MascotNum2016 conference - Local Pearson correlation coefficients for global sensitivity analysis

Peio LOUBIERE
Université de Paris-Est Créteil (UPEC)

Supervisor(s): Prof. Patrick Siarry (UPEC), Prof. Rachid Chelouah (EISTI) and Dr. Astrid Jourdan (EISTI)

Ph.D. expected duration: 2013-2016

Adress: plo@eisti.eu

Email: siarry@u-pec.fr, rc@eisti.eu, aj@eisti.eu

Abstract:

In this paper, we present a global sensitivity analysis method, based on Pearson correlation coefficients. These coefficients can detect the influence of each input parameter of a computer model, with a low calculation cost, but under linearity hypothesis. Inspired by the Morris' method calculation of mean and standard deviation of elementary effects, we propose to transform the global study of a non linear model into a local study of quasi-linear sub-parts of the model to evaluate the global influence of each input parameter on the model. Our method is designed to be included in various metaheuristics' search process.

In a previous work [2], Morris' OAT method has been integrated in a metaheuristic optimization algorithm, Artificial Bee Colony (ABC) algorithm, to improve the optimum search, with less evaluations. Morris' method fits particularly well to ABC algorithm because of its one-dimension neighborhood search process.

This is a special case, more improvement of this algorithm and other metaheuristics algorithms like evolutionary algorithms search a neighbor offsetting various dimensions at a time. So we need to use a sensitivity analysis method that can detect influence of input parameters in a multidimensional neighborhood. Our constraints are:

- process with complex functions in the context of hard optimization
- no constraint on experience plan, because the points are given by a multi-dimensional neighborhood search
- a reasonable number of evaluations

Finally, we need as a result, the identification of non influent parameters and a ranking of influence of each input parameter.

Among all existing methods, Sobol indices ([4]) can deal complex functions and leads to precise informations on the behavior of each input parameter (such as simple effects, interactions or total effects). But Sobol method, as its improvement (FAST), still requires a huge number of inputs and evaluations to be fulfilled. The use of a metamodel is taken away, due to its important numerical cost. With the same computational cost than Morris method ([3], [1]), there are linear and rank regressions. These two last method have a huge constraint on the model (linear, monotonous). We propose to combine Pearson correlation coefficients with Morris method to get a global sensitivity analysis of a non linear model.

For a linear D-dimensional model, considering a set of N points $X^i = (X^i_1, \ldots, X^i_D)$ and outputs Y^i, $i = (1, \ldots, n)$, The Pearson correlation coefficients method evaluates the linear correlation coefficient $\rho(.,.)$ between X_j and Y:

$$\rho_j \;=\; \rho(X_j, Y) \;=\; \frac{Cov(X_j, Y)}{\sqrt{Var(X_j)Var(Y)}} \tag{1}$$

This method is simple and do not need many points and evaluations: $N \geq D + 1$. But it can be applied for global sensitivity analysis only with linear model.

To use a global linear method in a non linear context, the idea is to evaluate the coefficients of a local neighborhood of points. We select a number k of points ($k \leq N$) and for each point, we select the p closest neighbors. For each of these k neighborhoods, the linear correlation coefficients (LCC) are evaluated according to equation (1). The LCC matrix (2) is built with all of the k points:

$$M_{LCC} \;=\; \begin{pmatrix} LCC_1^1 & ... & LCC_1^k \\ & ... & \\ LCC_d^1 & ... & LCC_d^k \end{pmatrix} \tag{2}$$

As done in Morris' method, normalized mean and standard deviation are computed, by dimension (ie lines) of the LCC matrix (2) as described by equations (3).

$$\mu_i^* \;=\; \frac{1}{k}\sum_{j=1}^{k} |LCC_i^j|, \qquad \sigma_i \;=\; \sqrt{\frac{1}{k}\sum_{j=1}^{k}(LCC_i^j) - \mu_i)^2} \tag{3}$$

The weight of the i^{th} input parameter is given by the distance d_i of the couple (μ_i^*, σ_i) to the origin. It is given by equation (4).

$$weightl_i = di/(\sum_{j=1}^{D} d_i) \tag{4}$$

These weights will help to guide metaheuristics' search process, to prioritize promising dimensions for fast convergence. The method has been validated on sensitivity analysis' usual benchmark functions (polynomial, Ishigami, Sobol g-function). This method can also be used as sensitivity analysis on given data.

References

[1] Francesca Campolongo, Jessica Cariboni, and Andrea Saltelli. An effective screening design for sensitivity analysis of large models. *Environ. Model. Softw.*, 22(10):1509–1518, 2007.

[2] Peio Loubière, Astrid Jourdan, Patrick Siarry, and Rachid Chelouah. A sensitivity analysis method for driving the artificial bee colony algorithm's search process. *Summited to Applied Soft Computing*, 2015.

[3] Max D. Morris. Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2):161–174, April 1991.

[4] I.M Sobol. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics and Computers in Simulation*, 55(13):271 − 280, 2001. The Second {IMACS} Seminar on Monte Carlo Methods.

Short biography – I have a Master 2 in applied mathematics. After having working in an IT service company as a software engineer for four years, in a I am employed at EISTI engineer school since 2008. In 2013, my employer allowed me to start a Phd around optimization using swarm intelligence metaheuristics, coupled with sensitivity analysis methods to improve the optimum search.