Empirical Study of Surrogate Models for Black Box Optimizations obtained using Symbolic Regression via Genetic Programming

Glen D. Rodriguez Rafael
Universidad Nacional de Ingenieria
Fac. Ind. and Systems Eng. & Fac. Sciences
Lima, Peru
grodriguez@uni.edu.pe

Carlos Javier Solano Salinas
Universidad Nacional de Ingenieria
Faculty of Sciences
Lima, Peru
jsolano@uni.edu.pe

ABSTRACT

A black box model is a numerical simulation that is used in optimization. It is computationally expensive, so it is convenient to replace it with surrogate models obtained by simulating only a few points and then approximating the original black box. Here, a recent approach, using Symbolic Regression via Genetic Programming, is compared experimentally to neural network based surrogate models, using test functions and electromagnetic models. The accuracy of the model obtained by Symbolic Regression is proved to be good, and the interpretability of the function obtained is useful in reducing the optimization's search space.

Categories and Subject Descriptors

I.2.2 [**Artificial Intelligence**]: Automatic Programming—*Program synthesis*

General Terms

Algorithms, Experimentation

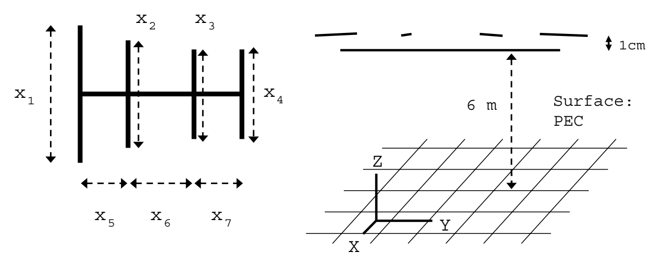
Keywords

Black box optimization, surrogate models, genetic programming, symbolic regression, neural networks

1. INTRODUCTION

Optimization of engineering systems using metaheuristics requires a lot of computationally expensive numerical simulations. Recently, numerical simulation have been replaced with a surrogate model created with the results of the simulation of a few points, using methods like Design and Analysis of Computer Experiments (DACE) and Artificial Neural Networks (ANN) [2]. DACE and ANN models replace an expensive black-box by a cheaper black box. It is not possible to gain any insight into the original black box with the new black box. SR via GP has been used by other researchers before [1], but this paper discusses for the first time the advantage of interpretability of the model generated by SR via GP; and it presents a performance study of SR via GP compared to ANN for a wider range of problems, proving that accuracy of surrogate models obtained SR via GP from a small number of samples is competitive with ANN models.

GECCO'11, July 12–16, 2011, Dublin, Ireland.
ACM 978-1-4503-0690-4/11/07.



(a) Top-down view (b) Lateral view

Figure 1: Geometry of the Yagi antenna study case

2. REVIEW OF RELATED METHODS

DACE or Kriging modeling uses functions made of Kriging basis, such as multi-variate Gaussian functions. It has trouble dealing with short scale variability, and they have no guarantee of accuracy near maxima and minima. ANNs are used to create surrogate models [2]. They also have trouble dealing with short scale variability and accuracy in maxima and minima. SR via GP is a method for obtaining mathematical expressions that match samples. Kordon [1] has already used it for building surrogate models. He recognizes 2 advantages: low development efforts and modeling with no assumptions. This paper suggest other advantage: it offers solutions that are interpretable; that means we can analyze the surrogate model and get better insight into the problem.

3. NUMERICAL EXPERIMENTS

Three study cases will be done. Surrogate models for two functions (Branin Function and Rastrigin Function) will be calculated. Equation (1) is Branin; eq.(2) is Rastrigin.

$$f(x_1, x_2) = (x_2 - \frac{5.1}{4\pi^2}x_1^2 + \frac{5}{\pi}x_1 - 6)^2 + 10(1 - \frac{1}{8\pi})cos(x_1) + 10 \quad (1)$$

$$f(x_1, x_2, x_3) = 30 + \sum_{i=1}^{3}[x_i^2 - 10cos(2\pi x_i)] \quad (2)$$

The third case is an electromagnetic problem:.the calculation of Forward Gain by a method of moments (MoM) simulation of a Yagi antenna with 4 elements, using the Numerical Electromagnetics Code 2 (nec2). The antenna is 6 m above a perfect electric conductor (PEC) ground. The elements' length are x_1,x_2,x_3,$x_4 \in [0.2, 0.4]$ m; the distances between elements are x_5,x_6,$x_7 \in [0.1, 0.2]$ m. The geometry of the antenna is depicted in fig.1. The driven element is the

Table 1: Comparison for surrogate Rastrigin against test set of 64000 samples

Set	MAE		RMSE	
	ANN	SR	ANN	SR
1	11.10367	5.93×10^{-6}	13.70285	7.36×10^{-6}
2	11.38934	16.6×10^{-6}	14.14982	20.7×10^{-6}
3	11.54664	6.13×10^{-6}	14.30546	7.60×10^{-6}
4	11.15623	8.92×10^{-6}	13.78459	11.1×10^{-6}
5	11.92689	8.96×10^{-6}	14.80023	11.2×10^{-6}
6	11.72848	2.90×10^{-6}	14.54511	3.60×10^{-6}
7	11.29536	5.93×10^{-6}	14.06576	7.36×10^{-6}
8	11.00931	5.80×10^{-6}	13.63077	7.21×10^{-6}
9	12.42965	20.0×10^{-6}	15.58501	24.9×10^{-6}
10	11.24940	8.81×10^{-6}	13.95887	10.9×10^{-6}

Table 2: Coefficients for Yagi's SR model

Cf.	Value	Cf.	Value	Cf.	Value
c_1	19.420713	c_9	34.592365	c_2	0.74041718
d_1	9.6399803	c_3	30.241159	d_2	16.280024
c_4	2.9810262	d_3	7.9505529	c_5	4.0442653
d_4	5.8341556	c_6	28.972141	d_5	74.196007
c_7	44.467743	d_6	14.458252	c_8	2.5384953

second starting from the left. There is a support beam 0.01 m below the elements. Material is aluminium with conductivity= $3.72 \times 10^7 S/m$. The objective is to maximize Forward Gain for frequency $f = 435$ Mhz.

Samples of the search space will be generated with latin hypercube. For each test function, 10 different sets of random points (samples) will be created, and for each set 1 surrogate model using ANNs and 1 using SR via GP will be obtained. For Branin, 100 samples per set are used, and for Rastrigin 216 samples. For Yagi, only 1 set of 300 samples will be created. For each set, NM different models will be built (NM=12 for test functions, NM=100 for Yagi), with different (random) initial conditions for ANN training, and the best model is chosen; "best" meaning the model with the largest R^2. For each set, 70% of points will be used for training, 15% for validation. A two-layer feed-forward network with sigmoid hidden neurons and linear output neurons, trained with Levenberg-Marquardt algorithm, is used. For Branin, there are 2 neurons on input layer, 12 on hidden layer and 1 on output. For Rastrigin, it is 3, 12 and 1. For Yagi, it is 7, 20 and 1. The symbolic regression models for the test functions are calculated after 2×10^6 generations, and for the Yagi problem after 2×10^6 generations. Eureqa [3] is used, running in parallel on 4 cores of Intel Xeon X3430 with 256 individuals in population; 70% of points are used for training, and the remaining 30% are used for validation. The alphabet chosen are constants, +, -, *, /, exp(), log(), sin() and cos(). Then, after having calculated the surrogate models, they will be used on M random points in the search space (M=10000 for Branin and Yagi, M=64000 for Rastrigin) and compare true values against predicted values. Mean absolute error (MAE) and root mean square error (RMSE) are used as criteria of comparison.

The results for Branin shown that ANN is superior to SR in 70% of the sets. Table 1 shows MAE and RMSE metrics for Rastrigin. Here, SR is superior to ANN. For the Yagi problem, the regression using SR via GP was done in two stages: the first stage gets a first approximation $\hat{y}_1$ for the Forward Gain; then $\Delta y = Gain - \hat{y}_1$ is obtained and a SR model $\hat{y}_2$ for Δy is evolved. The final model $\hat{y} = \hat{y}_1 + \hat{y}_2$ approximates the forward gain. The evolved functions are shown in eqs. (3)-(4), and the coefficients for these equations are presented in table 2.

Table 3: Comparison for surrogate Yagi model

Set	Model	R^2	MAE	RMSE
Testing set	ANN	0.62957	1.81097	2.38158
(10000 samples)	SR	0.79957	1.28872	1.75186

$$\hat{y}_1 = \frac{cos(c_7x_3)}{c_8x_6} + \frac{c_2 - sin(c_3x_1) - c_4x_4}{x_5} + c_1 + c_5sin(c_6x_1) - c_9x_3 \quad (3)$$

$$\hat{y}_2 = -d_1x_4sin(d_2x_4)^2 - d_3sin(d_2x_4)^3 - d_4x_3cos(d_5x_3)sin(d_2x_4)^2 - d_6x_3sin(d_2x_4)^2 \quad (4)$$

Table 3 shows all metrics for both ANN and SR models. The test againts the 10000 random samples indicates that SR is better than ANN. Knowing that $\nabla\hat{y} = 0$ for the maxima and minima:

$$\partial\hat{y}/\partial x_5 = -(c_2 - sin(c_3x_1) - c_4x_4)/x_5^2 = 0 \quad (5)$$

$$\partial\hat{y}/\partial x_6 = -cos(c_7x_3)/c_8x_6^2 = 0 \quad (6)$$

According eq.(6), in the region of interest there are 3 possible values of x_3, {0.2472708, 0.3179196, 0.3885684}. Following eq.(5), thre is a curve obeying $sin(c_3x_1) = c_2 - c_4x_4$. For each value of x_1 there is only one possible value of x_5, according to $\frac{\partial\hat{y}}{\partial x_1} = 0$. According to $\frac{\partial\hat{y}}{\partial x_3} = 0$, for each combination of variables (x_3, x_4) there is only one possible x_6. Therefore, allowing for some errors, the search space can be reduced from the original 7-D cube of $size = 0.2^7$ into 5-D and 6-D regions. After running 1600 numerical models in those areas, a candidate maximum was found in **x** =(0.3172737, 0.2499185, 0.3080215, 0.3047194, 0.2611489, 0.10451, 0.2310036), with Forward Gain=17.16 dBi. It is very close to the best Forward Gain =17.25 dBi found between the 10000 random samples.

4. CONCLUSIONS

In this paper, an empirical comparison between ANN and SR via GP for surrogate modeling has been presented. Two main advantages of this approach were shown here: the ability to exploit the function obtained by SR as a "white box", amenable to analysis by calculus (in the Yagi problem, this analysis helps to reduce the search space), and good accuracy (competitive with ANNs).

Acknowledgments to CONCYTEC, IdI-FC-UNI, II-FIIS-UNI.

5. REFERENCES

[1] A. Kordon and C. T. Lue. Symbolic regression modeling of blown film process effects. In *Congress on Evolutionary Computation 2004*, volume 1, pages 561–568, June 2004.

[2] J. E. Rodríguez, A. L. Medaglia, and C. A. Coello Coello. Design of a motorcycle frame using neuroacceleration strategies in moeas. *Journal of Heuristics*, 15:177–196, April 2009.

[3] M. Schmidt and H. Lipson. Distilling free-form natural laws from experimental data. *Science*, 324(5923):81–85, 2009.