

Bio-inspired Computing in R

Deliverable 1: Final Year Dissertation

Ryan Porteous

BSc (Hons) Computer Science

Supervisor: Dr Michael Lones

Second Reader: Dr Katrin Lohan

**Declaration of own work**

I, Ryan Porteous confirm that this work submitted for assessment is my own and is expressed in my own words. Any uses made within it of the works of other authors in any form (e.g., ideas, equations, figures, text, tables, programs) are properly acknowledged at any point of their use. A list of the references employed is included.

Signed:

Date:

**Abstract**

R has become the first-choice language for data scientists. However, it is typically not the first choice for people developing and implementing bio-inspired algorithms. Consequently, it can be hard for data scientists to make use of bio-inspired methods. This project will look at the current availability of bio-inspired algorithms in R, identify holes in the provision, and develop a package to fill in one of these holes.

Contents

[1 Introduction 1](#_Toc496917811)

[1.1 Purpose 1](#_Toc496917812)

[1.2 Aims and Motivations 1](#_Toc496917813)

[2 Bio-inspired Computing 3](#_Toc496917814)

[2.1 Genetic Algorithms 3](#_Toc496917815)

[3 Availability of Bio-inspired Algorithms in R 4](#_Toc496917816)

[3.1 Main Areas 4](#_Toc496917817)

[3.2 Areas for Improvement / Areas to be Developed 4](#_Toc496917818)

[4 Chosen Area 4](#_Toc496917819)

[5 Requirements Analysis 4](#_Toc496917820)

[6 R Packages 4](#_Toc496917821)

[6.1 Process of Package Creation 4](#_Toc496917822)

[6.2 Packages and Software Tools to Aid in the Process 4](#_Toc496917823)

[6.3 Possibly another section on versioning or other software? 4](#_Toc496917824)

[7 Evaluation Strategy 4](#_Toc496917825)

[8 Project Management 5](#_Toc496917826)

[8.1 Project Schedule 5](#_Toc496917827)

[8.1.1 Work Breakdown Structure 5](#_Toc496917828)

[8.1.2 Project Timetable 5](#_Toc496917829)

[8.2 Risk Analysis 5](#_Toc496917830)

[8.2.1 Risk Identification 5](#_Toc496917831)

[8.2.2 Risk Management 5](#_Toc496917832)

[8.3 Professional, Legal, Ethical and Social Issues 5](#_Toc496917833)

[9 References 5](#_Toc496917834)

[10 Appendices 5](#_Toc496917835)

# Introduction

## Purpose

The purpose of this document is to give an overview of the domain of the project; to identify and describe the motivations and objectives of the project; and to give initial plans of how to overcome foreseeable problems. Supporting technical literature relevant to the project domain is also discussed.

## Aims and Motivations

Bio-inspired computing is a field which takes inspiration for its algorithms from a variety of nature’s systems such as evolution, and the way populations of animals interact with each other in an environment. This collection of algorithms can be applied to various areas and can also be used as a searching method to solve NP-hard problems due to the way the search space can be explored. R is a programming language which has become more popular in recent years as can be seen in the TIOBE Index [1]. It is among the first-choice of languages for statisticians and data miners, with competition being mainly from Python, but is not the first-choice for those who are interested in implementing bio-inspired algorithms. Due to this, it can be hard for R users to apply bio-inspired methods to problem due to their limited availability.

The primary aims of this project can be defined as:

1. Investigate the availability of bio-inspired algorithms in R

I will begin by identifying the main areas of bio-inspired computing and searching for implementations of these algorithms available through the Comprehensive R Archive Network (CRAN). CRAN is a network which provides a central platform for R users to upload the software packages they have developed and provide access to them for other users.

1. Identify implementations to be improved or built upon, and areas where no solution exists

From the implementations found in the previous step, I will assess how the solution has been implemented, what the solution provides and if it can be improved upon. Where no solution exists, this will be identified as an area that can be developed.

1. Produce an R Package to improve the availability of bio-inspired tools for R

A package will be developed to improve upon an existing solution or to provide a solution where no solution exists. This package will be released on CRAN so that other R users may use the package.

# Bio-inspired Computing

Give an overview of main areas mentioning main applications of each area

Need to find main areas that I can actually talk about and reference. Make sure to get a good amount of references in here backing up all the applications I say.

## Genetic Algorithms

Genetic Algorithms (GAs) are an evolutionary search heuristic which takes inspiration from the process of natural selection [2]. The algorithm uses a population of solutions to the given problem where each solution is given a fitness value which defines how suitable this solution is in this domain. The fitness value which can either be maximised or minimised is given from a fitness function. This fitness value is used in the selection process which mimics natural selection [2]. Each solution has a probability relative to their fitness value of being chosen as a parent. Parent is a term used to refer to a solution from the current generation which will be used in the crossover process to produce a child. A child is a solution that will carry over to the next generation. Crossover is a process, or operator, where two parents are used to generate a child solution. The goal is to combine both parents while removing the negative characteristics of each parent so that the child will have an improved fitness. Another operator which is used is mutation which randomly alters the child solution and can help to explore the search space quickly [3]. For example, one implementation of this is to choose two random values in the solution and to swap them [4]. This process is repeated until a pre-defined number of generations have completed, or a certain number of generations have passed with no improvement.

Cellular Automata

Neural Networks

Emergent Systems / Swarm Intelligence

Genetic Programming

Particle Swarm

# Availability of Bio-inspired Algorithms in R

## Main Areas

Identify R Packages that provide access to the main areas mentioned in Section 3. Take into account what they do, how they are implemented, if they are still being built upon

## Areas for Improvement / Areas to be Developed

Select areas which can be improved or are absent from the above section giving ways to improve the existing implementations if one exists

# Chosen Area

Reasons as to why I have chosen this area

More in depth look at this areas packages

Main literature review

# Requirements Analysis

# R Packages

## Process of Package Creation

## Packages and Software Tools to Aid in the Process

## Possibly another section on versioning or other software?

# Evaluation Strategy

# Project Management

## Project Schedule

### Work Breakdown Structure

### Project Timetable

## Risk Analysis

### Risk Identification

### Risk Management

## Professional, Legal, Ethical and Social Issues

# References

[1] Tiobe.com. (2017). R | TIOBE - The Software Quality Company. [online] Available at: https://www.tiobe.com/tiobe-index/r/ [Accessed 27 Oct. 2017].

[2] Darwin, C. & Wallace, J., 1998. The origin of species / Charles Darwin., Ware: Wordsworth Editions.

[3] Sivanandam, S. and Deepa, S. (2008). Introduction to genetic algorithms. Berlin: Springer, pp.14-39.

[4] Moon, C., Kim, J., Choi, G. and Seo, Y. (2002). An efficient genetic algorithm for the traveling salesman problem with precedence constraints. European Journal of Operational Research, 140(3), pp.606-617.

# Appendices