On the use of Metaheuristics in Hyperparameters Optimization of Gaussian Processes

Pramudita Satria Palar
Faculty of Mechanical and Aerospace Engineering
Institut Teknologi Bandung
Bandung, Indonesia
pramsp@ftmd.itb.ac.id

Lavi Rizki Zuhal
Faculty of Mechanical and Aerospace Engineering
Institut Teknologi Bandung
Bandung, Indonesia
lavirz@ae.itb.ac.id

Koji Shimoyama
Institute of Fluid Science, Tohoku University
Sendai, Japan
shimoyama@tohoku.ac.jp

ABSTRACT

Due to difficulties such as multiple local optima and flat landscape, it is suggested to use global optimization techniques to discover the global optimum of the auxiliary optimization problem of finding good Gaussian Processes (GP) hyperparameters. We investigated the performance of genetic algorithms (GA), particle swarm optimization (PSO), differential evolution (DE), and covariance matrix adaptation evolution strategy (CMA-ES) for optimizing hyperparameters of GP. The study was performed on two artificial problems and also one real-world problem. From the results, we observe that PSO, CMA-ES, and DE/local-to-best/1 consistently outperformed two variants of GA and DE/rand/1 with per-generation-dither on all problems. In particular, CMA-ES is an attractive method since it is quasi-parameter free and it also demonstrates good exploitative and explorative power on optimizing the hyperparameters.

KEYWORDS

Gaussian Process Regression, Hyperparameters optimization, Metaheuristics, Likelihood function

ACM Reference Format:
Pramudita Satria Palar, Lavi Rizki Zuhal, and Koji Shimoyama. 2019. On the use of Metaheuristics in Hyperparameters Optimization of Gaussian Processes. In *Genetic and Evolutionary Computation Conference Companion (GECCO '19 Companion), July 13–17, 2019, Prague, Czech Republic.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3319619.3322012

1 INTRODUCTION

Surrogate models have found wide application in the field of science and engineering as tools that aid various tasks such as optimization [9, 10]. Of interest is the Gaussian processes regression (GP) model [6] (a.k.a. Kriging) which provides the extra uncertainty structure which is highly useful for error-based surrogate refinement or Bayesian optimization [7]. Metaheuristics methods are widely used for training GP models due to their capabilities in performing global optimization. Besides building a single GP model, there are some occasions where it is required to build multiple GP models. For example, it is necessary for surrogate-based memetic algorithms that use GP to train multiple GP models several times [3, 5]. The application of GP to big data regression also requires the construction of multiple GP models, especially in cases where the region is clustered into several subregions [8, 11]. Although one can choose any optimizer in hand, we realize that there are no existing studies regarding the comparison of metaheuristics for GP.

In this paper, we have an interest in studying the impact of the choice of metaheuristics to the optimization of GP hyperparameters. We performed this study in a MATLAB environment where we also used some built-in metaheuristics method from MATLAB. Four test problems were used in this paper: Sasena function, Hartmann-6 function, four-dimensional blended wing body (BWB) [4] and eight-dimensional transonic airfoil problem [1]. We added 5% simulated noises to the algebraic problems.

2 METAHEURISTICS FOR HYPERPARAMETERS OPTIMIZATION

The algorithms that we compared are: **(1) GA1**, GA from MATLAB global optimization toolbox, scattered crossover (default), crossover probability=0.8 (default), Gaussian mutation (default), **(2) GA2**, GA from MATLAB global optimization toolbox, arithmetic crossover, crossover probability=0.8 (default), Gaussian mutation (default), **(3) PSO**, PSO from MATLAB global optimization toolbox, global self adjustment weight = 1.49 (default), social adjustment weight = 1.49 (default), inertia range = [0.1-1.1] (default), **(4) DE1**, DE/rand/1 with per-generation-dither, implementation by Storn and Price, **(5) DE2**, DE/local-to-best/1, implementation by Storn and Price, and **(6) CMA**, CMA-ES, no parameters are adjusted. We set the number of solutions at each iteration and the maximum number of iterations to 200 and 1500, respectively.

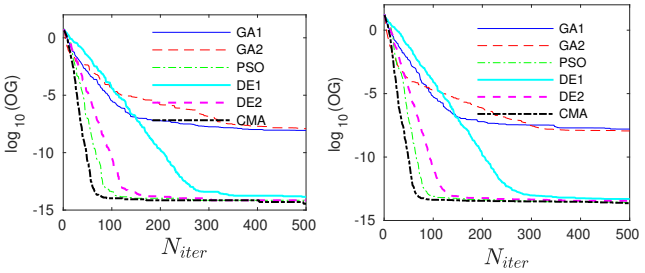
We use the zero-mean GP formulation to simplify our GP model; it is worth noting that this formulation is also widely used in the machine learning community. For noisy and real-world problems, we tune all possible hyperparameters (i.e., $\boldsymbol{\theta}$, σ_n^2, and σ_f^2). The length-scale is tunable in the range of 10^{-3} to 10^2. The σ_n^2 and σ_f^2 (i.e., noise and signal variance, respectively) are tuned in the range of 10^{-6} to 10^{-1} and 10^{-6} to 10^2, respectively. The function responses were normalized so that the mean equals to zero and the variance equals to unity.

3 RESULTS AND DISCUSSIONS.

The performance is measured through the log of the optimality gap (OG) metric. Our experiments reveal that PSO, CMA-ES, and

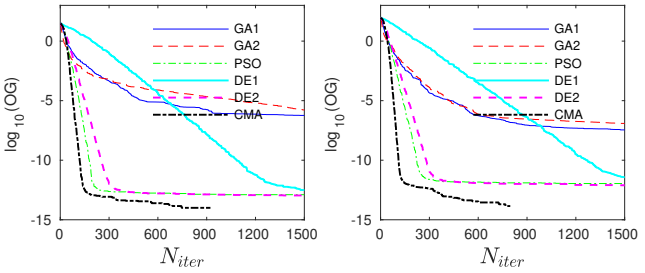
GECCO '19 Companion, July 13–17, 2019, Prague, Czech Republic

ACM ISBN 978-1-4503-6748-6/19/07…$15.00
https://doi.org/10.1145/3319619.3322012



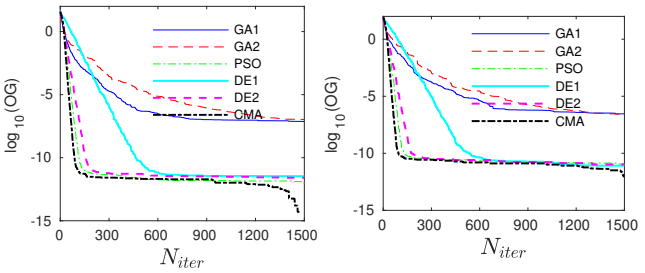
(a) $N_{samp} = 20$ **(Median).** **(b)** $N_{samp} = 40$ **(Median).**

Figure 1: Convergence plot of log-marginal likelihood for the noisy Sasena problem.



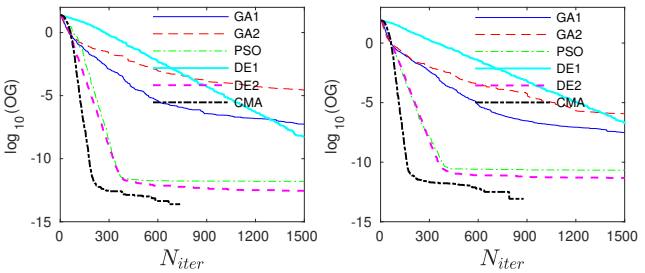
(a) $N_{samp} = 60$ **(Median).** **(b)** $N_{samp} = 120$ **(Median).**

Figure 2: Convergence plot of log-marginal likelihood for the noisy Hartmann 6 problem.



(a) $N_{samp} = 40$ **(Median).** **(b)** $N_{samp} = 80$ **(Median).**

Figure 3: Convergence plot of log-marginal likelihood for the BWB problem.



(a) $N_{samp} = 40$ **(Median).** **(b)** $N_{samp} = 80$ **(Median).**

Figure 4: Convergence plot of log-marginal likelihood for the airfoil problem.

DE2 are the best performers, in the sense that they converged to the global optimum faster compared to their competitors (ie., GA1, GA2, and DE1). Interestingly, the results show that the two GAs are slower in both exploration and exploitation phase compared to the other methods, with the exception of DE1 (however, eventually DE1 converged to high accuracy faster than the two GAs). We observe that CMA-ES, PSO, and DE2 converged to high precision typically just within less than 500 iterations. Thus, basically, the combination of less than 500 metaheuristic iterations and 100 individuals is enough to ensure high-quality GP. The quasi-parameter free nature of CMA-ES is one particular advantage for ones who wish to deploy gradient-free techniques for hyperparameters optimization without tweaking metaheuristics parameters; it also has strong theoretical properties [2].

Although PSO, CMA-ES, and DE/local-to-best/1 are highly performing methods, they can still get trapped in a local optimum even after a long run. Thus, we think that it is wise to restart the search several times to ensure that the global optimum is found. It is also worth noting that other implementations of GA (e.g. by changing different crossover, crossover and mutation probability) might work better for solving GP's auxiliary optimization problems; however, at least for the implementations that we have right now, they are not robust enough.

REFERENCES

[1] JHS de Baar, TP Scholcz, and RP Dwight. 2015. Exploiting adjoint derivatives in high-dimensional metamodels. *AIAA Journal* 53, 5 (2015), 1391–1395.

[2] Nikolaus Hansen, Sibylle D Müller, and Petros Koumoutsakos. 2003. Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). *Evolutionary computation* 11, 1 (2003), 1–18.

[3] Minh Nghia Le, Yew Soon Ong, Stefan Menzel, Yaochu Jin, and Bernhard Sendhoff. 2013. Evolution by adapting surrogates. *Evolutionary computation* 21, 2 (2013), 313–340.

[4] Rhea P Liem and Joaquim Martins. 2014. Surrogate models and mixtures of experts in aerodynamic performance prediction for mission analysis. In *15th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*. 2301.

[5] Pramudita Satria Palar, Takeshi Tsuchiya, and Geoffrey Thomas Parks. 2016. A comparative study of local search within a surrogate-assisted multi-objective memetic algorithm framework for expensive problems. *Applied Soft Computing* 43 (2016), 1–19.

[6] Carl Edward Rasmussen. 2004. Gaussian processes in machine learning. In *Advanced lectures on machine learning*. Springer, 63–71.

[7] Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P Adams, and Nando De Freitas. 2016. Taking the human out of the loop: A review of bayesian optimization. *Proc. IEEE* 104, 1 (2016), 148–175.

[8] Bas van Stein, Hao Wang, Wojtek Kowalczyk, Thomas Bäck, and Michael Emmerich. 2015. Optimally weighted cluster kriging for big data regression. In *International Symposium on Intelligent Data Analysis*. Springer, 310–321.

[9] Felipe AC Viana, Timothy W Simpson, Vladimir Balabanov, and Vasilli Toropov. 2014. Special section on multidisciplinary design optimization: metamodeling in multidisciplinary design optimization: how far have we really come? *AIAA journal* 52, 4 (2014), 670–690.

[10] G Gary Wang and Songqing Shan. 2007. Review of metamodeling techniques in support of engineering design optimization. *Journal of Mechanical design* 129, 4 (2007), 370–380.

[11] Hao Wang, Bas van Stein, Michael Emmerich, and Thomas Bäck. 2017. Time complexity reduction in efficient global optimization using cluster kriging. In *Proceedings of the Genetic and Evolutionary Computation Conference*. ACM, 889–896.