Research Excellence Framework Optimisation Problem

Project Plan

52103576

Bryn Elliott

u03be21@abdn.ac.uk

Department of Computing Science,
University of Aberdeen, Aberdeen AB24 3UE, UK

Introduction

The Research Excellence Framework (REF) is the main national system for assigning block-grant research funding to Higher Education Institutions (HEIs)[1]. While funding is assigned at the institution level, assessments are carried out at the research level by one of a large number of panels. Due to the fact that not all research is submitted, as well as so cross-functional research which can be submitted to one of several panels, there is significant opportunity for optimisation in terms of submission structure. While the final aim is to maximise funding, this is difficult to directly measure on a paper-to-paper basis, so the REF 3-part rating system (each of Output, Impact and Sustainability) is being optimised as an intermediary.

The project evaluates the REF submission as a constrained optimisation problem where a certain number of papers must be assigned to panels (Constraint applied by REF, not all research can be submitted) in order to maximise the average ratings of the institution as a whole.

Goals

The aim of this project is to answer the research questions, the hypotheses are:

H1: 'A Genetic Algorithm will have a statistically significant improvement in average rating for the institution when compared to historic submissions'

H2: 'A Genetic Algorithm will have a statistically significant improvement in average rating for the institution as a whole when compared to a greedy heuristic approach'

This will be accomplished by building a genetic algorithm that is able to optimise the total submissions of one Institution to maximise funding. To quantify success within the scope of this project (unable to wait until a system output can be submitted to REF 2029), it will be measured in 2 parts:

1. The first is to calibrate the ratings from the fitness rating system. This is vital to compare hypothetical submission structures. This would be regarded as successful if the average fitness rating (mean of the estimated ratings for each paper that are contained within a submission) is consistently within a 5% margin (standard practice for statistical significance) of the real ratings an institution's research received[2] across its submissions.
2. Once this is verified, our second is that the algorithm outputs a submission set that overachieves when compared to the real submission set in that verified fitness check. Both success criteria must be achieved across many submission sets to verify robustness with different input sets.

As a stretch goal, the inclusion of supplementary pieces to build the algorithm into a usable system that can be deployed to institutions. This would require several additional components including a SQL database for storing submissions, a remote web interface for non-local data input, among others.

Methodology

The majority of this development is in several. These will each be built iteratively, testing throughout with 2021 datasets.

Regular milestones are laid out in the Schedule section. There are, however, some key milestones that mark significant progress. These are:

- The creation of the fitness rating system.
- The creation of the genetic algorithm itself.
- The creation of the baseline greedy heuristic.
- Stretch: Unifying milestone 2 and 3 into a usable software for institutions.

The fitness rating system being built first creates the foundation for both the genetic algorithm and the greedy heuristic, allowing for a like-for-like comparison of the 2 approaches.

Once the systems are built, the two algorithms (genetic algorithm and greedy heuristic) are then run on the 2021 dataset, and their average final average rating from several runs is taken, this is done to offset the random nature of genetic algorithms with variance taken to assess consistency. These ratings are then compared to each other and the real average rating received in 2021. This will be repeated for several different institutions to verify repeatability of the various methods.

Resources Required

This project is not resource intensive, requiring only a PC for hardware. This is due to the manageable dataset size and moderate compute requirements.

In terms of software/languages used. The algorithm and heuristic themselves will be built in Python. This is due to the strong ecosystem for data analysis work and optimisation in Python. The fitness ranking system will also most likely be built in Python, to integrate more easily with the algorithms, subject to potential limitations of the language. If such limitations create roadblocks for the build process, the fitness ranking system can also be built in C# and integrated separately.

Deployment-related components (as listed in the stretch goals) would be built in C#. This is due to its diverse toolset in this area when compared to Python.

C# will only be used if stretch goals are pursued, the core research can be conducted in purely Python.

While genetic algorithm libraries do exist, building a custom one allows for much greater control over the form of input/output, which, due to the REF's complex rating system, offsets the extra work and time lost from the custom build.

Finally, a remote GitHub repository will be used for version control and synchronisation across devices to mitigate the risk of progress loss or data corruption as well as verify built software across device types for robustness. The final piece is the datasets of papers and ratings, which are both publicly available online.

Risk Assessment

The project is heavily dependent on the success of the genetic algorithm, so a failure to create it will result in a complete failure of the project goal. To counteract this, a proof-of-concept algorithm is being built first to verify the efficacy of the algorithm build before the full system build.

The only other potential risk of this project is, as with all projects, is progress loss. This is amplified by the intention to work across devices depending on location. To counteract this, I will be using a central GitHub repository to synchronise and back up data and code as it is created.

Timetable

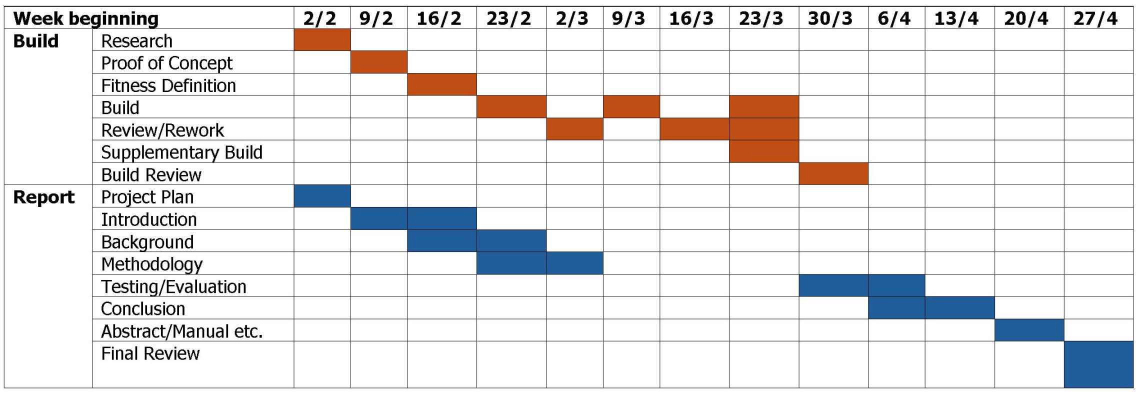

Figure 1: Timetable for Build, Evaluation and Reporting process at a weekly level, between 2/2/2026 and 3/5/2026

Deliverables

The output of the system is a layout of an Institution-wide submission, detailing which papers to submit and how to submit them, along with an estimated average rating for the whole submission set.

In addition to this, a report detailing the creation process of the system will be included in the final submission.

References

[1] Sofia Branco Sousa and John L. Brennan. *The UK Research Excellence Framework and the Transformation of Research Production*, pages 65–80. Springer Netherlands, Dordrecht, 2014.

[2] Gobinda Chowdhury, Kushwanth Koya, and Pete Philipson. Measuring the impact of research: Lessons from the uk's research excellence framework 2014, Jun 2014.