

## Abstract

R has become the first-choice language for data scientists but it is not typically the first-choice for people developing bio-inspired algorithms. This makes it hard for data scientists to make use of bio-inspired methods.

This project assessed the available bio-inspired algorithms in R and implemented a package to fill in a hole in the provision.

## Main Objectives

- Investigate the availability of bio-inspired algorithms in R
- Identify areas to be improved or built upon
- Produce a package to provide access to an identified area
- Release the package to improve the availability of bio-inspired methods

## Research

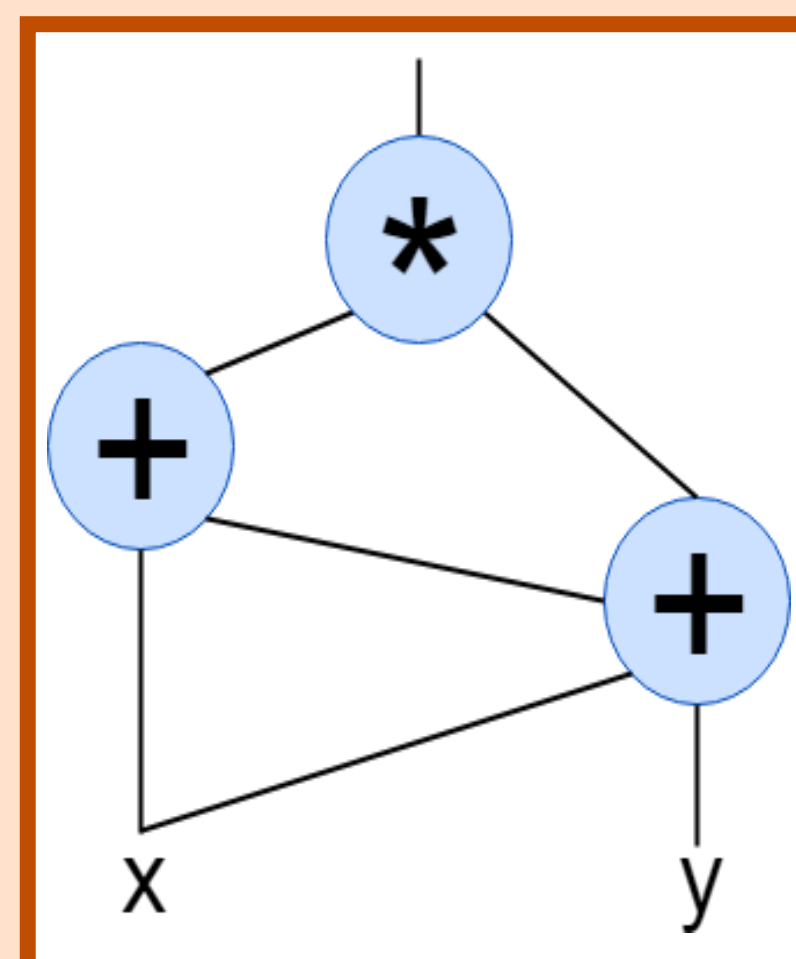
- Genetic Programming was a poorly represented area
- Only one package related to the area and is now deprecated
- The package implemented tree-based GP
- Chose to implement Cartesian GP which solves some problems tree-based suffers from
- Uses a fixed size grid so programs cannot grow to large lengths as with tree-based GP
- Graph-based allowing reuse of evaluated expressions through different branches

## Methodology

Genetic Programming is an Evolutionary Algorithm so it follows a structure similar to the flowchart on the right.

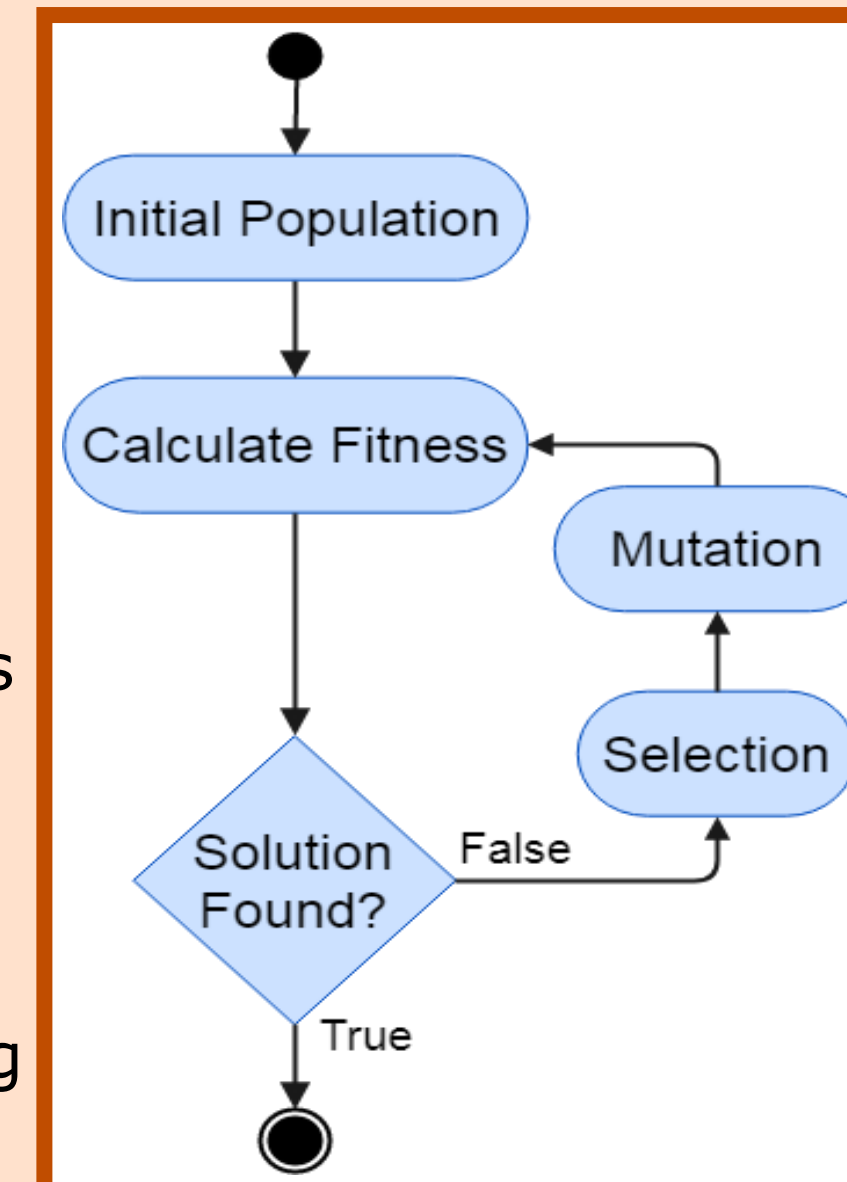
An additional step called Crossover may also be included between Selection and Mutation but has not been used in this case.

As the algorithm progresses, solutions are generated which progress closer to the desired result. The fitness is a measure of how close the solution is to this result.



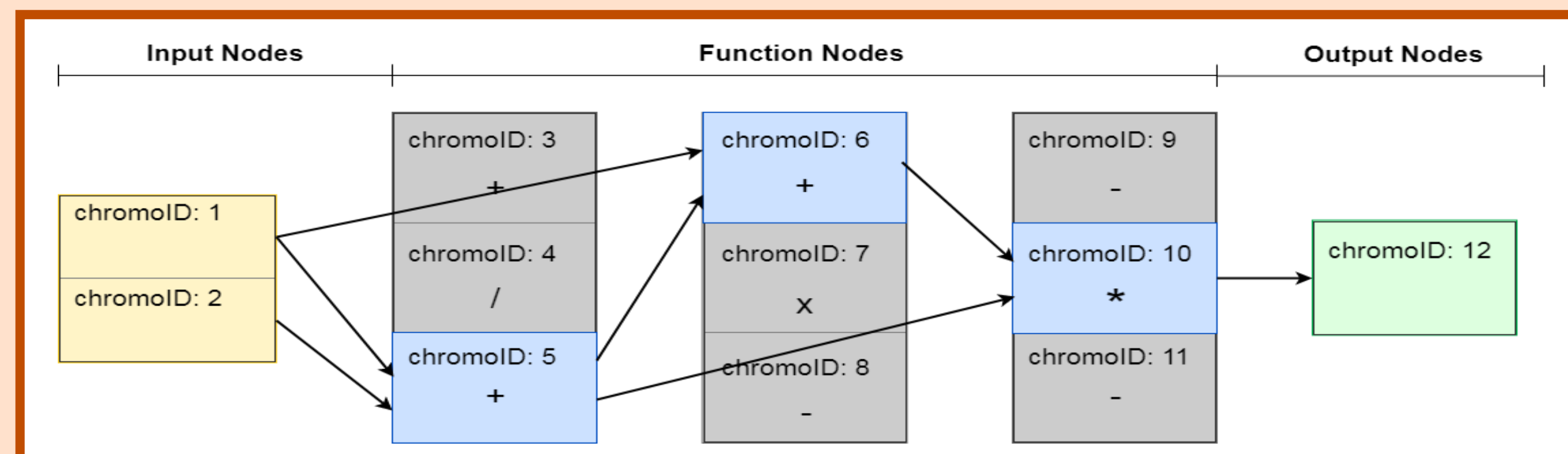
Cartesian Genetic Programming (CGP) represents programs using graphs.

The package I have produced known as "caRtesian" implements CGP and uses it to solve symbolic regression problems. The graphs are used to encode arithmetic expressions.



A dataset is loaded into the program with inputs and expected values. Attempts are made to find the function which connects these values.

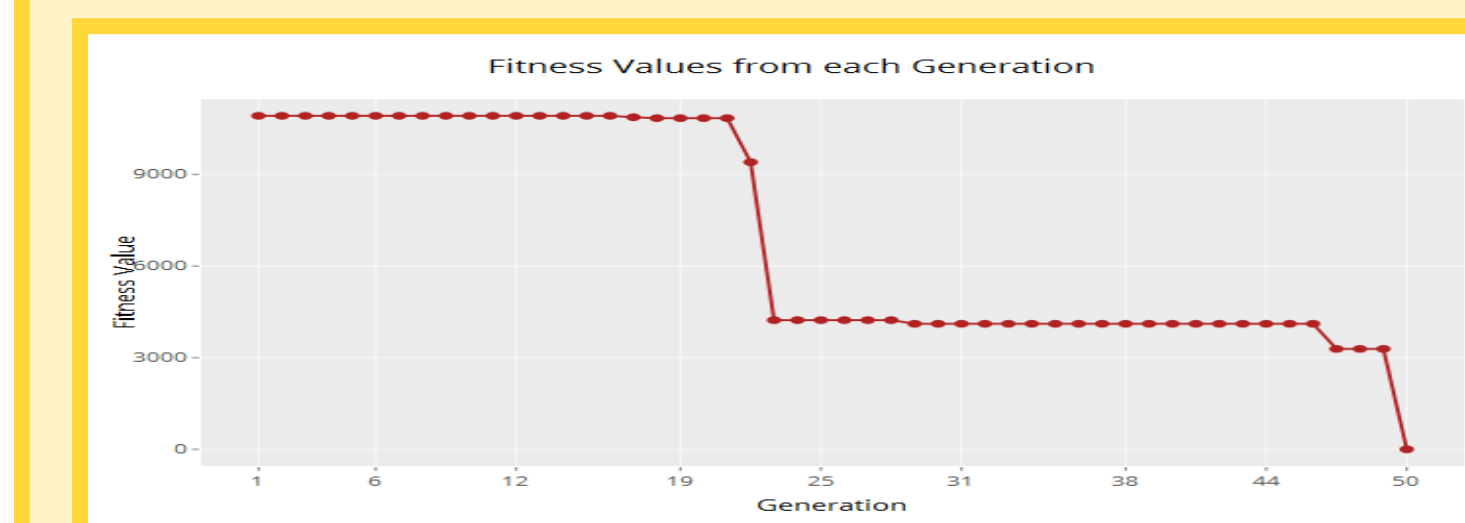
The graph displayed above is shown in the diagram but encoded as a grid of nodes. The greyed out nodes are not active meaning they are not part of the graph. These nodes also have connections but these have been omitted.



## Results

The package is able to find solutions to problems. The plot below shows the results of trying to find the function represented by the graph on the left.

Although a perfect solution was found in this experiment, this may not always be the case due to the stochastic nature of the program.



## Conclusions

"caRtesian" is a package which provides access to CGP for R users. The package can be improved in many areas but provides a starting point for anyone wishing to extend this implementation.

## Further Work

### Functionality

- Increased customisation of parameters
- Crossover operator to be investigated as it may produce beneficial results
- Ability to visualise resulting graphs

### Usability

- Address usability issues found during study

### Performance

- Improve performance of package overall by optimising demanding tasks