^* | \xi_t = x) P(\xi_t = x) \leq \beta_t (1-\mu_t)$

then the chain converges to X^* and the expected FHT is bounded by

$$\mathbb{E}[\tau] \leq \alpha_0 + \sum_{t=2}^{+\infty} t \alpha_{t-1} \prod_{i=0}^{t-2} (1-\alpha_i)$$

$$\mathbb{E}[\tau] \geq \beta_0 + \sum_{t=2}^{+\infty} t \beta_{t-1} \prod_{i=0}^{t-2} (1-\beta_i)$$

To use our theorem

- We analyze four EAs to demonstrate the usefulness of our theorem
- Four EAs using:
  - mutation
  - mutation with population
  - mutation with recombination
  - time-variant mutation

- On a specific <u>subset sum</u> problem
  - Most of existed works studied EAs on "easy" problems
  - We study the four EAs on this "hard" problem

The subset problem

Given a set of n positive integers $W = \{w_i\}_{i=1}^{n}$ and a constant c, solve:

$$\underset{W' \subseteq W}{\operatorname{argmax}} \sum_{w \in W'} w$$

$$\text{s.t. } \sum_{w \in W'} w \leq c$$

To find the subset of W with maximum sum but must not exceed c

A specific problem

Problem 1:

1. $w_i \geq 2 (\forall i = 1,2, \dots, n-1)$
2. $w_n = \left(\sum_{i=1}^{n-1} w_i\right) - 1$
3. $c = w_n$

The optimal subset of this specific problem is $\{w_n\}$ (contains only the last element) encoded as (000…01)

Common configuration of four EAs

- Encoding: Each solution is coded into an n-bit string
- Initialization: M randomly generated solutions
- Reproduction: Generate M new solutions
- Selection: Select the best M solutions
- Fitness: $fitness(x) = c - \theta \Sigma_{i=1}^{n} w_i x_i$

  where θ is 1 for feasible solutions and zero otherwise
- Stop Criterion: the lowest fitness = zero

Analysis on EA-1

EA-1:

- Mutation 1: flip each bit with probability $p_m \in (0,0.5]$
- Population size M=1

- Result:
  - EA-1 needs exponential time to solve the problem

Analysis on EA-1 (con't)

Proof:

1. Estimate $\{\alpha_t\}_{t=0}^{+\infty}$ and $\{\beta_t\}_{t=0}^{+\infty}$ according to

$$\beta_t \geq \sum_{x \notin X^*} P(\xi_{t+1} \in X^* | \xi_t = x) \frac{P(\xi_t = x)}{\sum_{x \notin X^*} P(\xi_t = x)} \geq \alpha_t$$

2. Assume $P_{\max} \geq P(\xi_{t+1} \in X^* | \xi_t = x) \geq P_{\min}$ for all x, then

$$P_{\max} \geq \sum_{x \notin X^*} P(\xi_{t+1} \in X^* | \xi_t = x) \frac{P(\xi_t = x)}{\sum_{x \notin X^*} P(\xi_t = x)} \geq P_{\min}$$

so, we can let $\beta_t = P_{\max}$ and $\alpha_t = P_{\min}$

3. The probability that a solution is mutated to another one with k bits difference is $p_m^k (1 - p_m)^{n-k}$, which has an upper bound $p_m (1 - p_m)^{n-1}$ and a lower bound p_m^n

4. Thus for all t: $p_m (1 - p_m)^{n-1} \geq P(\xi_{t+1} \in X^* | \xi_t = x) \geq p_m^n$
so let $\alpha_t = p_m^n$ and $\beta_t = p_m (1 - p_m)^{n-1}$

Ar

$$\beta_t \geq \sum_{x \notin X^*} P(\xi_{t+1} \in X^* | \xi_t = x) \frac{P(\xi_t = x)}{\sum_{x \notin X^*} P(\xi_t = x)} \geq \alpha_t$$

2. Assume $P_{\max} \geq P(\xi_{t+1} \in X^* | \xi_t = x) \geq P_{\min}$ for all x, then

   $$P_{\max} \geq \sum_{x \notin X^*} P(\xi_{t+1} \in X^* | \xi_t = x) \frac{P(\xi_t = x)}{\sum_{x \notin X^*} P(\xi_t = x)} \geq P_{\min}$$

   so, we can let $\beta_t = P_{\max}$ and $\alpha_t = P_{\min}$

3. The probability that a solution is mutated to another one with k bits difference is $p_m^k (1 - p_m)^{n-k}$, which has an upper bound $p_m (1 - p_m)^{n-1}$ and a lower bound p_m^n

4. Thus for all t: $p_m (1 - p_m)^{n-1} \geq P(\xi_{t+1} \in X^* | \xi_t = x) \geq p_m^n$
   so let $\alpha_t = p_m^n$ and $\beta_t = p_m (1 - p_m)^{n-1}$

5. So reach the conclusion
   $\mathbb{E}[\tau] \leq \alpha_0 + \sum_{t=2}^{+\infty} t \alpha_{t-1} \prod_{i=0}^{t-2} (1 - \alpha_i) = p_m^{-n}$
   $\mathbb{E}[\tau] \geq \beta_0 + \sum_{t=2}^{+\infty} t \beta_{t-1} \prod_{i=0}^{t-2} (1 - \beta_i) = p_m^{-1} (1 - p_m)^{1-n}$
   note that both bounds are exponential in n

Analysis on EA-2

EA-2:

- Mutation 1: flip each bit with probability $p_m \in (0,0.5]$
- Population size M $\geq$ 1

- Result:
  - Population can reduce the steps needed, but can not help to reduce the complexity

Learning And Mining from DatA
http://lamda.nju.edu.cn

Analysis on EA-2 (con't)

Proof:

1. Estimate $\{\alpha_t\}_{t=0}^{+\infty}$ and $\{\beta_t\}_{t=0}^{+\infty}$ according to

$$\beta_t \geq \sum_{x \notin X^*} P(\xi_{t+1} \in X^* | \xi_t = x) \frac{P(\xi_t = x)}{\sum_{x \notin X^*} P(\xi_t = x)} \geq \alpha_t$$

2. Assume $P_{\max} \geq P(\xi_{t+1} \in X^* | \xi_t = x) \geq P_{\min}$ for all x, then

$$P_{\max} \geq \sum_{x \notin X^*} P(\xi_{t+1} \in X^* | \xi_t = x) \frac{P(\xi_t = x)}{\sum_{x \notin X^*} P(\xi_t = x)} \geq P_{\min}$$

   so, we can let $\beta_t = P_{\max}$ and $\alpha_t = P_{\min}$

3. The probability that a solution is mutated to another one with k bits difference is $p_m^k(1-p_m)^{n-k}$, which has an upper bound $p_m(1-p_m)^{n-1}$ and a lower bound p_m^n

4. The probability that a population is mutated to the optimal population is upper bounded by $1-(1-p_m(1-p_m)^{n-1})^M$ and

3. The probability that a solution is mutated to another one with k bits difference is $p_m^k(1-p_m)^{n-k}$, which has an upper bound $p_m(1-p_m)^{n-1}$ and a lower bound p_m^n

4. The probability that a population is mutated to the optimal population is upper bounded by $1-(1-p_m(1-p_m)^{n-1})^M$ and lower bounded by $1-(1-p_m^n)^M$

5. Thus for all t: $1-(1-p_m(1-p_m)^{n-1})^M \geq$

$$P(\xi_{t+1} \in X^* | \xi_t = x) \geq 1-(1-p_m^n)^M$$

so let $\alpha_t = 1-(1-p_m^n)^M$ and $\beta_t = 1-(1-p_m(1-p_m)^{n-1})^M$

6. So reach the conclusion

$$\frac{1}{1-(1-p_m^n)^M} \geq \mathbb{E}[\tau] \geq \frac{1}{1-(1-p_m(1-p_m)^{n-1})^M}$$

7. To simply the conclusion, apply $(1-x)^n \sim 1-nx$ as $x \to 0$

$$\frac{1}{M}\frac{1}{p_m^n} \geq \mathbb{E}[\tau] \geq \frac{1}{M}\frac{1}{p_m(1-p_m)^{n-1}}$$

Analysis on EA-3

EA-3:

- Mutation 1: flip each bit with probability $p_m \in (0,0.5]$
- Recombination: exchange the first σ bits of two solutions
- Population size M=2

- **Result:**
  - The complexity is lower bounded by $\Theta\left(\left(\frac{1}{1-p_m}\right)^n\right)$
  - The reason why recombination is not very helpful is that the situation where recombination is useful in this task rarely occurs

Analysis on EA-3 (con't)

Proof:

1. The population set where recombination is useful is

$$X_c = \{x | x \in X \wedge P(\xi_{t+c} \in X^* | \xi_t = x) > 0\}$$

2. At the initial state:

$$P(X_c) \leq \sum_{\sigma=1}^{n-1} \left(\sum_{i=\sigma}^{n-1} \frac{1}{2^i} \right) \left(\sum_{j=1}^{\sigma} \frac{1}{2^{n-j}} \right) < \frac{3(n-1)}{2^{n+1}}$$

note that $P(X_c)$ will not increase, because feasible populations out of X_c have lower fitness than those in X_c

3. According to Corollary 1 in [He & Yu, JSA01]:

$$\beta_t = \beta_{D,t}\beta_{c,t} + \beta_{m,t} - \beta_{D,t}\beta_{c,t}\beta_{m,t}$$

where $\beta_{D,t} \geq P(X_c)$, $\beta_{c,t} \geq P(\xi_{t+c} \in X^* | \xi_t = x)$

and $\beta_{m,t} \geq P(\xi_{t+m} \in X^* | \xi_t = x)$

4. Let ... $3(n-1)$ ... $= 1$ ... $= 2n \; (1 - \ldots)^{n-1}$

Ar

om DatA

.edu.cn

$$P(X_c) \le \sum_{\sigma=1}^{n-1}\left(\sum_{i=\sigma}^{n-1}\frac{1}{2^i}\right)\left(\sum_{j=1}^{\sigma}\frac{1}{2^{n-j}}\right) < \frac{3(n-1)}{2^{n+1}}$$

note that $P(X_c)$ will not increase, because feasible populations out of X_c have lower fitness than those in X_c

3. According to Corollary 1 in [He & Yu, AIJ01]:

$$\beta_t = \beta_{D,t}\beta_{c,t} + \beta_{m,t} - \beta_{D,t}\beta_{c,t}\beta_{m,t}$$

where $\beta_{D,t} \ge P(X_c)$, $\beta_{c,t} \ge P(\xi_{t+c} \in X^* | \xi_t = x)$

and $\beta_{m,t} \ge P(\xi_{t+m} \in X^* | \xi_t = x)$

4. Let $\beta_{D,t} = \frac{3(n-1)}{2^{n+1}}$ $\beta_{c,t} = 1$ $\beta_{m,t} = 2p_m(1-p_m)^{n-1}$

5. Then $\beta_t = \beta_{D,t}\beta_{c,t} + \beta_{m,t} - \beta_{D,t}\beta_{c,t}\beta_{m,t}$

$$= \frac{3(n-1)}{2^{n+1}} + 2p_m(1-p_m)^{n-1} \in \Theta((1-p_m)^n)$$

6. Obtain

$$\mathbb{E}[\tau] \ge \Theta\left(\left(\frac{1}{1-p_m}\right)^n\right)$$

Analysis on EA-4

EA-4:

- Mutation 2: flip each bit with probability $(0.5 - d)e^{-t} + d$ where $d \in (0,0.5)$ and $t = 0,1,...$
- Population size M=1

- Result:
  - The complexity is lower bounded by $\left(\frac{2}{1-d}\right)^{n-1}$
  - Previous approaches such as the analytic approach [He & Yao, AIJ03] can hardly be used to analyze this EA

Analysis on EA-4 (con't)

Proof:

1. Estimate $\{\alpha_t\}_{t=0}^{+\infty}$ and $\{\beta_t\}_{t=0}^{+\infty}$ according to

$$\beta_t \geq \sum_{x \notin X^*} P(\xi_{t+1} \in X^* | \xi_t = x) \frac{P(\xi_t = x)}{\sum_{x \notin X^*} P(\xi_t = x)} \geq \alpha_t$$

2. Assume $P_{\max} \geq P(\xi_{t+1} \in X^* | \xi_t = x) \geq P_{\min}$ for all x, then

$$P_{\max} \geq \sum_{x \notin X^*} P(\xi_{t+1} \in X^* | \xi_t = x) \frac{P(\xi_t = x)}{\sum_{x \notin X^*} P(\xi_t = x)} \geq P_{\min}$$

   so, we can let $\beta_t = P_{\max}$ and $\alpha_t = P_{\min}$

3. The probability that a solution is mutated to another one with k bits difference is $\left((0.5-d)e^{-t}+d\right)^k (1-(0.5-d)e^{-t}-d)^{n-k}$, which is upper bounded by

$$((0.5-d)e^{-t}+d)(1-(0.5-d)e^{-t}-d)^{n-1} < 0.5(1-d)^{n-1}$$

4. Thus for all t: $P(\xi_{t+1} \in X^* | \xi_t = x) \leq 0.5(1-d)^{n-1}$

An

$$\beta_t \geq \sum_{x \notin X^*} P(\xi_{t+1} \in X^* | \xi_t = x) \frac{P(\xi_t = x)}{\sum_{x \notin X^*} P(\xi_t = x)} \geq \alpha_t$$

2. Assume $P_{\max} \geq P(\xi_{t+1} \in X^* | \xi_t = x) \geq P_{\min}$ for all x, then

$$P_{\max} \geq \sum_{x \notin X^*} P(\xi_{t+1} \in X^* | \xi_t = x) \frac{P(\xi_t = x)}{\sum_{x \notin X^*} P(\xi_t = x)} \geq P_{\min}$$

   so, we can let $\beta_t = P_{\max}$ and $\alpha_t = P_{\min}$

3. The probability that a solution is mutated to another one with k bits difference is $\left((0.5-d)e^{-t}+d\right)^k (1-(0.5-d)e^{-t}-d)^{n-k}$, which is upper bounded by

$$((0.5-d)e^{-t}+d)(1-(0.5-d)e^{-t}-d)^{n-1} < 0.5(1-d)^{n-1}$$

4. Thus for all t: $P(\xi_{t+1} \in X^* | \xi_t = x) \leq 0.5(1-d)^{n-1}$

   so let $\beta_t = 0.5(1-d)^{n-1}$

5. So reach the conclusion

$$\mathbb{E}[\tau] \geq \left(\frac{2}{1-d}\right)^{n-1}$$

m DatA

du.cn

Summary

LAMDA
Learning And Mining from DatA
http://lamda.nju.edu.cn

A new approach to estimating the expected FHT

by establishing a bridge between the expected FHT and the convergence rate

- ✓ Does not need extra information, e.g. distance to the target
- ✓ Able to analyze EAs that cannot be analyzed before, e.g. EAs with dynamic operators

As an illustration, we have applied the approach to analyze 4 kinds of EAs

in particular, EA-4 involves time-variant mutation, which can hardly be analyzed with previous approaches

The second part of the talk has not got published, which will be shared in the near future … …

Remarks

Although EC has achieved great success in many real-world tasks, its theoretical foundation are largely underdeveloped

Besides designing powerful EAs and building successful applications, theoretical aspects of EC should attract more attention

Thanks!

Q/A?