A Hybrid Neural Network Model for Noisy Data Regression

Eric W. M. Lee, Chee Peng Lim, Richard K. K. Yuen, and S. M. Lo

Abstract—**A hybrid neural network model, based on the fusion of fuzzy adaptive resonance theory (FA ART) and the general regression neural network (GRNN), is proposed in this paper. Both FA and the GRNN are incremental learning systems and are very fast in network training. The proposed hybrid model, denoted as GRNNFA, is able to retain these advantages and, at the same time, to reduce the computational requirements in calculating and storing information of the kernels. A clustering version of the GRNN is designed with data compression by FA for noise removal. An adaptive gradient-based kernel width optimization algorithm has also been devised. Convergence of the gradient descent algorithm can be accelerated by the geometric incremental growth of the updating factor. A series of experiments with four benchmark datasets have been conducted to assess and compare effectiveness of GRNNFA with other approaches. The GRNNFA model is also employed in a novel application task for predicting the evacuation time of patrons at typical karaoke centers in Hong Kong in the event of fire. The results positively demonstrate the applicability of GRNNFA in noisy data regression problems.**

Index Terms—**Fuzzy adaptive resonance theory (ART), general regression neural network (GRNN), general regression neural network with fuzzy ART clustering (GRNNFA), noisy data regression.**

I. INTRODUCTION

DATA regression is a major research topic in the area of function approximation. Different artificial neural network (ANN) models for data regression including multilayer perceptron (MLP) and radial basis function (RBF) have been developed. In MLP, the regression surface is constructed by nonlinear transformation of the combination of neurons outputs. RBF approximates the underlying function by the combination of nonlinear semi-parametric functions (i.e., kernel functions). However, the number of hidden neurons or kernels of these ANN models has to be pre-determined prior to network training. The general regression neural network (GRNN) model developed by Specht [1] is a powerful regression tool with a dynamic network structure. The network training speed is extremely fast. Due to the simplicity of the network structure and ease of implementation, it has been widely applied to a variety of fields including image processing [2], nonlinear adaptive control [3], machinery fault diagnosis [4], and financial prediction [5]. During the training process, each training sample is recruited as a kernel. Then, a regression surface can be established by using Parzen-window estimators [6] with all the kernel widths assumed to be identical and spherical in shape. The major drawback of the GRNN is the requirement of extensive computer resources for storing and processing all the training samples. Specht [1] proposed a clustering algorithm for the GRNN to reduce the number of kernels and hence computational burden. Instead of recruiting all the training samples as kernels, a clustering procedure is used to compress all training samples into fewer kernels.

Fuzzy adaptive resonance theory (FA), on the other hand, is a powerful unsupervised classifier. The architecture of FA was developed based on adaptive resonance theory (ART) [7]. Fuzzy subsethood [8] is utilized as a similarity measure between input patterns and network prototypes. FA has been proven to be a stable network that solves the stability-plasticity dilemma. It has also been applied to different fields. Brezmes [9] *et al.* employed FA as a classification tool in his electronic nose for determination of fruit quality. Araujo and Almeida [10], [11] used FA for building maps by signals of sensors for navigation of a mobile robot. Cinque *et al.* [12] applied FA with a modified choice function to image segmentation. FA was also used for customer grouping in e-commence by Park [13].

This paper presents a hybrid network, denoted as general regression neural network with fuzzy ART clustering (GRNNFA), that employs FA as a pre-processor for the GRNN to compress training data samples into fewer kernels. Comparing with other popular clustering methods (e.g., the kohonen self-organizing map [14] and fuzzy C-means clustering [15]), FA has been proven to be stable [16], and is extremely fast in learning. During the course of learning, the boundaries (i.e., hyper-rectangles) of the prototypes will be increased monotonically. In GRNNFA, the center and label of a kernel are determined, respectively, by computing centroids of the input and output of the data points being clustered to that kernel upon completion of the clustering process. Noise embedded in the data points inside each hyper-rectangle can possibly cancel each other if they are distributed symmetrically around the centroid of the data samples. For determining the kernel width, an optimization algorithm has been developed. Prior to the use of the algorithm, the kernel widths are initialized by the K-nearest neighboring kernels [17].

The organization of this paper is as follows. Section II describes the basic architectures of the GRNN and FA. Section III illustrates the fusion of the GRNN and FA into GRNNFA. A series of experiments on four benchmark datasets and a novel application of GRNNFA to predicting evacuation time during fire disasters are presented in Section IV. The results are compared and discussed. Section V concludes the GRNNFA and its possible enhancements.

E. W. M. Lee, R. K. K. Yuen, and S. M. Lo are with Department of Building and Construction, City University of Hong Kong, Hong Kong (e-mail: w.m.lee@student.cityu.edu.hk).

C. P. Lim is with School of Electrical and Electronic Engineering, University of Science Malaysia, Penang, Malaysia.

II. ARCHITECTURES OF THE GRNN AND FA

A. GRNN Architecture

Assuming that the underlying scalar function to be approximated is $y = f(\mathbf{x})$ where $\mathbf{x} \in \Re^p$ is a multi-dimensional independent variable and $y \in \Re$ is the dependent variable, regression in the GRNN is carried out by the expected conditional mean of y given $\mathbf{x}$ as shown in (1) where $g(\mathbf{x}, y)$ is the Parzen probability density estimator [6] with Gaussian function as defined in (2)

$$E[y|\mathbf{x}] = \frac{\int_{-\infty}^{\infty} y g(\mathbf{x}, y) dy}{\int_{-\infty}^{\infty} g(\mathbf{x}, y) dy} \tag{1}$$

$$g(\mathbf{x}, y) = \frac{1}{n(2\pi)^{p+1/2}}$$
$$\cdot \left\{ \sum_{i=1}^{n} \frac{1}{\sigma_i^{p+1}} \exp\left[-\frac{1}{2} \sum_{k=1}^{p} \left(\frac{x_k - x_{ik}}{\sigma_i} \right)^2 \right] \right.$$
$$\left. \cdot \exp\left[-\frac{1}{2} \left(\frac{y - y_i}{\sigma_i} \right)^2 \right] \right\}. \tag{2}$$

Equation (3) shows the predicted output of the GRNN model. The kernels are assumed to be hyper-spherical in shape but with different radii. The same was used in the modified GRNN model in [18] where

x_j jth component of the input vector;
x_{ik} kth component of the ith kernel position vector;
σ_i width of the ith kernel;
σ_{iy} width of the ith kernel label.

$$\hat{y}(\mathbf{x}) = \frac{\sum_{i=1}^{n} \frac{y_i}{\sigma_i^p} \cdot \exp\left[-\frac{1}{2\sigma_i^2} (\mathbf{x} - \mathbf{x}_i)^T (\mathbf{x} - \mathbf{x}_i) \right]}{\sum_{i=1}^{n} \frac{1}{\sigma_i^p} \exp\left[-\frac{1}{2\sigma_i^2} (\mathbf{x} - \mathbf{x}_i)^T (\mathbf{x} - \mathbf{x}_i) \right]}. \tag{3}$$

The main drawback of the GRNN is the requirement of extensive amount of computational resources for holding the information of the kernels. It also requires substantial computational time for kernel width optimization. Specht [19] addressed the basic concept of inclusion of clustering techniques in the GRNN model. Fig. 1 shows the architecture of the clustering version of the GRNN model. According to [1], (4) and (5) are used where $A_i(k)$ and $B_i(k)$ are incrementally updated each time when the output of the training sample for cluster i is encountered. This clustering version of GRNN is employed in GRNNFA.

$$\hat{y}(\mathbf{x}) = \frac{\sum_{i=1}^{n} \frac{A_i}{\sigma_i^p} \cdot \exp\left[-\frac{1}{2\sigma_i^2} (\mathbf{x} - \mathbf{x}_i)^T (\mathbf{x} - \mathbf{x}_i) \right]}{\sum_{i=1}^{n} \frac{B_i}{\sigma_i^p} \cdot \exp\left[-\frac{1}{2\sigma_i^2} (\mathbf{x} - \mathbf{x}_i)^T (\mathbf{x} - \mathbf{x}_i) \right]} \tag{4}$$

$$\begin{cases} A_i(k) = A_i(k-1) + y_j \\ B_i(k) = B_i(k-1) + 1 \end{cases}. \tag{5}$$

Different clustering approaches for reducing the number of kernels of the GRNN or probabilistic neural network (PNN) model were proposed, e.g., Kohonen's learning vector quantization technique [20], the k-means algorithm [21], and a mixture of Gaussian densities models [22]. These clustering schemes require determination of the number of kernels prior to the clus-
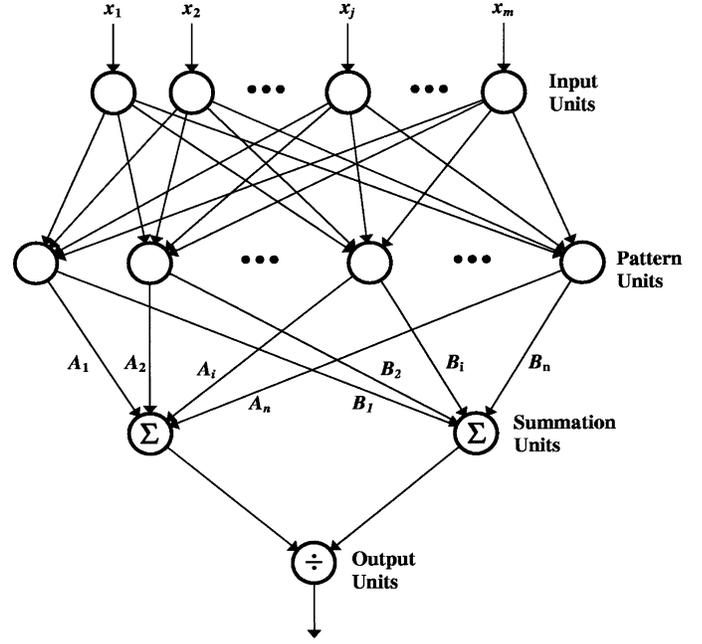
Fig. 1. Architecture of the clustering version of the GRNN model.
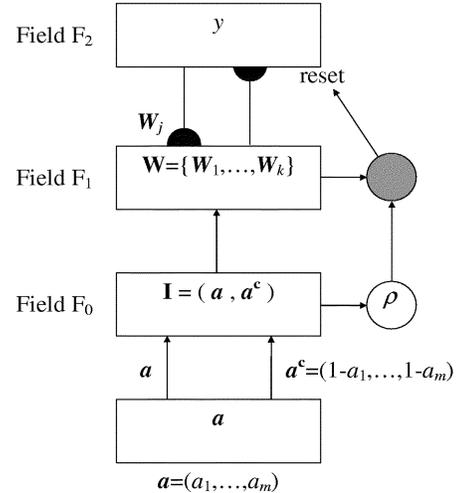
Fig. 2. Architecture of the fuzzy ART model.

tering process. In [16], the possible instability of the clustering techniques by the measurements of Euclidean distances was discussed. On the other hand, FA [16] with guaranteed convergence in network training has been proven to be stable. As a result, FA is used for clustering the training samples incrementally and stably into fewer numbers of kernels in the GRNNFA.

B. FA Architecture

Fig. 2 depicts the FA network architecture. The learning algorithm of FA is as follows. An m-dimensional input pattern, $\mathbf{a} = \{a_1, a_2, \ldots, a_m\}$ where $I_k \in [0, 1]$ and $k = 1, 2, \ldots, m$, is presented to the F_1 layer in a complement-coded format, $\mathbf{I} = (\mathbf{a}.\mathbf{a}^c) \in \Re^{2m}$ where $\mathbf{a}^c = \{a_1^c, a_2^c, \ldots, a_m^c\}$ of which $a_k^c \equiv 1 - a_k$ for $k = 1, 2, \ldots, m$. The complement-coded input pattern will be compared with the prototypes $\mathbf{W}_j \in \Re^{2m}$ in the F_2 layer by using a choice function as shown in (6). The proto-

type that has the highest degree of fuzzy subsethood is selected by (7). Note that $|\mathbf{x}| \equiv \sum_i x_i$, $\mathbf{x} \wedge \mathbf{y} \equiv \min(x_i, y_i)$ and the choice parameter, α, is a small positive number

$$T_j = \frac{|\mathbf{I} \wedge \mathbf{W}_j|}{|\mathbf{W}_j| + \alpha} \qquad (6)$$

$$J = \arg\max_j \{T_j\}. \qquad (7)$$

Resonance is said to occur if the prototype chosen by the choice function (i.e., $j = J$) also satisfies (8) where the vigilance parameter ρ is a threshold predetermined by users

$$\frac{|\mathbf{I} \wedge \mathbf{W}_J|}{|\mathbf{I}|} \geq \rho. \qquad (8)$$

A prototype fulfilling both (7) and (8) is selected as the winning prototype. If a prototype satisfies (7) but fails (8), it will be inhibited. Searching for the winning prototype will be repeated until a prototype that satisfies both (7) and (8) is found. Otherwise, a new prototype will be created to represent the current input pattern. Once the winning prototype is identified, it will be updated according to (9) where β is the learning rate parameter.

$$\mathbf{W}_J^{(\text{new})} = \beta \left(\mathbf{I} \wedge \mathbf{W}_J^{(\text{old})} \right) + (1 - \beta) \mathbf{W}_J^{(\text{old})}. \qquad (9)$$

It is called fast learning when $\beta = 1$. It has been shown in [16] that the maximum size of prototype j is controlled by the vigilance parameter, and $|\mathbf{W}_j|$ is a monotonically increasing quantity to guarantee convergence of the learning procedure.

III. FUSION OF GRNN AND FA

The GRNNFA architecture, as shown in Fig. 3, consists of two modules i.e., FA is employed for training whereas the GRNN is employed for prediction. The basic approach of combining the GRNN and FA models is to first cluster all training samples to fewer numbers of prototypes by FA. Then, the FA prototypes are converted into the GRNN kernels. Since each FA prototype is originally represented by two vertices of the hyper-rectangle, a scheme to obtain the three parameters of each Gaussian kernel (i.e., center, width, and label) from the respective hyper-rectangle is proposed as follows.

1) Kernel Center Estimation: FA is applied to establish prototypes in the input domain according to the distribution of input samples. However, the prototypes created by FA cannot be used directly as the GRNN kernels since these prototypes, in accordance with the FA learning algorithm, only represent the vertices of hyper-rectangles. As a result, the method proposed in [17] for estimating kernel centers of the prototypes created by Fuzzy ARTMAP (FAM) is adopted in GRNNFA. The kernel center $\mathbf{x}_J$ of cluster J is determined by (10) where $\mathbf{a}_i$ ($i = 1, 2, \ldots, N_J$) comprise all N_J samples belonging to J as follows:

$$\mathbf{x}_J = \frac{\sum_{i=1}^{N_J} \mathbf{a}_i}{N_J}. \qquad (10)$$

2) Kernel Label Estimation: A statistical regression model can be developed by taking the expected value over N numbers
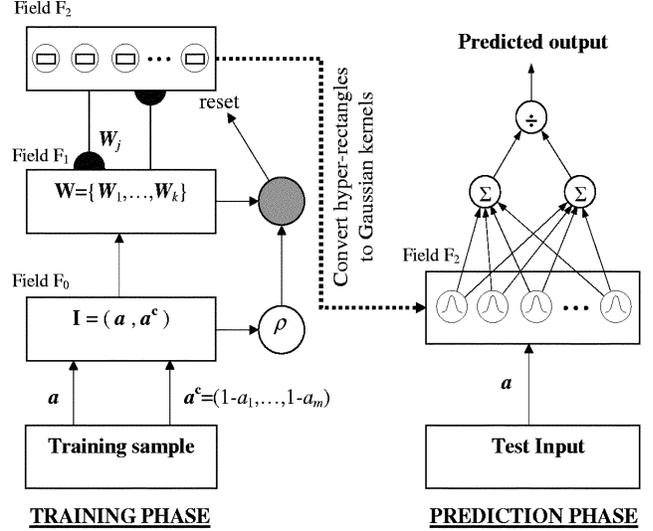
Fig. 3. Architecture of the GRNNFA model.

of kernels shown as follows in (11) where $\hat{y}$, ξ_i, and $P(K_i|\mathbf{x})$ are, respectively, the predicted output, the label of the kernel K_i and the probability of kernel K_i given the input vector $\mathbf{x}$:

$$\hat{y}(\mathbf{x}) = \sum_{i=1}^{N} \xi_i \cdot P(K_i|\mathbf{x}). \qquad (11)$$

By applying Bayesian theory to (11), the following regression model can be obtained where n_i is the number of samples belonging to kernel K_i:

$$\hat{y}(\mathbf{x}) = \frac{\sum_{i=1}^{C} n_i \xi_i p(\mathbf{x}|K_i)}{\sum_{i=1}^{C} n_i p(\mathbf{x}|K_i)}. \qquad (12)$$

The format of (12) is similar to (4). It can be observed that n_i in denominator of (12) (i.e., total number of samples of kernel K_i) is exactly equal to the value of B_i in (5). It is proposed to equate $n_i \xi_i$ in (12) and A_i in (5), i.e.,

$$\xi_i = \frac{A_i}{n_i}. \qquad (13)$$

According to (13), the centroid of the outputs vectors of the clustered input samples should be taken as the label of kernel K_i. The proposed compression scheme to obtain the kernel center and label is depicted in Fig. 4.

3) Kernel Width Estimation: Before applying the kernel width optimization scheme, each kernel width is initialized according to (14) which is similar to the scheme proposed in [17]

$$\sigma_j = \frac{1}{2K} \sum_{k=1}^{K} \|\mathbf{x}_j - \mathbf{x}_k\| \quad j \neq k \quad 1 \leq K \leq N - 1. \qquad (14)$$

The width of kernel j is first set to half of the average distance over K numbers of the nearest neighbors to kernel j. By using appropriate K-nearest neighbors and the vigilance parameter, a network structure with minimum validation error can be initialized.
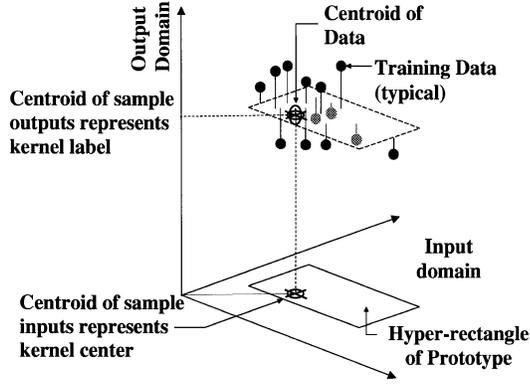
Fig. 4. Data compression scheme for noise removal. Samples are clustered by applying FA on the sample inputs. The centroids of the sample inputs and outputs represent the kernel center and label respectively.

The prediction accuracy can further be improved by applying a gradient-based kernel width optimization scheme. The error of a particular training sample is defined as

$$e_k \equiv \frac{1}{2}\delta_k^2 \tag{15}$$

where $\delta_k = \hat{y}(\mathbf{x}_k) - \tilde{y}_k$, $\hat{y}(\mathbf{x}_k)$ is the GRNNFA prediction and $\tilde{y}_k$ is the target value of the corresponding input vector $\mathbf{x}_k$. The error of all N training samples is

$$E = \sum_{k=1}^{N} e_k. \tag{16}$$

The gradient of the error surface can be obtained by differentiating (16) with respect to the kernel widths, i.e.,

$$\frac{\partial E}{\partial \sigma_j} = \sum_{i=1}^{N} \delta_k \frac{\partial \hat{y}(\mathbf{x}_k)}{\partial \sigma_j}. \tag{17}$$

Equation (18) is obtained by substituting (4) into (17), i.e.,

$$\frac{\partial E}{\partial \sigma_j} = \sum_{i=1}^{N} \delta_k \frac{\partial}{\partial \sigma_j}$$
$$\left\{ \frac{\sum_{i=1}^{n} \frac{A_i}{\sigma_i^p} \cdot \exp\left(-\frac{1}{2\sigma_i}(\mathbf{x}-\mathbf{x}_i)^T(\mathbf{x}-\mathbf{x}_i)\right)}{\sum_{i=1}^{n} \frac{B_i}{\sigma_i^p} \cdot \exp\left(-\frac{1}{2\sigma_i}(\mathbf{x}-\mathbf{x}_i)^T(\mathbf{x}-\mathbf{x}_i)\right)} \right\}. \tag{18}$$

By using $D_{kj}^2 = (\mathbf{x}_k - \mathbf{x}_j)^T(\mathbf{x}_k - \mathbf{x}_j)$, i.e., the squared Euclidean distance between $\mathbf{x}_k$ and $\mathbf{x}_j$, (19) is obtained

$$\frac{\partial E}{\partial \sigma_j} = \sum_{k=1}^{N} \left\{ \frac{\delta_k \hat{y}(\mathbf{x}_k)}{\sigma_j^{p+1}} \times \left[\left(\frac{D_{kj}}{\sigma_j}\right)^2 - p \right] \right.$$
$$\left. \times \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{D_{kj}}{\sigma_j}\right)^2 \right] \times \left[\frac{A_j}{h(\mathbf{x}_k)} - \frac{B_j}{s(\mathbf{x}_k)} \right] \right\}. \tag{19}$$

With the traditional gradient descent algorithm, the kernel width, σ_j is updated according to (20) as follows:

$$\sigma_j^{(\text{new})} = \sigma_j^{(\text{old})} - \alpha^2 \frac{\partial E}{\partial \sigma_j}. \tag{20}$$

To avoid the problem of over-fitting, the training data samples are divided into two i.e., the training and validation sets. The training process is stopped when the validation error reaches

a pre-set threshold, i.e., the fast-stop validation training procedure. One of the deficiencies of the traditional gradient descent algorithm is the slow rate of convergence. As a result, the elastic updating factor, as introduced in [23], is used, i.e.,

$$\alpha_{\text{new}}^2 = \begin{cases} \rho\alpha_{\text{old}}^2, & \text{if } \Delta E < 0 \\ \sigma\alpha_{\text{old}}^2, & \text{if } \Delta E > 0 \end{cases} \tag{21}$$

where ΔE is the change of training error from the last step. The values of ρ and σ are typically set to 1.1 and 0.5 [23]. If the training error increases or decreases, the updating factor will be decreased or increased correspondingly. This scheme can help to determine an appropriate updating factor to force the training error to reduce monotonically. The optimization algorithm is terminated if the validation error has no further improvement over several numbers of epochs. In this study, 2,000 epochs were used. The pseudo code of the GRNNFA training procedure is shown in Appendix A. The executable and instruction files of the GRNNFA model can be obtained at http://home.i-cable.com/ericello/grnnfa/grnnfa.zip.

IV. EXPERIMENTAL STUDIES

The performance of the GRNNFA model was evaluated by using five datasets, with the first four being benchmark problems. The noisy-two-intertwined-spirals problem is first used to view and compare the results reconstructed by the GRNNFA. It is a classification problem using synthetic data with Gaussian noise introduced. The second and third problems, i.e., Ozone and Friedman#1, comprise real and synthetic data (with Gaussian noise introduced), respectively. The fourth problem is a real, astrophysical dataset, i.e., Santa Fe Series-E, which is noisy, discontinuous and nonlinear in nature. The fifth problem is the prediction of evacuation time of patrons to karaoke centers. The data samples were generated by the spatial grid evacuation model (SGEM) [24] with Gaussian noise introduced.

It is expected that the randomization procedures used for generation of noise-corrupted data and selection of data for network training and testing might affect the prediction errors. In order to compare the results predicted by the GRNNFA model with other published results by other models irrespective to the effect of randomization, bootstrapping [25], [26] was employed to quantify the performance indicators statistically. Bootstrapping is a method for estimating statistical variations of a parameter in situations where the underlying sampling distribution of the parameter is unknown or difficult to estimate. It has been proven useful to compute population parameters statistically in problems with small data samples. Fig. 5 shows the procedure used to mitigate the effect of randomization in the experiments.

Instead of a single run, several experiments were conducted and bootstrapping was applied to obtain the prediction errors in form of means and 95% confidence intervals. The principle of bootstrapping for computing mean of a set of data samples is as follows.

Step 1) A set of data $\mathbf{X} = x_1, \ldots, x_n$ is collected. Suppose that n is the size of the sample observed from a completely unspecified probability distribution F, $\hat{\mu}$ is the mean of all the values in $\mathbf{X}$, and N is the number of repeated times of bootstrapping.
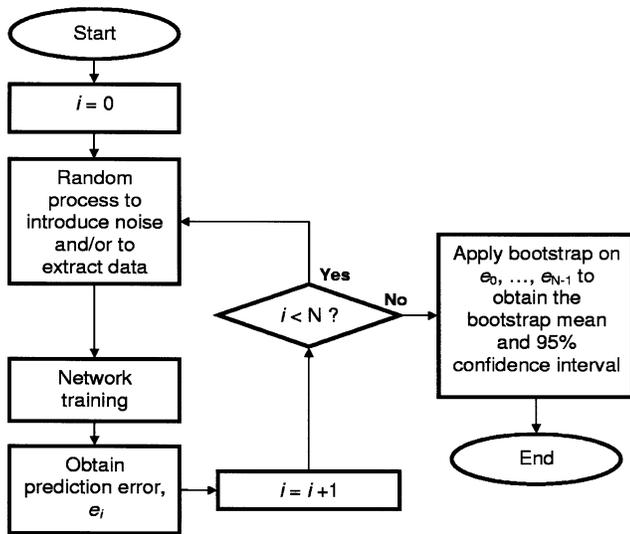
Fig. 5. Bootstrap algorithm for evaluation of mean and 95% confidence interval of the prediction errors.

Fig. 6. Reconstruction of noisy two intertwined spirals by GRNNFA and Gaussian ARTMAP (adapted from [27]).

Step 2) Draw a random sample of n data points independently, with replacement, from $\mathbf{X}$. The new set of data $\mathbf{X}^*$ is the bootstrap sample.

Step 3) The bootstrap sample mean of $\mathbf{X}^*$, $\hat{\mu}$, is calculated.

Steps 2 and 3 are repeated N times to obtain bootstrap estimates of $\hat{\mu}_1^*, \ldots, \hat{\mu}_N^*$.

A. Noisy Two-Intertwined Spirals

This synthetic benchmark problem was designed for noisy data classification. The input domain is a two-dimensional unit square (i.e., $[0,1]^2$) which contains two intertwined spirals. Each spiral consists of 97 isotropic Gaussian distributions centered along the spiral. The standard deviation of each Gaussian is 0.025. The training dataset was created according to the procedure in [27]. Note that classification is a kind of regression but with discrete outputs. The problem is selected because the ability of GRNNFA in reconstructing the two intertwined spirals based on noisy data can be viewed and compared with that from Gaussian ARTMAP, i.e., a supervised ART-based network for noisy classification tasks [27].

After several trials, the vigilance parameter and the K value of the GRNNFA model were set to 0.95 and 2, respectively. A total of twenty experiments were carried out with different random data samples. Note that the output value of each image pixel was obtained from bootstrapping with 2,000 numbers of re-samplings. They were discretized, by setting demarcation at the value of 0.5, to either 0 or 1. Fig. 6 shows the reconstructed images of the two intertwined spirals. It can be clearly seen that the reconstructed images from GRNNFA are more lucid than those from Gaussian ARTMAP (adapted from [27]) with different numbers of training samples.

B. Ozone

This dataset was obtained from University of California at Berkeley (ftp://ftp.stat.berkeley.edu/pub/users/breiman). It has 330 samples with eight inputs and one output. The input sam-ples comprised meteorological information such as humidity and temperature. The target output is the maximum daily ozone at a location in the Los Angeles basin. In accordance with [28], 250 samples were randomly selected from the dataset for network training of which 125 samples were used for validation. The remaining 80 samples were used for network testing.
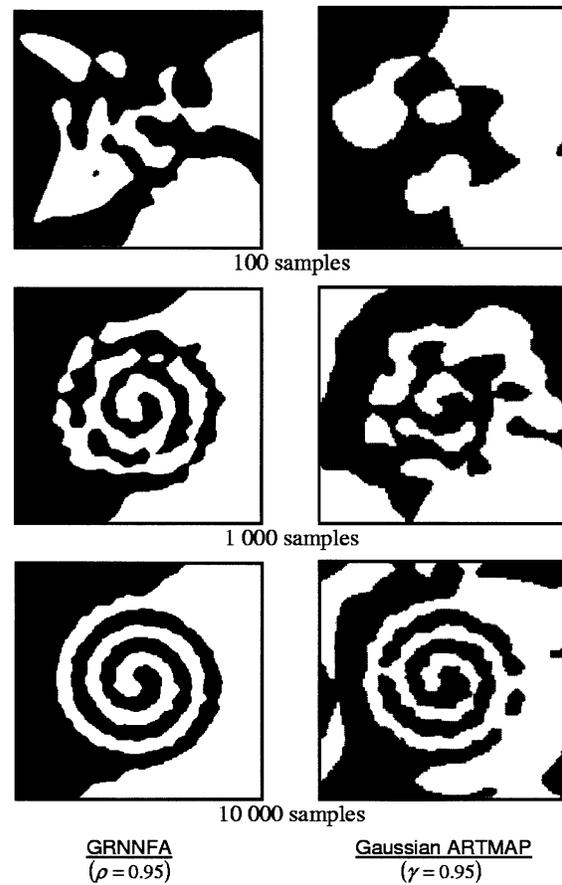
After several trials, the best values of the K-nearest neighbors and vigilance parameters were set to 3 and 0.9, respectively. A total of twenty experiments were performed. The average mean-squared-error (MSE) and its standard deviation obtained from the test set were calculated. The results of GRNNFA were compared with those from the neural-BAG (NBAG), Bench, and Simple models. Note that the Bench model [29] uses bagging to produce an ensemble of neural networks sub-models trained by different datasets resampled from the original dataset by the bootstrap technique. It takes the average of the predicted outputs of the neural network sub-models as the final predicted output. The Simple model is similar to Bench but equipped with a fast-stop training algorithm [28]. The NBAG model is similar to simple but with an algorithm to control the diversity among the neural network sub-models to increase the generalization performance of the overall model. These models were applied to the Ozone benchmark problem and the results are shown in Table I. The MSE of GRNNFA is 17.18 with a standard deviation of 2.57. These results are better than those from other models as reported in [28].

TABLE I
MSE OF DIFFERENT MODELS ON THE OZONE PROBLEM. STANDARD
DEVIATIONS CALCULATED FROM 20 RUNS ARE BRACKETED (RESULTS
OF NBAG, BENCH, AND SIMPLE ARE ADAPTED FROM [28])

Model	Ozone
NBAG	18.37 (3.59)
Bench	18.58 (3.40)
Simple	19.14 (3.21)
GRNNFA	17.18 (2.57)

TABLE II
BOOTSTRAP MEANS AND CONFIDENCE LIMITS OF MSE
FOR THE OZONE PROBLEM

No. of re-sampling	Lower Confidence Limit	Upper Confidence Limit	Mean
200	16.104	18.280	17.154
400	16.061	18.208	17.201
800	16.203	18.253	17.204
1600	16.058	18.317	17.165
3200	16.033	18.287	17.187
6400	16.074	18.294	17.178

TABLE III
MSE OF DIFFERENT MODELS ON THE FRIEDMAN#1 PROBLEM. STANDARD
DEVIATIONS CALCULATED FROM 20 RUNS ARE BRACKETED (RESULTS OF
NBAG, BENCH, AND SIMPLE ARE ADAPTED FROM [28])

Model	Friedman#1
NBAG	4.502 (0.268)
Bench	5.372 (0.646)
Simple	4.948 (0.589)
GRNNFA	4.563 (0.195)

To further examine the GRNNFA performance statistically, bootstrap means and 95% confidence limits of the MSE were computed, as shown in Table II. Since variations of the means and confidence limits of MSE for resamplings larger than 1600 is small (i.e., less than 0.2%), the results obtained from 1600 re-samplings were taken. These results justify that the performance of GRNNFA is significantly better than other models from the statistical point of view.

C. Friedman#1

This is a synthetic benchmark dataset proposed in [30]. Each sample consists of five inputs and one output. The formula for data generation is $t = 10\sin(\pi x_1 x_2) + 20(x_3 - 0.5)^2 + 10x_4 + 5x_5 + \varepsilon$ where ε is a Gaussian random noise $N(0, 1)$, and $x_1, \ldots, x_5$ are uniformly distributed over the domain $[0, 1]$. Similar to [28], 1,400 samples were created, of which 200 samples were randomly chosen for network training and 200 samples for validation. The remaining 1000 samples were used for network testing. After several trials, the best values of the K-nearest neighbors and vigilance parameter were 3 and 0.95, respectively. Table III summarizes the results predicted by the GRNNFA and the other models as listed in [28]. The MSE obtained by averaging the results of 20 runs is 4.563 with a standard deviation of 0.195. The MSE of GRNNFA is higher than that of NBAG, but lower than the rest. Besides, GRNNFA yielded the smallest standard deviation, indicating a stable performance. Bootstrapping was again applied to quantify

TABLE IV
BOOTSTRAP MEANS AND CONFIDENCE LIMITS OF MSE FOR THE
FRIEDMAN#1 PROBLEM

| No. of re-sampling	Lower Confidence Limit	Upper Confidence Limit	Mean
200	4.4843	4.6461	4.5631
400	4.4760	4.6494	4.5634
800	4.4743	4.6462	4.5613
1600	4.4773	4.6438	4.5627
3200	4.4761	4.6423	4.5619
6400	4.4780	4.6424	4.5635
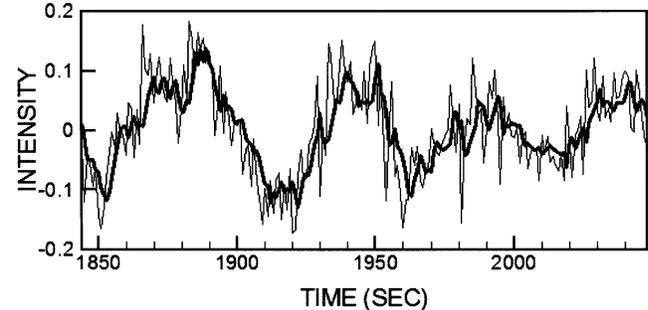
Fig. 7.   Comparison of the actual time series of the Sante Fe Series-E and time series predicted by the GRNNFA model.

the GRNNFA results statistically. Table IV summarizes the bootstrap mean and 95% confidence limits of the MSE.

The mean MSE obtained from 1600 resamplings (with variation less than 0.04%) is 4.5627 with 95% confident limits between 4.4773 and 4.6438. It can be observed that the MSE from NBAG (i.e., 4.502) is within the 95% confidence interval of the MSE distribution of GRNNFA estimated by bootstrap. This implies that the performance of the GRNNFA could be comparable to that of NBAG in this benchmark test.

D. Sante Fe Series E

This is the Series-E problem of the Sante Fe Time Series Competition [31]. The dataset can be downloaded from http://www-psych.stanford.edu/~andreas/Time-Series/. It is a univariate time series of astrophysical (variation in light intensity of a star) data samples which are noisy, discontinuous, and nonlinear in nature. In accordance with [32], 2048 samples were used, each with five inputs and one output, i.e., $x_1 = f(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5})$ where x_t is the intensity of the star at time t. The data presentation order was exactly the same as the original. The first 90% of the dataset were extracted for network training and validation. The last 10% were extracted for testing. After several trials, the best values of the K-nearest neighbors and vigilance parameters were 2 and 0.8, respectively. Twenty experiments were carried out. Fig. 7 shows the comparison between the test data (thin line) and the predicted outputs from GRNNFA (bolded line).

The average MSE is shown in Table V. The results reported in [32], i.e., pattern modeling and recognition system (PMRS), exponential smoothing (ES) and neural network (NN), are included for comparison. Note that PRMS is designed for noisy time series prediction by employing one-step forecasting, while

TABLE V
MSE OF DIFFERENT MODELS ON THE SANTE FE SERIES-E PROBLEM (RESULTS OF PMRS, ES AND NN ARE ADAPTED FROM [32])

Model	MSE
PMRS	0.015
ES	0.033
NN	0.078
GRNNFA	0.00326

TABLE VI
BOOTSTRAP MEANS AND CONFIDENCE LIMITS OF MSE FOR THE SANTE FE SERIES-E PROBLEM

No. of re-sampling	Lower Confidence Limit	Upper Confidence Limit	Mean
200	0.003349	0.003516	0.003429
400	0.003356	0.003514	0.003434
800	0.003353	0.003517	0.003433
1600	0.003357	0.003519	0.003432
3200	0.003351	0.003516	0.003431
6400	0.003352	0.003513	0.003431

ES is a regression method with an exponential smoothing parameter. The NN model is a feed-forward multilayer perceptron with one hidden layer and the number of hidden nodes was determined by using the procedure in [33] to achieve the minimum generalization error or maximum generalization performance. It can be seen that the MSE of the GRNNFA is much lower than those from other models. The results of bootstrapping are shown in Table VI.

Since variations of the means and the confidence limits of MSE for resamplings larger than 200 is small (i.e., less than 0.2%), the results obtained from 200 re-samplings were taken. It can be observed that the upper limit of the 95% confidence interval (i.e., 0.003 516) is lower than those MSE of other models as shown in Table V. Once again, the performance of GRNNFA is significantly better than other models from the statistical point of view.

E. Evacuation

The GRNNFA model was applied to a novel application in fire safety research. An experimental study was conducted for the prediction of evacuation time in the event of fire. Since evacuation is highly related to human behavior (e.g., choice of escape route, response to alarm, etc.), such problems are suitable to be handled by ANN models.

Karaoke centers are places of entertainment in which people sing in association with music. Fig. 8 shows the architectural layout of a typical karaoke center in Hong Kong.

Normally, a long corridor is designed to connect rooms to the lobby area at which an exit is provided. It is an interesting topic to investigate the relationship between the architectural parameters and evacuation time, which is defined as the time taken for the last evacuee coming out from the exit door. In this experimental study, three different architectural parameters (i.e., number of rooms, corridor width and lobby area) were varied to observe the effects on the evacuation time. Table VII shows variations of parameters pertaining to the training data.
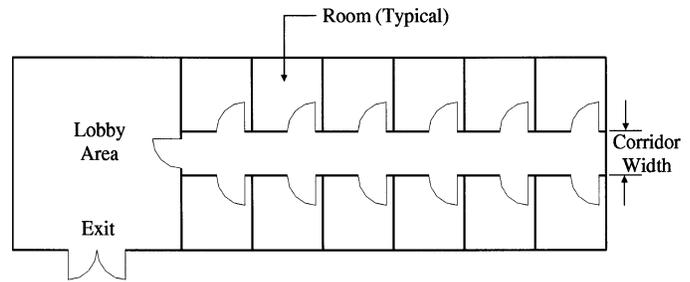

Fig. 8. Typical architectural layout of Karaoke centre in Hong Kong.

TABLE VII
RANGES OF PARAMETERS OF THE EVACUATION DATASET

Parameter	Minimum	Maximum
No. of Rooms	4	22
Corridor Width (mm)	1050	1400
Lobby Area (m^2)	15	35

The evacuation time was computed by the Spatial-Grid Evacuation Model (SGEM) [24]. In order to evaluate the robustness of the GRNNFA against noise, Gaussian noise with zero mean and standard deviation taken at 20% of the output range were introduced into the output data samples in the network training and validation sets.

A total number of 750 samples were created. The training and validation sets, respectively, contained 350 and 200 samples. The remaining 200 samples were used for testing the predictive power of GRNNFA. After several trials, the best vigilance parameter and K-nearest-neighbor value were set to 0.94 and 19, respectively. A GRNN model was created with global hyper-ellipsoidal kernel [34] and evaluated using the same training and test data. The adaptive network based fuzzy inference system (ANFIS) [35] model was also employed to benchmark this problem. It is a fuzzy inference system of which the membership functions are adaptively adjusted in the training process. After several trials, the best number of membership functions of each input component was determined to be three. The fast-stop training procedure with 10 000 epochs was used to construct the network structure.

A total of ten runs were carried out with different random initialization for extraction of the training, validation and test datasets. Bootstrap was applied to quantify the results statistically. The predictions from the three models (i.e., GRNNFA, GRNN and ANFIS) are summarized in Table VIII.

Owing to the small variations (i.e., less than 0.5%) of the means and confidence limits of Root Mean Squared Error (RMSE) for re-samplings larger than 3200 samples, the results obtained from 3200 re-samplings were taken. It can be seen that there is no overlapping among the 95% confidence intervals of the three models. The confidence intervals of ANFIS (21.8205 and 23.2144) are higher than others. The lower 95% confidence limit of GRNN (i.e., 16.3028) is also higher than the upper 95% confidence limit of GRNNFA (i.e., 16.2157). The results justify, statistically, the superior performance of GRNNFA in comparison with the other two models in noisy data regression.

TABLE VIII
BOOTSTRAP MEANS AND CONFIDENCE LIMITS OF ROOT MEAN SQUARED
ERROR FOR THE EVACUATION PROBLEM ("LOWER" AND "UPPER" ARE
LIMITS OF THE 95% CONFIDENCE INTERVALS)

No. of re-sampling	GRNNFA Mean (Lower, Upper)	GRNN Mean (Lower, Upper)	ANFIS Mean (Lower, Upper)
200	15.9164 (15.5818, 16.1985)	16.5935 (16.2391, 16.9031)	21.7161 (20.3690, 22.9388)
400	15.9342 (15.6078, 16.2258)	16.5937 (16.2932, 16.8852)	21.8427 (20.4143, 23.2183)
800	15.9359 (15.5834, 16.2386)	16.6090 (16.2835, 16.9061)	21.8279 (20.4546, 23.1756)
1600	15.9295 (15.5865, 16.2263)	16.6047 (16.3056, 16.8963)	21.7986 (20.4450, 23.1947)
3200	15.9348 (15.5936, 16.2157)	16.6057 (16.3028, 16.9120)	20.4901 (21.8205, 23.2144)
6400	15.9360 (15.6025, 16.2243)	16.6040 (16.3035, 16.8996)	21.8262 (20.4839, 23.2199)
12800	15.9361 (15.5996, 16.2238)	16.3017 (16.6030, 16.8556)	20.4808 (21.8197, 23.2067)

V. CONCLUSIONS

The GRNNFA model, based on the fusion of the GRNN and FA, has been introduced. It is a novel hybrid neural network with the properties of fast and stable learning. The FA model is used in network training to determine the number of prototypes, which can be interpreted as resolution of the prediction by the GRNN. The data compression scheme regulates the locations and labels of the kernels for prediction by the GRNN model. In this way, symmetrically distributed noise can possibly be eliminated. The K-nearest neighbors scheme was proposed to initialize the kernel widths prior to a gradient-based kernel width optimization algorithm.

The effectiveness of GRNNFA in noisy data regression problems has been demonstrated by using the Noisy Two-Intertwined Spirals, Ozone, Friedman#1, Sante Fe Series-E benchmark problems. In addition, GRNNFA has been applied to the prediction of evacuation time in the event of fire using noise-corrupted data samples. All the experimental results have been quantified using bootstrapping. The results positively demonstrate that the GRNNFA model is statistically comparable, if not superior, to the other neural network and regression models in the noisy data regression problems.

In terms of computational requirements, GRNNFA includes FA as a pre processor for data compression and a kernel optimization procedure for fine-tuning the smoothing parameter. However, since FA is an extremely fast clustering approach and the number of kernels in GRNNFA is fewer than that in the GRNN, the extra computational burden in GRNNFA can be minimized.

The next stage of development will focus on several potential modifications for the GRNNFA model. At present, GRNNFA employs FA, an unsupervised clustering method, as the pre-processor of the GRNN. A supervised clustering approach may be employed instead. Fuzzy ARTMAP (FAM) [36] could be a viable option. Since the GRNNFA model is developed for noisy data regression problems, solely relying on a deterministic prediction error feedback mechanism, as in FAM, to determine the optimum number of clusters may lead to the prototype proliferation problem [16], [36]. A suitable similarity measurement for error feedback is therefore necessary. A pruning procedure can be used to eliminate spurious nodes. In addition, GRNNFA

can be adapted for online learning by using an online kernel width tuning algorithm. These potential improvements of the GRNNFA model will be considered in further work.

APPENDIX
PSEUDO CODE FOR THE TRAINING PROCEDURE OF THE GRNNFA MODEL

```
Set vigilance parameter
Set K value of the K nearest neighbor
Ns = no. of training samples
N = no. of prototypes
// FA to determine kernel centers and outputs
for (i = 0, i < Ns, i + +)
{
 read sample from dataset
 normalize the input of the sample to range
[0, 1]
 complement coded the sample
 for (j = 0, j < N, j + +)
 {
  apply choice function to prototype j to get
Tj
 }
 J = -1
 exit = false
 k = 0
 do
 {
  k + +;
  apply vigilance test on prototype J
  // where J = arg(j, max Tj)
  if (vigilance test = pass)
  {
   update the prototype J
   update kernel_center (J)
   update kernel_output (J)
   exit = true
  }
  else set Tj = -1
  if (k = N) exit = true
 } while (!exit)
 if (J = -1)
 {
  create new node to code the input sample
  N + +
 }
}
// kernel width initialization and optimiza-
tion
for (i = 0, i < N, i + +)
{
 kernel_width (i) = average distance over
     K nos. of nearest neighbors
}
apply the gradient descent algorithm to opti-
mize kernel widths
End of training
```

REFERENCES

[1] D. F. Specht, "A general regression neural network," *IEEE Trans. Neural Networks*, vol. 2, pp. 568–576, Nov. 1991.

[2] E. J. Rzempoluck, "Neural network classification of EEG during camouflaged object identification," *Int. J. Med. Inform.*, vol. 44, pp. 169–175, 1997.

[3] C. Schaffner and D. Schroder, "An application of general regression neural network to nonlinear adaptive control," in *Proc. 5th Eur. Conf. Power Electronics Applications*, vol. 4, 1993, pp. 219–224.

[4] B. G. Hyun and K. Nam, "Faults diagnoses of rotating machines by using neural nets: GRNN and BPN," in *Proc. IEEE 21st Int. IECON Conf. Industrial Electronics, Control, Instrumentation*, vol. 2, 1995, pp. 1456–1461.

[5] M. T. Leung, A. S. Chen, and H. Daouk, "Forecasting exchange rates using general regression neural networks," in *Computers & Operations Research*. New York: Pergamon, 2000, vol. 27, pp. 1093–1110.

[6] E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Statist.*, vol. 33, pp. 1065–1076, 1962.

[7] S. Grossberg, "Adaptive pattern recognition and universal recoding II: Feedback, expectation, olfaction and illusions," *Biol. Cybern.*, vol. 23, pp. 187–202, 1976.

[8] B. Kosko, "Fuzzy entropy and conditioning," *Inform. Sci*, vol. 40, pp. 165–174, 1986.

[9] J. Brezmes, E. Llobet, X. Vilanova, J. Orts, G. Saiz, and X. Correig, "Correlation between electronic nose signals and fruit quality indicators on shelf-life measurements with pinklady apples," *Sens. Actuators*, vol. 80, pp. 41–50, 2001.

[10] R. Araujo and T. de Almeida, "Learning sensor-based navigation of a real mobile robot in unknown worlds," *IEEE Trans. Syst., Man, Cybern. B*, vol. 29, pp. 164–178, Apr. 1999.

[11] R. Araujo and A. T. de Almeida, "Fuzzy art based approach for real-time map building," in *Proc. 5th Int. Workshop Advance Motion Control, AMC-Coimbra*, 1998, pp. 623–628.

[12] L. Cinque, G. L. Foresti, A. Gumina, and S. Levialdi, "A modified fuzzy art for image segmentation," in *Proc. 11th Int. Conf. Image Analysis Processing*, 2001, pp. 102–107.

[13] S. Park, "Neural networks and customer grouping in e-commerce: A framework using fuzzy art," in *Proc. Academia/Industry Working Conf. Research Challenges*, 2000, pp. 331–336.

[14] T. Kohonen, "The self-organizing map," *Proc. IEEE*, vol. 78, pp. 1464–1480, Sept. 1990.

[15] J. B. Bezdek, "A convergence theorem for the fuzzy isodata clustering algorithms," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-2, pp. 1–8, 1980.

[16] G. A. Carpenter, S. Grossberg, and B. R. David, "Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system," *Neural Netw.*, vol. 4, pp. 759–771, 1991.

[17] C. P. Lim and R. F. Harrison, "An incremental adaptive network for on-line supervised learning and probability estimation," *Neural Netw.*, vol. 10, no. 5, pp. 925–939, 1997.

[18] D. Tomandl and A. Schober, "A modified general regression neural network (MGRNN) with new, efficient training algorithms as a robust 'black box' tool for data analysis," *Neural Netw.*, vol. 14, pp. 1023–1034, 2001.

[19] D. F. Specht, "Enhancements to probabilistic neural network," in *Proc. Int. Joint Conf. Neural Network*, vol. 1, 1992, pp. 761–768.

[20] P. Burrascano, "Learning vector quantization for the probabilistic neural network," *IEEE Trans. Neural Networks*, vol. 2, pp. 458–462, Nov. 1991.

[21] M. L. Tseng, "Integrating neural networks with influence diagrams for multiple sensor diagnostic systems," Ph.D. dissertation, Univ. California, Berkeley, 1991.

[22] H. G. C. Traven, "A neural network approach to statistical pattern classification by 'semiparametric' estimation of probability density functions," *IEEE Trans. Neural Network*, vol. 2, pp. 366–377, Nov. 1991.

[23] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, U.K.: Clarendon, 1998, pp. 269–269.

[24] S. M. Lo and Z. Fang, "A spatial-grid evacuation model for buildings," *J. Fire Sci.*, vol. 18, pp. 376–395, 2000.

[25] B. Efron, "Bootstrap methods: Another look at the jackknife," *Ann. Stat.*, vol. 7, pp. 1–26, 1979.

[26] ——, "Nonparametric standard errors and confidence intervals," *Can. J. Stat.*, vol. 9, pp. 139–172, 1981.

[27] J. R. Williamson, "Gaussian artmap: A neural network for fast incremental learning of noisy multidimensional maps," *Neural Netw.*, vol. 9, no. 5, pp. 881–897, 1996.

[28] J. Carney and P. Cunningham, "Tuning diversity in bagged ensembles," *Int. J. Neural Syst.*, vol. 10, no. 4, pp. 267–279, 2000.

[29] L. Breiman, "Bagging Predictors," Dept. Statist., Univ. California, Berkeley, Tech. Rep. 421.

[30] J. Friedman, "Multivariate adaptive regression splines (with discussion)," *Ann. Stat.*, vol. 19, pp. 1–141, 1991.

[31] A. S. Weigend and N. A. Gersehnfield, *Time Series Prediction: Forecasting the Future and Understanding the Past*. Reading, MA: Addison-Wesley, 1994.

[32] S. Singh, "Noise impact on time-series forecasting using an intelligent pattern matching technique," *Pattern Recognit.*, vol. 32, pp. 1389–1398, 1999.

[33] S. M. Weiss and C. A. Kulikowski, *Computer Systems That Learn*. San Mateo, CA: Morgan Kaufmann, 1991.

[34] T. Masters and W. Land, "A new training algorithm for the general regression neural network," in *IEEE Int. Conf. Systems, Man, Cybernetics—Computational Cybernetics Simulation*, vol. 3, 1997, pp. 1990–1994.

[35] J. S. R. Jang, "ANFIS: Adaptive-network-based fuzzy inference system," *IEEE Trans. Syst., Man, Cybern.*, vol. 23, pp. 665–685, June 1993.

[36] G. A. Carpenter, S. Grossberg, N. Markuzon, and J. H. Reynolds, "Fuzzy artmap: A neural network architecture for incremental supervised learning of analog multidimensional maps," *IEEE Trans. Neural Networks*, vol. 3, pp. 698–713, Sept. 1992.

Eric W. M. Lee received the B.Eng. degree, with first class honors in building services engineering and the Ph.D. degree fire engineering from the City University of Hong Kong, Hong Kong, in 2000 and 2003, respectively.

He worked in the field of building construction in Hong Kong for over 13 years and was an Executive Engineer in one of the leading building services engineering consultant firm. His major research interest is the application of artificial neural network on building fire, evacuation, and risk studies.

Mr. Lee was awarded the University Research Tuition Scholarships, Sir Edward Youde Memorial Scholarship and the Croucher Foundation Research Studentship.

Chee Peng Lim received the B.E.E. degree, with first class honors from the University of Technology, Skudai Johore, Malaysia, in 1992, the M.Sc. and Ph.D. degrees in engineering from the University of Sheffield, Sheffield, U.K., in 1993 and 1997, respectively.

He is currently Associate Professor at School of Electrical and Electronic Engineering, University of Science, Penang, Malaysia. His research interests include soft computing, pattern classification, medical diagnosis, fault diagnosis, condition monitoring, and intelligent manufacturing and control. He has published more than 70 technical papers.

Dr. Lim received three Best Paper awards at national and international conferences and is recipient of the Japan Society for the Promotion of Science (JSPS), Fulbright, and Commonwealth Fellowships, as well as The Outstanding Young Malaysians Award (Scientific and Technological Development), 2001.

Richard K. K. Yuen received the Ph.D. degree in fire dynamics and engineering from the University of New South Wales, Sydney, Australia.

He joined the Department of Building and Construction at the City University of Hong Kong in 1990, and is currently an Associate Professor. He has engaged in consultancies for both private and public sectors in the areas of fire safety engineering, environmental noise impact assessment, building energy studies, auditing, and building thermal system modeling studies. He has been active in the research and development, including fire safety and engineering, applications of computational fluid dynamics and neural network modeling in fire engineering, environmental noise impact, building energy, lighting and ventilation, heating, ventilation, and air-conditioning (HVAC) systems, and indoor air quality. He has also been involved in various professional activities including participation in Working Group on Fire Safety Engineering Approach, Buildings Department.

Dr. Yuen is a Chartered and Registered Professional Engineer and member of various professional bodies including the Hong Kong Institute of Engineers (HKIE), Chartered Institution of Building Services Engineers (CIBSE), and the Institution of Engineers Australia (IEAust).

S. M. Lo received the Ph.D. degree in architecture from the University of Hong Kong, Hong Kong.

He is an Associate Professor with the Department of Building and Construction, City University of Hong Kong. Prior to his teaching career, he has had over ten year practical experience in the construction industry. He is currently a nonofficial member of the Hong Kong Government's Fire Safety Committee. His main research interests include building design, evacuation modeling, fire risk assessment, etc. He holds many research grants supported by Hong Kong Research Grant Council for studying evacuation, fire risk analysis, wayfinding modeling, intelligent understanding of computer aided drawing (CAD) plans, and human behavior in fire, etc.