Interactive k-means clustering for investigation of optimisation solution data

Author

Rawlins, Tim, Lewis, Andrew, Hettenhausen, Jan, Mirjalili, Seyedali

Published

2012

Conference Title

CTAC 2012 Book of Abstracts

Rights statement

Downloaded from

Link to published version

Griffith Research Online

ANZIAM J. 0(2012), 1–2

Interactive k-means clustering for investigation of optimisation solution data

T. RAWLINS ✉ [1], A. LEWIS[2], J. HETTENHAUSEN[3] and S. MIRJALILI[4]

(Received xx Month 2012)

Abstract

1. Extended Abstract

Currently, various multi-objective optimisation techniques are used to solve complex problems in engineering design. These techniques often generate a great deal of data that is unused because it is difficult to analyse and interpret, and then infer the behaviour of the complex underlying systems. Although there are various techniques to find patterns in this data, finding patterns in the small subset of data that comprises the Pareto optimal set may obscure those present in the general data. This can be problematic when attempting to find solutions that are robust to larger changes in operating conditions. We aim to use interactive techniques to investigate larger amounts of data from the total dataset an experiment yields.

As a case study, we look at applying k-means clustering [1] to data obtained from an airfoil optimisation problem. The parameter space was 8-dimensional, corresponding to the change in x and y positions of 4 control points of the free form deformation technique used to manipulate the airfoil shape, and the objectives were lift and drag as calculated by Xfoil. Many solutions were deemed infeasible due to violating design constraints or not functioning at certain angles of attack.

From computational experiments 51432 valid results were acquired, of which 36 were in the Pareto optimal set. Because the values on the Pareto front are those of most interest, the k-means clustering algorithm with 2 clusters was applied to a 36 x 8 array, each row being the parameter values of one of the Pareto optimal solutions. This was seeded using the parameter values corresponding to the solutions with best

[1]Griffith University, Australia; e-mail: tim.rawlins@griffithuni.edu.au.
[2]Griffith University, Australia; e-mail: a.lewis@griffith.edu.au.
[3]Griffith University, Australia; e-mail: j.hettenhausen@griffith.edu.au.
[4]Griffith University, Australia; e-mail: seyedali.mirjalili@griffithuni.edu.au.

lift and best drag as the initial cluster centroids. This resulted in the Pareto front being approximately bisected, dividing the lift and drag values, which was promising.

Since such a small number of values was insufficient to study the behaviour across the extent of the parameter space, the same technique was then applied to all the data. However, clusters divided between high lift and low drag weren't found. The entire dataset contains a large "plume" of dominated solutions with very poor lift and drag values, so that any seeds chosen near the Pareto front caused clustering to discriminate solutions "near" and "far" from the front.

To address this difficulty, an interactive version of k-means was developed where the number of clusters and cluster seeds could be adjusted in real time. This allows dynamic exploration of the dataset, with the projection of derived clusters in parameter space being displayed using a "colour map" [2]. This interface was extended to allow interactive selection of a subset of the data, allowing progressive expansion of the set of solutions considered from the successfully clustered Pareto optimal set. Preliminary results show that a significant number of points can be retained while obtaining a lift/drag split, allowing investigation of characteristic parameter values for different solution categories. It appears that this simple, interactive approach to exploration of optimisation data can lead to increased insight into the behaviour of the underlying systems.

References

[1] J.B. MacQueen. (1967). 'Some methods for classification and analysis of multivariate observations'. *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, (1967), 281–297.

[2] J. Hettenhausen, A. Lewis and S. Mostaghim. 'Interactive multi-objective particle swarm optimisation with heatmap visualisation based user interface'. *Engineering Optimization*, **42** (2), (2010), 119–139.