INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

IMPROVEMENT IN POST PARETO ANALYSIS IN MULTI-OBJECTIVE OPTIMIZATION USING CLUSTERING TECHNIQUE

P.M.CHAUDHARI, Dr. R.V. DHARASKAR, Dr. V. M. THAKARE

1. Ph.D. Student,Post Graduate Department of Computer Science & Engineering, G. H. Raisoni College of Engineering, Nagpur, India (Currently at PIET, Nagpur, India)
2. Director, Matoshri Pratishthan's Group of Institutions (MPGI) Integrated Campus, Nanded, India.
3. Professor & Head, Post Graduate Department of Computer Science, Faculty of Engineering & Technology, S.G.B. Amravati University, Amravati, India.

Accepted Date:

27/02/2013

Publish Date:

01/04/2013

Keywords

Multi-objective problems, Clustering technique, Pareto-optimal set

Corresponding Author

Mrs. P.M. Chaudhari

Abstract

The data clustering is a classical activity in data mining. In this paper we propose a method to carry out data clustering using Evolutionary algorithms. We use evolutionary characteristics to define the data clustering procedure. Evolutionary clustering technique is proposed that opts for cluster centers straight way from the data set, further making it to speed up the fitness evaluation by estimating a data table in advance. It saves the distances among pairs of data points, and by using binary instead of string representation to encode a variable number of cluster centers.

INTRODUCTION

In many engineering optimization problems, it is not rare to face a challenge when there are several criteria or problem objectives to be satisfied simultaneously. Considering that generally such objectives are in conflict with each other, and then the problem becomes one of finding the best possible solution that satisfies the competing objectives under different tradeoff scenarios. With several multiple objectives and constraints taken into consideration, an accurate optimization formulation can be determined. This type of problems is known as *multi-criteria problems* [1]. Because of their nature, multi-objective optimization problems may not have one solution which is best (global minimum or maximum) with respect to all objectives. Instead, there may be a set of solutions which are superior to the rest of the solutions in the search space when all objectives are considered, but are inferior to other solutions in the search space in one or more objectives. These solutions are known as Pareto-optimal solutions or non-dominated solutions [2]. Although several methods for solving multi-objective optimization problems have been developed and studied, little prior work has been done on the evaluation of results obtained in multi-objective optimization. Value function is used to help the decision-maker identify the most preferred solution in multi-objective optimization problems. Greedy Algorithm (GR) is analyzed to obtain a sub-set of Pareto optima from a larger Pareto set. The selection of the sub-set was based on maximizing a function of the vector of percentile ordinal rankings of the Pareto optima within the large set. However, choosing a solution for system implementation from the Pareto-optimal set can be a difficult task, generally because Pareto-optimal sets can be extremely large or even contain an infinite number of solutions. [3]

This discussion makes clear that there is a need to achieve smaller practical sets of promising solutions. Thus, the motivation for the current work stems from challenges encountered during the post-Pareto analysis phase. A practical approach is proposed to help in the analysis of the solution of multi-objective optimization and provide the decision-maker a workable sized set of solutions to analyze [4].This method is based on an unsupervised cluster analysis technique, in which the solutions in the Pareto optimal set are clustered so that the Pareto optimal front is reduced to a set of *k* clusters [5]. Each cluster consists of solutions with similar properties, and therefore the decision maker only has to investigate one solution per cluster; in this case, the closest solution to each cluster centroid. Moreover, with this method, once the optimal number of clusters is identified, the decision maker can focus to the "knee" cluster, which contains the solutions that are likely to be more interesting to the decision maker [6].

To illustrate the method, some well-known optimization problems will be formulated as multiple objective problems. To solve them, the fast elitist non-dominated sorting genetic algorithm (NSGA-II) will be initially used to determine a set of Pareto solutions [7]. To reduce the size of the Pareto-optimal solutions, a partitional clustering algorithm will be used to directly decompose the Pareto-optimal set into a set of disjoint clusters. This method does not require *a priori* knowledge of the relative importance of the conflicting objectives, providing the decision-maker a smaller set of optimal tradeoffs [8].

2. CLUSTER ALGORITHM

Cluster analysis, also known as unsupervised learning, is one of the most useful methods in the cluster analysis process for discovering groups. Clustering aims to organize a collection of data items into clusters, such that objects within the same cluster have a high degree of similarity, while objects belonging to different clusters have a high degree of dissimilarity. Cluster analysis makes it possible to look at properties of whole clusters instead of individual objects. This is a simplification that is useful when handling large amounts of data [7].

```
Begin
1.  t=0
2.  initialize population P(t)
3.  compute fitness P(t)
4.  t = t+1
5.  if termination criterion achieved go to step 10
6.  select P(t) from P(t-1)
7.  crossover P(t)
8.  mutate P(t)
9.  go to step 3
10. Output best and stop
End
```

Fig. 1. Basic steps in GAs.

3. METHODOLOGY

A cantilever design problem is considered with two decision variables i.e. diameter (d) and length (l). the beam has to carry an end load P. Let us also consider two conflicting objectives of design , i.e. minimization of weight f1 and minimization of end deflection f2. the first objective will resort to an optimum solution having the smaller dimensions of d and l, so that the overall weight of the beam is minimum. Since the dimensions are small , the beam will not be adequately rigid and the end deflection of the beam will be large. On the other hand . if the beam is minimized for end deflection , the dimensions of the beam are expected to be large , thereby making the weight of the beam large .the left plot in Figure 1 marks the feasible decision variable space in the overall search space enclosed by 10 ≤ d ≤ 50 mm and 200≤ l ≤ 1000 mm. it is clear that not all solutions in the rectangular decision space are feasible . Every feasible solution in this space can be mapped to a solution in the feasible objective space shown in the right plot. The correspondence of a point in the left figure with that in the right figure is also shown.

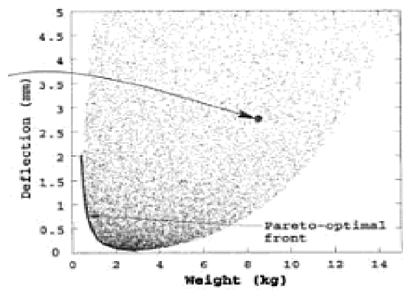

Fig.2 The feasible objective space (right)

This Fig 2 shows many solutions trading-off differently between the two objectives. Any two solutions can be picked from the feasible objective space and compared. For some pairs of solutions, it can be observed that one solution is better than the other in both objectives as given in Table 1.All solutions lying on this curve are special in the context of multi-objective optimization and are called Pareto-optimal solutions. The curve formed by joining these solutions is known as Pareto-optimal front.

Table 1 Five solutions for the cantilever design problem.

Solution	d (mm)	l (mm)	Weight (kg)	Deflection (mm)
A	18.94	200.00	0.44	2.04
B	21.24	200.00	0.58	1.18
C	34.19	200.00	1.43	0.19
D	50.00	200.00	3.06	0.04
E	33.02	362.49	2.42	1.31

This approach is suitable for decision-makers that do not have *a priori* knowledge of the relative importance of the conflicting objectives in Multicriteria optimization problem.

The developed approach is based on the following steps:

1. Obtain the entire Pareto-optimal set or sub-set of solutions by using a multiple-objective evolutionary algorithm (MOEA) or by another means.

2. Apply the GA based clustering algorithm to form clusters on the solutions contained in the Pareto set.

3. To determine the "optimal" number of clusters, *k*, in this set, silhouette plots are used. A value of the silhouette width, *s*(*i*), is obtained for several values of *k*. The clustering with the highest average silhouette width is selected as the "optimal" number of clusters in the Pareto-optimal set.

4. For each cluster, select a representative solution. To do this, the solution that is closest to its respective cluster centroid is chosen as a good representative solution.

5. Analyze the results. At this point, the decision-maker can either:

5.1 Analyze the "knee" cluster. The suggestion is to focus on the cluster that has solutions that conform to the "knee" region. The "knee" is formed by those solutions of the Pareto-optimal front where a small improvement in one objective would lead to a large deterioration in at least one other objective. Moreover, from this "knee" cluster the decision maker can select a promising solution for system implementation. This would be the solution closest to the ideal or utopian solution of the multiple objective problem, in a standardized space.

5.2 Analyze the *k* representative solutions and/or select the most promising solutions among this *k* set, selecting the solution closest to the ideal point. By applying the proposed technique, the Pareto-optimal front of a multiple objective problem can be reduced to the "knee cluster" as in 5.1, or to a set of *k* solutions as in 5.2. In both cases the decision maker can choose a good tradeoff for system implementation by selecting the closest solution to the ideal or utopian solution of the multiple objective problems, in a standardized space.

A Matlab code is developed to perform the steps of the proposed technique. From standardized data, the code will run the *clustering* algorithm and from two to a specified number of means it will calculate the average silhouette values and it will return the value of *k* suggesting the most optimal allocation. After this, it will also return the "knee cluster" of the optimal partition, the *k* representative solutions of the Pareto front, and in both cases, the solution closest to the ideal or utopian point.Fig.3 shows the solution sets.

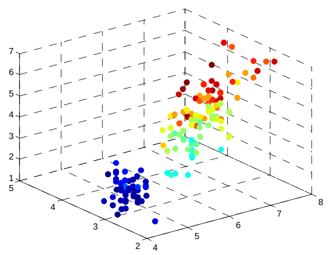

Fig.3 The Solution Set1

CONCLUSION

The proposed methodology based on cluster analysis is implemented to assist the decision-maker in the analysis of the solutions of multi-objective problems .Choosing a solution for system implementation from the Pareto-optimal set can be a difficult task, generally because Pareto-optimal sets can be extremely large or even contain an infinite number of solutions. The proposed method provides the decision-maker a smaller set of optimal tradeoffs [22].

Over the past few years, the research on evolutionary algorithms has demonstrated their role in solving multi-objective optimization problems, where the goal is to find a number of Pareto-optimal solutions in a single simulation run. Many studies have depicted different ways evolutionary algorithms can progress towards the true Pareto-optimal solutions with a widely spread distribution of solutions. However, none of the multi-objective evolutionary algorithms (MOEAs) has a proof of convergence to the true Pareto-optimal solutions with a wide diversity among the solutions [23].

Thus, the motivation for the current work stems from challenges encountered during the post-Pareto analysis phase. A practical approach is proposed to help in the analysis of the solution of multi-objective optimization and provide the decision-maker a workable sized set of solutions to analyze.

The proposed methodology is very much feasible taking into the account the work carried out by the researchers internationally.

REFERENCES

1. Hirotaka Nakayama, Multi-objective Optimization and its Engineering Applications, Proc. of third China-Japan-Korea Joint Symposium on Optimization of Structural and Mechanical System, pp.13-26 (2004)

2. Zitzler, E. and Thiele, L. (1999), Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. IEEE Transactions on Evolutionary Computation, 3, p. 257–271.

3. Kaufman L. and Rousseeuw P.J. (1990). Finding groups in data. An introduction to Cluster Analysis. Wiley-Interscience.

4. Kalyanmoy Deb and Shamik Chaudhuri. I-MODE: An Interactive Multi-Objective Optimization and Decision-Making using Evolutionary Methods. KanGAL Report Number 2007003.

5. A. Rakhlin. Stability of clustering methods. NIPS Workshop"Theoretical Foundations of Clustering", December 2005.

6. Guha, S., R. Rastogi & K. Shim. CURE: An Efficient Clustering Algorithm for Large Databases. In Proc. Of ACM SIGMOD Intl.Conf. on Management of Data, pp. 73-82, 1998.

7. Deb, K., Agarwal, S., Pratap, A. and Meyarivan, T. (2000a). A Fast Elitist Non-Dominated Sorting Genetic Algorithm for Multi-Objective Optimization: NSGA-II. KanGAL Report Number 200001, Indian Institute of Technology. Kanpur, India.

8. Davidson, I., Ravi, S.S.: Clustering with constraints: Feasibility issues and the k-means algorithm. In: Proceedings of the 2005 SIAM International Conference on Data Mining. (2005)

9. Kaufman, L. and P.J. Rousseeuw, 1990. Finding Groups in Data: An Introduction to Cluster Analysis. Wiley, New York.

10. Karypis G., Han E.-H. and Kumar V. (1999) Chameleon: A hierarchical clustering algorithm using dynamic modeling. IEEE Computer, 32(8) pp68-75.

11. P. S. Bradley, U. Fayyad, and C. Reina, "Scaling Clustering Algorithms to Large Databases", To appear,Proc. 4th International Conf. on Knowledge Discovery and Data Mining (KDD-98). AAAI Press, Aug. 1998.

12. Rousseeuw P. J. (1987): Silhouettes: A graphical aid to the interpretation and validation of cluster analysis.Journal of computational and applied mathematics, 20, 53- 65.

13. Rousseeuw P., Trauwaert E. an d Kaufman L. (1989): Some silhouette-based graphics for clustering interpretation. Belgian Journal of Operations Research, Statistics and Computer Science, 29(3).

14. Pradyumn Kumar Shukla and Kalyanmoy Deb.On Finding Multiple Pareto-Optimal Solutions Using Classical and Evolutionary Generating Methods. KanGAL Report Number 2005006.