

A New Variable Screening Method for Design Optimization of Large-Scale Problems

Kai Zheng
Shanghai Jiao Tong Univ.

Ren-Jye Yang
Ford Motor Co

Jie Hu
Shanghai Jiao Tong Univ.

ABSTRACT

Design optimization methods are commonly used for weight reduction subjecting to multiple constraints in automotive industry. One of the major challenges remained is to deal with a large number of design variables for large-scale design optimization problems effectively. In this paper, a new approach based on fuzzy rough set is proposed to address this issue. The concept of rough set theory is to deal with redundant information and seek for a reduced design variable set. The proposed method first exploits fuzzy rough set to screen out the insignificant or redundant design variables with regard to the output functions, then uses the reduced design variable set for design optimization. A vehicle body structure is used to demonstrate the effectiveness of the proposed method and compare with a traditional weighted sensitivity based main effect approach.

CITATION: Zheng, K., Yang, R., and Hu, J., "A New Variable Screening Method for Design Optimization of Large-Scale Problems," *SAE Int. J. Mater. Manf.* 8(3):2015, doi:10.4271/2015-01-0478.

1. INTRODUCTION

In design optimization problem, a set of design variables that describe the formulation needs to be identified [1]. Often, too many design variables are defined in the design optimization problem due to lack of information. A large number of design variables may result in insurmountable computational burden, especially for large-scale real-world applications. As a result, selecting a smaller and representative design variable set is imperative for large-scale optimization problems.

Variable screening is a useful method for design optimization as it plays a key role in selecting critical design variables. Craig et al. [2] applied the coefficients and their confidence intervals of a linear response surface to select key variables in automotive crashworthiness design. Wang [3] used a forward regression model to efficiently identify essential variables among candidate variables based on data. Izenman [4] exploited manifold learning to preserve the input information in the reduced dimension for efficient statistical analysis. Choi and Kim [5] analyzed the design sensitivity to select the vital variables based on the rate of change in the performance measure at the design point. Chen et al. [6] applied analysis of variance (ANOVA) decomposition to global sensitivity analysis to evaluate main effects and interaction effects. Wu and Mohanty [7] employed the local sensitivity of mean-response and cumulative

distribution function as measures to screen out the variables. Welch et al. [8] used maximum likelihood estimator (MLE) of the correlation parameters to identify important variables for deterministic problems.

Despite of the interest, most previous works have limitations for applications to the design optimization of large-scale problems. One is that the screening results will be different at different designs, and thus it is difficult to select the specific design point. The other one is that they cannot consider multiple responses effectively.

This paper proposes a fuzzy rough set based variable screening method for design optimization of large-scale problems. Rough set theory, first proposed by Pawlak [9], is a mathematical tool to deal with redundant information and seek for a reduct including the critical design variables. In this approach, the fuzzy similarity is first introduced to define the similarities between each two training samples, and the inconsistency matrix is exploited to compute the number of inconsistency pairs. Then an attribute reduction algorithm [10] of the fuzzy rough set method is used to find the critical variables. The remainder of this paper is arranged as follows: the proposed method for variable screening is presented firstly, as well as the introduction of several key steps including fuzzy similarity (FS) calculation, inconsistency matrix construction, attribute reduction

algorithm. An example is used to demonstrate the effectiveness of the proposed method and to compare with a traditional weighted sensitivity based main effect approach. Finally, summary is given.

2. FUZZY ROUGH SET BASED VARIABLE SCREENING METHOD

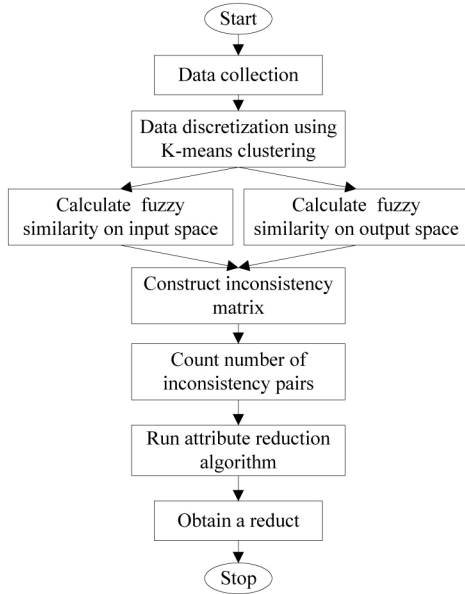


Fig. 1. Flowchart of fuzzy rough set based variable screening

In this part of this section, a novel fuzzy rough set based screening method together with fuzzification theory, inconsistency matrix is presented, and the procedure is described as Fig. 1.

The main process starts with data discretization with K-means clustering method, followed by fuzzy similarity (FS) calculation between all pairs of training samples in terms of design variables and output variables. Then the inconsistency matrix is constructed based on the fuzzy similarities. Afterwards, count the number of inconsistency pairs, and finally find a reduct including the critical design variables based on the attribute reduction algorithm of the fuzzy rough set. Some of these steps are described detailedly in the following subsections.

2.1. Fuzzy Similarity Calculation

For continuous variables, they need to be fuzzified, in which the number of fuzzy intervals n and the membership function should be determined firstly. Assuming there exist p design variables and q output variables in the large-scale optimization problems and l samples can be used for variable screening. Based on the fuzzification theory, the value v_{ij} of the i th ($1 \leq i \leq l$) sample regarding to the j th ($1 \leq j \leq p+q$) variable is expressed as follows:

$$v_{ij} = \{(I_j^1, m_{ij}^1), (I_j^2, m_{ij}^2), \dots, (I_j^n, m_{ij}^n)\} \quad (1)$$

where I_j^k ($1 \leq k \leq n$) represents the k th fuzzy interval of the j th variable, and m_{ij}^k represents the membership of the i th sample belongs to this interval. The triangular membership function [11] is selected as the membership function for the continuous data, shown as Equation (2, 3, 4) and Fig. 2.

$$m_{ij}^1 = \begin{cases} 1, & v_{ij} \leq v_j^1 \\ (v_j^2 - v_{ij}) / (v_j^2 - v_j^1), & v_j^1 < v_{ij} < v_j^2 \\ 0, & v_{ij} \geq v_j^2 \end{cases} \quad (2)$$

$$m_{ij}^2 = \begin{cases} (v_{ij} - v_j^1) / (v_j^2 - v_j^1), & v_j^1 < v_{ij} < v_j^2 \\ (v_j^3 - v_{ij}) / (v_j^3 - v_j^2), & v_j^2 \leq v_{ij} < v_j^3 \\ 0, & v_{ij} \leq v_j^1 \text{ or } v_{ij} \geq v_j^3 \end{cases} \quad (3)$$

$$m_{ij}^3 = \begin{cases} 0, & v_{ij} \leq v_j^2 \\ (v_{ij} - v_j^2) / (v_j^3 - v_j^2), & v_j^2 < v_{ij} < v_j^3 \\ 1, & v_{ij} \geq v_j^3 \end{cases} \quad (4)$$

where v_j^k ($1 \leq k \leq 3$) is computed with fuzzy C-means (FCM) clustering method [12], so here $I_j^1 = (-\infty, v_j^2]$, $I_j^2 = [v_j^1, v_j^3]$, $I_j^3 = [v_j^2, \infty]$.

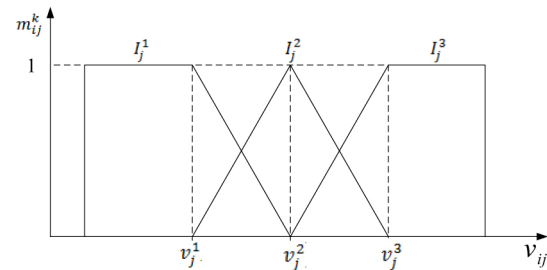


Fig. 2. The triangular membership function

Fuzzy similarity is calculated after data fuzzification. Usually distance function is applied to similarity quantification, mainly including Manhattan distance [13] and Euclidean distance [14]. Here Manhattan distance is selected to calculate the fuzzy similarity. The fuzzy similarity between the e th and f th ($1 \leq e, f \leq l$) samples regarding to the j th variable is expressed as follows:

$$r_j(t_e, t_f) = 1 - \frac{1}{2} \sum_{i=1}^n |m_{ej}^i - m_{fj}^i| \quad (5)$$

The above described is the fuzzy similarity calculation regarding to a single design variable or output variable, and the fuzzy similarities between the e th and f th samples regarding to the whole design variable set C and the output variable set D are expressed respectively as follows:

$$R_C(t_e, t_f) = \min_{1 \leq j \leq p} r_j(t_e, t_f)$$