Fitness Approximation in Evolutionary Computation

Yaochu Jin
Honda Research Institute Europe

Khaled Rasheed
University of Georgia

Tutorial, GECCO'05, Washington D.C.

Motivations

- No explicit fitness function exists: to define fitness quantitatively
- Fitness evaluation is highly time-consuming: to reduce computation 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

Tutorial, GECCO'05, Washington D.C.

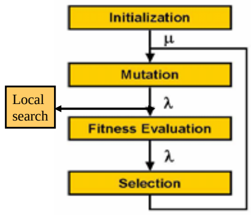
Fitness Approximation Methods

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

Tutorial, GECCO'05, Washington D.C.

How to Use Approximate Model

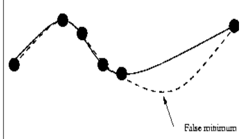
- Fitness evaluations
- Life-time learning (local search)
- Initialization, crossover, mutation
- Multiple populations



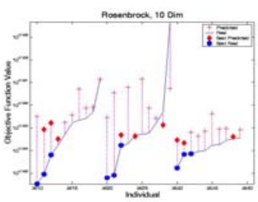
Tutorial, GECCO'05, Washington D.C.

Meta-model in Fitness Evaluations (I)

- Use meta-models only: Risk of false convergence



Inconsistent estimation error



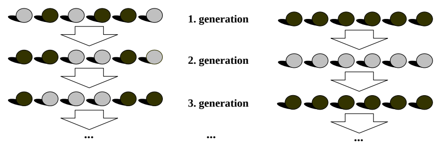
Example of inconsistent error

Tutorial, GECCO'05, Washington D.C.

Meta-model in Fitness Evaluations (II)

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

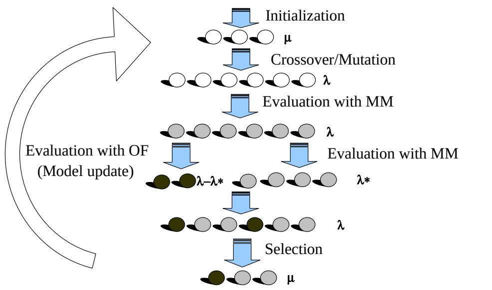
Individual-based control — Generation-based control



Individuals evaluated with the original fitness

Individuals evaluated with the meta-model

Individual-based Evolution Control

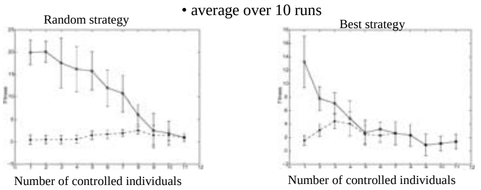


Evolution Control Methods (Model Management)

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

Random and Best Strategy

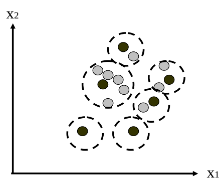
- 12-D Ackley function
- (3,12)-ES
- average over 10 runs



- 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

Choose Representatives Using 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)



Choose Most Uncertain Individuals

- Uncertainty measure (Branke et al, 2005)
  - Polynomials are used for meta-modeling
  - Define the following uncertainty measure: $u_j = \frac{1}{\sum_{i=1}^{k} 1/d_{ij}}$

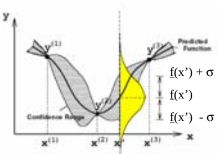
    d_{ij}: Euclidean distance between data point i and j;
    k: Number of data points in the neighborhood
  - Choose individuals for re-evaluation to minimize the following criterion: $r_j = \sum_{i \in Q \setminus \{j\}} u_i$
  - A combination of quality and uncertainty: $r'_j = \sum_{i \in Q \setminus \{j\}} p_i u_i$

    p_i: the probability of an individual selected

Choose (Potentially) Good Individuals

When a stochastic model is used,

- Mean fitness value: $f = \underline{f}\,(x')$;
- Lower confidence bound (Emmerich et al, 2002): $f = \underline{f}\,(x') - \alpha\sigma(x')$ $(\alpha>0)$

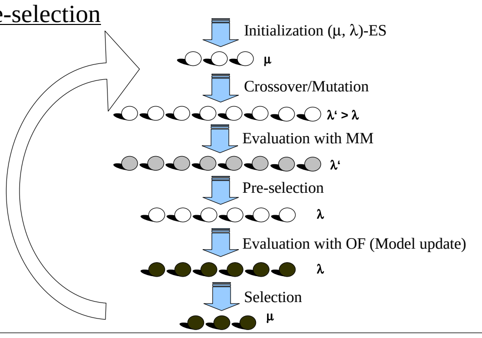


(Abstracted from Emmerich et al, 2005)

- PoI (Probability of Improvement) (Ulmer et al, 2003, Ong et al 2005)

  $PoI = \Phi((f_{min}-\underline{f}(x'))/\ \sigma(x'))$,
- Expected Improvement (Schonlau, 1998; Emmerich et al, 2005)

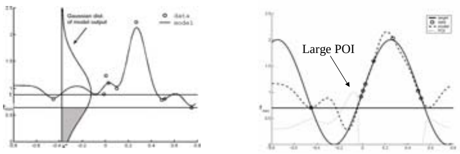
Pre-selection



Pre-selection Based on PoI

- Choose individuals having a large probability of improvement (PoI) (Ulmer et al, 2003)

POI = $\Phi((F-\underline{f}(x'))/\ \sigma(x))$, Φ -- a normal cumulative distribution function
$F \le f_{\mathbf{min}}$
POI: the probability of $Y < F$

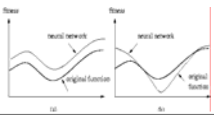


Adaptation in Individual-based Evolution Control

- 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

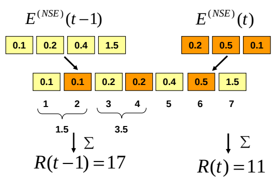
Performance Metrics

- 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)



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)



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

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

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

$$\rho^{(Sel)}_{\max} = \sum_{i=1}^{\mu} \left(\lambda^*_t - i\right)$$

- 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 ($\rho^{(sel)}_t > \rho^{(rand)}$):

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

Correlation

• Rank based correlation:

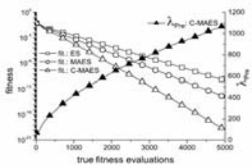
$$\rho^{(rank)} = 1 - \frac{6\sum_{l=1}^{\lambda_t^*}(r_l - \hat{r}_l)^2}{\lambda_t^*\left((\lambda_t^*)^2 - 1\right)}$$

• Fitness based correlation:

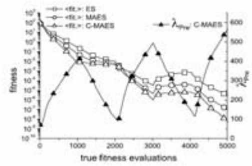
$$\rho^{(corr)} = \frac{\frac{1}{\lambda_t^*}\sum_{l=1}^{\lambda_t^*}\left(\phi_l^{(MM)} - \bar{\phi}^{(MM)}\right)\left(\phi_l^{(OF)} - \bar{\phi}^{(OF)}\right)}{\sigma^{(MM)}\sigma^{(OF)}}$$

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)

Generation-based Evolution Control

• Evolution control with a fixed frequency (Bull, 1999)

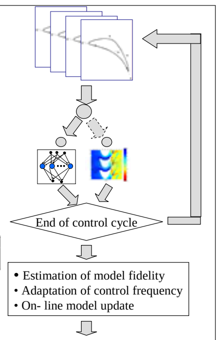
• Evolution control after convergence (Ratle, 1998)

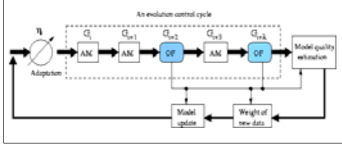
• Evolution control with an adaptive control frequency (Jin et al, 2000, 2002a)

• Evolution control after convergence combined with uncertainty measure (Büche et al, 2005)

Control Frequency Adaptation

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





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:

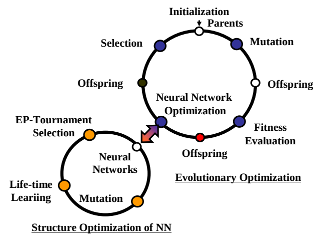
$$\mathbf{x}^{best} - \mathbf{d}/2 \leq \mathbf{x} \leq \mathbf{x}^{best} + \mathbf{d}/2,$$

$$d_i = \max_c(x_{c,i}) - \min_c(x_{c,i}), \quad x_{c,i} \in N_C \text{ closest neighbors}$$

• Train the model using the N_C closest data and N_R most recently evaluated data to prevent the model from getting stuck

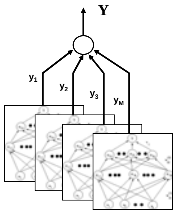
Meta-model Improvement (I)

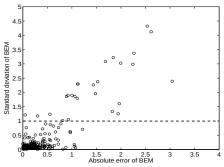
• Online structure optimization of neural networks (Hüsken et al, 2005)



Meta-model Improvement (II)

• Neural network ensemble (Jin et al, 2004)





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

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
 ➢ Adaptation of trust-region (Dennis and Torczon, 1997)

$$\rho^k = \frac{f(x_c^k) - f(x_{lo}^k)}{\hat{f}(x_c^k) - \hat{f}(x_{lo}^k)}$$

$$\Delta^{k+1} = \begin{cases} 0.25\,\Delta^k, & \text{if } \rho^k \le 0.25, \\ \Delta^k, & \text{if } 0.25 < \rho^k < 0.75, \\ \zeta\,\Delta^k, & \text{if } \rho^k \ge 0.75, \end{cases}$$

1) Calculate figure of merit 2) Adapt the trust-region size

Search for Robust Solutions

• Expected fitness (explicit averaging):

$$f_{\exp}(\mathbf{x}) = \int_{-\infty}^{\infty} f(\mathbf{x}+\delta)\cdot p(\delta)d\delta \qquad \hat{f}_{\exp}(\mathbf{x}^0) = \sum_{i=1}^{n} \frac{1}{n}\hat{f}(\mathbf{x}_i),$$

 ➢ Need additional fitness evaluations (n-1, n is the ample size)

 ➢ Use of approximate models could alleviate this difficulty

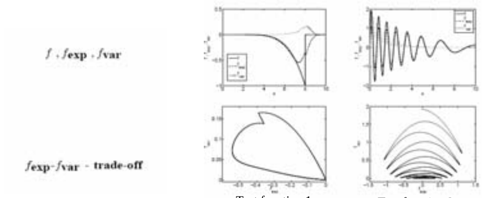
• Expected fitness and variance of the fitness (Multi-objective approach)

$$f_{\mathrm{var}}(\mathbf{x}) = \int_{-\infty}^{\infty} (f(\mathbf{x}+\delta) - f_{\exp}(\mathbf{x}))^2 \cdot p(\delta)d\delta. \qquad \hat{f}_{\mathrm{var}}(\mathbf{x}^0) = \sum_{i=1}^{n} \frac{1}{n}[\hat{f}(\mathbf{x}_i) - \hat{f}_{\exp}(\mathbf{x}^0)]^2$$

Meta-model for Search for Robust Solutions

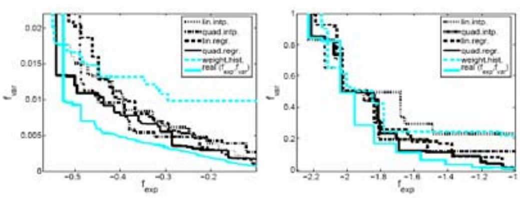
• Averaging based approach to robust solutions needs a large number of additional fitness evaluations

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



Test function 1 Test function 2

Meta-model for Search for Robust Solutions



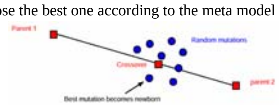
Test function 1 Test function 2

Meta-model in Initialization, Crossover, Mutations

• Generate a number 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 according to the meta-model

• Informed mutation (Rasheed et al, 2000; Abboud et al, 2002)
 ➢ Generate multiple individuals randomly
 ➢ Choose the best one according to the meta model



Ad hoc Methods: 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 $s_1=1$, $f=f_1$; if $s_2=1$, $f=f_2$)

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

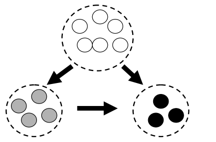
$r = [(s_1 r_1)^2 + (s_2 r_2)^2]/(s_1 r_1 + s_2 r_2)$

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

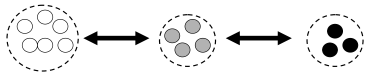
- To estimate the fitness of an individual from that of others in the same generation (Kim et al, 2001, Jin et al, 2004)

Use Meta-models: Multiple Populations

- Injection island model (Eby et al, 1998)



- Hierarchical model (Sefrioui and Periaux, 2000)



Legends

- Model of lowest complexity
- Model of medium complexity
- Model of highest complexity
- Uni-directional migration
- Bi-directional migration

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 incrementally

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

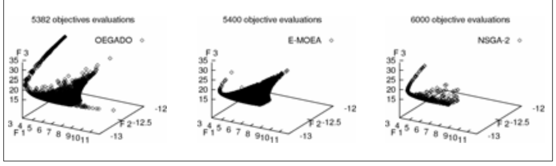
- 1. Both the GAs are run concurrently with each GA optimizing one of the two objectives while also forming a meta model of it.
- 2. At intervals equal to twice the population size, each GA exchanges its meta model with the other GA.
- 3. Informed operators are used. The IOs generate multiple children 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.

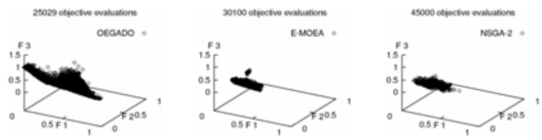
Results (3-objective problem Remy)

- 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
- Theoretical analysis of EA dynamics using meta-models is still missing

Further Information

- On-line bibliography: http://www.soft-computing.de/amec.html
- Survey papers:
  - Y. Jin. A comprehensive survey of evolutionary computation with fitness approximation. *Soft Computing*, 9(3), 2005
  - Y. Jin and J. Branke. Evolutionary optimization in uncertain environments. *IEEE Transactions on Evolutionary Computation*. 9(3):303-317, 2005
- Journal special issues:
  - Special issue on "Approximation and learning in evolutionary computation", *Soft Computing*, 9(1), 2005
  - Special issue on "Evolutionary optimization in the presence of uncertainties", *IEEE Transactions on Evolutionary Computation.* 2005
  - Special issue on "Evolutionary computation in dynamic and uncertain environments", Genetic Programming and Evolvable Machines (**Submission deadline**: July 31, 2005)

Acknowledgements

Yaochu Jin would like to thank his colleagues in the EL-TEC 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