Available online at www.sciencedirect.com

ScienceDirect

Procedia Social and Behavioral Sciences 2 (2010) 7737–7738

Procedia
Social and Behavioral Sciences

Sixth International Conference on Sensitivity Analysis of Model Output

Evaluation of MARS modeling technique for sensitivity analysis of model output

Davood Shahsavani[a]*, Stefano Tarantola[b], Marco Ratto[b]

[a] *Shahrood University of Technology, 3619955161 Shahrood - IRAN*
[b] *Joint Research Centre of the European Commission, 21027 Ispra (VA) - ITALY*

Abstract

Sensitivity analysis of computer models can typically require a large number of model runs. When these models are computationally expensive to run, it may be advantageous to invest in computationally cheaper surrogate models (emulators or meta-models) that can provide almost the same output as the original model and estimate the sensitivity indices for each input. In this abstract the MARS method is used to mimic the behavior of a nonlinear and non-additive test function. The results show that, overall, MARS provides acceptable estimates of total sensitivity indices at a much lower cost than using only runs of the original model.

Keywords: Surrogate model; Total sensitivity indices; Multivariate adaptive regression Splines (MARS)

1. Main text

The method of multivariate adaptive regression splines (MARS) is a nonparametric regression technique introduced by Friedman in 1991. It can be considered as a generalization of stepwise linear regression and a modification of the CART method (Hastie *et al.*). MARS uses a class of pairwise spline basis functions $(x-t)_+$ and $(t-x)_+$ for each input x_j where $t \in \{x_{1j}, x_{2j}, \ldots, x_{nj}\}$ $j = 1,2,\ldots,p$ is called the knot, and $(\ )_+$ shows the positive part. The general form of the MARS model can be represented by the following expression

$$\hat{y} = \beta_0 + \sum_{j=1}^{M} \beta_j B_j(\boldsymbol{x}) \tag{1}$$

where $\boldsymbol{x}=(x_1, x_2, \ldots, x_p)$ is the vector of inputs, B_j is the j-th basis function, which can be a single spline function or a product of two or more basis functions, and the coefficients β_j s are estimated by minimizing the sum of squared residuals (SSR). In fact, MARS uses a specific class of basis functions as predictors in place of the original input variables. In other words, the MARS regression model is constructed by fitting basis functions to distinct intervals

* Corresponding author. Tel.: +98-273-3392012; fax: +98-273-3392012.
E-mail address: dshahsavani@shahroodut.ac.ir

doi:10.1016/j.sbspro.2010.05.204

of the independent variables that are smoothly connected together at knots. MARS uses a forward-backward procedure to construct the final model. The forward step is very similar to stepwise regression, but instead of the x_j 's, MARS uses the basis functions. The complex model that is achieved after this step is fitted to the data well but this is an over-fitted model with poor ability to predict new untried points. Hence a pruning backward procedure is needed to remove the redundant basis functions. In each step of the backward procedure, a basis function that has less contribution in increasing the SSR is removed. By removing each basis function, a new model is estimated that could be a candidate for the final model. The performance of the new model is evaluated by Generalized Cross Validation (GCV). This process is continued until all basis functions are removed from the model. Finally, MARS chooses the best model with less GCV.

The performance of the MARS and classic methods (Saltelli et al., 2010) to estimate sensitivity indices is examined on the test function

$$y = \prod_{i=1}^{k} f(x_i) \qquad ; \; f(x_i) = \frac{b\left[0.25 + \frac{1}{2\pi}\arcsin(\cos(2\pi(x_i + c)))\right]^{\alpha} + a_i}{1 + a_i} \tag{2}$$

where $k = 8$, $[a_1, a_2, \ldots, a_8] = [\,0,1,4.5,9,99,99,99,99\,]$, c=0, α=0.25 and $b=(\alpha+1)2^{\alpha}$.

We used four groups of training sets at different sample sizes (128, 256, 512 and 1024 points). The training sets for MARS were generated using the quasi-random number generator of Sobol'. The outputs from the surrogate model were calculated on a test set of 36,864 untried points, on which the sensitivity indices were estimated. To quantify the confidence of sensitivity estimates, the whole exercise was repeated 50 times.
The results are summarized in Figure 1. For the most influential factor x_1 (left panel), the estimates of S_{T1} using MARS converge to the analytical value faster than the classic approach. For the non-important factor x_7 the classic method overestimates the analytical value at small sample size, while the approach based on MARS does not have such bias. Nonetheless, the estimates of S_T using MARS are larger than those of the classic method.

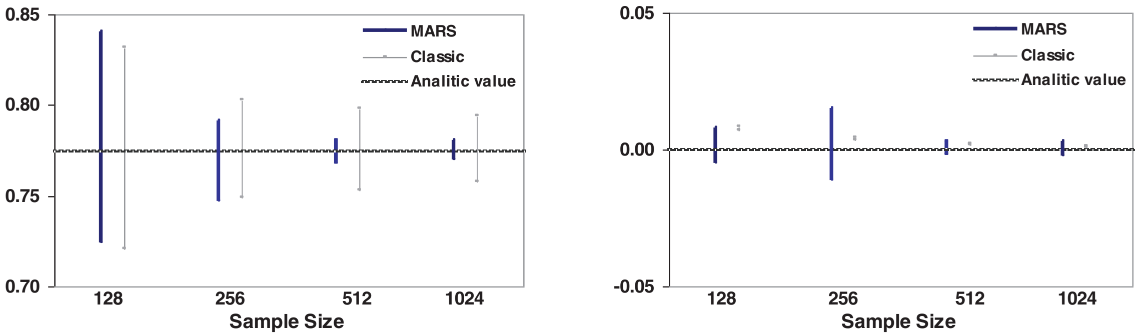

Figure 1. The confidence intervals (vertical lines) of estimated total sensitivity indices against different sample sizes for the influential factor x_1 (left panel) and for non-important factor x_7 (right panel). Both are based on 50 replicates.

2. References

Friedman, J. H., 1991: "Multivariate adaptive regression splines". *The Annals of Statistics*, 19(1), 1–67.
Hastie, T., Tibshirani, R., Friedman, J. 2001: "The Elements of Statistical Learning: Data mining, Inferences, and Prediction". *Springer*, New York.
Saltelli, A., P. Annoni, I. Azzini, F. Campolongo, M. Ratto, S. Tarantola, (2010) Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index, *Computer Physics Communications*, 181, 259–270.