Metaheuristics
Chapter 4.5-4.9

Johan Falkenjack

NLPLAB
Dept. of Computer and Information Science
Linköping University

Overview

- Diversity preservation
- Elitism
- Performance evaluation and Pareto front structure

Diversity preservation

- Diversity where?
  - Decision space
  - Objective space
- Kernel methods
- Nearest neighbor
- Histogram

Kernel methods

- Define a kernel function $K(d_{ij})$
- Estimate the density of a solution i by applying $\sum_{j=1}^{n} K(d_{ij})$
- In other words, the density of a solution is a function of its similarity to all other solutions weighted in some way by the closeness of solutions.
- The value of a solution can now be viewed as a function of its fitness and its similarity to other solutions according to some kernel.
- Example: Fitness sharing, p.344

Nearest-Neighbor

- Special case of kernel?
- Uses the distance to the k nearest neighbors to estimate the density of a solution.
- In other words, the density of a solution is a function of its similarity to its k nearest neighbors.
- Again, the value of a solution is viewed as a function of its fitness and and its similarity to other solutions.
- Example: Nondominated sorting genetic algorithm, p.346

HISTOGRAM

- Another special case of kernel?
- Fits a hypergrid to the search space.
- Density of a solution is estimated as the number of solutions in the same box in the grid.
- Example: Pareto archived evolution strategy

Elitism

- Using an archive to store “good” solutions.
- Passive elitism
  - Store Pareto optimal solutions in an archive to make sure the performance does not degrade over time.
  - Slow but safe, in that it does not influence the search process.
- Active elitism
  - Store Pareto optimal solutions in an archive and use these in some way to generate new solutions.
  - Faster but unsafe, in that it does influence the search process and might lead to premature convergence, i.e. getting stuck in a local optimum in objective space.

Designing the archive

- Size criteria
  - In theory, the size of the Pareto set might be infinite, storing all solutions won't be possible.
- Convergence criteria
  - Which solutions should be stored?
  - Dominance is the most used approach, but other approaches is possible, like a scalar function (fitness function in objective space?) or an indicator function selecting solutions with specific features in objective space.
- Diversity criteria
  - When updating the archive, should diversity be preserved, and if so, how?
  - Any approach discussed above would be appropriate.

Performance indicators

- Function I of one or a pair of Pareto front in an objective space.
- Aggregates the front, or a pair of fronts, into a scalar.
- Easy to use but lossy.
- Unary/binary indicators
  - Unary indicators can be likened to a fitness score of a Pareto front.
  - Binary indicators compare two Pareto fronts to decide which is better.
- Do we know the Pareto optimal set?
  - If yes, then specific indicators can be designed using this as a basis and compare candidate sets in to this.
  - Rarely useful in real world problems, mainly in research.
- Can we construct a reference set?
  - Using similar metrics to when the Pareto optimal set is known, but with a non-optimal set.
  - Found solutions might be better than the reference front, making some comparisons tricky.

Convergence based indicators

- Contribution
  - Binary measure between two approximations $PO_1$and PO_2.
  - The contribution of PO_1 relatively to PO_2 is the ratio of non-dominated solutions produced by PO_1 in the Pareto set of $PO_1 \cup PO_2$.

$$Cont(PO_1/PO_2) = \frac{\frac{||PO_1 \cap PO_2||}{2} + ||W_1|| + ||N_1||}{||PO^*||}$$

where$PO^*$is the set of all Pareto solutions in $PO_1 \cup PO_2$, $W_1$is the set of solutions in PO_1which dominate some solution in PO_2 and $N_1$is the set of non-comparable solutions of PO_1(see formula on p.352).

  - The sum of $Cont(PO_1/PO_2)$ and $Cont(PO_2/PO_1)$ is 1 and the approximation having the larger contribution is assumed to be better.

Convergence based indicators

- Generational distance I_{GD}
  - Unary measure on a pareto front A using a reference front R.
  - The average distance from each solution in a A to its closest solution in R.

$$I_{GD}(A,R) = \frac{(\sum_{u \in A} min_{v \in R}(||F(u) - F(v)||^2))^{\frac{1}{2}}}{|R|}$$

  - Distance here is Euclidian distance in objective space.

Convergence based indicators

- ϵ-Indicator $I^1_{\epsilon+}$
  - Unary measure on a Pareto front A using a reference front Z^*_N
  - The minimum factor ϵ by which A has to be translated in the objective space to weakly dominate Z^*_N

$$I^1_{\epsilon+}(A) = I_{\epsilon+}(A, Z^*_N)$$

where

$$I_{\epsilon+}(A, B) = min_{\epsilon \in \mathbb{R}}\{\forall z \in B, \exists z' \in A : z'_i - \epsilon \leq z_i, \forall 1 \leq i \leq n\}$$

Convergence based indicators

- If the optimal Pareto set is known the following two types of measures are suggested:
  - Cardinality based measures.
    - For instance, the proportion of solutions in a set A that belong to the optimal set.
  - Distance based measures.
    - For instance, using generational distance with the optimal set as reference.

Diversity based indicators

- Spread indicator I_S is a unary measure of the dispersion of an approximated set A in objective space.

$$I_S(A) = \frac{\sum_{u \in A} |\{u' \in A : ||F(u) - F(u')|| > \sigma\}|}{|A| - 1}$$

  - Where $\sigma > 0$ is a neighborhood parameter.
  - Values close to 1 indicate a good spread.

Diversity based indicators

- Extent indicator I_{ex} is a unary measure of the extent of an approximated set A in objective space.
  - We measure the distance between the upper and lower bounds found for each objective and sum over these

$$I_{ex}(A) = (\sum_{i=1}^{n} max_{u,u' \in A}(||f_i(u) - f_i(u')||))^{\frac{1}{2}}$$

where n is the number of objectives.

Diversity based indicators

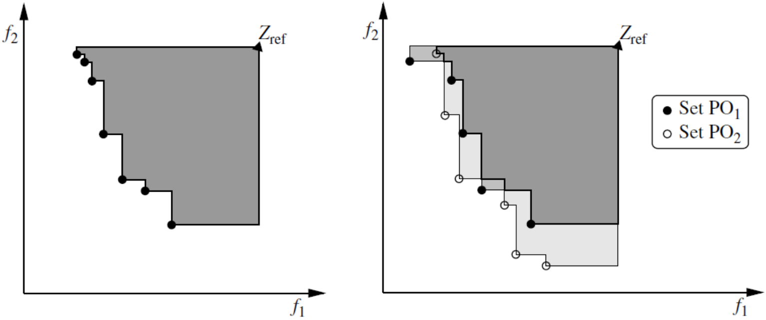
- Entropy, in the information theory sense
  - The objective space is partitioned along each dimension (each objective) into μ intervals, this partitions an n dimensional objective space into μ^n hypergrids, also called niches.
  - Solutions in the same niche are considered equivalent and the entropy is a measure of the dispersion of the solutions around the hypergrids.
  - Unary or binary, binary version presented in the book and here.

$$E(PO_1, PO_2) = \frac{-1}{log(\gamma)} \sum_{i=1}^{|PO_1 \cup PO^*|} \left(\frac{1}{N_i} \frac{n_i}{|PO_1 \cup PO^*|} log\left(\frac{n_i}{|PO_1 \cup PO^*|}\right)\right)$$

where $PO^*$is the set of all Pareto solutions in $PO_1 \cup PO_2$, N_i is the cardinality of solutions in $PO_1 \cup PO^*$ which are in the same niche as the i:th solution, n_i is the cardinality of solutions in PO_1 which are in the same niche as the i:th solution and $\gamma = \sum_{i=1}^{|PO_1 \cup PO_2|}(\frac{1}{N_i})$, that is, the sum of the coefficients of all solutions.

Hybrid indicators

- Hypervolume I_H is a unary or binary measure of the volume of the objective space weakly dominated by a set or sets.
  - Needs a reference point Z_{ref} used as an upper bound when calculating volume.
  - Z_{ref} can be fixed as $(1.05 \times Z_1^{max}, \ldots, 1.05 \times Z_n^{max})$ where Z_i^{max} is the upper bound of objective i in a reference set Z_N^*.



Hybrid indicators

- R-metrics is a combination of a number of utility functions produced by a decision maker.
  - Given a set of parameterized utility functions u_λ and corresponding parameters Λ an R-metric is some combination of these functions applied to a set, or pair of sets, of solutions.
  - Many different utility functions can be used, such as a weighted sums of objectives. Some examples given by the book is:
    - Weighted linear function, which uses an ideal point z_N^*, and measure a weighted Manhattan distance to this point from each solution in a set.
    - Weighted Chebychev distance, which works as above but use only the dimension in which the point is farthest from the ideal point z_N^*.
    - Augmented Chebychev distance is a combination of the two above.

Landscape Analysis of Pareto Structures

- Convex or concave
  - Weighted aggregates of objectives is more effective in convex structures.
  - Dominance based ranking is useful in the convex case.
  - Convexity is measured by proportion of supported solutions and their distribution along the front.
- Continuous or discontinuous
  - Discontinuous structures are troublesome for some S-metaheuristics.
  - Diversity preservation works better in continuous structures.
- Modality
  - Locally Pareto optimal solutions makes convergence difficult.

Multiobjectivization

- Making a monoobjective optimization problem multiobjective.
- Might transform a rugged search space into a smoother one.
- Might allow us to factor in additional features other than pure optimality, such as robustness or generalizability
- Strategies:
  - Objective function decomposition.
  - Adding helper objectives.