Comparative Study on Data Mining Methods in Structural Reliability Prediction

Willy Husada[1], I-Tung Yang[2], and Tri Joko Wahyu Adi[1]

Abstract–The goal of reliability-based design optimization (RBDO) is to find the optimal structure design with minimum cost subjected to maximum failure probability limit. Since failure probability is usually small, it takes a large amount of computation time for accurate estimation in reliability analysis. Surrogate models usually created to replace the time-consuming reliability analysis. In this empirical study, we use several data mining methods with focus on Classification and Regression Tree (CART), Artificial Neural Network (ANN) and Support Vector Machine (SVM) method to create the surrogate models on a empirical benchmark case study. We aim to find the best data mining method in predicting the failure probability which divided into two parts: classification and regression. The main findings of this study is that CART method performed better than ANN and SVM in both classification and regression. Support Vector Machine (SVM) method is the worst in both cases.

Index Terms – data mining, failure probability, reliability-based design optimization, surrogate model.

INTRODUCTION

Design quality is an important part in the structural construction project. A structural designs should produce a structure that is reliable enough subjected to uncertain conditions such as variability from construction process, material properties and external loads. Design optimization is used to improve the design quality so that the actual structure can have adequate safety with minimum cost. One of the most popular design optimization methods is Reliability-Based Design Optimization (RBDO). RBDO has two processes, design optimization and reliability analysis which aim to find the optimal design with minimum structure cost or weight subjected to maximum failure probability limit. In practical, RBDO involves highly non-linear limit state functions and non-normally distributes distributed random variables. These issues create challenges for accurate reliability analysis [1].

There are three integration frameworks of RBDO: double-loop, single-loop and decoupled. The double-loop method requires a full reliability analysis at every step of the design optimization process and too computationally expensive for practical application [2]. In single-loop method, a surrogate model is created to replace the time-consuming reliability analysis [3]. Despite the enhanced efficiency, the single-loop method may be inaccurate in estimating the structure failure probability because the surrogate model is associated with certain errors. Decoupled method divides double-loop method into sequential cycles and them improve the reliability by formulating a new optimization constraint in the next cycle for violated reliability constraints [4].

To improve the accuracy of the single-loop RBDO method, a better surrogate model is needed. Data mining through artificial intelligence (AI) based methods can provide a better surrogate model to predict the structure failure probability. This study attempts to implement several data mining methods to construct surrogate model in RBDO problem and has a main objective to find the best data mining method in predicting failure probability from a structural design. The prediction is divided into two parts, binary classification and regression. Binary classification model is focusing on minimization of cost while regression model is focusing on minimization of cost and failure probability. The proposed RBDO surrogate models is validated through a empirical benchmark case study which is ten-bar truss problem to demonstrate the prediction accuracy and computation time of proposed RBDO framework.

METHODS

The specific steps of this study are as follows:

1) Perform literature review related to data mining and RBDO.
2) Adopt the ten-bar plane truss to be the experimental case.
3) Use Monte Carlo Simulation (MCS) to conduct reliability analysis for preparing training data set.
4) Conduct preliminary experiments to select the best-three data mining methods among popular AI algorithms in the estimation of reliability.
5) Develop surrogate models based on the theories of CART, ANN and SVM.
6) Fine-tune the control parameters of the ANN and SVM data mining methods. CART method does not have any time-consuming parameter-tuning.
7) Evaluate the surrogate models performance using ten-fold cross validation in terms of prediction accuracy and computation time.

[1]Willy Husada and Tri Joko Wahyu Adi are with Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia. E-mail: willyhusada_srw@hotmail.com; trijoko_w@yahoo.com.

[2]I-Tung Yang is with National Taiwan University of Science and Technology (NTUST), Taiwan. E-mail: ityang@mail.ntust.edu.tw.

8) Compare the surrogate models and find the best data mining method in failure probability prediction.
9) Draw the study conclusions based on the experiment results.

CASE STUDY AND RESULT DISCUSSION

The benchmark case used in this study is a ten-bar plane truss problem. The shape, geometry and loading of the ten-bar truss structure are shown in Fig. 1. The ten-bar truss is pin-jointed and subjected to two external loads, P_1 and P_2. Every bar is made of hollow carbon steel pipes and may have different sizes. The selection of bars represents a discrete set with three features: pipe outside diameter (D), wall thickness (t) and cross-sectional area (A). In total, there are 36 discrete options that can be selected from the list and these options form a design space of 36^{10} discrete combinations which is more than 3.65×10^{15} options. This amount of possible options is considered huge for a relatively small RBDO problem.
We perform a series of preliminary experiments to choose the best surrogate models and conduct more detailed on these models. We use SPSS Clementine 12.0 to perform preliminary experiments because it is packaged with several popular data mining algorithms for both classification and regression. At first, there are seven methods for classification and four methods for regression. After conducting the preliminary experiments, the best three data mining methods for both classification and regression problem are CART, ANN and SVM.

We finally create surrogate models using three data mining methods: CART, ANN and SVM. The settings and types of each method in Matlab are as follows:

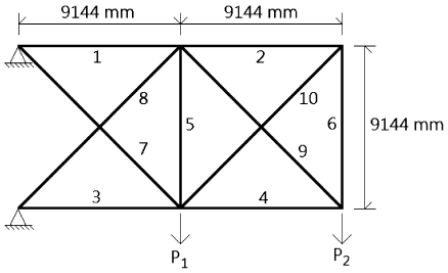

Figure 1. Ten-bar truss geometry and loading.

1) CART: Gini Index as Impurity Measurement
2) ANN: 3 Hidden Layers, 5 Neurons @ Hidden Layer, Scaled Conjugate Gradient Method (Classification), Levenberg-Marquardt Method (Regression), Log-Sigmoid Transfer Function
3) SVM: Least Square Support Vector Machine (LSSVM), RBF Kernel, Regularization Parameter Value (γ) = 0 to 1, Kernel Parameter Value (σ) = e^{-10} to e^{10}

There are eighteen classification models and eighteen regression models created during this process forming a total of thirty-six surrogate models. The main findings of this study is that CART performed better than ANN and SVM in both classification and regression. SVM method is always the worst in both cases. Table. 1 shows the accuracy of all models in classification and regression.

TABLE 1. ACCURACY OF ALL THE THREE MODELS IN CLASSIFICATION AND REGRESSION

Data Mining Method	The Accuracy and Performance of the Surrogate Models	
	Classification Model Accuracy	Regression Model MAPE (Error)
CART	93.86%	39.24%
ANN	92.89%	40.83%
SVM	87.03%	72.63%

REFERENCES

[1] K. Deb, *Multi-objective optimization using evolutionary algorithms.* New York: Wiley, 2001.

[2] R. Yang and L. Gu, "Experience with approximate reliability-based optimization methods," *Structural and Multidisciplinary Optimization*, vol. 26, pp.152–159, 2004.

[3] D. Steenackers et al., "Reliability-based design optimization of computation-intensive models making use of response surface models," *Quality and Reliability Engineering International*, vol. 27, no.4, pp.555–568, 2011.

[4] K. W. Liao and C. Ha, "Application of reliability-based optimization to earth-moving machine: hydraulic cylinder components design process," *Structural and Multidisciplinary Optimization*, vol. 36, pp.523–536, 2008.