Combating User Fatigue in iGAs: Partial Ordering, Support Vector Machines, and Synthetic Fitness

Xavier Llorà, Kumara Sastry, David E. Goldberg, Abhimanyu Gupta, and Lalitha Lakshmi

Illinois Genetic Algorithms Lab
University of Illinois at Urbana-Champaign

{xllora,kumara,deg,agupta21}@illigal.ge.uiuc.edu

Human and Social Aspects of GAs

Recombination agent	Selection agent: computational	Selection agent: human
computational	Standard Genetic Algorithms	Interactive Genetic Algorithms
human	Computer Aided Design (CAD)	Human-Based Genetic Algorithms
	computational	*human*

Selection agent

(Kosorukoff & Goldberg, 2002)

Motivation

- Interactive Genetic Algorithms
  - No quantitative fitness is available
  - Qualitative fitness is provided by the user
    - Sorting of several solutions
    - Choosing one of solution out of a subset
    - …
  - User fatigue (short time periods 1-2 hours)
  - Frustration (repeated evaluation of similar solutions)
  - Big time scale between user evaluations and the evolutionary mechanisms
- Efficiency enhancement for iGAs

Facets of an Interactive Genetic Algorithm

- ✖ Users need a clear idea of the outcome
  - *Need for a clear criteria of the goal to reach*
- ✖ A good visualization goes a long way
  - *A non-intuitive visualization may mislead user's evaluations*
- ✔ Lack of numerical fitness can be a problem
  - *No numeric form that can be optimize is available*
- ✔ User fatigue needs to be minimized
  - *User may only be able to provide reliable evaluations for short time periods (1-2 hours)*
- ✖ Users tend to change their criteria along the way
  - *Easy to maintain an unique criteria for short time periods*

Lack of a Real Fitness

- A minimal iGA scenario
  - Two solutions are presented for user evaluation
  - Three possible evaluation outcomes:
    1. The first is better than the second
    2. The second is better than the first
    3. Don’t know, don’t care
  - Evaluation by comparison
- The GA equivalent
  - Tournament selection s=2
- Can we compute a numerical fitness out of partial user evaluations?

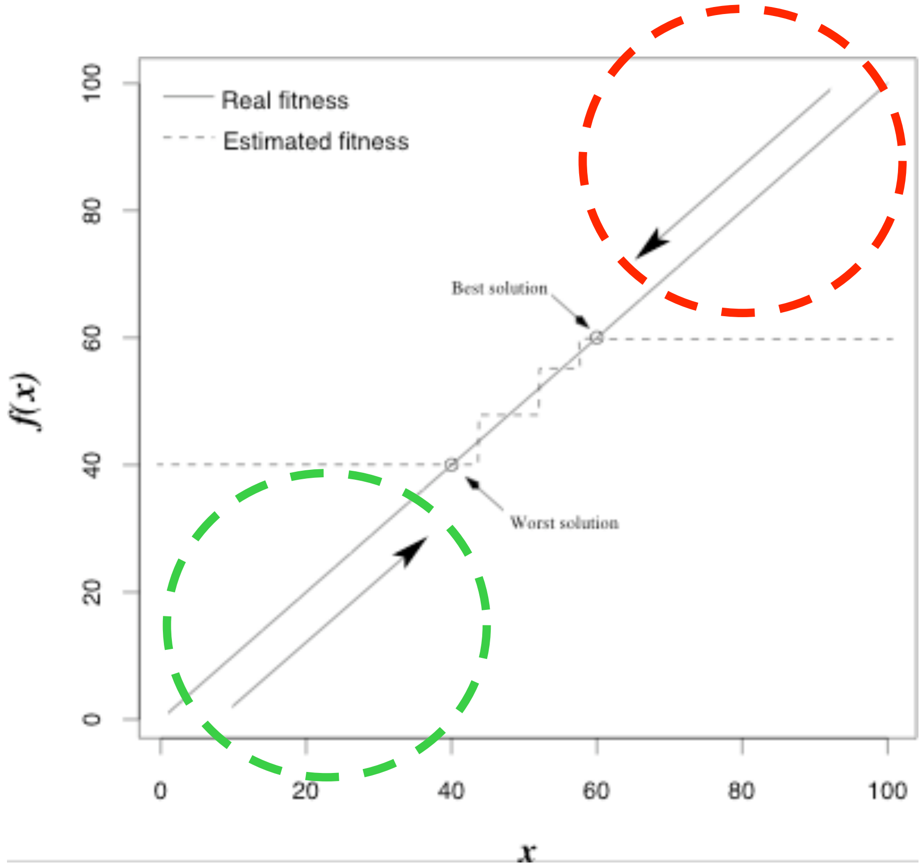
Properties of a Synthetic Fitness

- Solution-quality order should be maintained
  - Any synthetic fitness needs to maintain the solution ordering provided by the user

$$\text{If } s_1 \geq s_2 \geq \ldots \geq s_n \text{ then } f_s(s_1) \geq f_s(s_2) \geq \ldots \geq f_s(s_n)$$

- Synthetic fitness should allow extrapolation
  - Any synthetic fitness needs to be able to generalize the ordering relation beyond the available evaluations collected

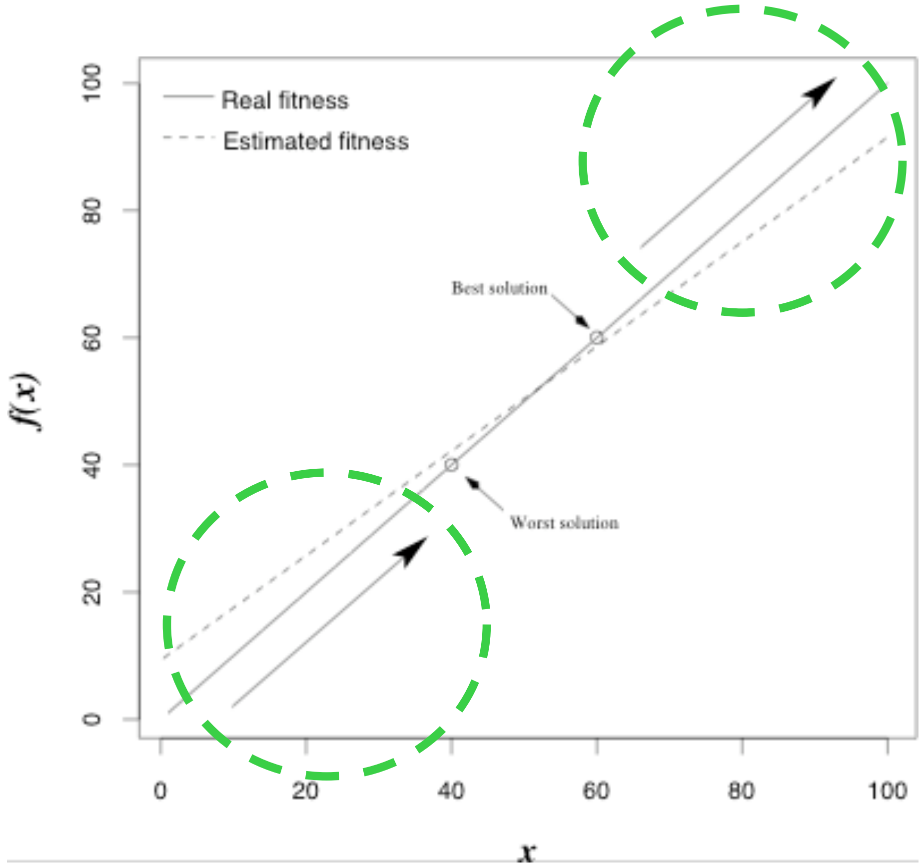
Fitness Extrapolation (I/II)



Nearest Neighbor

- Given a f_s: X → Y
  - X: problem attributes
  - Y: numeric value
- Set of evaluated solutions
- New solution
  - Compute the k-nearest neighbor
  - Assign the fitness of the weighted fitness of the k-nearest neighbors
- No extrapolation beyond the current limits

Fitness Extrapolation (II/II)

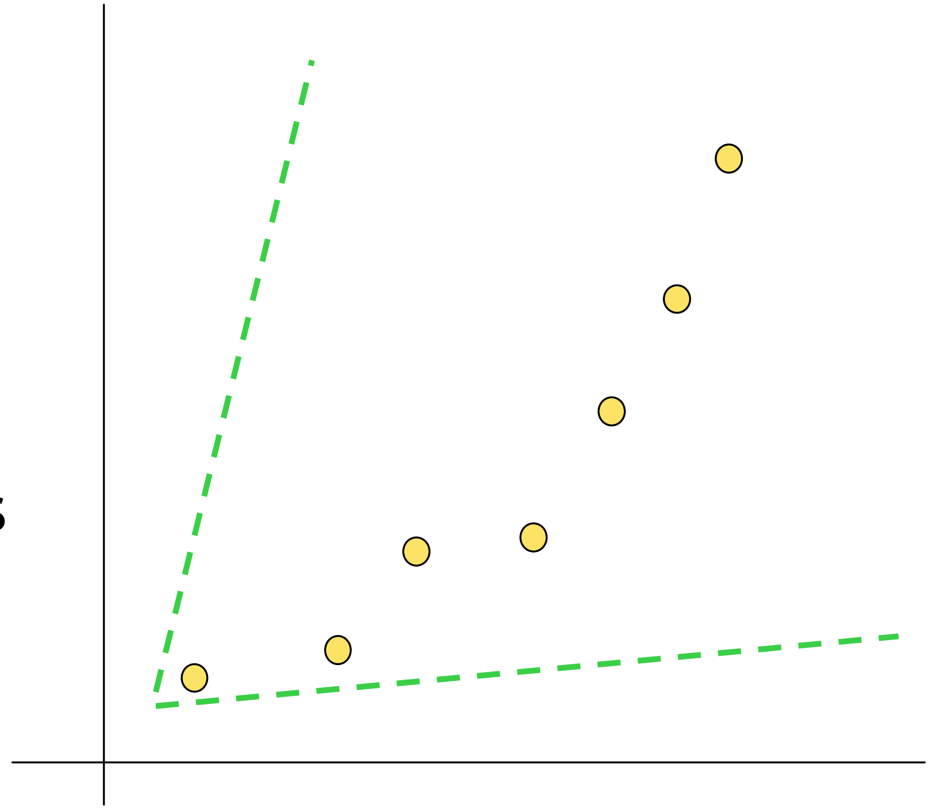


Support Vector Regression

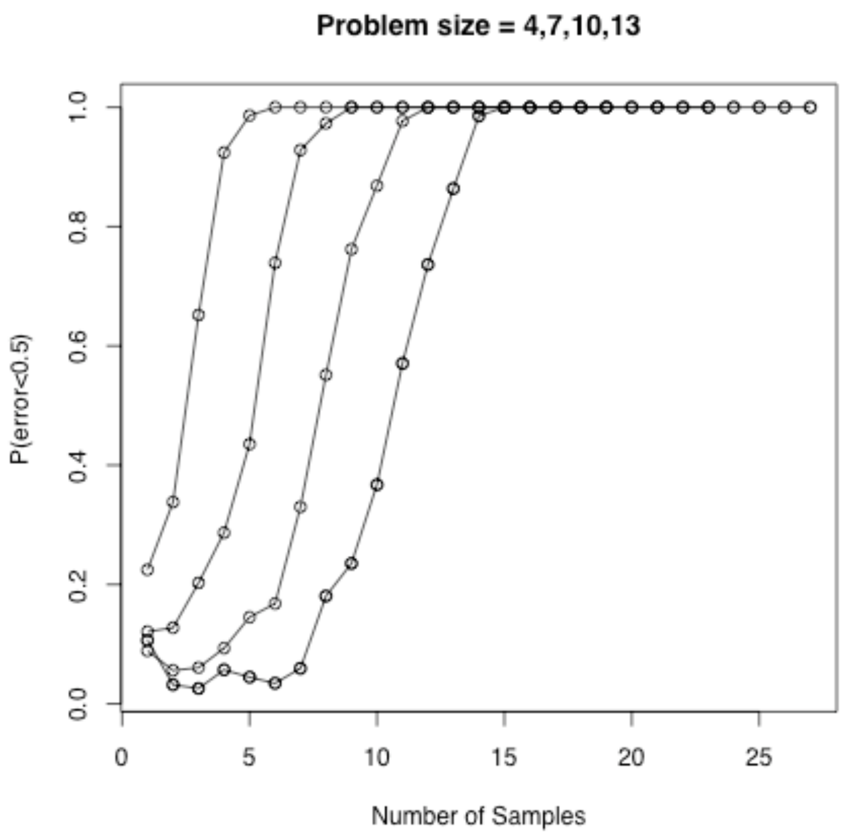
- Regression model
- Extrapolates beyond the evaluated solutions
- Provides a direction toward improvements
- How many solutions we need to have a reliable model?

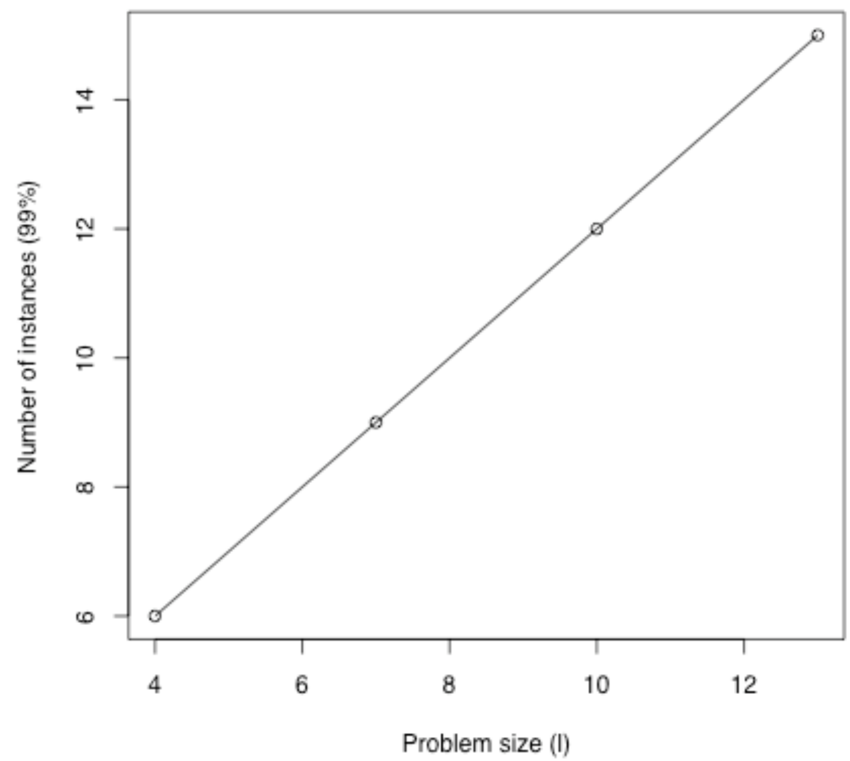
Order Maintenance

- $y = a*x + b$
- $a>1$ guaranties the order
- Tournament only cares about the partial order among solutions!
- Low cost, high error models



Accuracy of Fitness Extrapolation Using ε-SVM

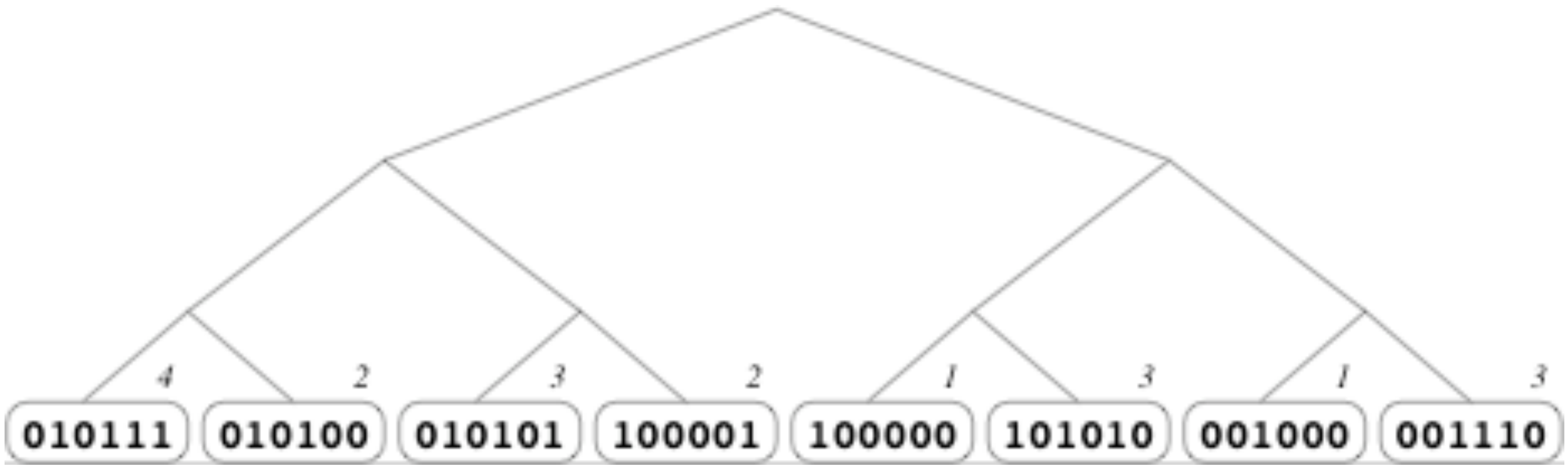




Yes, but...

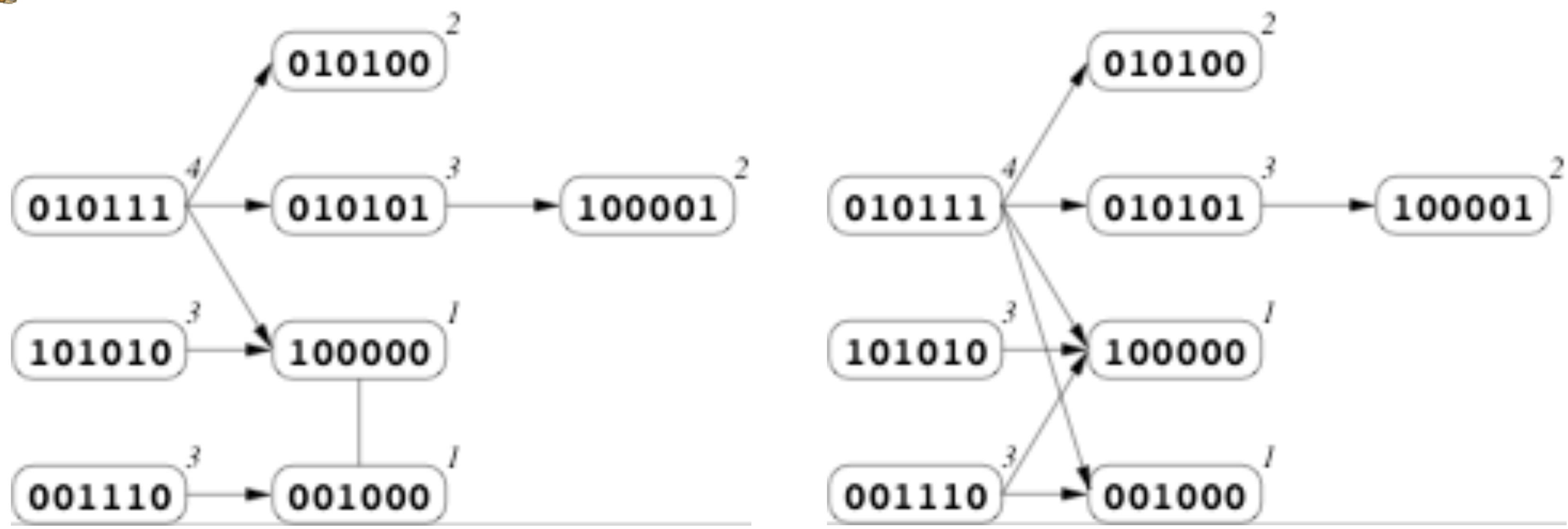
- If we have a numerical fitness:
  - We can build a regression model
  - We can use such model for combating fitness fatigue
- What is it available?
  - Partial ordering of solutions
  - Incremental refinement (new evaluations)
- The idea:
  - Use dominance measures on the graph of partially-ordered solutions
  - Build a map between the problem variables and a numeric ranking (Synthetic fitness)

Hierarchical Evaluation Order



- Given a set of solutions
  - They can be presented as a sequence of hierarchical tournaments
  - Obtain the partial ordering of the solutions
  - Such ordering can be expressed in a graph form

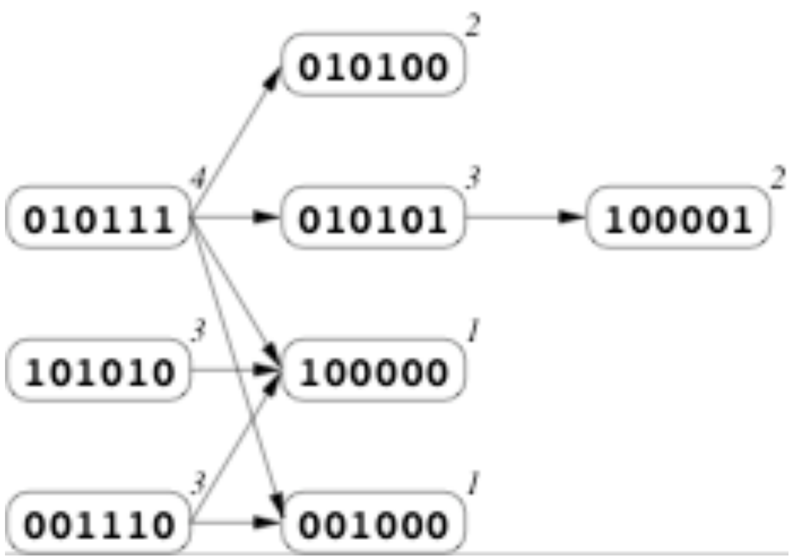
Partial Order Graph



- Nodes are solutions
- Edges represent the evaluation provided by the user
    1. The first is better than the second
    2. The second is better than the first
    3. Don't know, don't care
- Transformed to contain only 1 and 2 relations
- Property: cycles detect user contradictions in evaluations

The Numeric Fitness

- Dominance measure



Dominates — Dominated by — Difference

v	$f(v)$	$r(v)$	$\delta(v)$	$\phi(v)$	$\hat{f}(v)$	$\hat{r}(v)$
010111	4	1	5	0	5	1
010100	2	3	0	1	-1	4
010101	3	2	1	1	0	3
100001	2	3	0	2	-2	5
100000	1	4	0	3	-3	6
101010	3	2	2	0	2	2
001000	1	4	0	3	-3	6
001110	3	2	2	0	2	2

Real — Synthetic

Synthetic Fitness

1. Collect the user evaluations
2. Build the partial order graph
3. Compute the numeric fitness using the partial order graph
4. Use ε-SVM for creating a regression model
5. Use the learned regression as the synthetic fitness

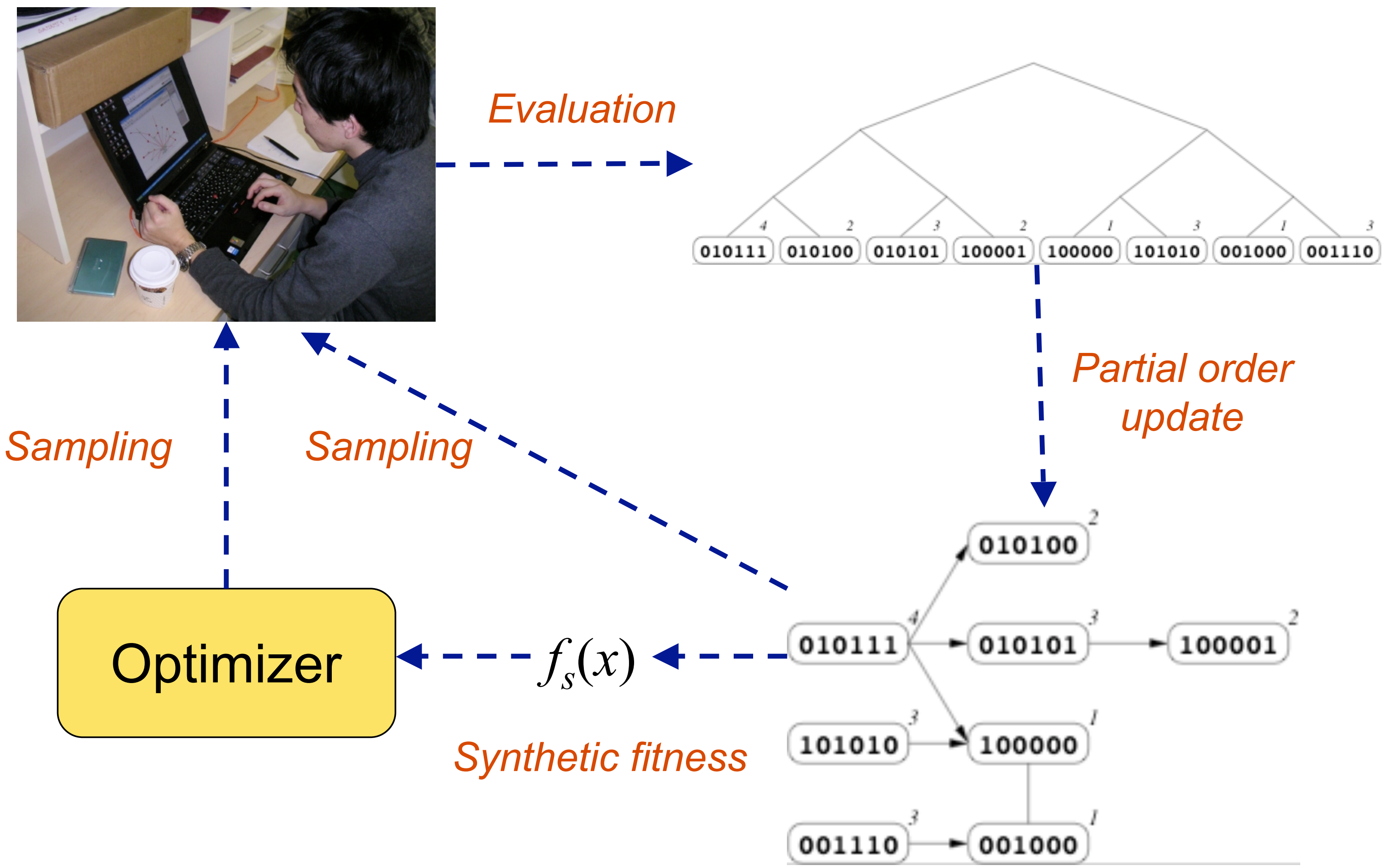
Fatigue

- Cut down the number of human evaluations

- Exploit synthetic fitness
  - Optimize it
  - Sample the best candidates
  - Show the best solutions to the user

- The sample best solutions help combating
  - Fatigue (educated guess of user preference)
  - Frustration (produce new eureka solutions)

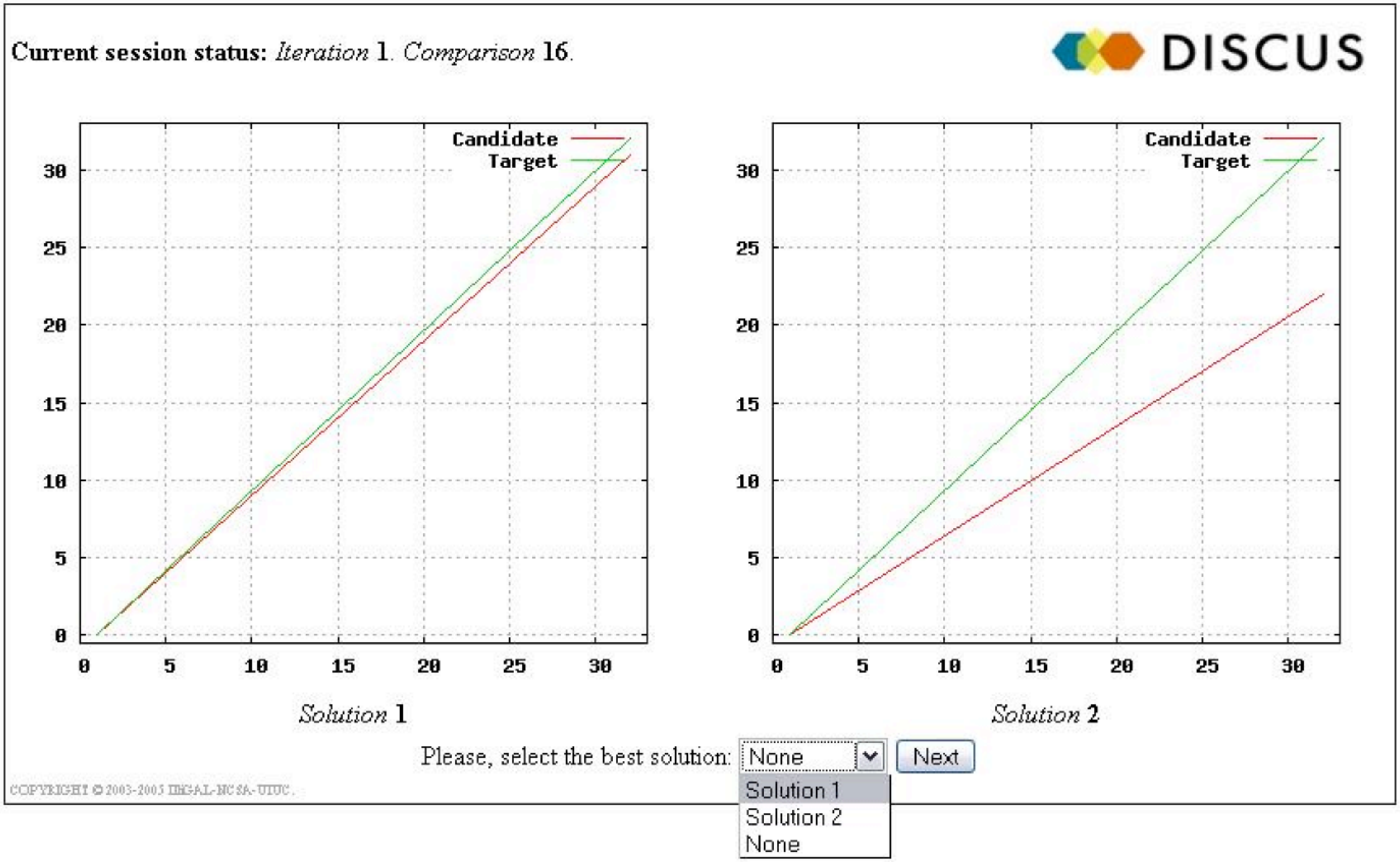
The Big Picture



A Real Experiment

- Considerations in the design process
  - Clear goal definition
  - Impact of problem visualization
  - Persistence of user criteria
- Focus
  - Lack of numeric fitness
  - User fatigue
- A simple controlled task
  - One Max

DISCUS & iGAs

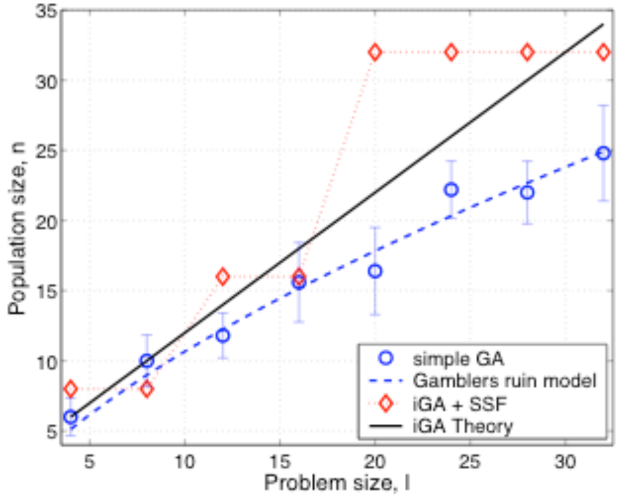


Current session status: Iteration 1. Comparison 16.
DISCUS
Candidate
Target
Solution 1
Solution 2
Please, select the best solution:
None
Next
Solution 1
Solution 2
None
COPYRIGHT © 2003-2005 IlliGAL-NCSA-UIUC.

Setting The Pieces

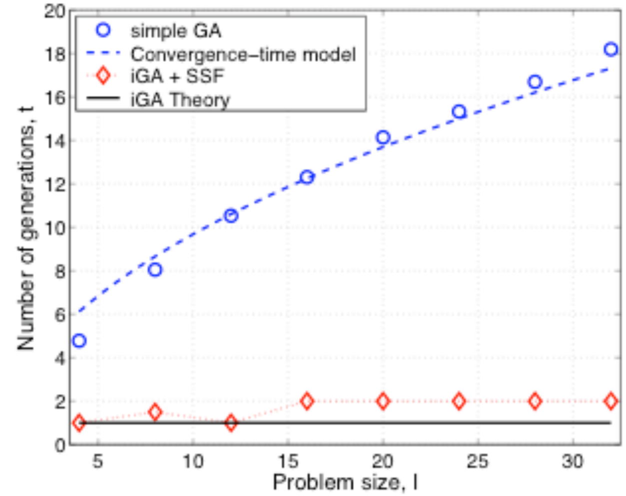
- The system
  - A simple web application
  - OneMax task
  - No linkage learning needed
  - ε-SVM and a linear kernel
  - The compact genetic algorithm (Harik, Cantú-Paz, Goldberg, & Miller, 1999)
- Set up
  - One user with no relation to the research
  - Repeated series of 10 independent runs for different problem sizes {4, 8, 12, 16, 20, 24, 28, and 32 variables}
  - Collect the data to compare it to a simple GA

DISCUS, IGA, & Research

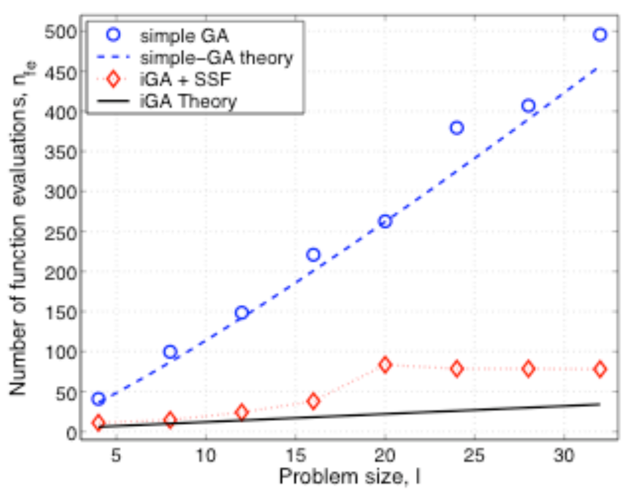


t

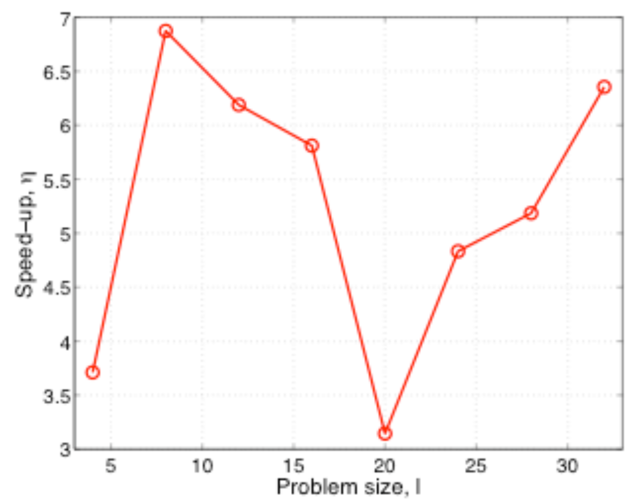
(a) Population Size



(b) Convergence Time



(c) Function Evaluations



(d) Speed up

Conclusions

- Address the lack of numerical fitness and fatigue
- Synthetic fitness model of user preferences
- Optimize such model to take the advantage of the timescale difference
- Sample new solutions out of the optimized model
- Inject these solutions in the evaluation process
- Remarkable speedups
- Real-work applications:
  - Emotional text-to-speech synthesis (two research groups)
  - Marketing campaign and product design (advertisement company)
  - Tuning of text mining analysis tools (chance discovery consortium)

Thank you

- DISCUS project web page

  http://www.i-discus.org

- IlliGAL web site

  http://www-illigal.ge.uiuc.edu

- IlliGAL blog

  http://illigal.blogspot.com