Assessment of Problem Modality by Differential Performance of Lexicase Selection in Genetic Programming

A Preliminary Report

Lee Spector
Cognitive Science
Hampshire College
http://hampshire.edu/lspector

Outline

- Modal problems
- Lexicase selection
- Measuring modality
- An example
- Conclusions

Modal Problems

- **Require successful programs to do something qualitatively different in different circumstances**
- "Circumstances" are assumed to be factored, at least in part, by fitness cases
- In general it may not be obvious what modes of response a problem requires, or how to tell when one or another is appropriate

Geometry Calculator

- Calculate formulae for perimeters, areas, surface areas, and volumes of various 2d and 3d shapes, depending on the value of a **`formula`** input
- A solution program will have to do qualitatively different things depending on **`formula`**
- Formulae will share structure, e.g. factors of π
- Harder: **`formula`** available only indirectly

Web Browser

- Display web pages, create bookmarks, display browsing histories, negotiate secure connections, launch helper applications, etc.
- Performance on one function (e.g. page display) largely independent of performance on others (e.g. bookmark creation)
- Many functions share sub-functions (e.g. HTML parsing, click handling)

Lexicase Selection

Select a parent by filtering the population on the basis of performance on one fitness case at a time, with the fitness cases considered in random order.

To select a parent for use in a genetic operation:

1. `Initialize`:
   (a) Set **`candidates`** to be *the entire population.*
   (b) Set **`cases`** to be a list of all of the fitness cases *in random order.*
2. `Loop`:
   (a) Set **`candidates`** to be the subset of the current **`candidates`** that have *exactly the best fitness* of any individual currently in **`candidates`** for the first case in **`cases`**.
   (b) If **`candidates`** or **`cases`** contains just a single element then return the first individual in **`candidates`**.
   (c) Otherwise remove the first case from **`cases`** and go to `Loop`.

Figure 1: Pseudocode for a simple version of the lexicase parent selection algorithm. A more complete description of this version of the algorithm is "global pool, uniform random sequence, elitist lexicase parent selection." Parts of the algorithm that can be varied to produce different versions of lexicase selection are indicated *in italics*; see text for details.

Table 1: Fitness components (errors by case, with smaller being better) and lexicase selection probabilities for a small population on a hypothetical problem.

Individual	Fitness case				Lexicase selection probability
	a	b	c	d	
#1	2	2	4	2	0.250
#2	1	2	4	3	0.000
#3	2	2	3	4	0.333
#4	0	2	5	5	0.208
#5	0	3	5	2	0.208

Measuring Modality

$$M(p,\lambda) = \frac{S(p,\lambda_\alpha) - S(p,\lambda)}{S(p,\lambda_\alpha) + S(p,\lambda)}$$

$M(p,\lambda)$: Modality of problem p relative to GP system λ

S: success rate for a system on a problem

λ_α: system λ modified to use lexicase selection

Related Work

- Multi-objective evolution (generally assumes objectives, which may not be factored by input, are known in advance)
- Multi-modal problems (generally refers to problems with multiple global optima)
- Lexicographic ordering in selection (but here we order fitness cases, in random order)
- Ensemble methods (but here we seek a single program perhaps with some code used for multiple modes)

Regression Example

- Problem A: $y = x^2 + x + 1$
- Problem B:
  - if $x < 0$, $y = x^2 + x + 1$
  - else $y = 7x$
- Problem C:
  - if $x < -5$, $y = x^2 + x + 1$
  - else *if* $x < 5$, $y = 7x$
  - else $y = 3x^2 - 3$

Successes

Table 2: Successful runs (out of 30 in each condition) for standard GP and GP with lexicase selection on problems with three levels of modality. See Section 5 for problem descriptions.

	Problem A	Problem B	Problem C
Standard	29	22	0
Lexicase	30	29	12

Problem Modalities

Table 3: Problem modalities computed from differential performance of lexicase selection, shown in Table 2, using Equation 1. See Section 5 for problem descriptions.

	Problem A	Problem B	Problem C
Modality	0.017	0.137	1

A Circular Definition?

Critic: You've developed a new problem-solving technique?
Author: Yes.
Critic: What's it good for?
Author: Modal problems.
Critic: How do you tell if a problem is modal?
Author: If my new technique works on it!

Conclusions

- Modal problems are ubiquitous and important
- Standard GP techniques do not handle them well
- Lexicase selection appears to help
- A measure of problem modality can be defined in terms of the degree to which it helps

Future Work

- Much more experimentation; the example presented here was only illustrative
- Compare to various multi-objective methods
- Improve efficiency of the selection algorithm
- Explore non-global, non-uniform, and/or non-elitist forms of lexicase selection

Thanks

This material is based upon work supported by the National Science Foundation under Grant No. 1017817. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation. Thanks also to Thomas Helmuth, Omri Bernstein, Kyle Harrington, Emma Tosch, Kwaku Yeboah Antwi, and Rebecca S. Neimark for discussions related to this work, to Josiah Erikson for systems support, and to Hampshire College for support for the Hampshire College Institute for Computational Intelligence.