RESEARCH PROPOSAL

Multiobjective genetic algorithms for improving breeding programme design

9th November 2020 - 1st February 2021

Fergus Currie
Supervisor : Yi Mei

The Problem

The New Zealand research institute Plant and Food are involved in a range of agricultural pursuits. One is farming fish. More specifically Chinook; a species of Pacific salmon. The main goal of farming is profit, as such biologists at Plant and Food want to breed the Chinook so that advantageous traits like taste and size are increased. Fish with more profitable genetic traits will create greater revenue. A difficulty is that if breeding is too selective, genetic diversity of the Chinook population becomes too small. A small gene pool makes it difficult to continue selection in the future.
The breeding program now has two objectives; increase profitable traits and keep genetic diversity large.

Genetic data is complex. As a computer scientist and not a geneticist there is a difficult barrier of entry. Coupled with my inexperience with multi-obejctive optimisation and genetic algorithms, Linley and Maren think it would be easier to work with simulated data. Using software called SLiM more simple data can be created. As this is an optimisation problem the primary concern is inputs or decision variables and outputs or objective functions, and SLiM can be treated as a function. This means no dataset will be made. Rather the algorithm will call SLiM with a set of decision variables, receiving the objective functions back.

Multi-objective optimisation have been studied a lot. Although there are many exisiting algorithms difficult problems need specific algorithms to solve them. It remains to be seen wether this problem will require novel techniques. The good thing about using SLiM simulation as a function is if the modelled problem is too easy it can be made more difficult. It's possible there will be several rounds of increasing simulation difficulty followed by several rounds of algorithm improvement.

Prior Research

A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-2
This paper covers an algorithm for MO which solved three main issues in the field at the time: (1) inefficiency (ON^3), (2) lack of elitism and (3) specification of sharing parameter. NSGA-2 outperformed the techniques it was compared to in almost all instances.

A Comprehensive Survey of Evolutionary-Based Multiobjective Optimisation Techniques
This paper reviews the most important evolutionary-based techniques for MO, the algorithms are summarised and their original paper referenced. This paper is written in 1998 and so does not include many advancements such as NSGA-2.

Multiobjective evolutionary algorithms: A survey of the state of art
This paper surveys advancements in MOEA and MOEA for complicated problems. It focus around developments in the field over the last eight years. Considering it is published in 2011 it is the most recent of these sources.

NSGA-2 and possibly SPEA-2 will be used to find a baseline performance on the dataset. It's likely that this will not sufficiently optimise the data. Therefore improvements will be made to these algorithms. From literature there are many applicable difficulties that could arise. Although much of the research on MO is theoretical, to be applied in the real world constraints are generally required. For example in relation to Salmon the population size must be in a realistic bounds. Noisy data may also be an issue. This is something that I can find out on my trip to Nelson. The data is gathered using image recognition and genetic analysis. The

degree of error may be relevant. It may be that techniques from multiple algorithms are combined to achieve the aim. Depending on the difficulty of the data novel techniques may be developed.

Aim

The aim of this project is to develop novel multi-criteria decision making algorithms/systems for informing breeding programme designs and to obtain a set of models of breeding programs with trade-offs between various potentially conflicting objectives, such as short-term and long-term gains.

Goals

1. Decide on metrics, constraints, parameters and decision variables
    1.1. This will require talking with Plant and Food as they have more of the biological understanding
2. Read relevant literature related to MO
    2.1. A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-2
    2.2. A Comprehensive Survey of Evolutionary-Based Multiobjective Optimisation Techniques
    2.3. Multiobjective evolutionary algorithms: A survey of the state of art
3. Collect data
4. Learn how to use relevant tools
    4.1. Matlab
    4.2. SLiM
5. Develop algorithm over several rounds
    5.1. Baseline with a range of prewritten algorithms such as NSGA-2 and SPEA-2

Each week I meet with plant and food. From our conversations an important aspect of performing this optimisation is the robustness of the algorithm. It needs to be applicable to more than the data we build it on. It remains to be seen wether this will be a requirement for future work due to the difficulty, which currently is uncertain.

Milestones

WEEK (Monday of week)	Milestones
9/11/2020	
16/11/2020	Reading Literature
23/11/2020	Finish Proposal and bench-marking
30/11/2020	
7/12/2020	
14/12/2020	Base line results + report sent to plant and food
BREAK	
11/1/2021	
18/1/2021	
25/1/2021	
1/2/2021	Paper sent to IEEE

Resources

SLiM
Matlab
Internet
Python
Papers/Readings
Victoria Computing, Printing and Work station