2.1 Hybrid GMDH-type algorithms and neural networks

Robust Pareto Design of GMDH-type Neural Networks for Systems with Probabilistic Uncertainties

N. Nariman-zadeh[1], F. Kalantary[2], A. Jamali[1], F. Ebrahimi[2]
[1]Department of Mechanical Engineering, University of Guilan
[2]Department of Civil Engineering, University of Guilan
PO Box 3756, Rasht, Iran

nnzadeh@guilan.ac.ir

Abstract. *In this paper, multi-objective evolutionary Pareto optimal design of Group Method of Data Handling (GMDH)-type neural networks have been used for modelling of systems using input-output data sets with probabilistic uncertainties. In this way, A Monte Carlo Simulation (MCS) is first performed to generate input-output data set using some probabilistic distributions. Multi-objective genetic algorithms (GAs) are then used for Pareto optimization of GMDH-type neural networks. The important conflicting objectives of GMDH-type neural networks that are considered in this work are, namely, the mean and variance of both Training Error (TE) and Prediction Error (PE) of such neural networks. It is shown that a robust GMDH-type neural network can be simply obtained using a criterion based on four values of means and variances of both TE and PE. The probabilistic evolved GMDH model exhibits much more robustness to the uncertainties involved within the input-output data sets than that of the deterministic evolved GMDH model. It is shown that GMDH-type neural networks can be successfully applied for input-output data set with uncertainties so that a robust polynomial neural network can be compromisingly obtained from some non-dominated optimum GMDH models.*

Keywords

Multi-objective optimization, Genetic algorithms, GMDH, Pareto, MonteCarlo, Uncertainties.

1 Introduction

System identification techniques are applied in many fields in order to model and predict the behaviors of unknown and/or very complex systems based on given input-output data [1]. In this way, soft-computing methods [2], which concern computation in an imprecise environment, have gained significant attention. Many research efforts have been expended to use of evolutionary methods as effective tools of soft-computing methods for system identification [2-3]. Among these methodologies, Group Method of Data Handling (GMDH) algorithm is a self-organizing approach by which gradually complicated models are generated based on the evaluation of their performances on a set of multi-input-single-output data pairs (i=1, 2, …, M). The GMDH was first developed by Ivakhnenko [4] as a multivariate analysis method for complex systems modelling and identification, which can be used to model complex systems without having specific knowledge of the systems. The main idea of GMDH is to build an analytical function in a feedforward network based on a quadratic node transfer function [5] whose coefficients are obtained using regression technique. In recent years, however, the use of such self-organizing networks leads to successful application of the GMDH-type algorithm in a broad range of areas in engineering, science, and economics [5-7].

There are two main concepts involved within GMDH-type neural networks design, namely, the parametric and the structural identification problems. In this way, some works present a hybrid GA and singular value decomposition (SVD) method to optimally design such polynomial neural networks [6-7]. A code (GEvoM) for the evolutionary design of GMDH type neural network has been developed by some of authors where more details about the code and general description of the technique may be found in the following web site: http://research.guilan.ac.ir/gevom.

With the aid of ever increasing computational power, there have been a great amount of research activities in the field of robust analysis and design devoted to the use of Monte Carlo simulation. In fact, Monte Carlo simulation (MCS) has also been used to verify the results of other methods in RDO or RBDO problems when sufficient number of sampling is adopted. MCS is a direct and simple numerical method but can be computationally expensive. In this method, random samples are generated assuming pre-defined probabilistic distributions for uncertain parameters. The system is then simulated with each of these randomly generated samples and both mean and variance of the performance metrics are then evaluated statistically. In Multi-objective optimization problems (MOPs), there are several objective or cost functions (a vector of objectives) to be optimized (minimized or maximized) simultaneously. These objectives often conflict with each other so that improving one of them will deteriorate another. Therefore, there is no single optimal solution as the best with respect to all the objective functions. Instead, there is a set of optimal solutions, known as Pareto optimal solutions or Pareto front for multi-objective optimization problems. Recently, an encoding method together with a new diversity preserving mechanism for multi-objective evolutionary optimization have been presented in [8] which will also be used in this work.

In this paper, EAs with a specific encoding scheme are used to evolutionary design the generalized structure GMDH-type (GS-GMDH) neural networks [8] in which the connectivity configuration in such networks is not limited to adjacent layers for modelling and prediction of soil shear strength, *Su* , based on 5 input parameters, namely, SPT number (Standard Penetration Test) *N'*, effective overburden stress σ'_0, moisture content percent *W* , *LL* liquid limit, and *PL* plastic limit of fine-graded clay soil. The data used in this study were gathered from the National Iranian Geotechnical Database, which has been set up in the Building and Housing Research Centre (BHRC) [9]. The database has been established under a mandate from the Management and Planning Organization (MPORG), which supervises the professional activities of all of the consultancy firms in Iran. In this way, multi-objective EAs [8] with a new diversity preserving mechanism are applied for Pareto optimization of such GS-GMDH-type neural networks. The important conflicting objectives of the GS-GMDH neural networks that are considered in this work are, namely, mean value of training error (TE) and prediction error (PE) of *Su* in deterministic modelling plus their statistical variances in stochastic modelling. Therefore, there are 2 and 4 objective functions in the multi-objective optimization of GMDH-type neural network modeling of soil shear stress in deterministic and probabilistic approaches, respectively.

2 Stochastic robust analysis

In real experimental engineering practice, there exist a variety of typical sources of uncertainty which have to be considered through robust modelling approach. Those uncertainties include plant parameter variations due to environmental condition, incomplete knowledge of the parameters, inaccuracies involved with measuring and experimental apparatus, and etc. In conventional optimum design of inductive modelling based on input-output data, uncertainties are not addressed and the optimum model of the process is often accomplished deterministically. In fact, it has been shown that optimization without considering uncertainty generally leads to non-optimal and potentially high risk solution. Therefore, it is very desirable to find robust model whose performance variation in the presence of uncertainties is not high. Generally, there exist two approaches addressing the stochastic robustness issue, namely, robust design optimization (RDO) and reliability-based design optimization (RBDO). Both approaches represent non deterministic optimization formulations in which the probabilistic uncertainty is incorporated into the stochastic optimal design process. In RBDO approach, some evaluated reliability metrics subjected to probabilistic constraints are satisfied so that the violation of design requirements is minimized. In RBDO approach, some evaluated reliability metrics subjected to probabilistic constraints are satisfied so that the violation of design requirements is minimized. In RDO approach, the stochastic performance is required to be less sensitive to the random variation induced by uncertain parameters so that the performance degradation from ideal deterministic behaviour is minimized.

Let *X* be a random variable, then the prevailing model for uncertainties in stochastic randomness is the probability density function (PDF), $f_X(x)$ or equivalently by the cumulative distribution function (CDF), $F_X(x)$, where the subscript *X* refers to the random variable. This can be given by

$$F_X(x) = \Pr(X \le x) = \int_{-\infty}^{x} f_X(x)dx \tag{1}$$

where Pr(.) is the probability that an event ($X \le x$) will occur. Some statistical moments such as the first and the second moment, generally known as mean value (also referred to as expected value) denoted by E(*X*) and variance denoted by $\sigma^2(X)$, respectively, are the most important ones. They can also be computed by

$$E(X) = \int_{-\infty}^{\infty} x dF_X(x) = \int_{-\infty}^{\infty} f_X(x) dx \tag{2}$$

and

$$\sigma^2(X) = \int_{-\infty}^{\infty} (x - E(X)) f_X(x) dx \tag{3}$$

In the case of discrete sampling, these equations can be readily represented as

$$E(X) \cong \frac{1}{N} \sum_{i=1}^{N} x_i \tag{4}$$

and

$$\sigma^2(X) \cong \frac{1}{N-1} \sum_{i=1}^{N} (x_i - E(X))^2 \tag{5}$$

where x_i is the i^{th} sample and N is the total number of samples.

3 Modeling Using GMDH Neural Networks

By means of GMDH algorithm a model can be represented as a set of neurons in which different pairs of them in each layer are connected through a quadratic polynomial and thus produce new neurons in the next layer. Such representation can be used in modeling to map inputs to outputs. The formal definition of the identification problem is to find a function $\hat{f}$ that can be approximately used instead of actual one, f in order to predict output $\hat{y}$ for a given input vector $X = (x_1, x_2, x_3, ..., x_n)$ as close as possible to its actual output y. Therefore, given *M* observations of multi-input-single-output data pairs so that

$$y_i = f(x_{i1}, x_{i2}, x_{i3}, ..., x_{in}) \quad (i=1, 2... M), \tag{6}$$

it is now possible to train a GMDH-type neural network to predict the output values $\hat{y}_i$ for any given input vector $X = (x_{i1}, x_{i2}, x_{i3}, ..., x_{in})$, that is

$$\hat{y}_i = \hat{f}(x_{i1}, x_{i2}, x_{i3}, ..., x_{in}) \quad (i=1, 2... M), \tag{7}$$

The problem is now to determine a GMDH-type neural network so that the square of difference between the actual output and the predicted one is minimized, that is

$$\sum_{i=1}^{M} [\hat{f}(x_{i1}, x_{i2}, x_{i3}, ..., x_{in}) - y_i]^2 \to \min , \tag{8}$$

General connection between inputs and output variables can be expressed by a complicated discrete form of the Volterra functional series in the form of

$$y = a_0 + \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j + \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} a_{ijk} x_i x_j x_k + ... , \tag{9}$$

which is known as the Kolmogorov-Gabor polynomial [9], [17-20]. This full form of mathematical description can be represented by a system of partial quadratic polynomials consisting of only two variables (neurons) in the form of

$$\hat{y} = G(x_i, x_j) = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_4 x_i^2 + a_5 x_j^2 . \tag{10}$$

There are two main concepts involved within GMDH-type neural networks design, namely, the parametric and the structural identification problems. In this way, works by some of authors present a hybrid GA and singular value decomposition (SVD) method to optimally design such polynomial neural networks [6-8]. The methodology in these references has been successfully used in this paper to obtain the robust polynomial model of the soil shear strength, *Su*, minimizing the mean values and variances of both training and testing of the geotechnical data table simultaneously. The table consists of 80 input-output data of which 50 are used for training whilst the rest are used for predicting purpose. In the deterministic approach, the mean square errors of training and prediction errors are chosen for the bi-objective Pareto optimization approach of GMDH models using the methodology explained in detail in [8]. Figure 1 depicts the Pareto front of both training and prediction errors of GMDH-type models. Points A and B stands for polynomial models having the least prediction and training errors, respectively. Design point C, however, depicts the tradeoff in such models which can be reasonably chosen as an optimal model compromisingly. The networks' structures together with their pertinent values of TE and PE of deterministic optimal design points A, B, and C are given in Table 1.

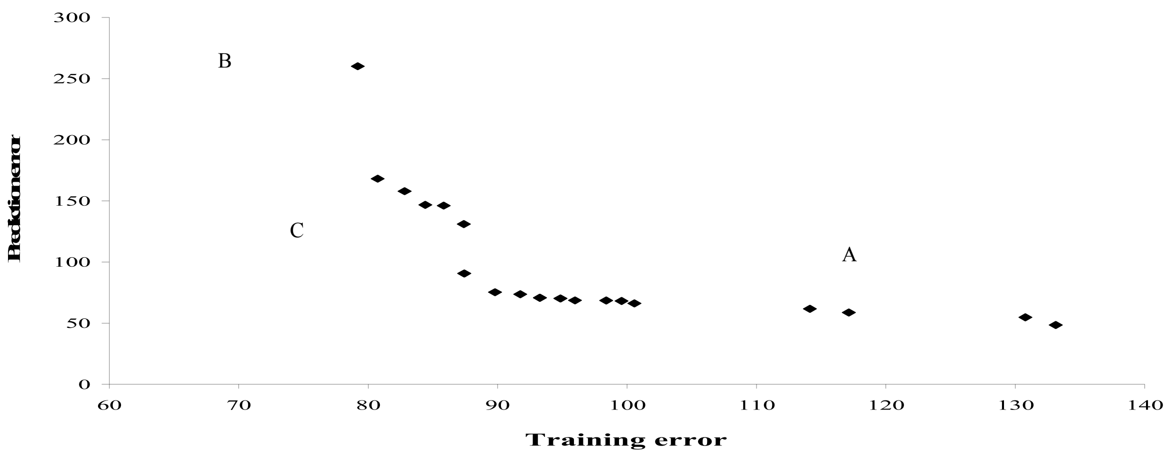

Fig. 1: Pareto front of TE and PE of GMDH models in deterministic design

In probabilistic design approach, a MCS is adopted to evaluate the values of means and variances of both TE and PE of the GMDH-type neural networks. In this way, 100 different data table is generated by MCS with 10 percent Gaussian probabilistic distribution of all six variables in the geotechnical data set. The multi-objective Pareto design of GMDH-type neural networks described in [8] is now used to obtain some non-dominated GMDH models using statistical measures given in equations 4-5. Consequently, 85 different non-dominated polynomial models have been obtained based on 4 objective functions of the mean and variance of both TE and PE. In order to compromise among the solutions obtained, all the values of these objective functions returned by the MCS are normalised. The minimum value of the sum of those normalised values is then simply selected as the robust GMDH model. The performance of this design point denoted by point D is given in Table 1 both for deterministic and probabilistic approaches using MCS for 2500 data sets. It is very evident from this table that the stochastic behaviour of the design point D in terms of mean and variance for both TE and PE is significantly superior to those obtained in deterministic approach (design points A, B, and C) and thus exhibits more robustness in the existence of probabilistic uncertainties. Figure 2 depicts the location of the design point D among other non-dominated design points in the plane of the variances of both TE and PE. It is clear that the selection criterion leads to the design point with almost least variances whose values are given in Table 1.

Tab. 1. Training and prediction errors of design points in both deterministic and probabilistic approaches

		Deterministic Values of TE &PE		Probabilistic measures of TE & PE			
Point	Network's structure	Training Error	Prediction Error	Mean of TE	Mean of PE	Variance of TE	Variance of PE
A	bbaebcacbcaeacee	133.12	48.49	323.76	161.49	174862.64	42019.59
B	bcaebacdbcbbadde	79.20	260.15	73785.2	17844.7	3.8E11	3.3E9
C	bcaebccdbdbcaccd	89.79	75.30	28366.5	709.8	3.7E10	2.6E6
D	abeecddd	132.79	237.59	243 .61	248.03	178.77	1174.283

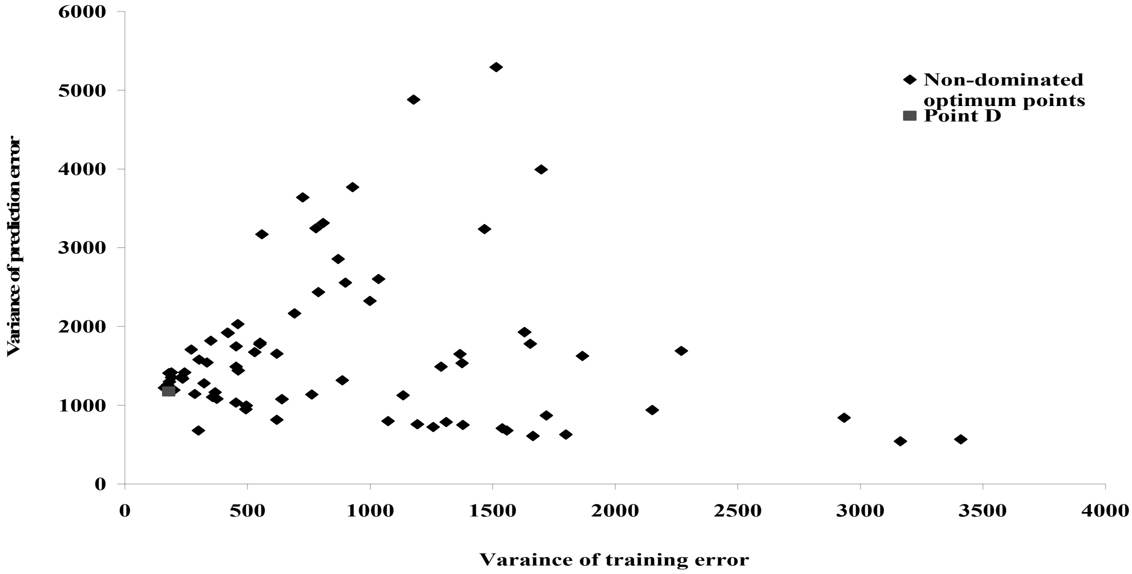

Fig. 2: Non-dominated design points and the selected one point D in the plane of variances of errors

4 Conclusion

A multi-objective genetic algorithm was used to optimally design GMDH-type neural networks from a robustness point of view in a probabilistic approach. The objective functions which often conflict with each other were appropriately defined using some probabilistic metrics. The multi-objective optimization of robust GMDH models led to the discovering some important trade-off among those objective functions. The framework of such hybrid application of multi-objective GAs and Monte Carlo Simulation of this work for the Pareto optimization of robust GMDH neural networks using stochastic objective functions is very promising and can be generally used in the optimum design of GMDH models in real-world complex systems with probabilistic uncertainties

References

[1] Astrom, K. J. and Eykhoff, P., System identification, a survey. Automatica 7-123-62, 1971.
[2] Sanchez, E., Shibata, T. and Zadeh L. A., Genetic Algorithms and Fuzzy Logic Systems. World Scientific, Riveredge, NJ, 1997.
[3] Kristinson, K. and Dumont, G., System identification and control using genetic algorithms. IEEE Trans. On Sys., Man, and Cybern, Vol. 22, No. 5, pp. 1033-1046, 1992.
[4] Ivakhnenko, A. G.,: Polynomial Theory of Complex Systems. *IEEE Trans. Syst. Man & Cybern*, SMC-1, 364-378, 1971.
[5] Farlow, S. J.,: Self-organizing Method in Modeling: GMDH type algorithm. Marcel Dekker Inc., 1984.
[6] Nariman-Zadeh, N., Darvizeh, A., Felezi, M. E. and Gharababaei, H.,: Polynomial modelling of explosive compaction process of metallic powders using GMDH-type neural networks and singular value decomposition. *Modelling and Simulation in Materials Science and Engineering*, Vol. 10, no. 6, pp. 727-744(18), 2002.
[7] Nariman-Zadeh, N., Darvizeh, A. and Ahmad-Zadeh, G. R.,: Hybrid Genetic Design of GMDH-Type Neural Networks Using Singular Value Decomposition for Modelling and Prediction of the Explosive Cutting Process. *Proceedings of the I MECH E Part B Journal of Engineering Manufacture*, Volume: 217, Page: 779 – 790, 2003.
[8] Nariman-zadeh, N. And Jamali,A., : Pareto Genetic Design of GMDH-type Neural Networks for Nonlinear Systems, In: *Proceeding of IWIM 2007*, Prague, Czech Rep., Sep. 2007.
[9] http://www.bhrc.ac.ir/geobank/