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.

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.

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

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

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

<table>
<thead>
<tr>
<th>Set</th>
<th>ANN</th>
<th>MAE</th>
<th>SR</th>
<th>ANN</th>
<th>RMSE</th>
<th>SR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11.1466</td>
<td>5.93</td>
<td>10^{-4}</td>
<td>13.70285</td>
<td>7.36</td>
<td>10^{-6}</td>
</tr>
<tr>
<td>2</td>
<td>11.38934</td>
<td>16.6</td>
<td>10^{-4}</td>
<td>14.1982</td>
<td>20.7</td>
<td>10^{-7}</td>
</tr>
<tr>
<td>3</td>
<td>11.54664</td>
<td>6.13</td>
<td>10^{-4}</td>
<td>14.30546</td>
<td>7.60</td>
<td>10^{-6}</td>
</tr>
<tr>
<td>4</td>
<td>11.95288</td>
<td>8.92</td>
<td>10^{-4}</td>
<td>13.78450</td>
<td>11.1</td>
<td>10^{-6}</td>
</tr>
<tr>
<td>5</td>
<td>11.92689</td>
<td>8.96</td>
<td>10^{-4}</td>
<td>14.80293</td>
<td>11.2</td>
<td>10^{-6}</td>
</tr>
<tr>
<td>6</td>
<td>11.72848</td>
<td>2.90</td>
<td>10^{-4}</td>
<td>14.95110</td>
<td>3.60</td>
<td>10^{-6}</td>
</tr>
<tr>
<td>7</td>
<td>11.29562</td>
<td>5.93</td>
<td>10^{-4}</td>
<td>14.06567</td>
<td>7.36</td>
<td>10^{-6}</td>
</tr>
<tr>
<td>8</td>
<td>11.00931</td>
<td>5.80</td>
<td>10^{-4}</td>
<td>13.63077</td>
<td>7.21</td>
<td>10^{-6}</td>
</tr>
<tr>
<td>9</td>
<td>12.42965</td>
<td>20.0</td>
<td>10^{-6}</td>
<td>15.58501</td>
<td>24.9</td>
<td>10^{-6}</td>
</tr>
<tr>
<td>10</td>
<td>11.24940</td>
<td>8.81</td>
<td>10^{-4}</td>
<td>13.95887</td>
<td>10.9</td>
<td>10^{-6}</td>
</tr>
</tbody>
</table>

Table 2: Coefficients for Yagi’s SR model

<table>
<thead>
<tr>
<th>Cl.</th>
<th>Value</th>
<th>Cl.</th>
<th>Value</th>
<th>Cl.</th>
<th>Value</th>
<th>Cl.</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>c₁</td>
<td>19.420714</td>
<td>c₂</td>
<td>34.972606</td>
<td>c₃</td>
<td>0.3172708</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d₁</td>
<td>9.639693</td>
<td>d₂</td>
<td>10.241195</td>
<td>d₃</td>
<td>10.280024</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d₄</td>
<td>2.9817062</td>
<td>d₅</td>
<td>2.9920629</td>
<td>c₅</td>
<td>4.9412653</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d₆</td>
<td>5.8341556</td>
<td>d₇</td>
<td>28.972411</td>
<td>d₈</td>
<td>74.196007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d₉</td>
<td>44.467743</td>
<td>d₁₀</td>
<td>14.458232</td>
<td>c₉</td>
<td>2.5384953</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

model \( \hat{y}_2 \) for \( \Delta y \) is evolved. The final model \( \hat{y}_2 = \hat{y}_1 + \hat{y}_2 \) approximates the forward gain. The evolved functions are shown in eqs. (3)-(4), and the coefficients for these equations are presented in table 2.

\[
\hat{y}_1 = \frac{\cos(c_7 x_3)}{c_8 x_6} + \frac{c_2 - \sin(c_3 x_1) - c_4 x_4}{x_5} + c_1 + c_5 \sin(c_6 x_1) - c_9 x_3 
\]

(3)

\[
\hat{y}_2 = -d_1 x_4 \sin(d_2 x_4)^2 - d_3 \sin(d_2 x_4)^3 
\]

(4)

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

\[ \frac{\partial \hat{y}}{\partial x_5} = -\left( c_2 - \sin(c_3 x_1) - c_4 x_4 \right) / x_5^2 = 0 \]

(5)

\[ \frac{\partial \hat{y}}{\partial x_6} = -\cos(c_7 x_3) / c_8 x_6^2 = 0 \]

(6)

According eq. (6), in the region of interest there are 3 possible values of \( x_3 \), \( 0.1272708, 0.3179196, 0.385684 \). Following eq.(5), there is a curve obeying \( \sin(c_3 x_1) = c_2 - c_4 x_4 \). For each value of \( x_3 \) there is only one possible value of \( x_5 \), according to \( \frac{\partial \hat{y}}{\partial x_5} = 0 \). According to \( \frac{\partial \hat{y}}{\partial x_6} = 0 \), 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 \( S = 0.2^7 \) into 5-D and 6-D regions. After running 1600 numerical models in those areas, a candidate maximum was found in \( x_5 = (0.3172737, 0.2499185, 0.3080215, 0.3047194, 0.2611489, 0.10451, 0.2310036) \), 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