A Fine-Grained View of Phenotypes and Locality in GP

James McDermott
[jmmcd@csail.mit.edu]
Edgar Galván-Lopéz
Michael O’Neill

GPTP IX, Ann Arbour, May 2011
Aims

- We would like to be able to:
  - Predict performance/measure difficulty
  - Detect bad representations
  - Design better representations
  - Say which aspect of a bad representation is to blame.
Aims

• We would like to be able to:
  • Predict performance/measure difficulty
  • Detect bad representations
  • Design better representations
    • Say which aspect of a bad representation is to blame.
In problem difficulty, behaviour of genotype-fitness mapping is important.

Idea: look at components of the mapping.

Intermediate steps may be phenotypes.

Locality can characterise behaviour of mappings.
Introduction

- Locality: neighbours map to neighbours

- Related to/also known as:
  - Continuity
  - Strong causality
  - ...

Thursday, May 12, 2011
Introduction

- Genetic Algorithm OneMax

Genotypes (bitstrings) → Fitness
Genetic Algorithm “Engineering”

Genotypes (bitstrings) → Phenotypes (designs?) → Fitness
Introduction

- Genetic Algorithm “Engineering”

![Diagram showing relationships between genotypes, phenotypes, and fitness.]

Genotypes (bitstrings) → Phenotypes (designs?) → Fitness
Introduction

- Genetic Programming

Genotypes (trees)  Phenotypes (semantics?)  Fitness
Introduction

- Grammatical Evolution

Genotypes (int arrays) → Derivation trees → Derived programs → Semantics → Fitness
Introduction

- Parallel case: Multiple chromosomes

Genotypes  ---  Phenotypes (p0)

Phenotypes (p1)  ---  Fitness
• General case (here, nodes represent spaces)
• Not needed in this paper!
Introduction

- The behaviour of the mapping from genotype to fitness affects problem difficulty (fitness-distance correlation, negative slope coefficient...)
- It can be decomposed into several component mappings
- *Locality* characterises the behaviour of a mapping.
Definition of Phenotypes

- Do we have one phenotype of multiple components, or multiple phenotypes?
- Any data structure which is created during calculation of fitness may be a (part of the) phenotype.
- Any data structure which depends on the genotype may be a (part of the) phenotype.
- Any data structure on which fitness depends may be a (part of the) phenotype.
Outline

- Introduction to Locality
- Artificial Ant
- Boolean Problems
- Discussion
Part 2: Artificial Ant
Two phenotypes, sequentially ordered.

(g: GP trees) (p0: binary decision diagrams) (p1: cell sequences) f
BDDs as Ant Phenotypes

- Binary decision diagram (BDD): like an acyclic finite state machine

- Genotype tree: (if food ahead move left)

- BDD-Phenotype
BDDs as Ant Phenotypes

- **Genotype**
  - 3
  - F
  - M
  - R
  - M
  - 2
  - L
  - F
  - L
  - F
  - M
  - R

- **BDD-Phenotype**
  - M
  - L
  - R
  - MR
BDDs as Ant Phenotypes

- Summarises ant’s behaviour
- Abstract
- Genotype to phenotype mapping is not a bijection
- Two-way translation is possible
- Some redundancy is removed
- Could be a phenotype for Cartesian GP, Grammatical Evolution, Evolutionary Programming...
BDDs as Ant Phenotypes

- BDD-phenotypes have a string representation
- Distance is just string-edit distance
- Phenotypic mutation is string-edit (syntactically restricted)

\[ <M, L<L, R>>MR \]
Cell-Sequence Phenotypes

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Ant 0**

- Distance is easy to define: sum of toroidal distances at corresponding time-steps.
- Could use Hamming distance over visited cells instead.

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Ant 1**
Artificial Ant

- Two phenotypes, sequentially ordered.

(g: GP trees) (p0: binary decision diagrams) (p1: cell sequences) f
Artificial Ant: Experiments

- Sample an individual
- Perform a mutation -- get a neighbour
- Calculate genotypic, phenotypic and fitness distances between original and neighbour
Artificial Ant: Results

- Box-plots of distance, categorised by mutation type

![Box-plots showing distance](image)

- Genotype distance (tree-alignment)
- BDD-phenotype distance
Artificial Ant: Results

- Box-plots of distance, categorised by mutation type

BDD-phenotype distance  
Fitness distance
Artificial Ant: Results

- Box-plots of distance, categorised by mutation type

- $g \rightarrow p_0$: depends on operator
- $g \rightarrow f$: non-local for any operator
- $p_0 \rightarrow f$: non-local
- $p_1 \rightarrow f$: local, by definition
- Therefore $p_0 \rightarrow p_1$ is at fault.
Artificial Ant: Conclusions

- The mapping from BDD-phenotype to cell-sequence phenotype is highly non-local.
- Alternative encodings can not avoid this mapping.
- An explanation for poor performance of many GP encodings (standard, Cartesian GP, Grammatical Evolution, Evolutionary Programming, etc) on ant problem.
- In the presence of the non-local mapping, random search performs relatively well.
- Fine-grained view with phenotypes seems useful.
Part 3: Boolean Problems
Boolean Problems: Phenotypes

- Genotype is a Boolean function
- Outputs constitute a truth table (phenotype)
- Phenotypic distance is Hamming distance

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Genotype</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>x0</td>
<td>x1</td>
<td>x0 AND x1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
### Boolean Problems: Phenotypes

- Genotype is a Boolean function
- Outputs constitute a truth table (phenotype)
- Phenotypic distance is Hamming distance

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Genotypes</th>
<th>Phenotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>x0</td>
<td>x1</td>
<td>x0 AND x1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Boolean Problems: Phenotypes

- Our model is $g \rightarrow p_0 \rightarrow f$

- $g$: GP trees
- $p_0$: semantics, i.e. truth tables
- $f$
Boolean Problems: Phenotypes

- Mapping from phenotype to fitness is highly local by definition.
- Eg Even-3:
  - 10010110 (fitness is 8)
  - 10010111 (fitness is 7)
- Small change in phenotype gives a small change in fitness.
Boolean Encodings and Fitness Functions

- Different encodings are possible:
  - And, Or, Not (AON)
  - And, Or, Nand, Nor (AONN)
  - And, Or, Not, If (AONI)
  - Nand (N)

- Different fitness functions are possible:
  - Even-6 Parity
  - 6-Multiplexer
  - Majority-6
  - True-6
Different Fitness Functions

Independent of fitness function

Highly local by definition

(g: GP trees)  (p0: semantics, ie truth tables)  f
Different Fitness Functions

- According to locality, prediction is that all fitness functions should be equally difficult...
- Results don’t agree:
  Even-6 Parity < (Majority-6 ≈ 6-Multiplexer) < True-6
Different Encodings

Locality depends on encoding

Independent of encoding

(g: GP trees) (p0: semantics, ie truth tables) f
Different Encodings

• According to locality measures:
  \[ N > AONN > AONI > AON \]

  (> means more highly local than)

• So prediction is that N will perform best, and AON worst…

• But results don’t agree:
  \[ AONI > AONN > AON > N \]

  (> means performs better than)
Boolean Problems: Conclusions

- Locality is not predicting relative performance of *encodings*, in practice.
- Locality **CANNOT** predict relative difficulty of different *fitness functions*, in principle.
- Fine-grained view and phenotypes have not helped.
Part 4: Discussion
• Breaking up the mapping into component parts is “artificial” -- algorithm never sees phenotypes!
• But useful?
  • Somewhat useful for artificial ant.
  • Not useful for Boolean problems.
Discussion

• To predict performance, need to consider:
  • Mapping behaviour (eg locality, fitness distance correlation, etc)
  • Size of search space
  • Size of solution space
  • Diameter of search space
  • Neutrality
  • Other features?
Definition of Phenotypes

• Any data structure which is created during calculation of fitness may be a (part of the) phenotype.
• Any data structure which depends on the genotypes may be a (part of the) phenotype.
• Any data structure on which fitness depends may be a (part of the) phenotype.

• Can we narrow this definition down?

• BDD phenotype is useful because abstract.
• Cell-sequence phenotype and truth-table phenotype are more typical “semantic” phenotypes: very close to fitness.
Finally

- See our papers at CEC & GECCO (both 2010) and forthcoming GPEM for an alternative approach.
- Ant-phenotype code is available: [www.skynet.ie/~jmmcd/representations.html](http://www.skynet.ie/~jmmcd/representations.html)
- Thanks to Science Foundation Ireland and Irish Research Council for Science, Engineering and Technology, for funding this work.
- Thanks to Colin Johnson & Alberto Moraglio
- Thanks to all in the NCRA.
- Thanks!
### Table 1-1. Structural step-size (mean $\mu$, standard deviation $\sigma$, and neutral ratio $n$).

<table>
<thead>
<tr>
<th>Encoding</th>
<th>AON</th>
<th></th>
<th>AONI</th>
<th></th>
<th>AONN</th>
<th></th>
<th>N</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu$</td>
<td>$\sigma$</td>
<td>$n$</td>
<td>$\mu$</td>
<td>$\sigma$</td>
<td>$n$</td>
<td>$\mu$</td>
<td>$\sigma$</td>
</tr>
<tr>
<td>OneMax</td>
<td>5.9</td>
<td>12.7</td>
<td>0.1</td>
<td>9.4</td>
<td>24.5</td>
<td>0.1</td>
<td>7.4</td>
<td>18.9</td>
</tr>
<tr>
<td>Majority</td>
<td>5.8</td>
<td>12.5</td>
<td>0.1</td>
<td>9.5</td>
<td>25.0</td>
<td>0.1</td>
<td>7.3</td>
<td>18.7</td>
</tr>
<tr>
<td>Multiplexer</td>
<td>5.8</td>
<td>12.6</td>
<td>0.1</td>
<td>9.6</td>
<td>25.0</td>
<td>0.1</td>
<td>7.3</td>
<td>18.6</td>
</tr>
<tr>
<td>EvenParity</td>
<td>5.8</td>
<td>12.5</td>
<td>0.1</td>
<td>9.5</td>
<td>24.8</td>
<td>0.1</td>
<td>7.2</td>
<td>18.6</td>
</tr>
<tr>
<td>NoSelection</td>
<td>7.3</td>
<td>13.4</td>
<td>0.1</td>
<td>11.8</td>
<td>25.4</td>
<td>0.1</td>
<td>7.1</td>
<td>17.4</td>
</tr>
</tbody>
</table>

### Table 1-2. Phenotypic step-size (mean $\mu$, standard deviation $\sigma$, and neutral ratio $n$).

<table>
<thead>
<tr>
<th>Encoding</th>
<th>AON</th>
<th></th>
<th>AONI</th>
<th></th>
<th>AONN</th>
<th></th>
<th>N</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu$</td>
<td>$\sigma$</td>
<td>$n$</td>
<td>$\mu$</td>
<td>$\sigma$</td>
<td>$n$</td>
<td>$\mu$</td>
<td>$\sigma$</td>
</tr>
<tr>
<td>OneMax</td>
<td>6.7</td>
<td>10.3</td>
<td>0.4</td>
<td>4.9</td>
<td>8.2</td>
<td>0.5</td>
<td>4.3</td>
<td>8.7</td>
</tr>
<tr>
<td>Majority</td>
<td>6.5</td>
<td>10.2</td>
<td>0.4</td>
<td>4.8</td>
<td>8.2</td>
<td>0.5</td>
<td>4.2</td>
<td>8.7</td>
</tr>
<tr>
<td>Multiplexer</td>
<td>6.5</td>
<td>10.2</td>
<td>0.4</td>
<td>4.8</td>
<td>8.2</td>
<td>0.5</td>
<td>4.1</td>
<td>8.6</td>
</tr>
<tr>
<td>EvenParity</td>
<td>6.6</td>
<td>10.2</td>
<td>0.4</td>
<td>4.9</td>
<td>8.3</td>
<td>0.5</td>
<td>4.2</td>
<td>8.7</td>
</tr>
<tr>
<td>NoSelection</td>
<td>12.2</td>
<td>13.2</td>
<td>0.3</td>
<td>10.2</td>
<td>11.8</td>
<td>0.3</td>
<td>8.5</td>
<td>11.6</td>
</tr>
</tbody>
</table>
### Boolean Problems Locality Results

#### Table 1-2. Phenotypic step-size (mean μ, standard deviation σ, and neutral ratio n).

<table>
<thead>
<tr>
<th>Encoding</th>
<th>AON</th>
<th></th>
<th></th>
<th>AONI</th>
<th></th>
<th></th>
<th>AONN</th>
<th></th>
<th></th>
<th>N</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>μ</td>
<td>σ</td>
<td>n</td>
<td>μ</td>
<td>σ</td>
<td>n</td>
<td>μ</td>
<td>σ</td>
<td>n</td>
<td>μ</td>
<td>σ</td>
<td>n</td>
</tr>
<tr>
<td>OneMax</td>
<td>6.7</td>
<td>10.3</td>
<td>0.4</td>
<td>4.9</td>
<td>8.2</td>
<td>0.5</td>
<td>4.3</td>
<td>8.7</td>
<td>0.6</td>
<td>5.2</td>
<td>7.6</td>
<td>0.4</td>
</tr>
<tr>
<td>Majority</td>
<td>6.5</td>
<td>10.2</td>
<td>0.4</td>
<td>4.8</td>
<td>8.2</td>
<td>0.5</td>
<td>4.2</td>
<td>8.7</td>
<td>0.6</td>
<td>5.2</td>
<td>7.7</td>
<td>0.4</td>
</tr>
<tr>
<td>Multiplexer</td>
<td>6.5</td>
<td>10.2</td>
<td>0.4</td>
<td>4.8</td>
<td>8.2</td>
<td>0.5</td>
<td>4.1</td>
<td>8.6</td>
<td>0.6</td>
<td>5.2</td>
<td>7.6</td>
<td>0.4</td>
</tr>
<tr>
<td>EvenParity</td>
<td>6.6</td>
<td>10.2</td>
<td>0.4</td>
<td>4.9</td>
<td>8.3</td>
<td>0.5</td>
<td>4.2</td>
<td>8.7</td>
<td>0.6</td>
<td>5.3</td>
<td>7.7</td>
<td>0.4</td>
</tr>
<tr>
<td>NoSelection</td>
<td>12.2</td>
<td>13.2</td>
<td>0.3</td>
<td>10.2</td>
<td>11.8</td>
<td>0.3</td>
<td>8.5</td>
<td>11.6</td>
<td>0.4</td>
<td>8.3</td>
<td>9.0</td>
<td>0.3</td>
</tr>
</tbody>
</table>

#### Table 1-3. Fitness step-size (mean μ, standard deviation σ, and neutral ratio n).

<table>
<thead>
<tr>
<th>Encoding</th>
<th>AON</th>
<th></th>
<th></th>
<th>AONI</th>
<th></th>
<th></th>
<th>AONN</th>
<th></th>
<th></th>
<th>N</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>μ</td>
<td>σ</td>
<td>n</td>
<td>μ</td>
<td>σ</td>
<td>n</td>
<td>μ</td>
<td>σ</td>
<td>n</td>
<td>μ</td>
<td>σ</td>
<td>n</td>
</tr>
<tr>
<td>OneMax</td>
<td>4.5</td>
<td>8.2</td>
<td>0.5</td>
<td>2.9</td>
<td>5.6</td>
<td>0.5</td>
<td>3.0</td>
<td>6.9</td>
<td>0.6</td>
<td>2.5</td>
<td>3.6</td>
<td>0.4</td>
</tr>
<tr>
<td>Majority</td>
<td>2.4</td>
<td>4.3</td>
<td>0.6</td>
<td>1.7</td>
<td>3.3</td>
<td>0.6</td>
<td>1.6</td>
<td>3.5</td>
<td>0.7</td>
<td>1.7</td>
<td>3.2</td>
<td>0.5</td>
</tr>
<tr>
<td>Multiplexer</td>
<td>1.7</td>
<td>2.8</td>
<td>0.6</td>
<td>1.4</td>
<td>2.4</td>
<td>0.6</td>
<td>1.1</td>
<td>2.3</td>
<td>0.7</td>
<td>1.6</td>
<td>2.4</td>
<td>0.5</td>
</tr>
<tr>
<td>EvenParity</td>
<td>0.1</td>
<td>0.3</td>
<td>0.9</td>
<td>0.1</td>
<td>0.4</td>
<td>0.9</td>
<td>0.1</td>
<td>0.3</td>
<td>0.9</td>
<td>0.2</td>
<td>0.5</td>
<td>0.8</td>
</tr>
<tr>
<td>NoSelection</td>
<td>6.2</td>
<td>9.5</td>
<td>0.5</td>
<td>4.8</td>
<td>7.6</td>
<td>0.5</td>
<td>4.9</td>
<td>9.1</td>
<td>0.6</td>
<td>3.4</td>
<td>5.0</td>
<td>0.4</td>
</tr>
</tbody>
</table>
## Boolean Problems Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>500</td>
</tr>
<tr>
<td>Generations</td>
<td>50</td>
</tr>
<tr>
<td>Initialisation</td>
<td>Ramped half &amp; half [1,7]</td>
</tr>
<tr>
<td>Maximum depth</td>
<td>7</td>
</tr>
<tr>
<td>Selection</td>
<td>Tournament [7]</td>
</tr>
<tr>
<td>Mutation</td>
<td>Subtree or One-point</td>
</tr>
<tr>
<td>Mutation probability</td>
<td>1.0</td>
</tr>
<tr>
<td>Crossover</td>
<td>None</td>
</tr>
<tr>
<td>Sampling</td>
<td>Metropolis-Hastings</td>
</tr>
<tr>
<td>Individuals</td>
<td>100,000 per problem</td>
</tr>
</tbody>
</table>
### Artificial Ant Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>500</td>
</tr>
<tr>
<td>Generations</td>
<td>50</td>
</tr>
<tr>
<td>Initialisation</td>
<td>Ramped half &amp; half [1,7]</td>
</tr>
<tr>
<td>Maximum depth</td>
<td>7</td>
</tr>
<tr>
<td>Selection</td>
<td>Tournament [7]</td>
</tr>
<tr>
<td>Mutation</td>
<td>Subtree</td>
</tr>
<tr>
<td>Mutation probability</td>
<td>0.01 or 1.0</td>
</tr>
<tr>
<td>Crossover</td>
<td>90/10</td>
</tr>
<tr>
<td>Crossover probability</td>
<td>0.7 or 0.0</td>
</tr>
<tr>
<td>Sampling</td>
<td>Metropolis-Hastings</td>
</tr>
<tr>
<td>Individuals</td>
<td>100,000 per problem</td>
</tr>
</tbody>
</table>

Thursday, May 12, 2011
## BDD-phenotype conversion

<table>
<thead>
<tr>
<th>Replace this:</th>
<th>... with this:</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>(prog2 X Y)</td>
<td>XY</td>
<td>(translation)</td>
</tr>
<tr>
<td>(prog3 X Y Z)</td>
<td>XYZ</td>
<td>(translation)</td>
</tr>
<tr>
<td>(if X Y)</td>
<td>&lt;X,Y&gt;</td>
<td>(translation)</td>
</tr>
<tr>
<td>&lt;X&lt;Y&gt;W,Z&gt;</td>
<td>&lt;XW,Z&gt;</td>
<td>(eliminate redundant if)</td>
</tr>
<tr>
<td>&lt;X,Y&gt;Z,W&gt;</td>
<td>&lt;X,ZW&gt;</td>
<td>(eliminate redundant if)</td>
</tr>
<tr>
<td>&lt;X,Y&gt;Z,W&gt;</td>
<td>&lt;X,ZW&gt;</td>
<td>(eliminate redundant if)</td>
</tr>
<tr>
<td>&lt;XY,ZY&gt;</td>
<td>&lt;X,Z&gt;Y</td>
<td>(bring out common term)</td>
</tr>
<tr>
<td>&lt;,&gt;</td>
<td></td>
<td>(eliminate empty if)</td>
</tr>
</tbody>
</table>