ANALYSIS OF ALTERNATIVE EVALUATION METHODS FOR MULTI-CRITERIA PROBLEMS

Ruiz-Torres, Alex, College of Engineering, Polytechnic University of Puerto Rico, San Juan, PR 00918, 787-622-8000, aruiz@pupr.edu
Lopez, Francisco J., College of Business, University of Texas at El Paso, El Paso, Texas, 79968-0544, 915-747-7741, fjlopez@utep.edu

ABSTRACT

The consideration of multiple criteria in decision making related research is highly relevant as this represents decision making today. However, there is no uniformly used method to evaluate the performance of heuristics for multi-criteria problems when the objective is to generate all the Pareto efficient solutions. This research analyzes four methods available in the literature to assess non-dominated solution sets by presenting cases and discussing the differences between the methods. The paper provides guidelines on how a method should be selected and proposes the use of multiple methods as a way to minimize the weaknesses of the individual methods.

INTRODUCTION

Decision making systems must consider multiple criteria in order to generate effective solutions. Organizations must be able to make decisions that provide *balance* between criteria such as cost, customer service, government regulations, and the environment to mention a few. Three types of multi-criteria problem formulations are used to represent relationships among criteria. The first combines the multiple criteria into a function, resulting in a single objective that is optimized. The second, called *hierarchical*, ranks and optimizes the criteria as follows: the first criterion is optimized; the second is then optimized subject to no deterioration in the first, and so on. The third case searches for the Pareto efficient solution set (non dominated solution set). This last case serves to model systems were the objectives cannot be directly combined into a single measure. When this occurs, the problem is known as a multi-criteria problem of the non-dominated form. These problems are often difficult to solve, thus finding the best approach to generate efficient solutions is important. For example, a job shop scheduling problem with multiple machines, with a hundred orders, and with two important objectives, minimizing total cycle-time and number of tardy jobs, may have a dozen or more non-dominated optimal solutions and millions of possible schedules. Finding all of these schedules may take several hours or days of computing time. Because of this, heuristics are often developed that can find good non-dominated solutions faster than an optimal search. Not only heuristics may require significant amounts of computation time, but also there may be a number of different heuristics for complex problems. Therefore it may be desirable to use only one or a few heuristics to solve a problem. The selection of which heuristics to use should be based on experimental performance over a significant number of test cases, choosing those that typically outperform all others. However, as expressed in [1], a critical issue remains unresolved: there is not a uniformly used and accepted evaluation method to compare sets of non-dominated solutions when there are no predisposed weights for the criteria in question. This paper presents part of a research project aimed at providing insights into the relative behavior of four comparison methods proposed in the literature for the evaluation of sets of non-dominated solutions. The behavior relates to the order of the heuristic solution sets and the selection of the best set.

THE EVALUATION METHODS

Four methods proposed to evaluate the output (non-dominated solution sets) of heuristics for the multi-criteria problem of the non-dominated form are: the Distance 2 method [2]; the Best Deviation method [3]; the IPF method [1]; and the FDH formulation of DEA, proposed by [4] for a scheduling problem. For a more extensive list of other proposed methods see [1]. Consider a multi-criteria problem and let X be the set $\{x_1, x_2, \ldots, x_{|X|}\}$ of criteria to minimize (if the objective is to maximize, the inverse can be considered) where $|X|$ is the number of criteria in X. Let $x_i[u]$ be the value of criterion x_i for solution (or schedule) u. Assume that there are N heuristics H1, H2, …, HN or algorithms that generate Pareto-efficient solutions, so there are N sets $S_1, \ldots, S_N$ populated only by non-dominated solutions. Let B be the optimal Pareto-efficient solution set for the problem, also called the benchmark solution set. If the optimal set is not know, let B be the set of efficient solutions of the combination of all N sets of solutions. Suppose that ND indicates the elimination of non-efficient solutions from a set. Then $B = ND[S_1 \cup S_2 \ldots S_{N-1} \cup S_N]$. Let $|B|$ be the number of solutions in B. We describe and illustrate the four evaluation methods considered in this article with an example, based on the data in Table 1. As in is [5] the data is normalized for each scale. In this example, set B is assumed to be an optimal set.

Table 1. Example Data.

Heuristic	Non-Dominated Solutions (S_i)	
	Raw Data	Normalized Data
H1	{(91,5), (84,8), (78,22)}	{(0.87,0.23), (0.8,0.36), (0.74,1)}
H2	{(100,5), (95,6), (90,9), (88,16)}	{(0.95,0.23), (0.91,0.27), (0.86,0.41), (0.84,0.73)}
H3	{(80,11), (78,16)}	{(0.76,0.5), (0.74,0.73)}
H4	{(105,7), (98,12), (78,16)}	{(1,0.32), (0.93,0.55), (0.74,0.73)}
B	{(90,5), (84,7), (80,11), (78,16)}	{(0.86,0.23), (0.8,0.32), (0.76,0.5), (0.74,0.73)}

The Distance 2 (Dist2) Method

The Dist2 method [2] is based on the maximum of the minimum distances from all solutions v in the benchmark set B to all solutions u in the set of solutions S_H. That is, for each solution v in set B, this approach identifies the solution u in set S_H that is closest to it. For each of these pairs of solutions, it is possible to measure the distance between the two solutions with respects to each criterion. The maximum of these distances, among all pairs of solutions is the Dist2 measure, calculated as: $D = max_{v \in B} \{ min_{u \in SH} \{ max_{j \in X} \; abs\,(x_j[v] - x_j[u]) \}\}$, where *abs* indicates absolute value. The Dist2 scores for H1, H2, H3, and H4 are 0.273, 0.095, 0.273, and 0.200 respectively. The lower the Dist2 score the better ranking of the corresponding heuristic, thus the best heuristic is H1 followed by H4 and then a tie between H1 and H3. The ranking is: H2, H4, (H1-H3).

The Best Deviation Method (DEV)

The DEV method [3] measures the "deviation" from the solutions in set S_H to the solutions in set B. This method utilizes the normalized data by criterion for all calculations. The normalized deviation $d^{u\text{-}v}$ between a solution u in S_H and a solution v in B is: $d^{u\text{-}v} = \sum_{j \in X} [max \{ 0, (x_j[v] - x_j[u]) \}] / |X|$. Note that the first super-index refers to a solution from set S_H and the second to a solution from set B. The DEV method is based on the notion that, for the purposes of evaluation, a solution can be "worsened" so it falls within the dominance area of each solution in set B. For example a solution from set S_H with bi-criteria values (1000, 340) is being compared to solution (950, 450) in *B*. These two solutions cannot be compared since (1000, 340) is not in the dominance area of solution (950, 450). A worsened score for solution (1000, 340) is (1000, 450), which falls within the dominance area of solution (950, 450). Based on this concept, the smallest deviation of set S_H with respect to solution v in the benchmark set *B* is expressed as: $\boldsymbol{sd}^{v} = min_{u \in SH} \{ \boldsymbol{d}^{u\text{-}v} \}$. The overall DEV score is the average of the smallest deviations for each of the points in the benchmark set *B* expressed as: DEV = $\sum_{v \in B} [sd^{v}] / |B|$. The DEV scores for H1, H2, H3, and H4 are 0.0376, 0.0971, 0.1136, and 0.1627 respectively. The smaller the DEV score, the better ranking, thus, the DEV ranking is H1, H2, H3, and H4.

The Integrated Preference Functional Method (IPF)

The Integrated Preference Functional method [1] is based on a function that evaluates the quality of sets of near-Pareto-optimal solutions for bi-criteria optimization problems. This approach incorporates weights to estimate the "expected" utility of each heuristic. The evaluation process requires estimating the optimal weight range for each solution. The weights are obtained from a weight function that contains or projects the preferences of the decision maker. The authors in [1] explain that this function can be thought of as "the probability of decision maker's preference for the weight of each objective." The IPF method generates a convex efficient frontier by removing non-dominated solutions that break convexity. This implies that the IPF method may use fewer solutions than other methods when evaluating heuristics. For example, this method does not take into account the normalized solution (0.933, 0.545) from S_4 because otherwise the efficient frontier would not be convex. For the sake of brevity the process is not included and the interested reader is directed towards the original article. The IPF scores for H1, H2, H3, and H4 are 0.5337, 0.5709, 0.6302, and 0.6095 respectively. The smaller the IPF score, the better the heuristic, thus, the IPF ranking is H1, H2, H4, and H3.

The Free Disposal Hull Method (FDH)

The Free Disposal Hull (FDH) formulation of [6] is proposed by [3] as a means to compare heuristics for a scheduling problem. The FDH method compares scheduling solutions on a one-to-one basis and classifies them as "efficient" or "inefficient." When no solution dominates the solution being evaluated the latter is considered efficient. As discussed, the benchmark set *B* serves as the reference set generated with solutions from all sets, when *B* is not a known optimal set. The "FDH efficiency score" of efficient scheduling solutions is one (indicating 100% efficiency) while that of inefficient solutions is less than one (less than 100% efficiency). The FDH-based method evaluates heuristics employing the FDH (DEA) "degree of efficiency" of the corresponding solutions. The degree of efficiency of an inefficient solution (not necessarily the same as its FDH efficiency score) can be computed in terms of the "slacks or surpluses" of the inefficient solution in relation to a solution that dominates it. As in the case of IPF, for the sake of brevity the equations have been omitted and the interested reader is refereed to [3]. The FDH scores for H1, H2, H3, and H4 are 0.932, 0.896, 1.00, and 0.922 respectively. As can be seen, heuristic H3 has an efficiency of 100% (best performer) while heuristic H2 has an efficiency of 89.6% (the worst performer). The ranking of the heuristics is H3, H1, H4, and H2.

COMPARISONS AMONG METHODS

We will summarize in our presentation the rankings of heuristics H1, H2, H3, and H4 according to the evaluation methods described above.