Global sensitivity analysis by HDMR combining with the improved GMDH algorithm

Lu Wang[a], Shufang Song[b], Shiyu Liu[c]

(School of Aeronautics, Northwestern Polytechnical University, Xi'an, Shaanxi, 710072, PR China)

[a]louisewanglu@gmail.com [b]shufangsong@nwpu.edu.cn [c]shiyu_liu22@163.com

Abstract

Global sensitivity analysis (GSA) is a very useful tool to evaluate the influence of input variables in the whole distribution range. The interactive influences have been taken into consideration in GSA. The variance-based GSA was proposed by Sobol' according to ANOVA decomposition [1]:

$$f(\boldsymbol{x}) = f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum_{1\le i<j\le n} f_{ij}(x_i, x_j) + \sum_{1\le i<j<k\le n} f_{ij}(x_i, x_j, x_k) + \cdots \tag{1}$$

where the input variables are $\boldsymbol{x}=(x_1, x_2, \cdots, x_n)^{\mathrm{T}}$. Furthermore, through a complete basis set of orthonormal polynomials function, random sampling high dimensional model representation (HDMR) can be used to solve the Sobol' first and second order global sensitivity indices [2].

$$f_i(x_i) \approx \sum_{p=1}^{m_1} \alpha_p^i \varphi_p(x_i), \qquad f_{ij}(x_i, x_j) \approx \sum_{p=1}^{m_2}\sum_{q=1}^{m_3} \beta_{pq}^{ij} \varphi_{pq}(x_i, x_j) \tag{2}$$

where m_1, m_2, m_3 are suitable maximum orders. And coefficients $\alpha_p^i, \beta_{pq}^{ij}$ are calculated by random sampling as follows:

$$\alpha_p^i \approx \frac{1}{N}\sum_{s=1}^{N} f\left(\boldsymbol{x}^{(s)}\right)\varphi_p\left(x_i^{(s)}\right), \quad \beta_{pq}^{ij} \approx \frac{1}{N}\sum_{s=1}^{N} f\left(\boldsymbol{x}^{(s)}\right)\varphi_{pq}\left(x_i^{(s)}, x_j^{(s)}\right) \tag{3}$$

However, the flaws of classical HDMR method cannot be ignored. For instance, it needs a large number of samples N to calculate the decomposition coefficients and cannot calculate high order sensitivity indices. The group method of data handling(GMDH) is a family of inductive algorithms for computer-based mathematical modeling of multi-parametric datasets that features fully automatic structural and parametric optimization of models. In order to improve the GMDH, neural network algorithm are integrated to generate the assured function description of the nonlinear model with high precision using a relative small samples[3].Besides, the IGMDH algorithm has a fixed number of layers which cannot exceed 3.

Thus, we consider to combine the improved group method of data handling (IGMDH) with HDMR for calculating the coefficients more efficiently. The main procedures of the IGMDH-HDMR method that we proposed are presented in Figure 1, and the structure of it is shown in Figure 2. The metamodeling of IGMDH-HDMR is:

$$f(\boldsymbol{x}) = f_0 + \sum_p \alpha_p^i \varphi_p(x_i) + \sum_{p,q} \beta_{pq}^{ij}\varphi_{pq}(x_i, x_j) + \sum_{p,q,r} \gamma_{pqr}^{ijk}\varphi_{pqr}(x_i, x_j, x_k) + \sum_{p,q,r,s} \delta_{pqrs}^{ijkl}\varphi_{pqrs}(x_i, x_j, x_k, x_l) \tag{4}$$

As a result, Sobol' sensitivity indices can be calculated by the derived coefficients:

$$\hat{S}_i = \sum_p \left(\alpha_p^i\right)^2 / \hat{D}, \quad \hat{S}_{ij} = \sum_{p,q}\left(\beta_{pq}^{ij}\right)^2 / \hat{D}, \quad \hat{S}_{ijk} = \sum_{p,q,r}\left(\gamma_{pqr}^{ijk}\right)^2 / \hat{D}_n, \quad \hat{S}_{ijkl} = \sum_{p,q,r,s}\left(\delta_{pqrs}^{ijkl}\right)^2 / \hat{D}_n \tag{5}$$

$$\hat{S}_i^{tot} = \hat{S}_i + \sum_{j\ne i}\hat{S}_{ij} + \sum_{k\ne j\ne i}\hat{S}_{ijk} + \sum_{l\ne k\ne j\ne i}\hat{S}_{ijkl} \tag{6}$$

where total variance is calculated as:

$$\hat{D} = \sum_{p} \left(\alpha_{p}^{i}\right)^{2} + \sum_{p,q} \left(\beta_{pq}^{ij}\right)^{2} + \sum_{p,q,r} \left(\gamma_{pqr}^{ijk}\right)^{2} + \sum_{p,q,r,s} \left(\delta_{pqrs}^{ijkl}\right)^{2} \tag{7}$$

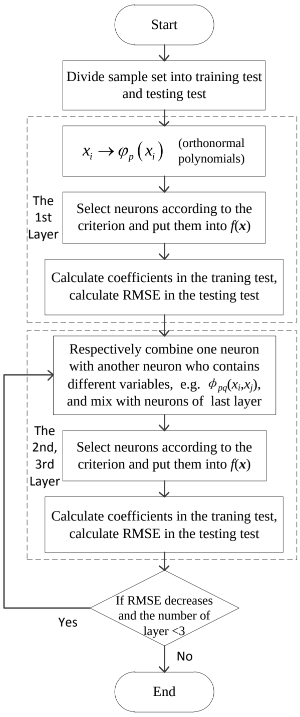

Fig. 1 The flowchart of GMDH-HDMR

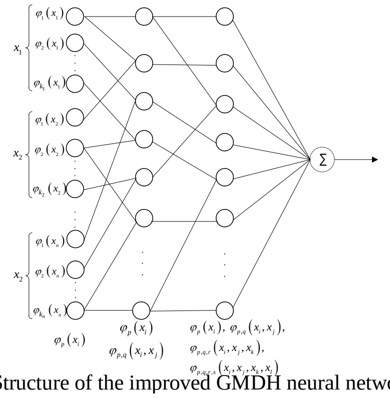

Fig. 2 Structure of the improved GMDH neural network

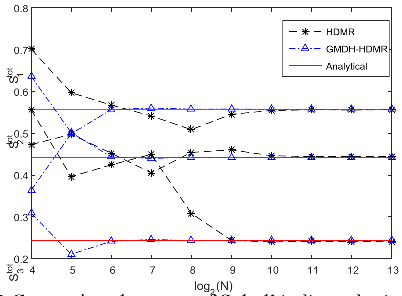

Fig. 3 Comparison between of Sobol' indices obtained from GMDH-HDMR with RS-HDMR

Here are the details of how to choose the maximal order. Assume $p_{max} = m_1$ and calculate the global sensitivity at different order. Thus we can get a series of $\hat{S}_{i}^{p}, \hat{S}_{ij}^{pq}, \hat{S}_{ij}^{pq}, \hat{S}_{ijkl}^{pqrs}$.

When $\left|\hat{S}_{i}^{p^*+1} - \hat{S}_{i}^{p^*}\right| < \varepsilon_1$, $p_{max}=p^*$. When $\left|\hat{S}_{ij}^{p^*+1,q^*+1} - \hat{S}_{ij}^{p^*,q^*}\right| < \varepsilon_2$, $(p_{max}, q_{max})=(p^*, q^*)$.

When $\left|\hat{S}_{ijk}^{p^*+1,q^*+1,r^*+1} - \hat{S}_{ijk}^{p^*,q^*,r^*}\right| < \varepsilon_3$, $(p_{max}, q_{max}, r_{max}) = (p^*, q^*, r^*)$.

When $\left|\hat{S}_{ijkl}^{p^*+1,q^*+1,r^*+1,s^*+1} - \hat{S}_{ijkl}^{p^*,q^*,r^*,s^*}\right| < \varepsilon_4$, $(p_{max}, q_{max}, r_{max}, s_{max}) = (p^*, q^*, r^*, s^*)$

The neurons whose order exceed the maximal value will be deleted.

Finally, the Ishigami function is considered, which is a highly nonlinear function of three inputs: $f(\boldsymbol{x}) = \sin(x_1) + 7\sin^2(x_2) + 0.1x_3^4 \sin(x_1)$, where x_i (i=1,2,3) are uniformly distributed on the interval $[-\pi, \pi]$. And the comparisons of Sobol' indices obtained from IGMDH-HDMR with HDMR are listed in Figure 3. It is obvious that both precision and convergence of the IGMDH-HDMR method are higher than those of the RS-HDMR method.

Keywords: global sensitivity analysis; high dimensional model representation(HDMR); group method of data handling(GMDH) algorithm; Sobol' sensitivity indices

References

[1] I. M. Sobol. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and Computers in Simulation, 2001, 55(1): 271–280.

[2] B. Feil, S. Kucherenko, N. Shah. Comparison of Monte Carlo and quasi Monte Carlo sampling methods in high dimensional model representation. First International Conference on Advances in System Simulation, Portugal, 2009: 12-17.

[3] AG. Ivakhnenko. Heuristic self-organization in problems of engineering cybernetics. Automatica, 1970, 6(2): 207-219.