Investigation of Gaussian Processes and Random Forests as Surrogate Models for Evolutionary Black-Box Optimization

Lukáš Bajer
Institute of Computer Science, Academy of Sciences of the Czech Republic
and
Faculty of Mathematics and Physics
Charles University in Prague
Malostranské nám. 25
Prague, Czech Republic
bajer@cs.cas.cz

Zbyněk Pitra
National Institute of Mental Health, Czech Republic
and
Faculty of Nuclear Sciences and Physical Engineering
Czech Technical University in Prague
Trojanova 13
Prague, Czech Republic
z.pitra@gmail.com

Martin Holeňa
Institute of Computer Science, Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 2
Prague, Czech Republic
martin@cs.cas.cz

ABSTRACT

This paper introduces two surrogate models for continous black-box optimization, Gaussian processes and random forests, as an alternative to the already used ordinal SVM regression. We employ the CMA-ES as the reference optimization method with which the surrogate models are combined and also compared on subset of the noisless BBOB testing set.

Categories and Subject Descriptors

G.1.6 [**Numerical Analysis**]: Optimization—*global optimization, unconstrained optimization*

Keywords

Black-box optimization; Surrogate model; Gaussian process; Random forest

1. INTRODUCTION

Evaluation of the real-world black-box objective functions is often very time-consuming and/or costly in numerous optimization problems. Surrogate modelling is an approach to decreasing the number of expensive function evaluations via fitting and using a regression model of the objective function [3]. This model is trained on the already gathered input–output-value pairs $(\mathbf{x}_i, y_i), i = 1, \dots, N$ and is used instead of the original expensive fitness to evaluate some of the points needed by the optimization algorithm. Because the surrogate-model evaluations are always affected by some error, which could easily mislead the whole optimization process, the original fitness has to be used for some part of the points, too.

This paper investigates two surrogate models based on *Gaussian processes* (GP) [4] and *random forests* (RF) [1], which have partly similar properties (e.g. they inherently provide estimation of the prediction error). The two models are used as surrogate models for the CMA-ES, comparing the speed-up of different settings of surrogate-assisted version against the CMA-ES without the surrogate model.

GECCO '15 July 11-15, 2015, Madrid, Spain

ACM ISBN 978-1-4503-3488-4/15/07.
DOI: http://dx.doi.org/10.1145/2739482.2764692

2. SURROGATE CMA-ES

The surrogate models were integrated in the CMA-ES algorithm in its fitness-evaluating part: the original fitness evaluation is replaced with the Algorithm 1 forming our modified algorithm, temporarily named Surrogate CMA-ES (S-CMA-ES).

The algorithm follows up the standard CMA-ES sampling of the new offspring (noted as step 1 in the pseudocode). It is, in fact, a *generation evolution control* of combining the original- and surrogate-evaluated samples: the whole offspring in each generation is either evaluated by the original fitness (and a new GP/RF model is trained in this generation), or the offspring is evaluated by the previously trained model. Model estimated y-values are linearly transformed in order not to be lower than the so-far optimum (step 15).

3. EXPERIMENT

GP and RF were tested[1] on the BBOB benchmark functions [2] $f_{1-3}, f_{5-6}, f_8, f_{10-14}, f_{20-21}$ in dimensions 2, 5, and 10.

First, GP and RF parameters resulting in the best regression performance were identified. These parameters for GP include the type of covariance function *cov* and starting values for noise variance σ_n^2, characteristic length-scale ℓ and signal variance σ_f^2. RF comprised 100 trees, each containing at least p_l training points in each leaf, growing on the

[1]the source code is freely available at https://github.com/bajeluk/surrogate-cmaes

Algorithm 1 Surrogate CMA-ES Algorithm

Input: g (generation), g_m (number of model generations), σ, λ, $\mathbf{m}$, $\mathbf{C}$ (CMA-ES internal variables), r (maximal distance between training points and $\mathbf{m}$), n_{REQ} (minimal number of points for model training), n_{MAX} (maximal number of points for model training), $\mathcal{A}$ (archive), $f_{\mathcal{M}}$ (model), f (original fitness function)

$\mathbf{x}_k \sim \mathcal{N}\left(\mathbf{m}, \sigma^2\mathbf{C}\right) \quad k = 1, \ldots, \lambda \quad$ {*CMA-ES sampling*}
if g is original-evaluated **then**
  $y_k \leftarrow f(\mathbf{x}_k) \qquad k = 1, \ldots, \lambda \qquad$ {*fitness evaluation*}
  $\mathcal{A} = \mathcal{A} \cup \{(\mathbf{x}_k, y_k)\}_{k=1}^{\lambda}$
  $(\mathbf{X}_{\text{tr}}, \mathbf{y}_{\text{tr}}) \leftarrow \{(\mathbf{x}, y) \in \mathcal{A} | (\mathbf{m} - \mathbf{x})^{\top} \sigma \mathbf{C}^{-1/2} (\mathbf{m} - \mathbf{x}) \leq r\}$
  if $|\mathbf{X}_{\text{tr}}| \geq n_{\text{REQ}}$ **then**
    $(\mathbf{X}_{\text{tr}}, \mathbf{y}_{\text{tr}}) \leftarrow$ choose n_{MAX} points if $|\mathbf{X}_{\text{tr}}| > n_{\text{MAX}}$
    $f_{\mathcal{M}} \leftarrow \text{trainModel}(\mathbf{X}_{\text{tr}}, \mathbf{y}_{\text{tr}})$
    mark $(g + 1)$ as model-evaluated
  else
    mark $(g + 1)$ as original-evaluated
  end if
else
  $y_k \leftarrow f_{\mathcal{M}}(\mathbf{x}_k) \qquad k = 1, \ldots, \lambda \qquad$ {*model evaluation*}
  {*shift y_k values if $(\min y_k) <$ best y from $\mathcal{A}$*}
  $y_k = y_k + \max\{0, \ \min_{\mathcal{A}} y - \min y_k\} \qquad k = 1, \ldots, \lambda$
  if g_m model generations passed **then**
    mark $(g + 1)$ as original-evaluated
  end if
end if

Output: $f_{\mathcal{M}}$, $\mathcal{A}$, $(y_k)_{k=1}^{\lambda}$

RF	GP
$p_l \in \{\mathbf{2}, 5, 8\}$	$cov \in \{K_{\text{SE}}, K_{\text{Matérn}}^{\nu=3/2}, \mathbf{K}_{\mathbf{Matérn}}^{\nu=5/2}\}$
$\varphi \in \{0.5, 0.8, \mathbf{1}\}$	$\exp(\sigma_f^2, \ell) \in \{(0.01, 0.05), (0.05, 0.25), (0.25, 1), (\mathbf{0.5}, \mathbf{2}), (1, 5)\}$
$p_r \in \{0, 1, 3, \mathbf{5}\}$	$\exp(\sigma_f^2, \ell) \in \{(0.01, 0.1), (0.1, 10), (1, 10^3), (5, 10^4)\}$ (*for* K_{SE})
	$\exp \sigma_n^2 \in \{\mathbf{0.01}, 0.1, 1\}$

Table 1: **Model parameter settings for regression testing.** The best-observed values are typeset in bold and were used for BBOB optimization experiments.

proportion φ of training points sampled with replacement from the training data. In addition, only trees predicting p_r training points with the best fitness value in the right order were included in forests. A grid search was performed through the parameter values shown in Table 1.

The best found model settings (see Table 1) were used in the BBOB optimization experiments in which the best number of model generations $g_m \in \{1, 2, 3\}$ was identified as $g_m = 3$ for GP and $g_m = 1$ for RF.

Based on the data from the BBOB optimization experiments, Figure 1 shows S-CMA-ES speed-up compared to the original CMA-ES. Both the S-CMA-ES models mostly outperform the CMA-ES in initial phases of the optimization process, especially on functions $f_1, f_5, f_{10}, f_{13}, f_{20}$. However, the speed-up is decreasing with the increasing number of evaluations.

4. ACKNOWLEDGMENTS

This work was supported by the the Czech Science Foundation (GAČR) grant P103/13-17187S, by the Czech Technical University grant No. SGS14/205/OHK4/3T/14, and by the project "National Institute of Mental Health (NIMH-CZ)", grant number CZ.1.05/2.1.00/03.0078 (and the European Regional Development Fund).

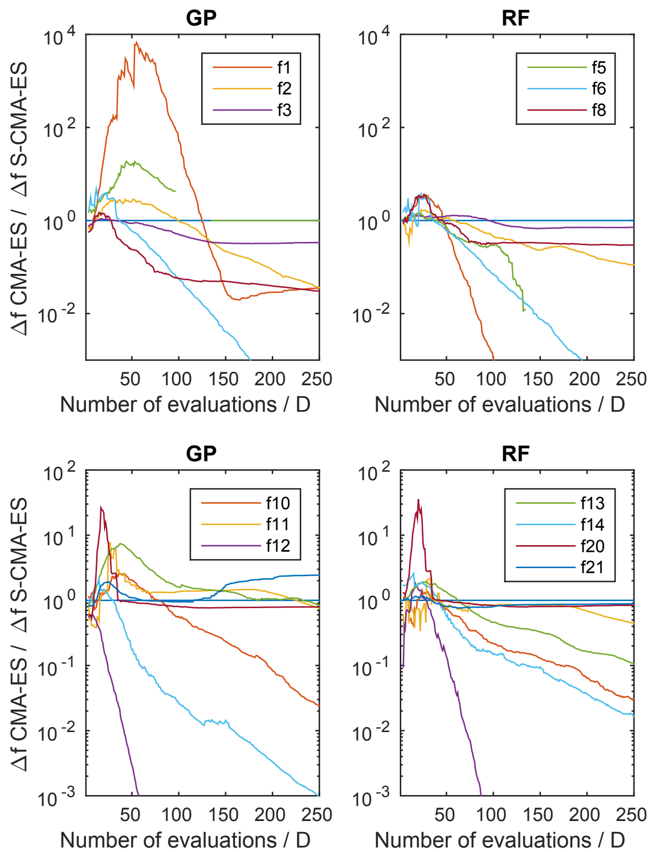

Figure 1: **Speed-up of the GP- and RF-based S-CMA-ES compared to the CMA-ES.** Ratios of distances to optimum according to the expected number of fitness evaluations divided by dimension for f_{1-3}, f_{5-6} and f_8 (first row) and f_{10-14}, f_{20-21} (second row) in dimensions 2, 5, and 10. Medians were taken from 20 independent runs for each function/dimension combination.

5. REFERENCES

[1] L. Breiman. *Classification and regression trees.* Chapman & Hall/CRC, 1984.

[2] N. Hansen, S. Finck, R. Ros, and A. Auger. Real-parameter black-box optimization benchmarking 2009: Noiseless functions definitions. Technical Report RR-6829, INRIA, 2009. Updated February 2010.

[3] Y. S. Ong, P. B. Nair, and A. J. Keane. Evolutionary optimization of computationally expensive problems via surrogate modeling. *AIAA Journal*, 41(4):687–696, 2003.

[4] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning.* Adaptative computation and machine learning series. MIT Press, 2006.