Expert Systems with Applications 42 (2015) 2703–2709

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

# Effect of error metrics on optimum weight factor selection for ensemble of metamodels

Erdem Acar *

*Department of Mechanical Engineering, TOBB University of Economics and Technology, Sogutozu, Ankara 06560, Turkey*

ARTICLE INFO

*Article history:*
Available online 20 November 2014

*Keywords:*
Ensemble
Error metrics
Metamodeling
Surrogate modeling

ABSTRACT

Optimization of complex engineering systems is performed using computationally expensive high fidelity computer simulations (e.g., finite element analysis). During optimization these high-fidelity simulations are performed many times, so the computational cost becomes excessive. To alleviate the computational burden, metamodels are used to mimic the behavior of these computationally expensive simulations. The prediction capability of metamodeling can be improved by combining various types of models in the form of a weighted average ensemble. The contribution of each models is usually determined such that the root mean square cross validation error (RMSE-CV) is minimized in an aim to minimize the actual root mean square error (RMSE). However, for some applications, other error metrics such as the maximum absolute error (MAXE) may be the error metric of interest. It can be argued, intuitively, that when MAXE is more important than RMSE, the weight factors in ensemble should be determined by minimizing the maximum absolute cross validation error (MAXE-CV). Interestingly, it is found that the ensemble model based on MAXE-CV minimization is less accurate than the ensemble model based on RMSE-CV minimization even if the MAXE is the metric of interest. The reason is found to be that MAXE-CV is mostly related with the geography of the DOE rather than the prediction ability of metamodels.



## 1. Introduction

Optimization of complex engineering systems is performed using computationally expensive high fidelity computer simulations (e.g., finite element analysis). During optimization these high-fidelity simulations are performed many times, so the computational cost becomes excessive. To alleviate the computational burden, metamodels are used to mimic the behavior of these computationally expensive simulations.

There exists a vast of metamodeling methods developed in literature. The commonly used metamodel types include but not limited to the polynomial response surface approximations, PRS (Box, Hunter, & Hunter, 1978; Myers & Montgomery, 2002), Kriging, KR (Sacks, Welch, Mitchell, & Wynn, 1989; Simpson, Mauery, Korte, & Mistree, 2001), radial basis functions, RBF (Buhmann, 2003; Dyn, Levin, & Rippa, 1986), Gaussian process, GP (MacKay, 1998; Rasmussen & Williams, 2006), neural networks (Bishop, 1995; Smith, 1993), and support vector regression, SVR (Clarke, Griebsch, & Simpson, 2005; Gunn, 1997). A good review of metamodeling methods can be found in Queipo et al. (2005), Wang and Shan (2007) and Forrester and Keane (2009).

Even though most research on metamodels focus on determining the most accurate metamodel for the problem at hand, there exist other studies that focus on merging multiple metamodels into a weighted average ensemble model (Acar & Rais-Rohani, 2009; Acar, 2010; Goel, Haftka, Shyy, & Queipo, 2007; Hamza & Saitou, 2012; Muller & Piche, 2011; Sanchez, Pintos, & Queipo, 2008; Zhou, Ma, Tub, & Feng, 2012). It is observed in these studies that the generated ensemble model has a better prediction ability than the individual metamodels that contribute to the ensemble.

The weight factors in an ensemble are chosen such that an error metric is optimized. The error metric can be a local error metric (Acar, 2010; Sanchez et al., 2008) or a global error metric (Acar & Rais-Rohani, 2009; Goel et al., 2007; Hamza & Saitou, 2012; Muller & Piche, 2011; Zhou et al., 2012). In this paper, we consider global error metrics. The most popular error metric used for selecting the weight factors in an ensemble is the root mean square cross validation error (RMSE-CV). Selecting the weight factors based on RMSE-CV aims at constructing the ensemble such that the mean square error over design space is minimized. However, for some applications, other error metrics may be of interest. For instance, in design of safety critical components, minimization of MAXE may be more important than minimization of RMSE. For these problems, the weight factor selection based on RMSE-CV minimization may not be appropriate and weight factor selection should

* Tel.: +90 312 292 4257; fax: +90 312 292 4091.
*E-mail address:* acar@etu.edu.tr

http://dx.doi.org/10.1016/j.eswa.2014.11.020
0957-4174/

**Nomenclature**

| | |
|---|---|
| **C** | mean square error matrix |
| $E_i$ | root mean square cross validation error of the *i*th metamodel |
| $EN_{MAXE}$ | ensemble model obtained through MAXE-CV minimization |
| $EN_{RMSE}$ | ensemble model obtained through RMSE-CV minimization |
| KR0, KR1 | Kriging models obtained by using zeroth-order and first-order trend models, respectively |
| MAXE | maximum absolute error (computed at a large number of test points) |
| MAXE-CV | maximum absolute cross validation error |
| $N_M$ | number of models of the ensemble |
| PRS2 | polynomial response surface of the second-order |
| RBF | radial basis functions |
| RMSE | root mean square error (computed at a large number of test points) |
| RMSE-CV | root mean square cross validation error (computed at training points) |
| $w_i$ | contribution of the *i*th model in the ensemble |
| $\hat{y}_{ens}$ | prediction of response obtained from the ensemble model |
| $\hat{y}_i$ | prediction of response obtained from the *i*th model of the ensemble |

be revised. The main objective of this paper is to explore the effects of error metrics on weight factor selection in an ensemble of metamodels.

The paper is organized as follows. The formulation for weighted average ensemble along with determination of the contribution of metamodels is explained in the next section. Section 3 presents the error metrics considered in this study. The mathematical and engineering example problems used in this study is presented in Section 4. Details of ensemble model generation is provided in Section 5. The results obtained from the example problems are discussed in Section 6. Finally, the paper culminates with a list of important conclusions presented in Section 7.

## 2. Ensemble of metamodels

In metamodel based optimization studies, first many different types of metamodels are constructed, and then the most accurate metamodel is selected to be used further whereas the other constructed metamodels are discarded. There are two major drawbacks of this practice. First, information obtained through building various different metamodels is not fully acknowledged. Second, the accuracies of the constructed metamodels depend on the current training data set, and a different metamodel than the selected one may become the most accurate with a new data set. These shortcomings can be addressed by using ensemble of metamodels.

Suppose that there exists a data set $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ that consists of $N$ observations of a $D$-dimensional variable $\mathbf{x}$, together with the corresponding observations of the response of interest $\{y_1, y_2, \ldots, y_N\}$. The predictions of the response corresponding to different types of stand-alone metamodels can be combined in the form of an ensemble method. The most commonly used ensemble method is the weighted average ensemble, where various different metamodels are combined as

$$\hat{y}_{ens}(\mathbf{x}) = \sum_{i=1}^{N_M} w_i \hat{y}_i(\mathbf{x}) \tag{1}$$

where $\hat{y}_{ens}$ is the response prediction obtained from the ensemble model, $N_M$ is the number of different models in the ensemble, $w_i$ is the contribution (or weight factor) of the *i*th model in the ensemble and $\hat{y}_i$ is the response prediction obtained from the *i*th model of the ensemble. To have an unbiased response estimation, the following equation must be satisfied by the weight factors:

$$\sum_{i=1}^{N_M} w_i = 1 \tag{2}$$

The weight factors, $w_i$, for the metamodels are usually chosen such that the root mean square cross validation error (RMSE-CV) is minimized in an aim to minimize the actual root mean square error (RMSE). However, for some applications, minimization of other error metrics may be more important. In that case, one may intuitively argue that the cross validation versions of these metrics should be minimized while selecting the weight factors. In this paper, the validity of this argument is questioned.

## 3. Error metrics

Prediction accuracy of metamodels can be measured using different metrics, and these metrics can be used for multiple purposes including (i) assessing the goodness of the approximation to be used for analysis and optimization studies, (ii) identifying the regions of high uncertainty in design space and performing additional sampling (adaptive sampling) at these regions, (iii) selecting the best metamodel among alternative models, and (iv) determining the weight factors of stand-alone metamodels in an ensemble of metamodels (Acar & Rais-Rohani, 2009; Goel et al., 2007). The most commonly used metrics are (i) root mean square error (RMSE), (ii) mean absolute error (MAE), (iii) coefficient of multiple determination ($R^2$), (iv) maximum absolute error (MAXE). The relative, normalized or adjusted versions of these metrics are also frequently used.

The most popular error metric is the RMSE, which measures the square root of the average value of the squared deviations of the predictions from the observed values. RMSE can be computed from

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n_v} (y_i - \hat{y}_i)^2}{n_v}} \tag{3}$$

where $n_v$ is the number of out-of-sample validation points. In design of safety critical components, MAXE may be more important. MAXE measures the absolute value of the maximum deviation of the predictions from the observed values. MAXE can be computed from

$$MAXE = \max_{n_v} |y_i - \hat{y}_i| \tag{4}$$

## 4. Example problems

Overall nine example problems are considered. The first seven example problems are well-known mathematical benchmark problems used in optimization studies. These are followed by two structural mechanics problems.

### 4.1. Mathematical benchmark problems

#### 4.1.1. Branin–Hoo function

$$y(x_1, x_2) = \left(x_2 - \frac{5.1x_1^2}{4\pi^2} + \frac{5x_1}{\pi} - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos(x_1) + 10 \tag{5}$$