Sampling Strategies in Ordinal Regression for Surrogate Assisted Evolutionary Optimization

Helga Ingimundardottir & Thomas Philip Runarsson

School of Engineering and Natural Sciences, University of Iceland

November 24th, 2011

Overview

Motivation

During evolution, different regions of space are sampled and as a consequence the surrogate ranking model may be insufficiently accurate for new regions of the search space.

Hence if the surrogate is not updated to reflect the original fitness function it is very probable that the ES converges a false optimum.

Goal

- General goal is how to validate goodness of fit for surrogate models during search
- Introducing a novel validation/updating policy
- Illustrated on classical numerical optimisation problems

Previous work

Methods previously proposed for validating surrogate models:

- Validating on entire vs. subset of candidate population (Ratle, 1999), e.g. accurate ranking of potential parent individuals (Runarsson, 2004)
- Generation based or individual based (Jin, 2005)

Ordinal regression

- Mapping of points to ranks: $\{h(\cdot) : X \mapsto Y\}$
  - $\mathbf{x}_i \succ \mathbf{x}_j \quad \Leftrightarrow \quad h(\mathbf{x}_i) > h(\mathbf{x}_j)$
- Ordinal regression: obtain function h^* that can for a given pair $(\mathbf{x}_i, y_i)$ and $(\mathbf{x}_j, y_j)$ distinguish between two different outcomes: $y_i > y_j$ and $y_j > y_i$.
- Problem of predicting the relative ordering of all possible pairs of examples

The surrogate considered may be defined by a linear function in the feature space (of dimension n):

$$h(\mathbf{x}) = \sum_{k=1}^{n} w_k x_k = \langle \mathbf{w} \cdot \mathbf{x} \rangle$$

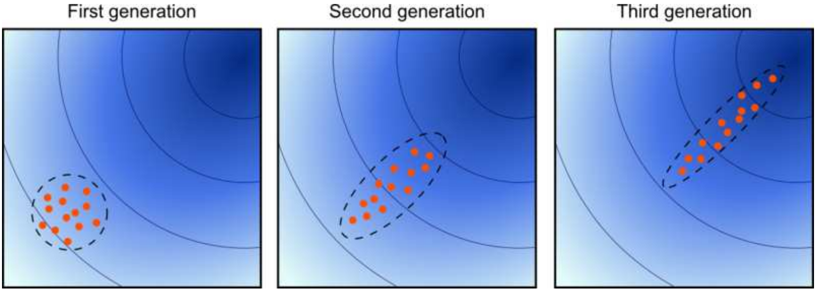
Kendall's τ

- Kendall's τ is a quality measure for the rank correlation between the surrogate ranking and the true ranking of the training data.
- Pair is concordant if the relative ranks of $h(\mathbf{x}_i)$ and $h(\mathbf{x}_j)$ are the same for $f(\mathbf{x}_i)$ and $f(\mathbf{x}_j)$, otherwise discordant.
- Kendall's τ is the normalized difference in the number of concordant and discordant pairs.
- Two rankings are the same when $\tau = 1$, completely reversed if $\tau = -1$, and uncorrelated for $\tau \approx 0$.

CMA-ES

- Implement a Covariance Matrix Adaptation Evolution Strategy (CMA-ES)
  - population size: $\lambda = 4 + \lfloor 3 \ln(n) \rfloor$
  - parents: $\mu = \lambda/4$
  - stopping criteria: $1000n$ function evaluations or fitness less than 10^{-10}

CMA-ES



Size of training set

- Small sample of training individuals of known fitness are needed to generate an initial surrogate.
- Size of training set is denoted $\overline{\ell}$. Note if training size is unlimited then updating the surrogate becomes computationally expensive, and therefore needs to be pruned.

Anova plot for different validation strategies:

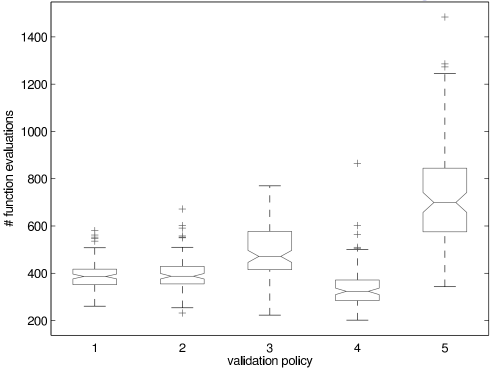
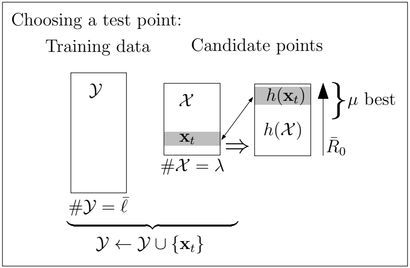


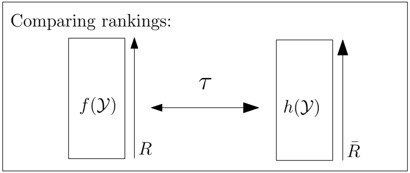
Figure: 1) prune old individuals, 2) prune bad individuals, 3) adding a pseudo mean candidate individual 4) correctly rank μ best ranked candidate individuals 5) update on every other generation for Rosenbrock's function for dimension $n = 2$

Completely concordant ranking

- If the training accuracy is not 100% then $\tau < 1$
  - resulting in additional training individuals to be forced for evaluation.
- However, since the search is stochastic this is deemed to strict, thus a sufficiently accurate surrogate would be if $\tau > 0.999$.

Schema for the sampling strategy





Classical Optimisation Problems

- Sphere model, for dimension $n = 2, 5, 10, 20$:

$$f(\mathbf{x}) = \sum_{k=1}^{n} x_k^2$$

- Rosenbrock's function, for dimension $n = 2, 5, 10, 20$:

$$f(\mathbf{x}) = \sum_{k=2}^{n} 100(x_k - x_{k-1}^2)^2 + (1 - x_{k-1})^2$$

Main statistics of experimental results for sphere model.

		Function eval.			Generations		
	n	mean	median	sd	mean	median	sd
all	2	130.59	132	18.33	49.02	49	6.51
μ	2	**81.53**	81	9.53	**48.11**	48	5.02
all	5	702.02	702	67.57	145.15	145	14.96
μ	5	**545.25**	547	54.27	**132.60**	132	11.03
all	10	1563.58	1553	117.09	241.83	240	18.47
μ	10	**1161.03**	1158	79.98	**226.60**	224	13.86
all	20	3383.83	3377	135.52	423.14	424	20.42
μ	20	**2795.28**	2804	132.77	**372.86**	372	16.56

Main statistics of experimental results for Rosenbrock's function.

		Function eval.			Generations		
	n	mean	median	sd	mean	median	sd
all	2	389.85	386	63.85	**132.31**	130	31.25
μ	2	**344.91**	336	78.58	172.16	170	49.95
all	5	2464.22	2280	748.55	**514.59**	492	105.77
μ	5	**1724.89**	1729	295.60	520.66	520	82.79
all	10	6800.50	6495	1258.68	**1079.82**	1052	177.76
μ	10	**6138.48**	6143	1398.15	1177.71	1103	310.11
all	20	19968.80	20004	234.66	**2494.00**	2500	49.60
μ	20	**19645.90**	20002	1086.37	2687.25	2748	230.50

Main Conclusions

- The new validation approach reduces the number of fitness evaluations, without a loss in performance although it might take a few more iterations in CMA-ES.
- Sampling used for validation of the accuracy of the surrogate can stop once the μ best ranked candidate individuals have been evaluated
  - since they are the only candidate individuals who will survive to become parents in the next generation
- In some cases the sampling could have stopped sooner, when the surrogate ranking was sufficiently concordant.
- No statistical difference between omitting oldest or lowest ranking individuals from training set.

Future Work

- Further investigation on the fitness landscape to determine more effectively points of interest (and no longer of interest).
  - e.g. disregarding points with the greatest euclidean distance from the current candidate individuals
- The allowable range for $\tau \in [0.999, 1]$ might be too narrow an interval, resulting in an excess of expensive function evaluations needed.
- This is a simple case study, and the framework needs to be implemented on a greater range of test functions, e.g. job shop scheduling problem.