University of Exeter
Department of Computer Science

Ordering and Visualisation of Many-objective Populations

David John Walker

September 2012

Supervised by Dr Jonathan Fieldsend and Prof Richard Everson

Submitted by David John Walker, to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Computer Science, September 2012.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

....

for my family

Abstract

In many everyday tasks it is necessary to compare the performance of the individuals in a population described by two or more criteria, for example comparing products in order to decide which is the best to purchase in terms of price and quality. Other examples are the comparison of universities, countries, the infrastructure in a telecommunications network, and the candidate solutions to a multi- or many-objective problem. In all of these cases, visualising the individuals better allows a decision maker to interpret their relative performance. This thesis explores methods for understanding and visualising multi- and many-criterion populations.

Since people cannot generally comprehend more than three spatial dimensions the visualisation of many-criterion populations is a non-trivial task. We address this by generating visualisations based on the dominance relation which defines a structure in the population and we introduce two novel visualisation methods. The first method explicitly illustrates the dominance relationships between individuals as a graph in which individuals are sorted into Pareto shells, and is enhanced using many-criterion ranking methods to produce a finer ordering of individuals. We extend the power index, a method for ranking according to a single criterion, into the many-criterion domain by defining individual quality in terms of tournaments. The second visualisation method uses a new dominance-based distance in conjunction with multi-dimensional scaling, and we show that dominance can be used to identify an intuitive low-dimensional mapping of individuals, placing similar individuals close together. We demonstrate that this method can visualise a population comprising a large number of criteria.

Heatmaps are another common method for presenting high-dimensional data, however they suffer from a drawback of being difficult to interpret if dissimilar individuals are placed close to each other. We apply spectral seriation to produce an ordering of individuals and criteria by which the heatmap is arranged, placing similar individuals and criteria close together. A basic version, computing similarity with the Euclidean distance, is demonstrated, before rank-based alternatives are investigated. The procedure is extended to seriate both the parameter and objective spaces of a multi-objective population in two stages. Since this process describes a trade-off, favouring the ordering of individuals in one space or the other, we demonstrate methods that enhance the visualisation by using an evolutionary optimiser to tune the orderings.

One way of revealing the structure of a population is by highlighting which individuals are extreme. To this end, we provide three definitions of the "edge" of a multi-criterion mutually non-dominating population. All three of the definitions are in terms of dominance, and we show that one of them can be extended to cope with many-criterion populations.

Because they can be difficult to visualise, it is often difficult for a decision maker to comprehend a population consisting of a large number of criteria. We therefore consider criterion selection methods to reduce the dimensionality with a view to preserving the structure of the population as quantified by its rank order. We investigate the efficacy of greedy, hill-climber and evolutionary algorithms and cast the dimension reduction as a multi-objective problem.

I would like to thank several people for their help and support during the course of my PhD. My parents have supported me in many ways over the last four years, and it is very much appreciated. I would also like to thank my many colleagues, particularly: Andrew Clark, Jacqueline Christmas, Max Dupenois, Kent McClymont and Zena Wood. I am also grateful to Antony Galton and Joshua Knowles for agreeing to examine this thesis.

Special thanks are due to my supervisors, Jonathan Fieldsend and Richard Everson, who have been a constant source of sound advice and encouragement.

Contents