Rank-based Dimension Reduction for Many-criteria Populations

David J. Walker*
Computer Science
University of Exeter, UK
D.J.Walker@ex.ac.uk

Richard M. Everson
Computer Science
University of Exeter, UK
R.M.Everson@ex.ac.uk

Jonathan E. Fieldsend
Computer Science
University of Exeter, UK
J.E.Fieldsend@ex.ac.uk

ABSTRACT

Interpreting individuals described by a set of criteria can be a difficult task when the number of criteria is large. Such individuals can be ranked, for instance in terms of their average rank across criteria as well as by each distinct criterion. We therefore investigate criteria selection methods which aim to preserve the average rank of individuals but with fewer criteria. Our experiments show that these methods perform effectively, identifying and removing redundancies within the data, and that they are best incorporated into a multi-objective algorithm.

Categories and Subject Descriptors

I.5.2 [**Computing Methodologies**]: Design Methodology—*Feature evaluation and selection*

General Terms

Algorithms

Keywords

Feature selection, multi-criteria decision making, visualisation

1. INTRODUCTION

In many populations the individuals attempt to optimise their performance on a set of K *criteria*, for example the solutions of a many-objective evolutionary optimisation, or the example considered here, universities which strive to optimise their performance on $K = 8$ key performance indicators such as student satisfaction and research quality. In order to facilitate understanding of the structure of the population in a smaller space we investigate methods of selecting the most informative criteria.

The inspiration for our dimension reduction methods is drawn from the comparison of different rankings, and we quantify the structure of the population by the *average rank* [1] of individuals, computed by averaging for each individual the result of ranking the population by each criterion in turn. This enables comparison between the original full-criteria set and criterion subsets by average rank, which we wish to preserve. We demonstrate the use of a multi-objective evolutionary algorithm (MOEA) for selecting the low-dimensional subset.

*David Walker was supported by an EPSRC CASE award with Motorola during the preparation of this paper.

GECCO'11, July 12–16, 2011, Dublin, Ireland.
ACM 978-1-4503-0690-4/11/07.

D		15	14	13	12	11	10	9	8	7	6	5	4	3	2
Duplicated Criteria	ρ	13	10	11	14	12	9	16	15	5	2	3	6	4	1
	τ	15	10	11	14	16	9	12	13	5	2	4	1	8	6
Averaged Criteria	ρ	-	-	-	-	-	-	9	10	3	5	4	1	7	8
	τ	-	-	-	-	-	-	9	10	3	4	1	5	8	6

Table 1: The order in which criteria are removed by a greedy backward algorithm for each ρ and τ on both augmented GUG09 datasets. Both metrics remove the synthetic criteria, 9-16 in the duplicated data, 9 and 10 in the averaged data, before any of the original criteria.

We illustrate these methods with the Times Good University Guide 2009 (GUG09) dataset which reports on the performance of 113 UK universities in 2008 [4, 5]. In order to evaluate how well rank-based dimension reduction can remove redundancy, the GUG data is modified to produce two synthetic datasets: one in which each criterion is duplicated; and the other in which two new criteria have been created by averaging two pairs of highly correlated criteria from the original data.

2. RANK-BASED DIMENSION REDUCTION

A ranking is a permutation of individuals, and we describe two well known metrics for the comparison of different permutations, Spearman's *footrule* [2] and Kendall's τ metric [3].

Spearman's footrule [2] is the city block distance between two rankings $\mathbf{r}$ and $\mathbf{r}'$ of N individuals, $\rho(\mathbf{r}, \mathbf{r}') = \sum_{i=1}^{N} |r_i - r'_i|$ where r_i is the rank of the ith individual. Kendall's τ metric [3] compares permutations by counting the number of times that the order of pairwise individuals in the two permutations is reversed, $\tau(\mathbf{r}, \mathbf{r}') = \sum_{ij} \tau_{ij}(\mathbf{r}, \mathbf{r}')$ where $\tau_{ij}(\mathbf{r}, \mathbf{r}') = 1$ if the ordering of the individuals i and j is different in $\mathbf{r}$ and $\mathbf{r}'$, and $\tau_{ij}(\mathbf{r}, \mathbf{r}') = 0$ otherwise. Although this formulation of the τ metric does not account for ties, it may straightforwardly be modified to do so.

We rank the original population by average rank [1], producing a permutation $\mathbf{r}$ of the population with respect to all K criteria. Then, as new subsets are considered we re-rank the population accordingly, producing the permutation $\mathbf{r}'$, and compute the distance between $\mathbf{r}$ and $\mathbf{r}'$ to identify the criterion subset that minimises the distance between permutations, and thus most closely preserves the structure of the population.

Initial work incorporated the permutation comparisons into a greedy criteria selection algorithm to reduce the dimensionality of the of the modified GUG datasets. Figure 1 shows the distance $\delta(\bar{\mathbf{r}}, \bar{\mathbf{r}}')$ between the original (all criteria) population average rank

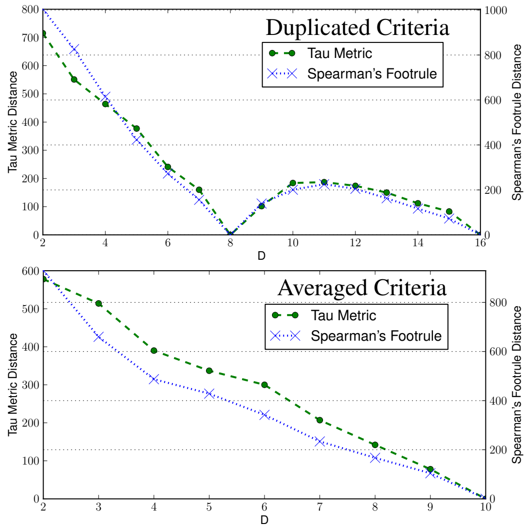

Figure 1: Information loss for criterion subsets of increasing size. Note that the two metrics are plotted on different ordinates, but for both datasets follow a similar trend. When $D = 16$ in the duplicated criterion set, and $D = 10$ in the averaged criterion set, all criteria are selected. In the duplicated set, when $D = 8$, information loss is 0 as all redundant copies have been removed leaving a single copy of each original criterion.

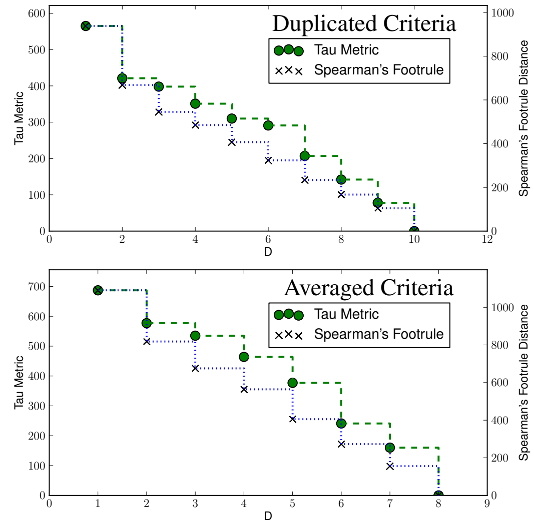

Figure 2: Attainment surfaces produced by optimising the number of criteria against the information loss measured by ρ and τ. Note, that since the metrics are plotted on different scales, they cannot be compared in terms of dominance.

$\bar{\mathbf{r}}$ and the average rank of the selected subset as criteria are removed. The two plots show the number of remaining criteria D along with the corresponding information loss $\delta(\bar{\mathbf{r}}, \bar{\mathbf{r}}')$ by both Kendall's τ metric and Spearman's footrule. Figure 1 shows results for the synthetic criterion sets. These figures, in conjunction with Table 1, show that the algorithms identify the redundant criteria before removing any of the original criteria.

2.1 Multi-objective Criterion Selection

Empirically we have found that a stochastic hill climber can approximate the information loss of the deterministic greedy algorithm in fewer steps than exhaustive search would require for a given D. An alternative is to relax the constraint governing D to an objective, and employ a MOEA to simultaneously trade-off information loss and the number of remaining criteria.

The problem is therefore defined as follows. The solution $\boldsymbol{\theta}$ is represented as a K-bit string coding which criteria are selected, which maps to a pair of objectives:

$$f_1(\bar{\mathbf{r}}, \bar{\mathbf{r}}', \boldsymbol{\theta}) = \delta(\bar{\mathbf{r}}, \bar{\mathbf{r}}'), \qquad f_2(\bar{\mathbf{r}}, \bar{\mathbf{r}}', \boldsymbol{\theta}) = \textstyle\sum_{k=1}^{K} \theta_k.$$

The first objective is the distance between permutations as before, and the second objective counts the number of remaining criteria.

Figure 2 shows *attainment surfaces* for the optimisation. Results for 20 independant runs of each metric have been merged to produce a single archive containing all of the non-dominated solutions found during the 20 runs.

By comparing the Pareto optimal criterion subsets, we can compare the solutions identified by the algorithm. In the duplicated dataset (Figure 2, top), both Spearman's footrule and Kendall's τ identify the same criterion subset in most cases. In the case $D = 8$, where the information loss is minimised, the subset contains exactly one instance of each criterion, either the original criterion or the copy.

For the averaged dataset (Figure 2, bottom), the smaller subsets prefer to include the additional criteria formed by averaging original criteria. We infer that this is because they were constructed from two of the original criteria, so that retaining a composite criterion preserves more more information about the rank structure of the data than either individual criterion. As the size of the criterion subsets increases, there is sufficient structural influence from the original criteria, and the averaged criteria are no longer included. When $D = 8$, the subset comprises all 8 original GUG09 criteria.

3. CONCLUSION

Applying the criterion selection process to the GUG09 data without synthetically redundant criteria finds that the two most significant criteria, contributing the most to the overall structure, are research quality and entry standards.

We are currently extending this work by investigating the efficacy of incorporating criterion selection into a MOEA, so that it can optimise problems consisting of a larger number of objectives by dynamically selecting the most relevant subset for optimisation.

4. REFERENCES

[1] P. J. Bentley and J. P. Wakefield. Finding acceptable solutions in the Pareto-optimal range using multiobjective genetic algorithms. In *Soft Computing in Engineering Design and Manufacturing*, pages 231–240, 1998.

[2] P. Diaconis and R. L. Graham. Spearman's footrule as a measure of disarray. *Journal of the Royal Statistical Society, Series B*, 39(2):262–268, 1977.

[3] M. G. Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93, 1938.

[4] J. O'Leary. *The Times Good University Guide 2009*. Times Books, 2009.

[5] D.J. Walker, R.M. Everson, and J.E. Fieldsend. Ordering and visualising many-objective populations. In *2010 Congress on Evolutionary Computation*, pages 3664–3671, 2010.