Diagnosing Island Supplemental Material

Jose Guadalupe Hernandez

2023-06-12

Contents

Chapter 1

Introduction

This is the supplemental material for experiments with basic diagnostics.

1.1   About our supplemental material

This supplemental material is hosted on GitHub using GitHub pages. The source code and configuration files used to generate this supplemental material can be found in this GitHub repository. We compiled our data analyses and supplemental documentation into this nifty web-accessible book using bookdown.

Our supplemental material includes the following paper figures and statistics:

- Exploitation rate results (Section 2)
- Ordered exploitation results (Section 3)
- Contradictory objectives results (Section 4)
- Multi-path exploration results (Section 5)

1.2   Contributing authors

- Jose Guadalupe Hernandez
- Alexander Lalejini
- Charles Ofria

1.3   Computer Setup

These analyses were conducted in the following computing environment:

```
print(version)
```

```
##                 _
```

```
## platform        x86_64-pc-linux-gnu
## arch            x86_64
## os              linux-gnu
## system          x86_64, linux-gnu
## status
## major           4
## minor           3.0
## year            2023
## month           04
## day             21
## svn rev         84292
## language        R
## version.string R version 4.3.0 (2023-04-21)
## nickname        Already Tomorrow
```

1.4   Experimental setup

Setting up required variables variables.

```r
# libraries we are using
library(ggplot2)
library(cowplot)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(PupillometryR)
```

```
## Loading required package: rlang
```

```r
# data diractory for gh-pages
DATA_DIR = '/opt/ECJ-2023-Suite-Of-Diagnostic-Metrics-For-Characterizing-Selection-Sch

# data diractory for local testing
# DATA_DIR = 'C:/Users/jgh9094/Desktop/Research/Projects/SelectionDiagnostics/ECJ-2023

# graph variables
SHAPE = c(5,3,1,2,6,0,4,20,1)
cb_palette <- c('#332288','#88CCEE','#EE7733','#EE3377','#117733','#882255','#44AA99',
```

```
TSIZE = 26
p_theme <- theme(
  text = element_text(size = 28),
  plot.title = element_text( face = "bold", size = 22, hjust=0.5),
  panel.border = element_blank(),
  panel.grid.minor = element_blank(),
  legend.title=element_text(size=22),
  legend.text=element_text(size=23),
  axis.title = element_text(size=23),
  axis.text = element_text(size=22),
  legend.position="bottom",
  panel.background = element_rect(fill = "#f1f2f5",
                                  colour = "white",
                                  size = 0.5, linetype = "solid")
)
```

```
## Warning: The `size` argument of `element_rect()` is deprecated as of ggplot2 3.4.0.
## i Please use the `linewidth` argument instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
# default variables
REPLICATES = 50
DIMENSIONALITY = 100

# selection scheme related stuff
ACRO = c('tru','tor','lex','gfs','pfs','nds','nov','ran')
NAMES = c('Truncation (tru)','Tournament (tor)','Lexicase (lex)', 'Genotypic Fitness Sharing (gfs
```

Chapter 2

Exploitation rate results

Here we present the results for **best performances** found by each selection scheme on the exploitation rate diagnostic. 50 replicates are conducted for each scheme explored.

2.1   Analysis dependencies

```
library(ggplot2)
library(cowplot)
library(dplyr)
library(PupillometryR)
```

2.2   Data setup

```
# dir = paste(DATA_DIR,'EXPLOITATION_RATE/','over-time.csv', sep = "", collapse = NULL)
over_time_df <- read.csv(paste(DATA_DIR,'EXPLOITATION_RATE/','over-time.csv', sep = "", collapse
over_time_df$scheme <- factor(over_time_df$scheme, levels = NAMES)
```

2.3   Performance over time

Best performance in a population over time.

```
# data for lines and shading on plots
lines = over_time_df %>%
  group_by(scheme, gen) %>%
  dplyr::summarise(
    min = min(pop_fit_max) / DIMENSIONALITY,
    mean = mean(pop_fit_max) / DIMENSIONALITY,
```
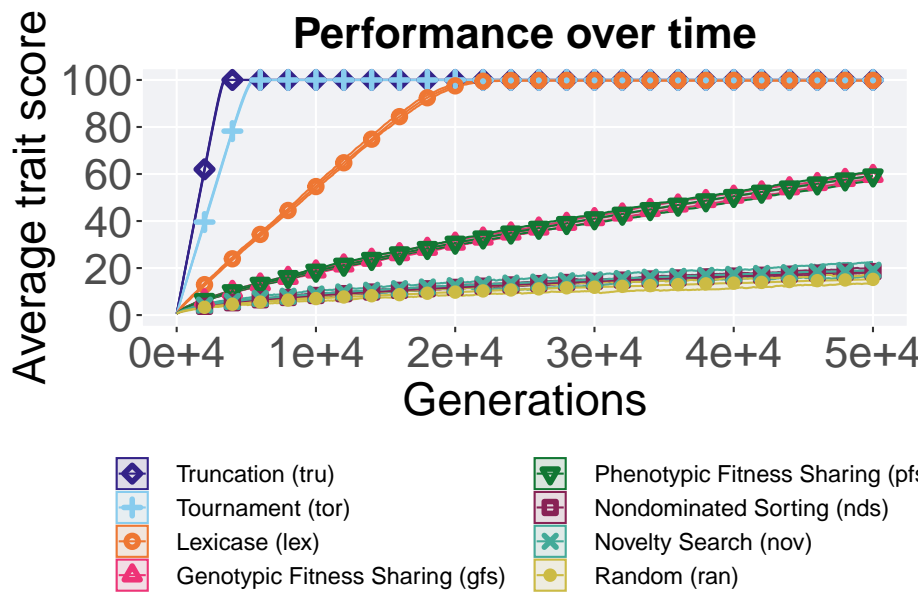
```
    max = max(pop_fit_max) / DIMENSIONALITY
 )
```

```
## `summarise()` has grouped output by 'scheme'. You can override using the
## `.groups` argument.
ggplot(lines, aes(x=gen, y=mean, group = scheme, fill =scheme, color = scheme, shape =
  geom_ribbon(aes(ymin = min, ymax = max), alpha = 0.1) +
  geom_line(size = 0.5) +
  geom_point(data = filter(lines, gen %% 2000 == 0 & gen != 0), size = 1.5, stroke = 2
  scale_y_continuous(
    name="Average trait score",
    limits=c(0, 100),
    breaks=seq(0,100, 20),
    labels=c("0", "20", "40", "60", "80", "100")
  ) +
  scale_x_continuous(
    name="Generations",
    limits=c(0, 50000),
    breaks=c(0, 10000, 20000, 30000, 40000, 50000),
    labels=c("0e+4", "1e+4", "2e+4", "3e+4", "4e+4", "5e+4")

  ) +
  scale_shape_manual(values=SHAPE)+
  scale_colour_manual(values = cb_palette) +
  scale_fill_manual(values = cb_palette) +
  ggtitle('Performance over time')+
  p_theme + theme(legend.title=element_blank(),legend.text=element_text(size=12)) +
  guides(
    shape=guide_legend(ncol=2, title.position = "bottom"),
    color=guide_legend(ncol=2, title.position = "bottom"),
    fill=guide_legend(ncol=2, title.position = "bottom")
  )
```

Performance over time

Legend:
- Truncation (tru)
- Tournament (tor)
- Lexicase (lex)
- Genotypic Fitness Sharing (gfs)
- Phenotypic Fitness Sharing (pfs)
- Nondominated Sorting (nds)
- Novelty Search (nov)
- Random (ran)

Chapter 3

Ordered exploitation results

Here we present the results for **best performances** found by each selection scheme on the ordered exploitation diagnostic. 50 replicates are conducted for each scheme explored.

Chapter 4

Contradictory objectives results

Here we present the results for **activation gene coverage** and **satisfacotory trait coverage** found by each selection scheme on the contradictory objectives diagnostic. 50 replicates are conducted for each scheme explored.

Chapter 5

Multi-path exploration results

Here we present the results for **best performances** found by each selection scheme on the multi-path exploration diagnostic. 50 replicates are conducted for each scheme explored.