

CSCI6506 - Genetic Algorithms and Programming

Jéssica Pauli de C. Bonson (B00617515)

Project Proposal

For the class project I will develop an advanced version of the sandbox. The system is an implementation for canonical GP for multi-class classification using a linear representation. The features that the system already have are:

- selection operators: proportional selection and steady tournament
- sampling heuristics: unbalanced and balanced
- variation operators: crossover and mutation
- stop criterion: run all generations
- variable programs size
- evolve modular behaviors

From these features, the selection operator used for the project will be proportional selection, the sampling heuristic will be balanced, and both variation operators will be used.

The new features that will be added to the system are:

- more complex solutions:
 - non-linear instructions (if, log, exp, sin)
- diversity:
 - fitness sharing
 - pareto (dominance ranking and/or dominance count)
 - objective switching
- improved performance:
 - introns removal

The system will be applied to at least two of the following datasets:

- <http://archive.ics.uci.edu/ml/datasets/Adult>
- [https://archive.ics.uci.edu/ml/datasets/Statlog+\(Shuttle\)](https://archive.ics.uci.edu/ml/datasets/Statlog+(Shuttle))
- <https://archive.ics.uci.edu/ml/datasets/Gisette>

In case there is enough time, the system will be applied to all three datasets.

The main metrics used to evaluate the program solutions are:

- accuracy
- class-wise detection rate

The projects' goal is to analyse how the new features impact the classification metrics for the datasets, and to compare the diversity features results.