Use review template for what point to cover

State hypothesis . . . run experiment to disprove your hypothesis

References for problem,

Our target audience is masters students in our class who have not taken the NC module

Am I trying to use genetic programming, algorithm or evolution . . .

Think about presentation and visuals

Timeline of NC, GA

# Abstract

Contents

[Abstract 1](#_Toc486091998)

[Introduction 3](#_Toc486091999)

[Natural Computing 3](#_Toc486092000)

[Evolutionary Computation 3](#_Toc486092001)

[Genetic Algorithms 3](#_Toc486092002)

[Genetic Programming 3](#_Toc486092003)

[Trading Strategies 3](#_Toc486092004)

[Application of Genetic Evolution to Trading 3](#_Toc486092005)

[Motivation and Objective 4](#_Toc486092006)

[PonyGE2 4](#_Toc486092007)

[Paper Layout 4](#_Toc486092008)

[Experimentation 4](#_Toc486092009)

[Assumptions and Simplifications 5](#_Toc486092010)

[Problem Description 5](#_Toc486092011)

[Experiment Results 5](#_Toc486092012)

[Discussion 5](#_Toc486092013)

[Conclusion 5](#_Toc486092014)

[References 6](#_Toc486092015)

# Introduction

This paper aims to capture the output of a project involving the application of recent advances in machine learning to an interesting finance problem. The first few sections of this introduction will bring a computer software student or IT professional who has not researched natural computing up to a base knowledge level where they will understand the project objective and outcome. For those initiated in the natural computing area, they may skip forward to the Trading Strategies section within the introduction.

## Natural Computing

Natural computing is a branch of computer science inspired by nature. One area of natural computing involves utilising materials other than silicone to carry out computational tasks. A great example of this is in biological computing where DNA has been used to successfully store information (Church et al. 2012). Another area of natural computing is in the machine learning domain where algorithms are derived from key characteristics of processes found in nature and are deployed to solve difficult problems (Brabazon & O'Neill, 2009). The application of a computing algorithm inspired by a nature forms the basis of this paper while the Natural Computing journal may provide further reading.

Examples of natural computing algorithms include exploration of shortest paths using swarm techniques such as ant colony optimisation or finding optimal points by modelling bird flocking behaviour (Banks et al. 2007). Another popular research area is artificial neural networks where the components and workings of a human brain are mimicked with software to artificially learn the solution to a range of problems such as developing game strategies. The algorithm implemented in this project was inspired by the process of natural evolution and a brief overview is presented in the next few sections.

## Evolutionary Computation

The concept of machine learning dates back as far as Turing and has been the subject of many sci-fi works since the beginning of the digital revolution (Turing, 1950). Friedberg (1958) hypothesised that computers could perform new tasks if the computer could create a new program through trial and error. The widely accepted Darwinian evolution theory is the evolution of species through slight variations in each generation where traits that strengthen the species are passed on with a random but probabilistic degree. One means of facilitating a trial and error methodology in computation is through evolutionary programming where an algorithm is designed to generate and test numerous candidate solutions to a problem. This algorithm is a metaphoric implementation of evolution and the trial and error simulation mimics the survival of the fittest aspect of nature (Beyer, 2002).

## Genetic Programming

Genetic programming works by creating an initial population of programs from a defined set of programming features and constraints such as code sections, inputs, outputs and transformations. This random initialisation means we are never guaranteed to converge on a solution however it does have the benefit of producing unexpected programs. For each generation of population, parent programs pass on traits to their children stochastically through two mechanisms (PolR et al. 2008):

* Crossover: Randomly selected traits from two parents are preserved.
* Mutation: A parent is randomly changed to produce a child.

This process continues until a convergence criteria or iteration limit has been reached by the algorithm. The final output program will be the program producing the best fitness result. An informative graphic showing the information flow is shown in figure 1 below.

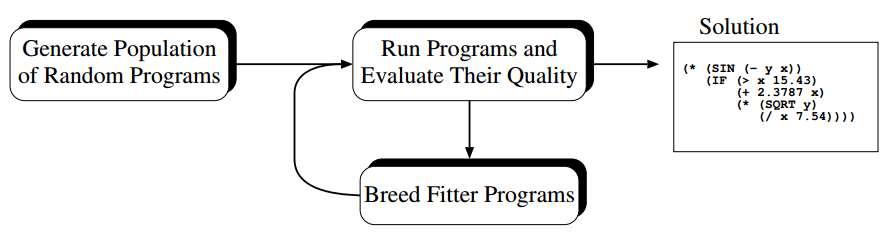


Figure 1 - GP Control Flow (PolR et al. 2008)

As Koza (2009) notes, the usefulness of genetically evolved programs has been proportional to improvements in computing power due to the computation required in this domain. Genetic programming has successfully produced software that rivals that produced by humans in a wide variety of fields such as software creation, game theory where strategies have been developed, electronic design where digital and analog circuits have been designed and mechanical engineering for systems design (Koza, 2009). Lastly and most importantly for this project, the finance industry has also employed genetic programming in many specific problem areas (Brabazon & O'Neill, 2009).

## Trading Strategies

Long term – buy and hold

Short term – beat the market

Profits vs variability

There are sites to try out/execute: <https://www.quantopian.com>

Can a market be predicted? No but a strategy can work without deterministic market

Efficiencies - As the financial markets become more efficient, can trading strategies be developed from natural computing methods

## Application of Genetic Evolution to Trading

What algorithm and why? (Finance can be a noisy environment just like nature – MO book)

Random unexpected program insights etc.

We can see the output in if statements etc

Nueral network will not show workings - intended to be a continually developing AI rather than a single use program. Intended as a continually evolving strategy that uses latest data rather than outputting a strategy to be analysed and implemented. Designed to be left alone - Argue that it can be left alone. Neural network selected as it may be more advanced . . .

Controversy – overfitting training data

## Motivation and Objective

Question – could a simple PonyGE2 GE beat the stock market

This has been tried before here and here . . . mine is the same/different in this way/etc

Continuing advances in computing power available to not just finance professionals but also consumers

What is the problem

Question being asked - would it be feasible

Objective – beat return on if you had bought one share and held for 1 years

An extension of the classic portfolio optimisation problem – without variance

Just need to code enough to show promise

Morgan Stanley using AI – becoming more and more popular

problem description, proposed variation

What role would a ge or NC alg play in bus an

Significance of the problem . . . and in bus an

Why that algorithm

Problem statement/objective

Method/Alg/Variation

Limitations of GP algorithms

## PonyGE2

Particular algorithm – positives, negatives and limitations of this one

Specific GE algorithm that Pony uses?

How GE works: Fitness, codon, genome, grammar etc mutation, crossover (match in experiment section)

## Paper Layout

# Experimentation

Question being asked - would it be feasible

Just need to code enough to show promise

Assumptions is that S&P companies do not change over time (they are updated . . .)

Fitness function

Simple is better

No cost of placing a trade

Not including variance but this could be included

my trades do not influence market

We can buy partial shares

Trade on opening price everyday

Removed NaN cells from dataset

Setup and results

Can use any input dataset, can add on other variables such as high/low/recommendations etc.

Frist experiment . . . . random buy/sell with probability 50% . .. . . Cash after 16 years:

Final result, cash remaining after 10000: 49376.2923

Site run times as a constraint and reasoning for simplifying

## Assumptions and Simplifications

We only look back one year of data

Cover the point that model does not consider global/outside indicators and influencers

We do not need to show workings to execute . . . some banks may not like this

Single objective – max profit ignoring variance

## Problem Description

## Experiment Results

# Discussion

Analyses of results

# Conclusion

Restate objective

Further research – add other variables, add ability to buy/sell portions not all, add hold option, two objectives – minimise variance

Refer back to objective . . . was it achieved

# References

1. Church, G.M., Gao, Y. & Kosuri, S. 2012, "Next-Generation Digital Information Storage in DNA", Science, vol. 337, no. 6102, pp. 1628.
2. Brabazon, A. & O'Neill, M. 2009, Natural computing in computational finance, Springer-Verlag Berlin and Heidelberg GmbH & Co. K, Berlin.
3. Banks, A., Vincent, J. & Anyakoha, C. 2007, "A review of particle swarm optimization. Part I: background and development", Natural Computing, vol. 6, no. 4, pp. 467-484.
4. Turing, A.M. 1950, "Computing Machinery and Intelligence", Mind, vol. 59, no. 236, pp. 433-460.
5. Friedberg, R.M., 1958. A learning machine: Part I. IBM Journal of Research and Development, 2(1), pp.2-13.
6. Beyer, H. & Schwefel, H. 2002, "Evolution strategies – A comprehensive introduction", Natural Computing, vol. 1, no. 1, pp. 3-52.
7. PolR, Langdon W B, McPhee N F (2008) A Field Guide to Genetic Programming. <http://www.gp-field-guide.org.uk>
8. Koza, J.R., 2010. Human-competitive results produced by genetic programming. Genetic Programming and Evolvable Machines, 11(3-4), pp.251-284.