

DR MENGJIE ZHANG

mengjie.zhang@vuw.ac.nz

<http://homepages.ecs.vuw.ac.nz/~mengjie/>

Genetic Programming for Data Mining

Background

Data mining tasks arise in a wide variety of practical situations, ranging from classification to regression, clustering, and optimisation tasks. The applications range from the military domain such as detecting F-15 helicopters and tanks from a set of satellite images, through the economic domain such as finding associate rules at retail sellers and predicting GDP or CPI of a nation/region, the engineering domain such as network intrusion detection and pattern matching in signal processing, and to daily life such as postal code recognition, human face detection and security control.

Genetic programming (GP) is an emergent but promising approach to building reliable data mining models quickly and automatically. GP uses ideas analogous to biological evolution to search the space of possible models to evolve a good solution for a particular task. GP has been applied to data mining and achieved some success. However, there are still limitations in program structures, search algorithms and theoretical foundations in GP that restrict GP for difficult data mining tasks.

The Project

The project aims to develop and investigate new methods and techniques to address the methodological/theoretical issues in GP described above for data mining tasks with the goal of improving performance of data mining or successfully solving real-world engineering data mining applications. Specifically, at least one of the following research topics will be considered in a PhD thesis:

- Develop and explore new representations and structures of computer programs in the population that GP can more effectively evolve and that are more suitable for multi-class classification
- Develop and investigate new methods and algorithms for automatically selecting important features from a large number of dimensions of low-level features and constructing a small number of high-level features from the relevant low-level features for data mining tasks
- Develop new fitness functions that are effective for classification for the problems where the numbers of training examples for different classes are highly unbalanced
- Develop new operators and techniques for reducing the computational cost of GP
- Explore new techniques for reuse of programs and subprograms for regression, classification, prediction and/or optimisation problems.
- Apply GP to engineering applications to successfully evolve good models and hidden patterns for real-world problems.

The Students

A strong background in C/C++/Java programming and a basic background in Artificial Intelligence and statistics are required. A good background in machine learning, statistics and operations research is desired.

Evolutionary Computer Vision

Background and Project

Computer vision and image processing is a major unsolved problem in computer science and engineering. Over the last decade there has been increasing interest in using evolutionary computation approaches to solving vision problems. Computer vision provides a range of problems of varying difficulty for the development and testing of evolutionary algorithms.

Since the late 1990s, evolutionary algorithms have been shown to be tools which can be used effectively in the development of systems (software or hardware) for computer vision and image/signal processing in complex domains of high industrial and social relevance.

The proposed project aims to use evolutionary algorithms for solving computer vision and image processing problems. The evolutionary algorithms will consider genetic programming, genetic algorithms, particle swarm optimisation, artificial immune systems, etc. Other computational intelligence techniques particularly neural networks will also be considered. Topics of interest will be focused on:

- Apply evolutionary algorithms to real-world vision and image processing problems
- Apply evolutionary algorithms to image recognition and analysis
- Apply hybrid approaches related to evolutionary algorithms to image analysis and processing
- Investigate theoretical developments
- Compare between different evolutionary techniques and between evolutionary and non-evolutionary techniques in image and vision applications

The students

A good background of programming, computer vision and image processing, and evolutionary computation is desired. A basic background of computer science and artificial intelligence is required.

Particle Swarm Optimisation for Classification

Background and Project

Introduced in the mid-1990s, particle swarm optimisation (PSO) is a relatively new paradigm in the area of evolutionary computation. Similarly to other EC paradigms

such as genetic algorithms, PSO is also a stochastic, population-based evolutionary computer algorithm for problem solving. Unlike other EC paradigms, it is a kind of swarm intelligence that is based on social-psychological principles and provides insights into social behaviour, as well as contributing to engineering applications.

Although a relatively new EC paradigm, PSO has been applied to a variety of tasks, such as the training of artificial neural networks and for finite element updating. Recently, PSO has been also applied in combination with grammatical evolution to create a hybrid optimization paradigm called "grammatical swarms".

However, there are only a very small number of applications of PSO to classification tasks. The current PSO techniques are still limited for classification.

This project aims to investigate and explore new methods and algorithms within PSO with the goal of improving system performance for classification tasks. The focus will be mainly on the following aspects:

- Explore and develop new fitness functions of within the canonical PSO form for classification
- Develop effective distance measures for evaluating individual particles for classification
- Investigate the relationship between the global best, neighbourhood best and the local best, and investigate more effective representations and schemes for classification
- Apply PSO to image recognition and analysis tasks

The students

A basic background of computer science and artificial intelligence is required. A good background of programming, pattern recognition and machine learning, and evolutionary computation is desired.