Report 3: Behavioral and Data Exploration

Due: Middle of week 5

Purpose: To explore the *behavior* of algorithms in context and identify directions for research questions

Overview

Two kinds of explorations help researchers ultimately create explanations: behavior exploration and data exploration.

As we have discussed, it is often useful to let the data inform you about what types of models may be appropriate for performing inferences, finding explanatory causes, etc. More specifically, many important types of computing research questions can be generated by informal observation of an algorithm's behavior in a "naturalistic" or laboratory setting. Such an approach can identify cases where an algorithm does not perform as expected or it can generate new questions about previously unexplored types of behavior. For example, the decision by Oates & Jensen (ICML 1997) to examine the relationship among training set size, tree size, and accuracy is an example of this latter phenomenon.

This type of investigation occupies a middle ground between theorizing (which often does not use empirical data) and formal experimentation (which often focuses on examining a predetermined question). The goal here is exploratory, so we will use an informal approach to generating and analyzing data.

Identify, measure, and record the values of several key variables from your system, task, and environment.

When appropriate, run an implementation of the system (or some portion it) on multiple instances of your selected task and environment, varying specific aspects of the algorithm, task, and environment in isolation and in combination. Measure and record the values of key variables.

Apply exploratory analysis tools to the data you have gathered or generated in an effort to examine:

- how the behavior of your algorithm varies with aspects of the task and environment, and/or
- how the aspects of the task and environment relate in ways that may inform a potential model or algorithm.

In particular, try to identify unexpected behavior or new combinations of variables that you had not previously considered important.

The length of this report should be at least two pages of single-spaced text with plenty of *additional* space for diagrams and citations. Figures and/or tables are expected. (Use standard 1" margins, rather than the generous LaTeX document class defaults.)

Details

You should produce single, collaboratively-generated report for each team. While this report will most likely not appear directly as part of your final product, you should find that thinking (and writing) clearly now will be extraordinarily helpful to your productivity and future work.

Matlab is very good at producing figures that may easily be incorporated into LyX. If the current figure window contains the plot you want to store, you can use the command

```
>> print('-depsc2','file.eps');
```

to store a color EPS (encapsulated PostScript) file that will scale and print well. Do not use PNG or another image format unless you are saving an image with `imwrite`. Be sure to title your figures (in addition to the captions), and include labels on axes (with units, as appropriate), and legends (if necessary).

To capture a simple image (rather than a plot in a figure window), you can save it using `imwrite`. You should consider using the data repository and Matlab scripts to generate final versions of your graphics programmatically.

You should use Figure (Table, resp.) floats in *LyX* to add figures (tables, resp.) and corresponding captions. Override the default placement to intelligently and *intentionally* locate your figures in the stream of text (right click on the grey float tab and choose "Settings...", using the "Here definitely" option), the same way you would use math equations as part of the narrative structure.
