Fitness Approximation in Evolutionary Computation

Yaochu Jin
Honda Research Institute Europe

Khaled Rasheed
University of Georgia

Motivations

- No explicit fitness function exists: to define fitness quantitatively
- Fitness evaluation is highly time-consuming: to reduce computational time
- Fitness is noisy: to cancel out noise
- Fitness is highly rugged: to smoothen the fitness landscape
- Search for robust solutions: to avoid additional expensive fitness evaluations

How To Estimate Fitness

- *Ad hoc* methods
  - Fitness inheritance (from parents)
  - Fitness imitation (from brothers and sisters)
- Problem approximation
  - To replace experiments with simulations
  - To replace full simulations with simplified simulations
- Data-driven functional approximation (*meta-model*)
  - Polynomials
  - Neural networks, e.g., multilayer perceptrons (MLPs), RBFN
  - Gaussian processes, Kriging models
  - Support vector machines

Fitness Inheritance and Imitation

- To estimate the fitness of an individual from that of its parents
  - plain average of the parents' fitness values (Smith et al, 1995, Sastry et al, 2002, Chen et al, 2002)
  - weighted sum of parents' fitness values (Smith et al, 1995, Salami et al, 2003)

$f = (s_1 r_1 f_1 + s_2 r_2 f_2)/(s_1 r_1 + s_2 r_2)$ (if $r_1=1$, $f=f_1$; if $r_2=1$, $f=f_2$)

s_1, s_2: similarity between the offspring and parents 1, 2
f_1, f_2: fitness of parents 1, 2
r_1, r_2: reliability of parents 1, 2

$r = [(s_1 r_1)^2 + (s_2 r_2)^2]/(s_1 r_1 + s_2 r_2)$

$r=1$ if an individual is evaluated with the original fitness function

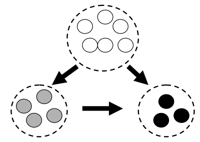
- To estimate the fitness of an individual from that other individuals of the same generation (Kim et al, 2001, Jin et al, 2004)

Issues in Choosing Meta-models

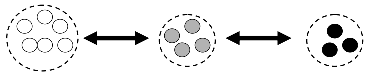
- Global model or local model
  - ➢ For high-dimensional problems, local models are more practical
- Deterministic or stochastic models
  - ➢ Stochastic models such as Gaussian models are ale to provide an estimate and an error bound of the estimate
  - ➢ Ensemble based error bound estimation
- Online learning capability
  - ➢ model is able to update recursively

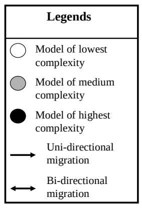
Use Meta-models: Multiple Populations

- Injection island model (Eby et al, 1998)



- Hierarchical model (Sefrioui and Periaux, 2000)



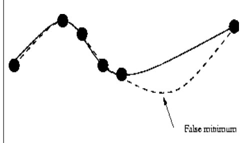


Meta-model in Initialization, Crossover and Mutations

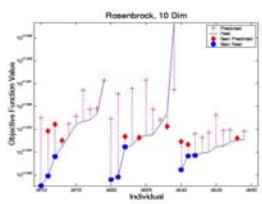
- Generate a plurality of individuals randomly and choose the best ones according to the meta-model for the initial population
- Informed crossover (Rasheed and Hirsch, 2000)
  - ➢ Generate multiple offspring individuals (instead of 2) using crossover
  - ➢ Keep the best two ones according to the meta-model
- Informed mutation (Rasheed et al, 2000; Abboud et al, 2002)
  - ➢ Generate multiple individuals randomly
  - ➢ Choose the best according to the meta model

Meta-model in Fitness Evaluations (I)

- Use meta-models only: Risk of false convergence



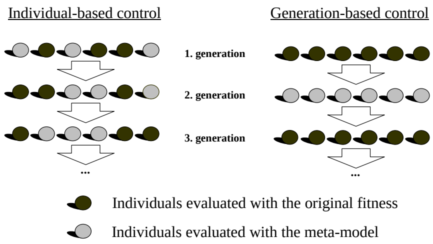
Inconsistent estimation error



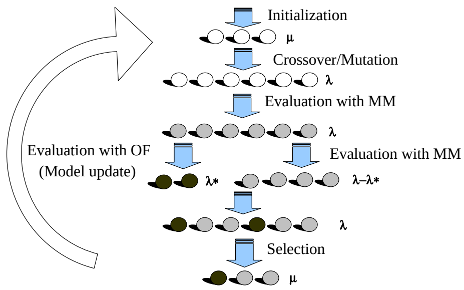
Example of inconsistent error

Meta-model in Fitness Evaluations (II)

- Use meta-models together with the original fitness (if available)



Individual-based Evolution Control

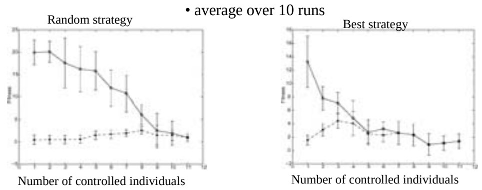


Evolution Control Strategies

- Choose individuals randomly (Jin et al, 2000)
- Choose the best individuals according to the model (Jin et al, 2001; Jin et al, 2002a)
- Choose the potentially best individuals with the help of estimated error bound (Emmerich et al, 2002; Ulmer et al, 2003)
- Choose the most uncertain individuals (Branke, 2005)
- Choose the representative individuals (Kim et al 2001; Jin et al, 2004)

Random and Best Strategy

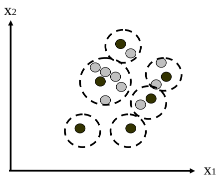
- 12-D Ackley function
- (3,12)-ES
- average over 10 runs



- The best strategy is more efficient than the random strategy
- In the best strategy, about half of the individual should be controlled to guarantee correct convergence

Clustering Method

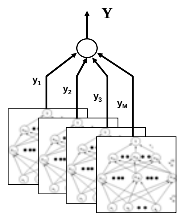
- Group the population
- Choose the individual closest to the cluster center (Jin et al, 2004)
- Choose the best individual of each cluster (Gräning et al, 2005)

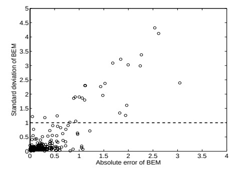


Tutorial, GECCO'05, Washington D.C.

Meta-model Improvement

- Neural network ensemble (Jin et al, 2004)





- Good correlation between error and variance
- Accept prediction only if the variance is below a certain threshold

Tutorial, GECCO'05, Washington D.C.

Choose Uncertain Individuals

- Uncertainty measure (Branke et al, 2005)
  - Polynomials are used for meta-modeling
  - Define the following uncertainty measure:

    $u_j = 1/\Sigma^{K}_{i=1} 1/d_{ij}$

    d_{ij}: Euclidean distance between data point i and j;

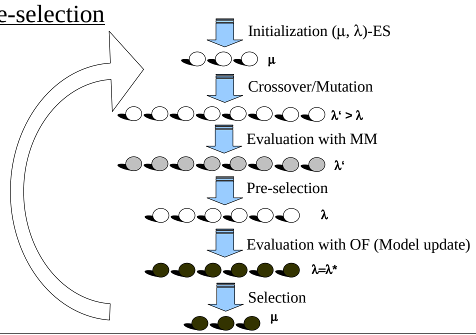
    K: Number of data points in the neighborhood
  - Choose individuals for re-evaluation to minimize the following criterion

    $r = \Sigma u_i$
  - A combination of quality and uncertainty

    $r'_j = \Sigma p_i u_i$, p_i: the probability of an individual to be selected

Tutorial, GECCO'05, Washington D.C.

Pre-selection



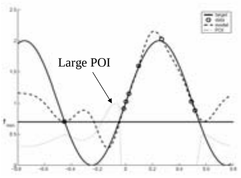
Tutorial, GECCO'05, Washington D.C.

Pre-selection Variants

- Use RBF network and pre-selection based on fitness only (Won et al, 2003)
- Use Gaussian model, considering both fitness and uncertainty (Emmerich et al, 2002)

$\underline{f} = f(x_i) - \alpha\ \sigma(x_i)$

- Choose individuals having a large probability of improve (POI) (Ulmer et al, 2003)



(Extracted from Ulmer et al, 2003)

Tutorial, GECCO'05, Washington D.C.

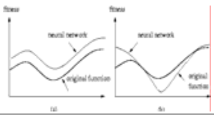
Adaptation in Individual-based Evolution Control

- Adaptation of **λ' (λ' ≥ λ)** in pre-selection
- Adaptation of **λ* (1 ≤ λ* ≤ λ)** in best strategy or clustering methods
- Adaptation based on the quality of the meta-model
  - Increase **λ* / λ'** if the quality of the meta-model becomes better or better than a certain criterion
  - Decrease **λ* / λ'** if the quality of the model becomes worse, or worse than a certain criterion

Tutorial, GECCO'05, Washington D.C.

Performance Metrics

- Approximation error: How good the meta-model approximates the fitness function
- Selection-based measure: For a (μ, λ) –ES, a meta-model is "perfect" if the same individuals are selected using meta-model or the original fitness function (Jin et al, 2002b; Hüsken et al, 2005)
- Calculate the correlation between the fitness value of the original fitness function and the meta-model (Jin et al, 2002b)
- Calculate the rank correlation (Jin et al, 2002b)



Tutorial, GECCO'05, Washington D.C.

Approximation Error

- Compare the average error of the controlled individuals (Jin et al, 2002a)
- Compare the distribution of the error of the individuals (Gräning, 2005)

$E^{(NSE)}(t-1)$: 0.1 | 0.2 | 0.4 | 1.5

$E^{(NSE)}(t)$: 0.2 | 0.5 | 0.1

0.1	0.1	0.2	0.2	0.4	0.5	1.5
1	2	3	4	5	6	7

$\underbrace{1 \quad 2}_{1.5} \quad \underbrace{3 \quad 4}_{3.5}$

$\downarrow \Sigma$ $R(t-1)$ $\qquad$ $\downarrow \Sigma$ $R(t)$

$$\rho^{(NERD)} = \frac{R(t-1)}{\lambda^*_{t-1}} - \frac{R(t)}{\lambda^*_t}$$

$$\lambda'_{t+1} = \lambda'_t + \rho^{(NERD)} \cdot \Delta\lambda$$

$$\lambda^*_{t+1} = \lambda^*_t - \rho^{(NERD)} \cdot \Delta\lambda$$

Correct Selection

- Assign a grade of λ-*i* if the *i*-th best individual is correctly selected (Jin et al, 2002b; Hüsken et al, 2005)

$$\rho^{(Sel)} = \sum_{i=1}^{m} \left(\lambda_t^* - i\right),$$

- The maximal grade that can be obtained is

$$\rho_{\max}^{(Sel)} = \sum_{i=1}^{\mu} \left(\lambda_t^* - i\right)$$

- If μ individuals are selected randomly:

$$\left\langle \rho^{(rand)} \right\rangle = \frac{\mu^2}{\lambda_t^*} \cdot \frac{2\lambda_t^* - \mu - 1}{2}$$

- If the model is better than random guess ($\rho_t^{(sel)} > \rho^{(rand)}$):

$$\lambda_{t+1}^* = \lambda_t^* - \frac{\rho_t^{(sel)} - \left\langle \rho^{(rand)} \right\rangle}{\rho^{(\max)} - \left\langle \rho^{(rand)} \right\rangle} \cdot \Delta\lambda$$

HRI Europe Honda Research Institute

Correlation

- Fitness based correlation:

$$\rho^{(rank)} = 1 - \frac{6\sum_{l=1}^{\lambda_t^*} \left(r_l - \hat{r}_l\right)^2}{\lambda_t^* \left(\left(\lambda_t^*\right)^2 - 1\right)}$$

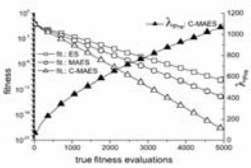
- Rank based correlation:

$$\rho^{(corr)} = \frac{\frac{1}{\lambda_t^*} \sum_{l=1}^{\lambda_t^*} \left(\phi_l^{(MM)} - \bar{\phi}^{(MM)}\right)\left(\phi_l^{(OF)} - \bar{\phi}^{(OF)}\right)}{\sigma^{(MM)} \sigma^{(OF)}}$$

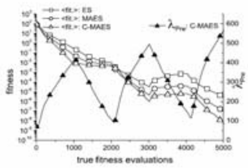
HRI Europe Honda Research Institute

Adaptive Pre-selection (I)

- Impressive results have been obtained in adaptive pre-selection using selection based criterion (Ulmer et al, 2004)



10D Sphere function, no noise



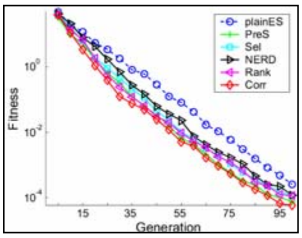
10D Sphere function, changing noise level

(Abstracted form Ulmer et al, 2004)

HRI Europe Honda Research Institute

Adaptive Pre-selection (II)

- Not that successful results (Gräning, 2005)

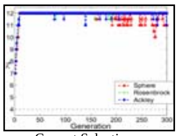


Sphere function

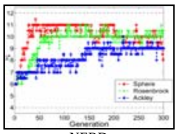
HRI Europe Honda Research Institute

Tutorial, GECCO'05, Washington D.C.

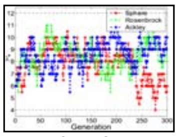
Adaptive Pre-selection (III)



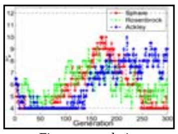
Correct Selection



NERD



Rank Correlation



Fitness correlation

HRI Europe Honda Research Institute

Tutorial, GECCO'05, Washington D.C.

Generation-based Evolution Control

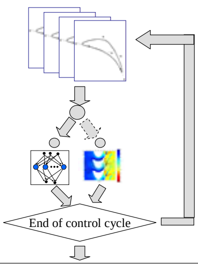
- Evolution control with a fixed frequency (Bull, 1999)
- Evolution control after convergence (Ratle, 1998)
- Evolution control with an adaptive control frequency (Jin et al, 2002a)
- Evolution control after convergence combined with uncertainty measure (Büche et al, 2005)

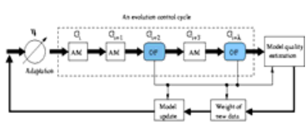
HRI Europe Honda Research Institute

Tutorial, GECCO'05, Washington D.C.

Control Frequency Adaptation

- Adaptation of control frequency based on error estimation (Jin et al, 2001; Jin et al, 2002a)





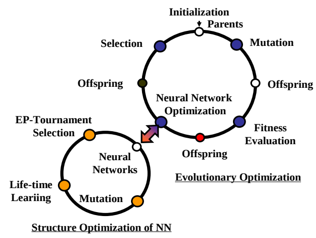
- Estimation of model fidelity
- Adaptation of control frequency
- On- line model update

HRI Europe Honda Research Institute

Tutorial, GECCO'05, Washington D.C.

Meta-model Improvement

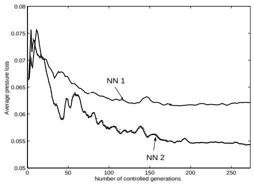
- Online structure optimization of neural networks (Hüsken et al, 2005)



HRI Europe Honda Research Institute

Application Example

- 2-D aerodynamic design optimization
- Bspline representation with 23 control points (46 parameters)
- Evolution strategy (2, 11) CMA-ES
- NN1: fully connected NN; NN2: structurally optimized NN



Surrogate Approach

- Generate a surrogate with initial data (Büche et al, 2005)

- Search on the meta-model until converges, restricting the search within the neighborhood of the current best solution:

$x^b - d/2 \leq x \leq x^b + d/2$
$d_i = \max(x_i) - \min(x_i),\ x_i \in$ NC closest neighbors

- Train the model using the NC closest data and NR most recently evaluated data to prevent the model from getting stuck

Local Search Using Meta-model

- Meta-model is used and updated only in local search (Ong et al, 2003)

- The trust-region framework for managing meta-models is applied to each individual during local search
  - The "trust-region" is a range in which the meta-model is trustful, first initialized as the range of the data used to construct the meta-model

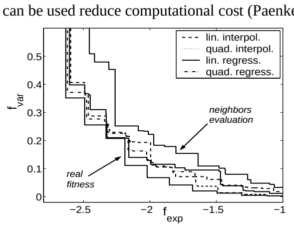
  - The radius of the trust-region is reduced if the accuracy of the meta-model is low; enlarged if the accuracy is high, otherwise not changed (Dennis and Torczon, 1997)

  - The result of local search is coded back to the chromosome using the Lamarckian mechanism

Meta-model in Search for Robust Solutions

- Averaging based approach to robust solutions need a large number of additional fitness evaluations

- Meat-model can be used reduce computational cost (Paenke et al, 2005)

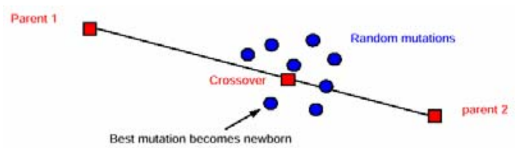


Informed Operators

•Idea: replace randomness with decisions informed by the surrogate model
•Examples:
- ➢Informed initialization
- ➢Informed Crossover (parents, type)
- ➢informed mutation (type, amplitude)

Informed mutation

-Crossover is done, followed by several random mutations
-Random mutations are evaluated using meta model, best becomes newborn



OEGADO (Objective Exchange GADO) for two objectives (Chafekar *et al*., 2005)

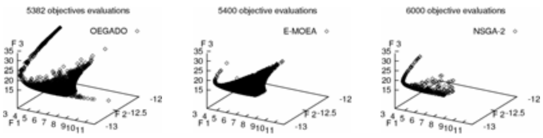
- 1. Both the GAs are run concurrently with each GA optimizing one of the two objectives while also forming a meta model of it.
- 2. At intervals equal to twice the population size, each GA exchanges its meta model with the other GA.
- 3. Informed operators are used. The IOs generate multiple children and use the meta model to compute the approximate fitness of these children. The best child is selected to be the newborn.
- 4. The true fitness function is then called to evaluate the actual fitness of the newborn corresponding to the current objective.
- 5. The individual is then added to the population using the replacement strategy.
- 6. Steps 2 through 5 are repeated till the maximum number of evaluations is exhausted.

OEGADO for three or more objectives

- 1. Each GA optimizes its objective and forms its own surrogate model.
- 2. After a given interval of evaluations each GA offers its meta model to one of the other GAs and obtains its meta model to use by its informed operators.
- 3. After the second interval each GA exchanges its meta model with one of the other remaining GAs.
- 4. This process continues and the GAs continue to exchange their meta models in a round-robin fashion.

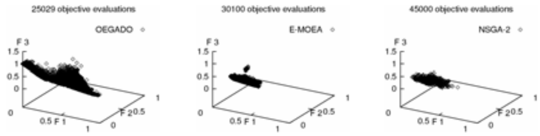
Results (3-objective problem Remy)

- OEGADO did better with a good distribution over the covered area
- NSGA II & εMOEA did not achieve a good distribution



Results (more difficult 3-objective DTLZ8)

- OEGADO did well with a good distribution over the covered area
- NSGA II & εMOEA did not achieve a good distribution



Summary

- Meta-modeling and other fitness approximation techniques have found a wide range of applications
- Proper control of meta-models plays a critical role in the success of using meta-models
- Proper choice of a meta-model: with/without error estimation, local/global
- Application of meta-models to multi-objective optimization, dynamic optimizations, search for robust solutions poses many challenging problems
- Stringent theoretical analysis of EA dynamics using meta-models is very desirable

Further Information

- On-line bibliography: http://www.soft-computing.de/amec.html
- Edited Book: Knowledge Incorporation in Evolutionary Computation, Springer, Berlin, 2005
- Special issues:
  - Special issue on "Approximation and Learning in evolutionary computation", *Soft Computing*, 9(1), 2005
  - Special issue on "Knowledge incorporation and extraction in evolutionary computation", *IEEE Transactions on Systems, Man, and Cybernetics, Part C – Applications and Reviews*, 2005
  - Special issue on "Evolutionary optimization in the presence of uncertainties", *IEEE Transactions on Evolutionary Computation* (to appear)
  - Special issue on "Evolutionary computation in dynamic and uncertain environments", Genetic Programming and Evolvable Machines (submission deadline:

Acknowledgements

Yaochu Jin would like to thank his colleagues in the ELTEC group at Honda Research Institute Europe,

Bernhard Sendhoff
Markus Olhofer
Martina Hasenjäger
Stefan Menzel

former collaborator,

Michael Hüsken

and former students

Ingo Paenke
Lars Gräning

Khaled Rasheed would like to thank his students in the Computer Science Dept. at the University of Georgia,

Deepti Chafekar
Liang Shi