NU Science Journal 2013; 10(1): 1 - 9

A Comparison of Artificial Neural Network and Kriging Model for Predicting the Deterministic Output Response

Anamai Na-udom[1*] and Jaratsri Rungrattanaubol[2]

[1]Department of Mathematics, Faculty of Science, Naresuan University, Phitsanulok, 65000, Thailand.
[2]Department of Computer Science and Information Technology, Faculty of Science, Naresuan University, Phitsanulok, 65000, Thailand.
*Corresponding author. E-mail: anamain@nu.ac.th

ABSTRACT

The aim of this paper is to compare the prediction accuracy between two popular approximation model methods namely, artificial neural network (ANN) and Kriging model for modeling the output response from computer simulated experiments (CSE). The natures of CSE are time-consuming and computationally expensive to run. Hence, many efforts have focused on developing inexpensive and reliable surrogate models to replace the CSE. Kriging model along with Latin hypercube designs (LHD) have been widely used for developing an accurate surrogate model in the context of CSE. The performance of Kriging model is based on the estimation of the unknown parameters. The most popular method to estimate these parameters is the maximum likelihood estimate (MLE) method. The MLE method is normally time consuming and fails to obtain the best set of parameters due to numerical instability and ill-conditioning of the model structure. Due to the popularity of ANN in modeling high and complex problem, this paper presents an application of ANN in the context of CSE and the comparison with Kriging model is employed. The results indicate that ANN performs well in terms of prediction accuracy and can be used as an alternative of Kriging model in some features of problem under this study.

Keywords: computer simulated experiments, artificial neural network, Kriging model, optimal Latin hypercube designs.

INTRODUCTION

Nowadays computer simulated experiments (CSE) have been extensively used to investigate the complex physical phenomena, especially when the physical experiments are not feasible due to cost and time constraints or the limitation of the experimental materials. Some examples of CSE are the use of reservoir simulator to predict ultimate recovery of oil (Cheong, 2005), use of finite element codes to predict behavior of metal structure under stress and bio-mechanical models to predict protein in sheep wool (Koehler and Owen, 1996), and so on. Generally, these computer codes comprise of a system of complex differential equations, which, for given setting of input variable conditions (X), can be solved numerically to obtain the value of output response (Y). For example, a reservoir simulator, for

given values of the field characteristic (like Gross rock volume, porosity, gas cap etc.), can be run to quantify the ultimate recovery of oil (Y) from the field (Cheong, 2005).

Running of computer codes at various settings of input variables to study output response is referred to as computer simulated experiment.Setting of input variables at which code is operated is referred to as a design (X) with each setting being referred to as a run. CSE is deterministic in nature; hence identical settings of input variables always produce an identical set of output response and typically the process of CSE is not known a priori. Therefore, space filling designs that aim to spread design points over a region of interest are very useful (Bates *et al*., 1996). Often, CSEs are computationally expensive to run and may require a large number of input factors. Bates et al. (1996) stated that the combination of system decomposition experimental design and developing of a surrogate model need to be performed in order to decrease the complexity of the problem, therefore the approximation model could be effective. In the view of complexities of the system, it is often more desirable to create cheaper surrogate models for the computer codes (Simpson *et al*., 2001), capable of predicting output with high accuracy. Hence many efforts have gone to the development of accurate surrogate models based on handful of runs.

There are several statistical models have been proposed for use in CSE. Sacks et al. (1989) and Guinta and Watson (1998) compared the prediction accuracy between RSM and Kriging model. Simpson et al. (2001) investigated the prediction accuracy of various statistical models. According to the results presented in the published work, there is no certain conclusion on which statistical model is best for any specific problems. However, Kriging model seems to be the most popular method in modeling response from CSE due to its interpolation property which is completely accurate when the untried input is nearby the design point (Welch *et al*., 1992). The drawback of Kriging model is that the estimation of all unknown parameters is so complex and sometimes fails to obtain the best set of parameters and hence the prediction accuracy of Kriging model deteriorates. Some researchers focus on enhancement of the parameter estimation method for estimating unknown parameters in Kriging model. For instance, Welch et al. (1992) proposed an efficient algorithm to estimate the parameters using the maximum likelihood method.

While Kriging model has received a wide attention in developing the surrogate models, there are many other types of statistical models such as Multivariate adaptive regression splines (MARS), Radial basis function (RBF) and Artificial neural network (ANN) have been adopted to use in the context of CSE. De Veaux et al. (1993) did a comparison between MARS and ANN using various complexity test problems. The results indicate that MARS models perform better than ANN in some cases especially when the dimension of the problem is small whereas ANN is simple to implement and its performance can be improved by adapting its architecture. Hence the aim of this paper is to compare the prediction accuracy between the two popular modeling methods namely Kriging and ANN. For a specific approximation model, the optimal Latin hypercube designs (OLHD) are used. The prediction accuracy of each model is implemented by using root mean square error (RMSE).For the sake of completeness, the percentage improvement

over ANN (PI) is also presented along with various nonlinear features of test problems.

THEORY AND RELATED WORKS

KRIGING MODEL

The first approach of developing a surrogate model for computer simulated experiments called Kriging model, was proposed by Sacks et al. (1989). This method is based on the idea that the response y can be modeled as a polynomial function of input variables and whatever is left can be regarded as a realization of stochastic process, *Z(x)*,with mean zero and some form of correlation function. Typically y is written as,

$$y = \sum_{j=1}^{k} \beta_j f_j(x) + Z(x) \tag{1}$$

, where k is the number of terms used in the model.

In order to make the model simple, in most of the practical problems, the polynomial function part in (1) is taken as a constant (Welch *et al*.,1992, Sacks *et al*., 1989),

$$y = \beta + Z(x) \tag{2}$$

Moreover the results obtained from the empirical studies revealed that there is no effect in terms of prediction capability (Sacks *et al*., 1989). The second part on the right of equation (1), *Z(x)* is considered as a Gaussian correlation function (Morris and Mitchell, 1995, Welch *et al*., 1992, Sacks *et al*., 1989), the most frequently used form can be written as,

$$R(X_{i.}, X_{j.}) = \prod_{j=1}^{d} \exp(-\theta_j \left| X_{i.} - X_{j.} \right|^{p_j}) \tag{3}$$

, where d is the number of input variables, $0 \le p_j \le 2$ and $\theta_j > 0$.

Normally Kriging model is fitted using the idea of generalized least squares method, and the problem of estimating all unknown parameters reduces to that of estimating the parameters of the correlation function which can be done by the method of maximum likelihood estimation (MLE) (Welch *et al*., 1992, Sacks *et al*., 1989). The maximum likelihood estimators can be obtained by maximizing the log likelihood function,

$$l(\beta, \sigma^2, \theta, p) = -\frac{1}{2}[n \ln \sigma^2 + \ln |R| + (y - 1\beta)^T R^{-1} (y - 1\beta) / \sigma^2] \tag{4}$$

Given the correlation parameters θ and p in (3), the generalized least estimate of β is $\hat{\beta} = (1^T R^{-1} 1)^{-1} 1^T R^{-1} y$, and the MLE of σ^2 is

$$\hat{\sigma}^2 = 1/n (y - 1\hat{\beta})^T R^{-1} (y - 1\hat{\beta}) \tag{5}$$

Substituting $\hat{\beta}$ and $\hat{\sigma}^2$ into the likelihood function in equation (4), the problem is to numerically maximize

$$-\frac{1}{2}(n\ln\hat{\sigma}^2+\ln|R|) \quad (6)$$

, which is a function of only the correlation parameters and the data from the design used in data collection step.

After all unknown parameters are obtained, the next step is to build a predictor, $\hat{y}(x)$, of $y(x)$ to act as a surrogate model for the complex computer simulation code. The best linear unbiased predictor (BLUP) at an untried input x is

$$\hat{y}(x)=\hat{\beta}+r^T(x)R^{-1}(y-1\hat{\beta}) \quad (7)$$

, where *r(x)* is the vector of correlation function between error ($Z(x)$) at n design runs and untried input variables (x).

Kriging model has received wide attention in many applications of computer simulated experiments due to its interpolation property. Simpson et al. (2001) reported that Kriging is very flexible because of the wide range of choices of the correlation functions to be chosen. However, the estimation of correlation parameters by maximizing likelihood function in equation (6) is not straightforward as the estimated parameters are very sensitive to initial values and selected design points (Welch *et al*., 1992).

ARTIFICIAL NEURAL NETWORK

Artificial neural network (ANN) is commonly used in sophisticated and complex problems (Bozdogan, 2003). Unlike a usual statistical approximation model, ANN does not require any assumptions of the model. This benefits ANN to be simple and easy to use in many applications such as science, engineering and health science (Ripley, 1993). The inspiration for neural networks was the recognition that complex learning systems in animal brains consisted of closely interconnected sets of neurons. A particular neural may be relatively simple in structure but dense networks of interconnected neurons could perform complex learning tasks such as pattern recognitions and approximation models. ANN consists of input (*p*), a data set, which is combined through a combination function such as summation ($\sum$) then pass such information into an activation function (*f*) to produce an output response (*y*) and *b* is a bias as shown in Figure 1.

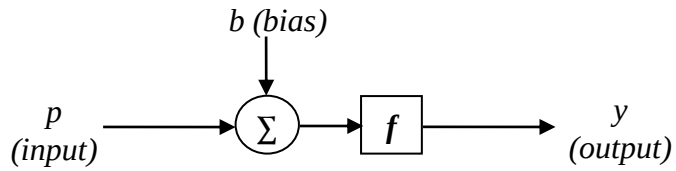

Figure 1 A simple layout of ANN

The summary of ANN process can be rewrite as

$$y = f(wp + b) \tag{8}$$

, where w is the weight of each input variable.

Typically ANN is formed by multiple nodes (in this case symbolized by $\sum$) as shown in Figure 2. An activation function can be the same or different functions, such as linear, sigmoid and symmetrical hard limit.

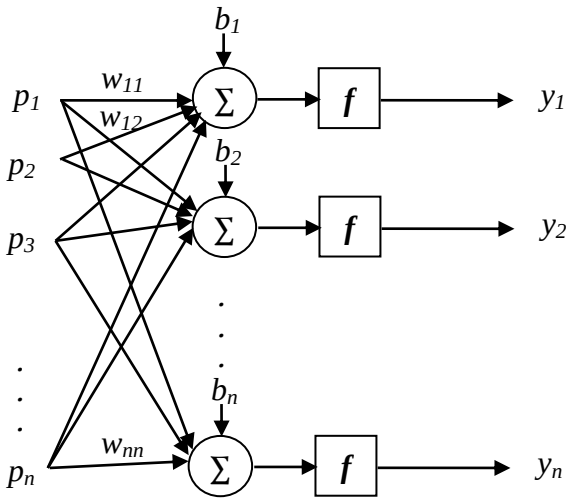

Figure 2 ANN with multiple nodes

The entire process can be rewritten as,

$$\begin{aligned} y^1 &= f^1(w^1 p + b^1) \\ y^2 &= f^2(w^2 p + b^2) \\ &\vdots \\ y^N &= f^N(w^N p + b^N) \end{aligned} \tag{9}$$

, where N is the number of nodes.

RESEARCH METHODS

As mentioned before that the aim of this study is to compare the prediction accuracy of Kriging and ANN models with respect to the prediction accuracy for any untried input variables. The prediction accuracy is implemented by using various test problems selected from the literature (Allen *et al*., 2003, Ye *et al*., 2000). There are four different test problems have been used to compare the prediction accuracy between Kriging and ANN models. The test problems are classified into two groups with respect to the feature of the response, namely non-

complex (smooth response) and complex (highly nonlinear and sharp change in some area of the response). The test problems consist of two-dimension problems, seven-dimension problem and ten-dimension problem. The 121 grid points were used as test points in order to ensure the coverage of design space for two-dimension test problems. In the case of seven and ten dimensional test problems, the 500 random test points were used to validate the prediction accuracy of those three statistical models. The details of all test problems are given in Table 1.

Table 1 Problems with main and interaction effect

Problem	*d*	Function
RM2	2	$f(x_1,x_2)=0.5(x_1-x_2^{-1})-x_1-5x_2^{-1}$ $1\le x_1,x_2\le 100$
Branin function	2	$f(x_1,x_2)=\left(x_2-5.1\left(\frac{x_1}{2\pi}\right)^2+x_1\left(\frac{5}{\pi}\right)-6\right)^2+10\left(1-\frac{1}{8\pi}\right)\cos(x_1)+10$ $-5\le x_1\le 10, 0\le x_2\le 15$
Cyclone model	7	$f(x_1,x_2,...,x_7)=174.42\left(\frac{x_1}{x_5}\right)\left(\frac{x_3}{x_2-x_1}\right)^{0.85}\times\sqrt{\frac{1-2.62\left\{1-0.36(x_4/x_2)^{-0.56}\right\}^{3/2}(x_4/x_2)^{1.16}}{x_6x_7}}$ $;0.09\le x_1\le 0.11, 0.27\le x_2\le 0.33, 0.09\le x_3\le 0.11,$ $0.09\le x_4\le 0.11, 1.35\le x_5\le 1.65, 14.4\le x_6\le 17.6,$ $0.675\le x_7\le 0.825$
10D Function	10	$y=\sum_{i=1}^{10}\left[\frac{3}{10}+\sin\left(\frac{16}{15}x_i-1\right)+\sin^2\left(\frac{16}{15}x_i-1\right)\right]$ $-1\le x_i\le 1, i=1,2,...,10$

In order to implement the accuracy of the models, the OLHD designs are generated by the enhanced of simulated annealing algorithm (SA) under ϕ_p optimality criteria (Morris and Mitchell, 1995). For two-dimension problems, the designs with 9 runs were used, the designs with 99 runs were used for seven-dimension problem, and the designs with 201 runs were used for ten-dimension problem.

From all test problems stated above, we fit Kriging and ANN models by using the DACE and ANN toolbox in MATLAB. After Kriging and ANN models for all cases is fitted, the prediction accuracy is implemented by using the Root mean squared error (RMSE), computed as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{k} (y_i - \hat{y}_i)^2}{k}} \quad (10)$$

, where k is the number of random test points, y_i is the actual response of the i^{th} test point and $\hat{y}_i$ is the predicted response from Kriging and ANN models for the i^{th} test point. Further, the percentage improvement over ANN, defined as

$$PI = \frac{RMSE(ANN) - RMSE(Kriging)}{RMSE(ANN)} \times 100\% \quad (11)$$

, is also computed in order to compare the performance of Kriging and ANN models.

RESULTS AND DISCUSSION

In this section the Kriging and ANN models are compared on the basis of RMSE and PI values. For each dimension and test problem, the 10 different OLHD designs were generated. The average of RMSE and PI values obtained from 10 OLHD designs are presented in Table 2.

As can be seen from Table 2, RMSE values generated from Kringing are considerably larger than that of ANN, especially for 10D function test problem. This indicates that using ANN as an approximation model leads to higher prediction accuracy. Table 2 also provides the scaled measurement of error, called percentage improvement over ANN (PI). This scaled measurement benefits in the ignorance of the differences in error magnitude of different test problems. The PI values also confirm that ANN is superior to Kringing as PI is approximately 52.0% improvement for RM2 test problem, 16.25% for Cyclone model and 61.03% for 10D function. In contrast, Kriging performs much better than ANN for Branin function (PI = 66.0%). This indicates that ANN fails to capture the features of the sharp change in some areas of the problem. Therefore, the more complex structure of ANN like multi layer perceptron could be performed to conduct a surrogate model.

Table 2 Comparison of RMSE from all test problems

Test Problem	RMSE	
	Kring	ANN
RM2	7.812	***5.138***
PI	-52.00	
Branin function	***29.390***	86.412
PI	66.00	
Cyclone	0.042	***0.016***
PI	-16.25	
10D	0.248	***0.154***
PI	-61.03	

CONCLUSIONS

As presented in the results and discussion section, ANN performs well in terms of prediction accuracy and can replace Kriging in some cases of problems. The advantage of ANN is that it is free of assumptions and hence a model adequacy checking is not required. In the case that ANN could not perform well, there is probably a need of more complex architecture for ANN to conduct the better-performance approximation model. Our empirical studies are limited to specific test problems. Hence, larger dimensional problems could be future investigated and studied to observe the additional conclusions. Furthermore, it should be noted that a success of Kriging and ANN modeling methods normally depend on an underlying design that is used to develop a surrogate model. From this empirical study, it indicates that an optimal orthogonal LHD is an appropriate design choice for both ANN and Kriging modeling methods. Hence, selecting an optimal LHD critically affects the implementation of the surrogate model.

REFERENCES

Allen, T.T., Bemshteyn, M.A. and Kabiri-Bamoradian, K. (2003). Constructing Meta-Models for Computer experiments. *Journal of Quality Technology*, *35*, 264-274.

Bates, R.A., Buck, R.J., Riccomagno, E.and Wynn, H. P. (1996). Experimental design and observation for large systems. *Journal of the royal statistical society, Series B, 58*, 77-94.

Box, G.E.P., Hunter, W.G. and Hunter, J.S. (2005) Statistics for Experimenters, *John Wiley & Sons*, New York.

Bozdogan, H. (2003). Statistical Data Mining and Knowledge discovery, *Chapman & Hall/CRC*, New York.

Cheong, Y.P. (2005). Experimental design and analysis methods in reservoir uncertainty studies, *Doctoral dissertation, PH.D. Thesis*. Curtin University of Technology.

De Veaux, R.D., Psichogios, D.C., Ungar, L.H. (1993). A comparison of two nonparametric estimation schemes: MARS and neural networks. *Computers Chemical Engineering*, *17*(8), 2-54.

Guinta, A.A. and Watson, L.T.(1998). A comparison of approximation modeling techniques: polynomial versus interpolating models. *Proceeding of the 7th IAA/USAF/NASA/ISSMO symposium on multidisciplinary analysis and optimization, 1*; 392-404.

Koehler, J. and Owen, R.J. (1996). *Handbook of Statistics*. *13*, 261-308.

Morris M.D. and Mitchell T.J. (1995). Exploratory design for computational experiments. *Journal of Statistical planning and inference*. *43*, 381-402.

Mullur, A.A. and Messac, A. (2006). Metamodeling using Extended Radial Basis Functions: A Comparative Approach. *Engineering with Computers*, *21*, 203-217.

Ripley, B.D., Statistical aspects of neural networks. In: Barndoff-Nielsen, O.E., Jensen, J.L., Kendall, W.S., editors. Networks and chaos-statistical and probabilistic aspects, *Chapman & Hall*, New York, 1993, 40-123.

Sacks, J., Welch, W.J., Mitchell, T.J.and Wynn, H.P. (1989). Design and Analysis of Computer Experiments, *Statistical Science, 4*(4), 409-435.

Simpson, T.W., Peplinski, J.D. and Koch, P.N., Allen, J.K. (2001). Metamodels for computer-based engineering design: Survey and recommendations. *Engineering with Computers, 17*(2), 129-150.

Welch, W.J., Buck, R.J., Sacks, J., Wynn, H.P., Mitchell, T.J. and Morris, M.D. (1992). Screening, predicting, and computer experiments, *Technometrics, 34*, 15-25.34, 15-25.

Ye K.Q., Li, W. and Sudjianto, A. (2000). Algorithmic construction of optimal symmetric Latin hypercube designs. *Journal of Statistical planning and Inference, 90*, 145-159.