Decision Support Systems 46 (2008) 411–419

Contents lists available at ScienceDirect

Decision Support Systems

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

Sales forecasting using extreme learning machine with applications in fashion retailing

Zhan-Li Sun, Tsan-Ming Choi *, Kin-Fan Au, Yong Yu

Institute of Textiles and Clothing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

A R T I C L E I N F O

Article history:
Received 10 August 2007
Received in revised form 22 July 2008
Accepted 31 July 2008
Available online 13 August 2008

Keywords:
Fashion sales forecasting
Extreme learning machine
Artificial neural network
Backpropagation neural networks
Decision support system

A B S T R A C T

Sales forecasting is a challenging problem owing to the volatility of demand which depends on many factors. This is especially prominent in fashion retailing where a versatile sales forecasting system is crucial. This study applies a novel neural network technique called extreme learning machine (ELM) to investigate the relationship between sales amount and some significant factors which affect demand (such as design factors). Performances of our models are evaluated by using real data from a fashion retailer in Hong Kong. The experimental results demonstrate that our proposed methods outperform several sales forecasting methods which are based on backpropagation neural networks.

1. Introduction

Sales forecasting refers to the prediction of future sales based on past historical data. Owing to competition [41,42] and globalization, sales forecasting plays a more and more prominent role in a decision support system [26] of a commercial enterprise. An effective sales forecasting can help the decision-maker calculate the production and material costs and determine the sales price. This will result in lower inventory levels, quick response and achieve the objective of just-in-time (JIT) delivery [2,5–7,12]. However, sales forecasting is usually a highly complex problem due to the influence of internal and external environments, especially for the fashion and textiles industry [25–27]. Thus, nowadays, how to develop more accurate and timely sales forecasting methods becomes an important research topic. Some retailers improve their stocking decisions by acquiring market information and revising their forecast in multiple stages [8–10]. A good forecasting method can help retailers reduce over-stocking and under-stocking costs [12]. Thus sales forecasting becomes one crucial task in supply chain management under uncertainty and it greatly affects the retailers and other channel members in various ways [31,43]. In this paper, we propose a new method which employs extreme learning machine (ELM) for sales forecasting in fashion retailing [32].

Recently, artificial neural networks (ANNs) have been applied extensively for sales forecasting [4,13,34,35,44–46] because they have very promising performance in the areas of control, prediction, and pattern recognition [15,21,22,30,33,38,40]. Many studies conclude that ANN is better than various conventional methods [1,3,28,29,39]. In [13], the statistical time-series model and the ANN based model were investigated for forecasting women's apparel sales. Chakraborty et al. [3] presented an ANN approach based on multivariate time-series analysis, which can accurately predict the flour prices in three cities in USA. Lachtermacher and Fuller [28] developed a calibrated ANN model. In the model, the Box–Jenkins methods are used to determine the lag components of the input data. Moreover, it employed a heuristics method to choose the number of hidden units. In Kuo and Xue [27], the authors reported that the ANNs are better than many conventional statistical forecasting methods (see [3,16] for more details). However, most ANN based sales forecasting methods use gradient-based learning algorithms, such as the backpropagation neural network (BPNN), and problems such as over-tuning and long computation time still arise. A relatively novel learning algorithm for single-hidden-layer feedforward neural networks (SLFN) called extreme learning machine (ELM) has been proposed in [20,47] recently. In ELM, the input weights (linking the input layer to the hidden layer) and hidden biases are randomly chosen, and the output weights (linking the hidden layer to the output layer) are analytically determined by using the Moore–Penrose (MP) generalized inverse. ELM not only learns much faster with a higher generalization performance than the traditional gradient-based learning algorithms but it also avoids many difficulties faced by gradient-based learning methods such as stopping criteria, learning rate, learning epochs, local minima, and the over-tuned problems [16–18,36].

To the best of our knowledge, the application of ELM for fashion sales forecasting has not been studied in the literature. In this paper, the ELM is selected to analyze fashion sales forecasting on the data provided by a Hong Kong fashion retailer. In this method, some design factors (size, color, etc.) and sales factors (price, etc.) of the fashion apparels are chosen as the input variables of the ELM. Although ELM has many

* Corresponding author. Tel.: +86 852 27666450; fax: +86 852 27731432.
E-mail address: tcjason@inet.polyu.edu.hk (T.-M. Choi).

0167-9236/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.dss.2008.07.009

advantages compared to those traditional gradient-based learning algorithms, a shortcoming is that its solution is usually different from time to time because the input weights and hidden biases are randomly chosen, which is also a common existing problem in ANNs that parameters are initialized randomly. As a result, we don't know exactly on which time the initiation will give a good result when we want to predict a future sales amount. Considering the randomness for the selection of weights and hidden biases [24], we propose to predict the future sales amount by an integration of ELMs. For this method, the average series of multiple ELMs' outputs is used as the finally predicted sales amount. Our findings indicate that this extension usually has a smaller predicting error when the fluctuation of ELM outputs is larger.

In addition, the data sets are usually normalized before training so that they fall in a specifically given interval. This measure considerably accelerates weight learning and avoids saturation or overflow of the hidden and output neurons whose activation values generally fall in the [0, 1] or [1, 1] interval. Finally, for the outputs of ELM, an unnormalization step is necessary to convert the data back into unnormalized units.

Based on the above analyses, we propose a sales forecasting method using ELM for fashion retailing in this paper. The rest of the paper is organized as follows. The fundamental principle of our proposed method is introduced in Section 2. Experimental results and related discussions are presented in Sections 3 and 4, respectively.

2. Methodology

In this section, we present the model for the fashion sales forecasting using extreme learning machine (ELM) algorithm. For this method, we first extract the sales data of one kind of fashion clothes from the raw data, which include all factors affecting the sales amount. Then the most significant factors are selected to be the inputs of the ELM. The output of the ELM is the sales amount. Subsequently, the data composed of these input/output pairs are divided into training, testing, and predicting sets. Before training, the training data and testing data are normalized respectively so that the inputs fall into a specific range. After training, the unnormalization step converts the data back into unnormalized units. Based on the input and output weights obtained by training data and testing data, the predicted sales series for the predicting data can be computed directly through the established ELM. We will give a specific example on how each step is implemented in the model in the experiment section. In the following, we give a concise review of ELM [20,37,47], and the normalization and unnormalization procedures [11,23].

2.1. Extreme learning machine

As shown in Fig. 1, ELM is a single hidden-layer feedforward neural network (SLFN). It randomly chooses the input weight matrix **W** and

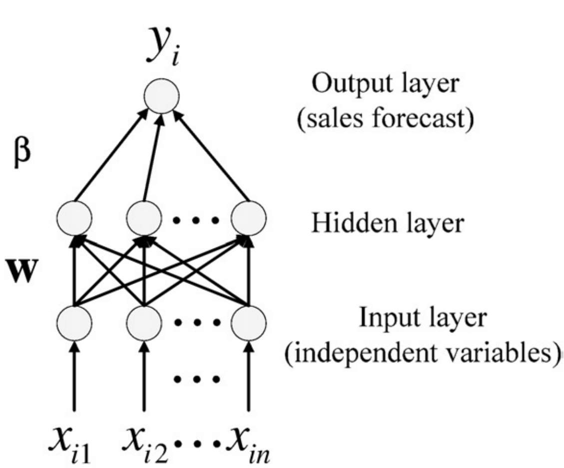


Fig. 1. The structure of ELM model.

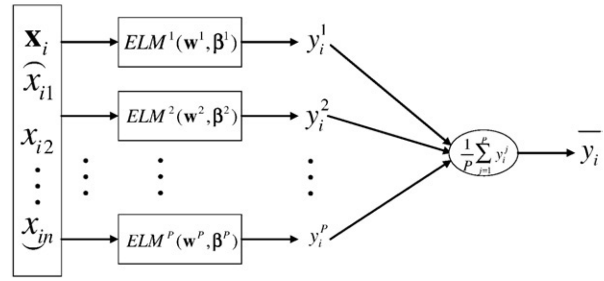


Fig. 2. The scheme of the ELM integration system.

analytically determines the output weight matrix $\boldsymbol{\beta}$ of SLFN. Suppose that we are training a SLFN with K hidden neurons and an activation function vector $\mathbf{g}(x)=(g_1(x),g_2(x),\ldots,g_K(x))$ to learn N distinct samples (x_i, t_i), where $x_i=[x_{i1}, x_{i2},\ldots,x_{in}]^T \in R_n$ and $t_i=[t_{i1}, t_{i2},\ldots,t_{in}]^T \in R_m$. If the SLFNs can approximate these N samples with a zero error, then we have

$$\sum_{j=1}^{N} ||\mathbf{y}_j - \mathbf{t}_j|| = 0, \tag{1}$$

where $\mathbf{y}$ is the actual output value of the SLFN. There also exist parameters $\boldsymbol{\beta}_i$, $\boldsymbol{w}_i$ and $\mathbf{b}_i$ such that

$$\sum_{i=1}^{K} \beta_i g_i(\boldsymbol{w}_i \cdot \mathbf{x}_j + b_i) = \mathbf{t}_j, j = 1,\cdots,N, \tag{2}$$

where $\boldsymbol{w}_i=[w_{i1},\ldots,w_{m}]^T$ is the weight vector connecting the ith hidden neuron and the input neurons, $\boldsymbol{\beta}_i=[\beta_{i1},\ldots,\beta_m]^T$, $i=1,\ldots,K$ is the weight vector connecting the ith hidden neuron and the output neurons, and b_i is the threshold of the ith hidden neuron. The operation $\boldsymbol{w}_i \cdot \mathbf{x}_j$ in Eq. (2) denotes the inner product of $\boldsymbol{w}_i$ and $\mathbf{x}_j$. The above N equations can be written compactly as:

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{T}, \tag{3}$$

where $\mathbf{H}=\{h_{ij}\}$ ($i=1,\ldots,N$ and $j=1,\ldots,K$) is the hidden-layer output matrix, $h_{ij}=g(w_j \cdot x_i+b_j)$ denotes the output of the jth hidden neuron with respect to $\mathbf{x}_i$, $\boldsymbol{\beta}=[\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_K]$ is the matrix of output weights, $\mathbf{T}=[\mathbf{t}_1, \mathbf{t}_2,\ldots, \mathbf{t}_N]^T$ is the matrix of targets.

In ELM, the input weights and hidden biases are randomly generated instead of tuned. Thus, the determination of the output weights (linking the hidden layer to the output layer) is as simple as finding the least-square solution to the given linear system. The minimum norm least-square (LS) solution to the linear system (3) is

$$\hat{\beta} = \mathbf{H}^{\dagger}\mathbf{T}, \tag{4}$$

where $\mathbf{H}^{\dagger}$ is the MP generalized inverse of the matrix $\mathbf{H}$. The minimum norm LS solution is unique and has the smallest norm among all the LS solutions. As analyzed by [10], ELM tends to obtain a good generalization performance with a dramatically increased learning speed by using this MP inverse method.

2.2. Normalization and unnormalization

The normalization method for the input variables and output variables can be described as follows [13,17]:

$$xn_{pi} = 2 \times (x_{pi} - \min\{x_{pi}\}) / (\max\{x_{pi}\} - \min\{x_{pi}\}), i = 1, 2,\cdots,n, p = 1,\ldots,N, \tag{5}$$

$$yn_p = 2 \times (y_p - \min\{y_p\}) / (\max\{y_p\} - \min\{y_p\}), p = 1,\cdots,N. \tag{6}$$

Table 1
A part of raw data provided by a Hong Kong fashion retailer

Month	Date	Code number	Color	Size	Price	Sales amount
11/2005	25	424101160	54	026	445	5
11/2005	26	424101160	54	029	89	1
11/2005	27	424101160	99	026	267	3
–	–	–	–	–	–	–

The unnormalization method for the normalized data is given as follows:

$$xun_{pi} = 0.5 \times xn_{pi} \times \left(\max\{xn_{pi}\} - \min\{xn_{pi}\}\right) + \min\{xn_{pi}\}, \quad (7)$$
$$i = 1, 2, \ldots, n, p = 1, \ldots, N,$$

$$yun_p = 0.5 \times yn_p \times \left(\max\{y_p\} - \min\{y_p\}\right) + \min\{y_p\}. \quad (8)$$

The operators min{·} and max{·} in Eqs. (5) to (8) are used to select the minimum and maximum values from the given data series, respectively.

2.3. Steps of fashion sales forecasting using extreme learning algorithm

Assume that the dimension of the input variables is n and the number of the samples is N. Define the input variables as $x_i=(x_{i1},x_{i2},\ldots,x_{in})$, the observed values as t_i,$i=1,2,\ldots,N$. In the following, we give the specific steps of the fashion sales forecasting method using ELM.

Step 1 Extract the sales data of one kind of fashion clothes from the raw data according to its code number;
Step 2 Observe the sales amount in a given time interval, select the factors that have a significant effect on the sales as the inputs of ELM;
Step 3 Divide the input/output data ($\mathbf{x}_i$, t_i) into training data (*TRD* of N_t), testing data (*TED* of N_c), and predicting data (*PRD* of N_p) randomly, where $N=N_i+N_e+N_p$;
Step 4 Normalize the training data and testing data using Eqs. (5) and (6), respectively;
Step 5 Select the activation function of hidden neuron and the neuron number of hidden layer of ELM;
Step 6 Input training data and testing data, compute the outputs of ELM, unnormalized the outputs, then obtain the predicted sales series y_i, $i=1,\ldots,N$ of training data and testing data;
Step 7 Based on the input and output weights obtained by Steps 5 and 6, compute the predicted sales series of predicting data and the corresponding predicting error.
Step 8 Repeat Steps 6 and 7 for P times for the same data, then obtain P predicting sales series y_i^j, $i=1,2,\ldots,N$, $j=1,2,\ldots,P$; compute the average predicting sales series $\bar{y}_i = \frac{1}{P}\sum_{j=1}^{P} y_i^j$, $i=1,2,\ldots,N$ and its predicting error.

As shown in Step 8 and Fig. 2, in order to obtain a higher prediction accuracy, a regression integration method is proposed in this paper as an extension of ELM (ELME) model. To compute the average predicting sales series $\bar{y}_i = \frac{1}{P}\sum_{j=1}^{P} y_i^j$, $i = 1, 2, \cdots, N$, we must first repeatedly run the ELM for P times with the same data set. It is well known that the mean value is closer to the expectation when the parameter P becomes larger. Since the expectation is a single value, the results obtained by the ELME will become more stable when the parameter P becomes larger. Unfortunately, the computation time also increases when the parameter P becomes larger. Considering both the computation time and the stability of ELME, the parameter P is selected from the interval [100, 1000].

As a remark, for the P trials, the structure of ELM, including the number of layers and the input, are all the same. The input weight matrices of P trials are different from each other because

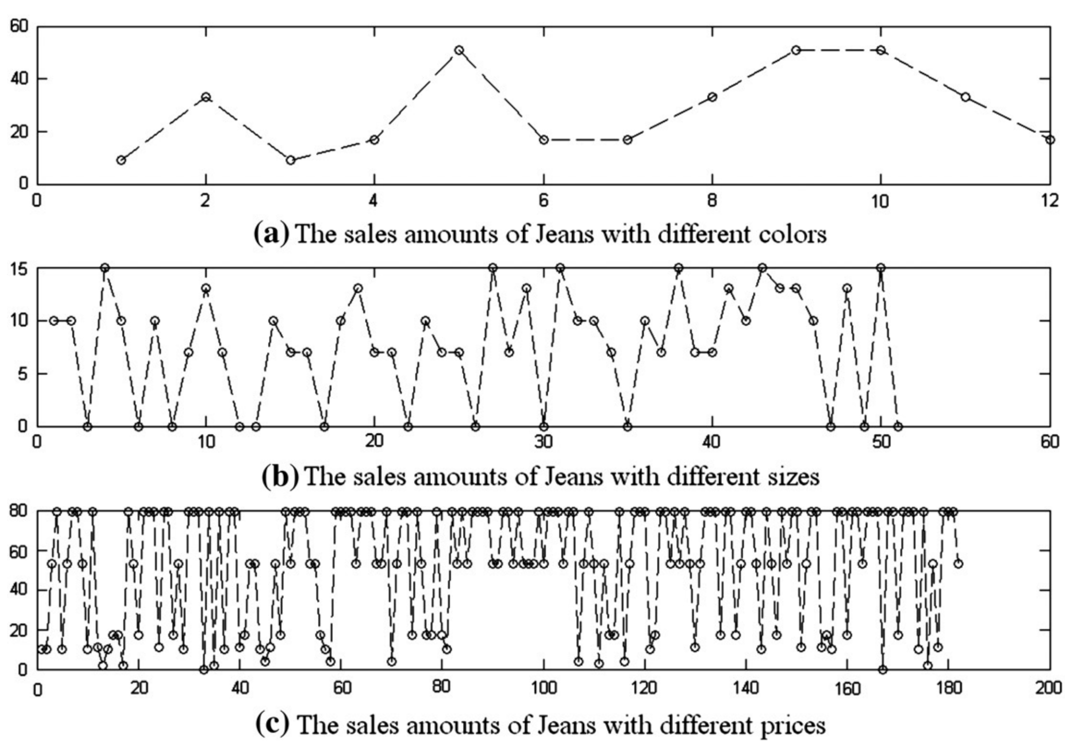


Fig. 3. The sales amount of Jeans in a month for different colors, sizes, and prices (the x-axes of panels (a), (b) and (c) denote the number of color, the number of size, and the number of price for the corresponding series, respectively. The y-axes denote the respective sales amounts).

Table 2
The sales mean (μ) and standard deviation (σ) of Jeans in one month for different colors, sizes and prices

Cases	Number	μ	σ
Case 1 (size code: 27; price: 356)	12 (colors)	28.1667	16.1461
Case 2 (color code: 54; price: 356)	51 (sizes)	7.6863	5.2096
Case 3 (color code: 54; size code: 27)	182 (prices)	52.8516	29.2236

Table 3
The coefficients of variation (cvs) of the three cases in Table 2

Cases	cv
1	0.5732
2	0.6778
3	0.5529

of the random initialization. The P predicted sales series are obtained by using ELM with the P input weight matrices. So the P predicted sales series are different from each other. The final predicted sales series are only the mean of the P predicted sales series. It should be pointed out that the structure and parameters of each ELM in the integration system should be retained when the trained system is used to predict a new sales series in real application.

3. Simulation studies

In order to test the validity and performance of the proposed algorithm, we present in this section three experimental results on three sets of real fashion sales data provided by a Hong Kong fashion retailer. Specifically, as an example, we give a concise explanation for each step in Section 2.3 combining with the data

Table 4
The comparisons of GDA, GDX, ELM and ELME in experiment 1

	GDA	GDX	ELM	ELME
$\mu_{\mathrm{mse}}^{\mathrm{tr}}$	1.1450	2.8538	0.0001	2.1124e−005
$\mathrm{std}_{\mathrm{mse}}^{\mathrm{tr}}$	0.5296	0.44053	0.0001	
$\mu_{\mathrm{mse}}^{\mathrm{te}}$	1.6903	2.138	2.0640	2.0629
$\mathrm{std}_{\mathrm{mse}}^{\mathrm{te}}$	0.4527	0.45464	0.0228	
$\mu_{\mathrm{mse}}^{\mathrm{pr}}$	1.0710	2.6304	0.0002	4.9872e−005
$\mathrm{std}_{\mathrm{mse}}^{\mathrm{pr}}$	0.5034	0.39369	0.0002	

processed in experiment 1. These explanations are not repeated in experiments 2 and 3 since they are all similar. The batch steepest descent backpropagation algorithm with an adaptive learning rate (GDA), and the gradient descent momentum and adaptive learning ratio backpropagation (GDX) are two typical back-propagation algorithms [15]. As a comparison, we also give the experimental results obtained by the GDA and GDX algorithms on the same data set.

All simulations are conducted in MATLAB running on an ordinary personal computer with a dual core CPU (1.73 GHZ and 0.97 GHZ) and 1 G memory. In the following experiments, the activation function is the sigmoidal function:

$$g(x) = \frac{1}{1+e^{-x}}. \tag{9}$$

The performance index is the mean squared error (mse) between the predicted sales amount and actual sales amount:

$$\mathrm{mse} = \frac{1}{N}\sum_{i=1}^{N}(y_i - t_i)^2, \tag{10}$$

where y_i and t_i are the predicted sales amount and actual sales amount, respectively.

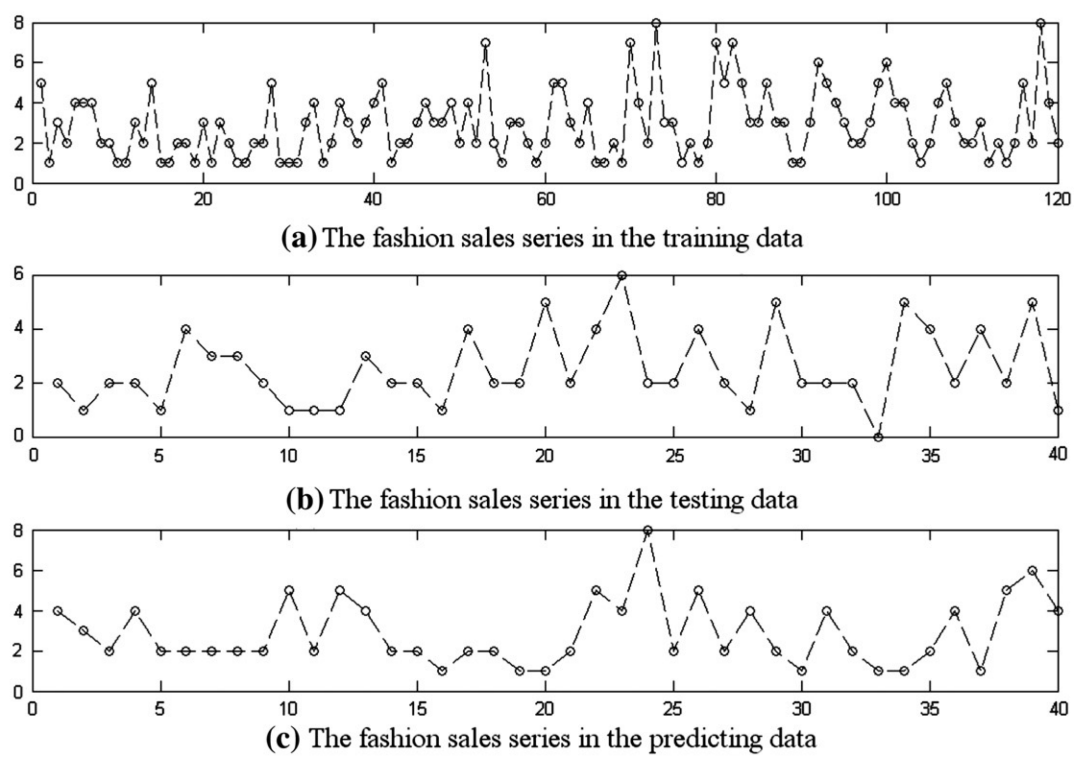


Fig. 4. The fashion sales series in the training, testing, and predicting sets in experiment 1 (The x-axes of panels (a), (b), and (c) denote the number of samples of the corresponding series; the y-axes denote the sales amounts).

Table 5
The performance index ratios in experiment 1

	$\frac{\text{GDA}}{\text{ELM}}$	$\frac{\text{GDX}}{\text{ELM}}$	$\frac{\text{ELME}}{\text{ELM}}$
$\mu_{\text{mse}}^{\text{tr}}$	11,450	28,538	0.1604
$\text{std}_{\text{mse}}^{\text{tr}}$	5296	4405.3	
$\mu_{\text{mse}}^{\text{te}}$	0.81894	1.0359	0.9994
$\text{std}_{\text{mse}}^{\text{te}}$	19.855	19.94	
$\mu_{\text{mse}}^{\text{pr}}$	5355	13,152	0.2296
$\text{std}_{\text{mse}}^{\text{pr}}$	2517	1968.4	

Table 6
The ratios between std_{mse} and μ_{mse} for ELM

	Training data	Testing data	Predicting data
$\frac{\text{std}_{\text{mse}}}{\mu_{\text{mse}}}$	1.0388	0.0111	1.1075

To measure the stability of the algorithms, we define the coefficient of variation cv as the ratio between standard deviation (σ) and mean (μ):

$$\text{cv} = \frac{\sigma}{\mu}. \tag{11}$$

The standard deviation std_{mse} of P trials is defined as:

$$\text{std}_{\text{mse}} = \sqrt{\frac{1}{P}\sum_{i=1}^{P} (\text{mse}_i - \mu)^2}, \tag{12}$$

where mse_i, $i=1,2,\ldots,P$ is the mean squared error obtained when we run the algorithm for the ith time with the same data set, μ_{mse} is the mean value of mse_i:

$$\mu_{\text{mse}} = \frac{1}{P}\sum_{j=1}^{P} \text{mse}_j. \tag{13}$$

3.1. Experiment 1

According to the code number of the commodity, the Jeans data is extracted from the raw data in Table 1 and used as the experimental data. There are 7 attributes for this kind of clothes: month, date, code number, color, size, price, sales amount. Since the goal of this study is to investigate the relationship between the fashion sales amount and the most significant factors affecting the sales amount, and owing to the features of the data sets, the attributes of month, date and code number are not selected.

According to the expert knowledge, the remaining factors, color, size, and price, have a significant impact on the sales amount. To analyze their impacts further, as shown in Fig. 3, we give the sales amount of Jeans with different colors, sizes and prices in March, 1999. As shown in Table 2, there are 12, 51, 182 kinds of colors, sizes, and prices, respectively for the Jeans in March. Fig. 3(a) shows the sales amount of Jeans with the same size (size number: 027) and price (356). Each point represents the sales amount in a month for one color. Fig. 3(b) shows the sales amount of Jeans with same color (color number: 54) and price (356) but with different sizes. The sales amount of Jeans with the same color (color number: 54) and size (size number: 027) but with different prices is shown in Fig. 3(c).

For the three cases in Fig. 3, the mean (μ) and standard deviation (σ) of the sales amount are given in Table 2. The coefficients of variation (cvs) are given in Table 3. From Fig. 3 and Tables 2–3, we can see that the factors of color, size, and price have significant impacts on the sales amount. We use the coefficients of variation (cvs) to measure the significance of factors. As shown in Table 3, three factors color, size, and price have similar cvs so they are all selected. Therefore, they are used as the inputs of the ELM. If more factors are available, we can sort

Table 7
The comparisons of GDA, GDX, ELM and ELME in experiment 2

	GDA	GDX	ELM	ELME
$\mu_{\text{mse}}^{\text{tr}}$	0.9859	2.1254	0.1556	0.1548
$\text{std}_{\text{mse}}^{\text{tr}}$	0.2817	1.8719	0.0010	
$\mu_{\text{mse}}^{\text{te}}$	0.6015	1.5050	1.4094	1.4077
$\text{std}_{\text{mse}}^{\text{te}}$	0.1665	1.3556	0.0425	
$\mu_{\text{mse}}^{\text{pr}}$	0.7810	1.6818	0.6515	0.649
$\text{std}_{\text{mse}}^{\text{pr}}$	0.2978	1.0578	0.0366	

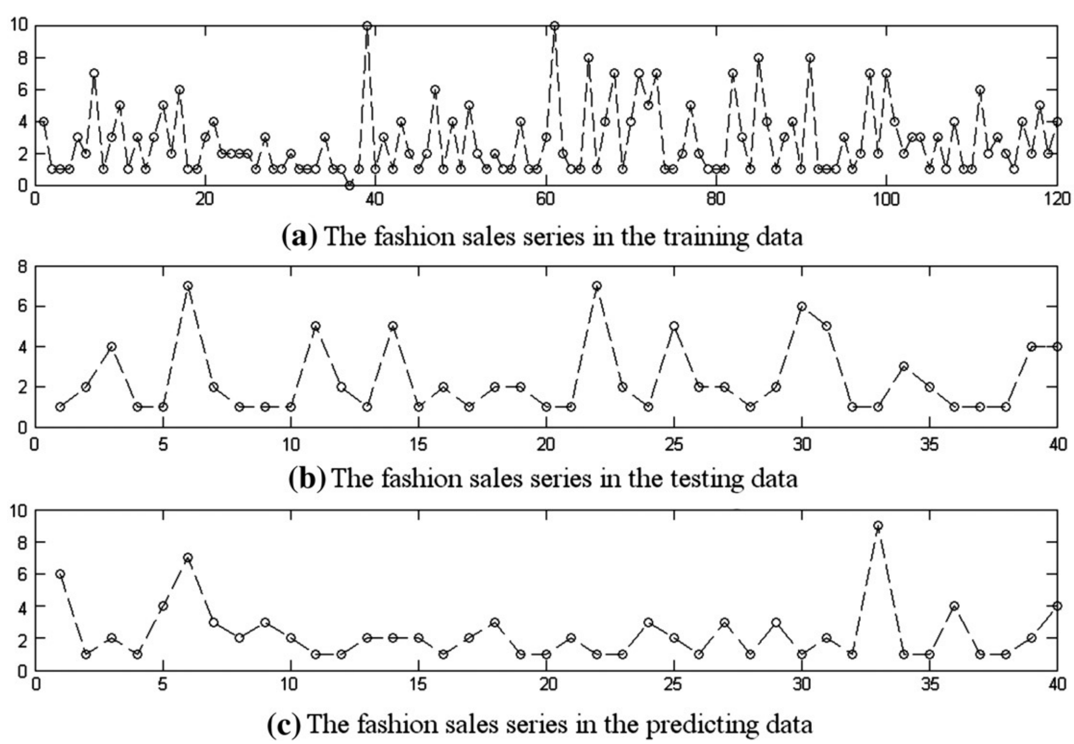


Fig. 5. The fashion sales series in the training, testing, and predicting sets in experiment 2 (The x-axes of panels (a), (b), and (c) denote the number of samples of the respective series; the y-axes denote the sales amounts).

Table 8
The performance index ratios in experiment 2

	$\frac{GDA}{ELM}$	$\frac{GDX}{ELM}$	$\frac{ELME}{ELM}$
μ_{mse}^{tr}	6.3361	13.659	0.9951
std_{mse}^{tr}	281.7000	1871.9	
μ_{mse}^{te}	0.4268	1.0678	0.9988
std_{mse}^{te}	3.9176	31.896	
μ_{mse}^{pr}	1.1988	2.5814	0.9962
std_{mse}^{pr}	8.1366	28.902	

Table 9
The ratios between std_{mse} and μ_{mse} for ELM

	Training data	Testing data	Predicting data
$\frac{std_{mse}}{\mu_{mse}}$	0.0066	0.0301	0.0561

Table 10
The comparisons of GDA, GDX, ELM and ELME in experiment 3

	GDA	GDX	ELM	ELME
μ_{mse}^{tr}	0.3368	0.4213	0.4930	0.4046
std_{mse}^{tr}	0.1346	0.1530	0.0979	
μ_{mse}^{te}	0.4893	0.3534	0.4297	0.2795
std_{mse}^{te}	0.2538	0.1982	0.1337	
μ_{mse}^{pr}	0.7308	0.6100	0.4952	0.3888
std_{mse}^{pr}	0.3336	0.2327	0.1493	

Table 11
The performance index ratios in experiment 3

	$\frac{GDA}{ELM}$	$\frac{GDX}{ELM}$	$\frac{ELME}{ELM}$
μ_{mse}^{tr}	0.67172	0.8546	0.82069
std_{mse}^{tr}	1.0338	1.5628	
μ_{mse}^{te}	1.0464	0.8224	0.65045
std_{mse}^{te}	1.2202	1.4824	
μ_{mse}^{pr}	1.4561	1.2318	0.78514
std_{mse}^{pr}	1.4625	1.5586	

them according to the criterion cvs and then select the factors with the large cvs. The sales amount is the output of the ELM. The data processed are composed of these input/output pairs.

Here, 200 input/output pairs are used as the experimental data, in which 60% of the data points are used for the training set, 20% for the testing set, and 20% for the predicting set. The sales amounts in the training, testing, and predicting sets are shown in Fig. 4(a), (b), and (c), respectively. The training data and testing data are first normalized by using Eqs. (5) and (6). As a result, the training data and testing data all fall into the interval [−1,1]. Then we perform the experiment on training and testing data according to steps 3–7 depicted in Section 2.3. As done in [19], the average mse (μ_{mse}) and standard deviation (std_{mse}) of 100 trials are used to evaluate the performance of the algorithms. For simplicity, the symbols tr, te, and pr are added into μ_{mse} and std_{mse} as superscripts to denote the experimental results for training, testing, and predicting data, respectively.

When the number of hidden neurons is increased from 1 to 30, the number that gives the smallest validation error is chosen for the ELM and the two backpropogation algorithms. The experimental comparisons for the batch steepest descent backpropagation algorithm with an adaptive learning rate (GDA), gradient descent momentum and adaptive learning ratio backpropagation (GDX), extreme learning algorithm (ELM) and its extension (ELME) are given in Table 4.

The performance index ratios between GDA, GDX, ELME and ELM are given in Table 5. It should be pointed out that the performance index std of the ELME is not filled in Tables 4 and 5 because there is only one result after the 100 trials of ELM. (For the stability of the ELME, we have given a discussion in Section 2.3.) From Tables 4 and 5, it can be seen that ELM and ELME generally have smaller training, testing, and predicting errors than GDA and GDX. In Table 5, the training, testing, and predicting errors of ELME are only 16.04%, 99.94% and 22.96% of the corresponding performance indices of ELM, respectively.

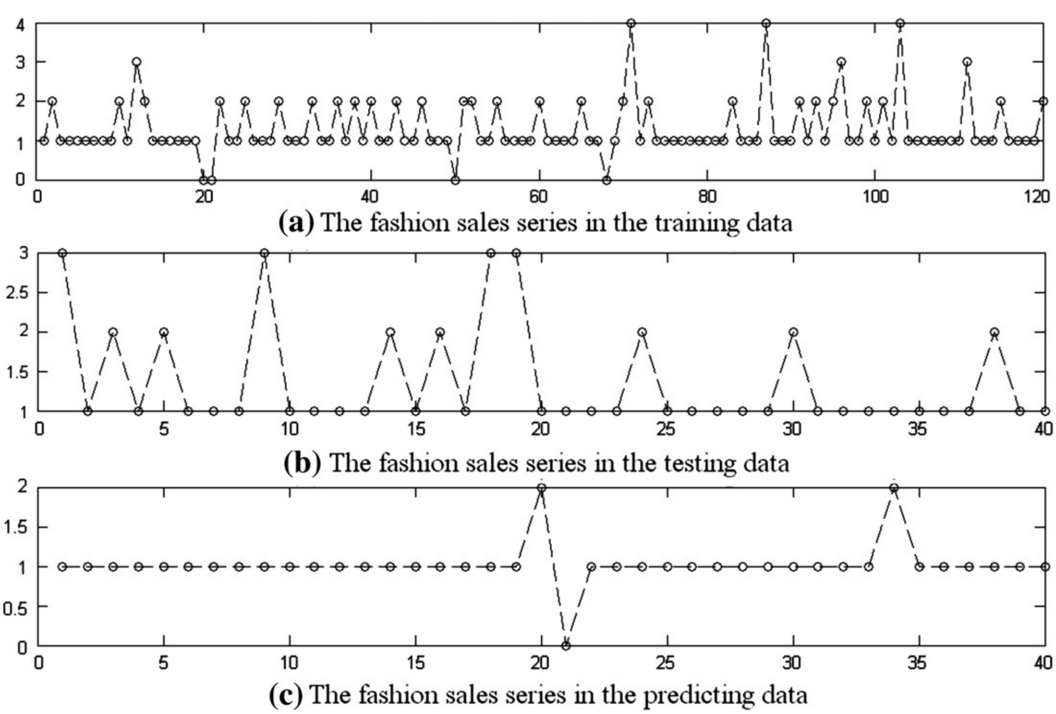


Fig. 6. The fashion sales series in the training, testing, and predicting sets in experiment 3 (The *x*-axes of panels (a), (b), and (c) denote the number of samples of the corresponding series; the *y*-axes denote the sales amounts).

Table 12
The ratios between std_{mse} and μ_{mse} for ELM

	Training data	Testing data	Predicting data
$\frac{\mathrm{std}_{mse}}{\mu_{mse}}$	0.1985	0.3112	0.3016

Table 13
The mean (μ) and standard deviation (σ) of sales amounts for three products in one year

Product	1	2	3
μ	1.8907	2.8979	1.4040
σ	1.3476	1.4040	0.8969

To analyze the fluctuation of ELM, the ratios between std_{mse} and μ_{mse} are given in Table 6. The standard deviation std_{mse} is about one time of the average error μ_{mse} in Table 6. Therefore, the training and predicting errors of ELM have a large variation from time to time due to the random initiation of input weights and hidden bias. That is to say, the fluctuation of ELM is large in experiment 1.

Based on the above analyses, we can conclude that ELM and its extension ELME are better than the other two methods for fashion sales forecasting. And the ELME is better than ELM when the fluctuation of ELM is large.

3.2. Experiment 2

In experiment 2, the sales data of sock are selected as the experimental data. In the data, there is only one kind of size for the sock. Thus the size factor is not selected as an important factor (cvs = 0). Only the color and sales price of the clothes are selected as the significant factors. Here, 200 samples are used as the experimental data, in which 60% of the data points are used for the training set, 20% for the testing set, and 20% for the predicting set. The sales amounts in the training, testing, and predicting sets are shown in Fig. 5(a), (b) and (c), respectively.

To compare ELM and ELME with the other two methods, the same data and inputs are used for the four methods. The experimental comparisons for GDA, GDX, ELM and ELME are given in Table 7. The performance index ratios between GDA, GDX, ELME and ELM are given in Table 8. From Tables 7 and 8, we can conclude that the ELM and ELME are better than the GDA and GDX algorithms when we consider the training, testing, predicting errors, and the standard deviations.

To analyze the fluctuation of ELM, the ratios between std and $\bar{\mathrm{mse}}$ are given in Table 9. From Table 9, we can see that the fluctuation of the training, testing and predicting errors in experiment 2 is much less than the corresponding results in experiment 1. In Table 8, the training, testing, and predicting errors of ELME are 99.51%, 99.88%, and 99.62% of the corresponding performance indices of ELM, respectively. The extent that the errors are decreased by the ELME is not as large as the result in experiment 1.

Table 14
The cvs of sales amounts for three products in one year

Product	1	2	3
cvs	0.7128	0.4845	0.6388

Table 15
The three experimental results for ELM

	1	2	3
μ_{mse}^{tr}	0.0001	0.1556	0.4930
std_{mse}^{tr}	0.0001	0.0010	0.0979
μ_{mse}^{te}	2.0640	1.4094	0.4297
std_{mse}^{te}	0.0228	0.0425	0.1337
μ_{mse}^{pr}	0.0002	0.6515	0.4952
std_{mse}^{pr}	0.0002	0.0366	0.1493

Table 16
The training time of four methods with different number of samples

NS	Δt(GDA)	Δt(GDX)	Δt(ELM)	Δt(ELME)
200	0.5934	0.5353	0.0122	1.2190
1000	0.8814	0.6414	0.0195	1.9530
5000	3.2447	2.2552	0.0786	7.8590
10,000	9.9348	6.4550	0.1528	15.2820

3.3. Experiment 3

In this experiment, the Jacket data is used as the experimental data. The color, size and price of the clothes are selected as the significant factors. 200 samples are used as the experimental data, in which 60% of the data points are used for the training set, 20% for the testing set, and 20% for the predicting set, and the respective sales amounts are shown in Fig. 6(a), (b), and (c).

The experimental comparisons for GDA, GDX, ELM and ELME are given in Table 10. The performance index ratios between GDA, GDX, ELME and ELM are given in Table 11. From Tables 10 and 11, the ELM and ELME are better than the GDA and GDX algorithms when we consider the training, testing, predicting errors, and the standard deviations.

In Table 11, the training, testing, and predicting errors of ELME are 82.07%, 65.05% and 78.51% of the corresponding performance indices of ELM, respectively. The ratios between std_{mse} and μ_{mse} are given in Table 12.

In general, considering the predicting accuracy and stability, ELM and its extension ELME are better than GDA and GDX. From the above simulation results, we can see that ELME can have a higher accuracy than ELM especially when the experimental results of ELM have a larger fluctuation.

4. Further discussions

The forecasting accuracy of an approach is often influenced by the inherent nature of a product and its sales pattern. In this section, the effect of sales amount fluctuation on the prediction's accuracy of ELM is investigated. The sales amount fluctuations are measured by the coefficients of variation (cvs). Sales data of three products are studied, each with a different sales feature. Table 13 shows the mean (μ) and standard deviation (σ) of sales amounts for three products in one year.

The cvs of sales amounts for three products in one year is given in Table 14. Table 15 shows three experimental results for ELM.

From Tables 14 and 15, it can be observed that there is a relationship between the cvs and the forecasting errors. When the cv is larger (i.e., the fluctuation of the product's demand is larger), the training, testing, and predicting errors of ELM are generally lower, and vice versa. This result shows that the fluctuation in demands of product has a great impact on the forecasting accuracy, and the ELM is especially accurate when forecasting for products with a large cv. The reason is that the neural network tends to be biased with data of relatively small variance (small cv) [14], and the bias will decrease the forecasting accuracy. Intuitively, the ANN is a soft computing method. It learns the pattern (sale amount) by training samples and remembering the knowledge using its structure. When the patterns are too close (with a small cv), the ANN can't remember and distinguish these patterns. On the contrary, when the pattern has a larger cv, it is easier to be learned by the ANN.

Table 17
The training time ratios between GDA, GDX and ELM

NS	$\frac{\Delta t(\mathrm{GDA})}{\Delta t(\mathrm{ELM})}$	$\frac{\Delta t(\mathrm{GDA})}{\Delta t(\mathrm{ELM})}$
200	48.639	43.877
1000	45.2	32.892
5000	41.281	28.692
10,000	65.018	42.245

To compare the computation efficiency of the four methods, Table 16 gives the training time of four methods (Δt) with different number of samples (NS). It should be pointed out that the training time of ELME is the sum of training time that is consumed in 100 trials of ELM.

The training time ratios between GDA, GDX and ELM are given in Table 17. From Tables 16 and 17, we can see that ELM has a very shorter training time than GDA and GDX.

5. Conclusion

Owing to market competition and globalization, sales forecasting plays a more and more prominent role in a decision support system of a commercial enterprise. It is especially true in fashion business. How to develop more accurate and timely sales forecasting methods becomes an important research topic. In this paper, we apply a relatively novel neural network technique, extreme learning machine (ELM) and its extension, to fashion sales forecasting. It is known that ELM not only has a higher generalization performance than the traditional gradient-based learning algorithms but it also avoids many difficulties faced by gradient-based learning methods such as stopping criteria, learning rate, learning epochs, local minima, and the over-tuned problem. Therefore, ELM is selected to analyze fashion sales forecasting on the data provided by a Hong Kong fashion retailer in this paper. Using this method, the most significant factors affecting the sales amount are selected as the inputs of ELM. As an extension, the arithmetic mean value of multiple trials is used as the final predicted sales forecasting amount. Our experiments have successfully demonstrated that both ELM and its extension can be employed in sales forecasting for fashion retailing and they can produce smaller predicting errors than some other sales forecasting methods based on two well-established backpropagation neural networks (BPNN). The ELM is thus a promising tool in sales forecasting for fashion retailers. Moreover, in our approach, by using the statistical mean value of multiple trials as the final forecasting result, the ELM forecasting result is more stable than the BPNN algorithms. This makes the ELM approach a better choice when employing to the practical forecasting of fashion sales, in which the BPNN result can be very unstable. This study also provides a guide for the selection of the important sales forecasting factors such as design factors (size, color, etc.) and the price factor.

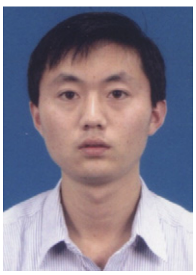
Acknowledgments

We sincerely thank the editor, and the anonymous reviewers for their many helpful suggestions and kind advice. Tsan-Ming Choi's research is partially supported by the Research Grants Council of Hong Kong, grant's number PolyU5145/06E, and the research funding by The Hong Kong Polytechnic University. Kin-Fan Au's research is partially supported by the Research Grants Council of Hong Kong, Grant's number PolyU5101/05E. Thanks are given to Tiaojun Xiao, Chun-Hung Chiu, Hye Kyung Im, Danqin Yang for their comments on the earlier draft of this paper.

References

[1] E. Alfaro, N. García, M. Gámez, D. Elizondo, Bankruptcy forecasting: an empirical comparison of AdaBoost and neural networks, Decision Support Systems 45 (1) (2008) 110–122.
[2] K.F. Au, N.Y. Chan, Quick response for Hong Kong clothing suppliers: a total system approach, Proceedings of the 13th Annual Conference of the Production and Operations Management Society (San Francisco, USA, 2002.
[3] K. Chakraborty, K. Mehrotra, C.K. Mohan, Forecasting the behavior of multivariate time series using neural networks, Neural Networks 5 (1992) 961–970.
[4] P.C. Chang, Y.W. Wang, Fuzzy Delphi and back-propagation model for sales forecasting in PCB industry, Expert Systems with Applications 30 (4) (2006) 715–726.
[5] T.M. Choi, Quick response in fashion supply chains with dual information updating, Journal of Industrial and Management Optimization 2 (2006) 255–268.
[6] T.M. Choi, Pre-season stocking and pricing decisions for fashion retailers with multiple information updating, International Journal of Production Economics 106 (2007) 146–170.
[7] T.M. Choi, P.S. Chow, Mean-variance analysis of quick response program, International Journal of Production Economics 114 (2) (2008) 456–475.
[8] T.M. Choi, D. Li, H. Yan, Optimal two-stage ordering policy with Bayesian information updating, Journal of the Operational Research Society 54 (2003) 846–859.
[9] T.M. Choi, D. Li, H. Yan, Optimal single ordering policy with multiple delivery modes and Bayesian information updates, Computers and Operations Research 31 (2004) 1965–1984.
[10] T.M. Choi, D. Li, H. Yan, Quick response policy with Bayesian information updates, European Journal of Operational Research 170 (2006) 788–808.
[11] H. Demuth, M. Beale, M. Hagan, Neural Network Toolbox for Use with MATLAB, The Mathworks, Inc., Natick, MA, 2006.
[12] G.D. Eppen, A.V. Iyer, Improved fashion buying with Bayesian updates, Operations Research 45 (1997) 805–819.
[13] C. Frank, A. Garg, L. Sztandera, A. Raheja, Forecasting women's apparel sales using mathematical modeling, International Journal of Clothing Science and Technology 15 (2) (2003) 107–125.
[14] S. Geman, E. Bienenstock, R. Doursat, Neural networks and the nias/variance dilemma, Neural Computation 4 (1) (1992) 1–58.
[15] D.S. Huang, Systematic Theory of Neural Networks for Pattern Recognition, Publishing House of Electronic Industry of China, Beijing, 1996.
[16] G.B. Huang, Learning capability and storage capacity of two-hidden-layer feedforward networks, IEEE Transactions on Neural Networks 14 (2) (2003) 274–281.
[17] G.B. Huang, H.A. Babri, Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions, IEEE Transactions on Neural Networks 9 (1) (1998) 224–229.
[18] G.B. Huang, L. Chen, C.K. Siew, Universal approximation using incremental constructive feedforward networks with random hidden nodes, IEEE Transactions on Neural Networks 17 (4) (2006) 879–892.
[19] G.B. Huang, Q.Y. Zhu, C.K. Siew, Extreme learning machine: a new learning scheme of feedforward neural networks, Proceedings of the International Joint Conference on Neural Networks (IJCNN2004) (Budapest, Hungary, July 2004, pp. 985–990.
[20] G.B. Huang, Q.Y. Zhu, C.K. Siew, Extreme learning machine: theory and applications, Neurocomputing 70 (2006) 489–501.
[21] C.L. Hui, T.W. Lau, S.F. Ng, K.C.C. Chan, Neural network prediction of human psychological perception of fabric hand, Textile Research Journal, 74 (2004) 375–383.
[22] C.L. Hui, S.F. Ng, A new approach for prediction of sewing performance of fabrics in apparel manufacturing using artificial neural networks, Journal of the Textile Institute 96 (2005) 401–406.
[23] J.S.R. Jang, N. Gulley, Fuzzy Logic Toolbox for Use with MATLAB, The Mathworks, Inc., Natick, MA, 2006.
[24] J. Kittler, M. Hatef, P.W. Duin, J. Matas, On combining classifiers, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (3) (1998) 226–239.
[25] R.J. Kuo, A sales forecasting system based on fuzzy neural network with initial weights generated by genetic algorithm, European Journal of Operational Research 129 (2001) 496–517.
[26] R.J. Kuo, K.C. Xue, A decision support system for sales forecasting through fuzzy neural networks with asymmetric fuzzy weights, Decision Support Systems 24 (2) (1998) 105–126.
[27] R.J. Kuo, K.C. Xue, Fuzzy neural networks with application to sales forecasting, Fuzzy Sets and Systems 108 (1999) 123–143.
[28] G. Lachtermacher, J.D. Fuller, Backpropagation in time-series forecasting, Journal of Forecasting 14 (1995) 381–393.
[29] M. Lam, Neural network techniques for financial performance prediction: integrating fundamental and technical analysis, Decision Support Systems 37 (4) (2004) 567–581.
[30] W. Leigh, R. Purvis, J.M. Ragusa, Forecasting the NYSE composite index with technical analysis, pattern recognizer, neural network, and genetic algorithm: a case study in romantic decision support, Decision Support Systems 32 (4) (2002) 361–377.
[31] W.Y. Liang, C.C. Huang, Agent-based demand forecast in multi-echelon supply chain, Decision Support Systems 42 (2006) 390–407.
[32] Z.L. Sun, K.F. Au, T.M. Choi, A hybrid neuron-fuzzy inference system through integration of fuzzy logic and extreme learning machines, IEEE Transactions on Systems, Man and Cybernetics-Part B: Cybernetics 37 (5) (2007) 1321–1331.
[33] Z.L. Sun, D.S. Huang, C.H. Zheng, L. Shang, Optimal selection of time lags for temporal blind source separation based on genetic algorithm, Neurocomputing 69 (7–9) (2006) 884–887.
[34] L.M. Sztandera, C. Frank, B. Vemulapali, Predicting women's apparel sales by soft computing, Lecture Notes in Artificial Intelligence 3070 (2004) 1193–1198.
[35] L.M. Sztandera, C. Frank, B. Vemulapali, Prediction of women's apparel sales using soft computing methods, Lecture Notes in Artificial Intelligence 3215 (2004) 506–512.
[36] S. Tamura, M. Tateishi, Capabilities of a four-layered feedforward neural network: four layers versus three, IEEE Transactions on Neural Networks 8 (2) (1997) 251–255.
[37] S. Thomassey, A. Fiordaliso, A hybrid sales forecasting system based on clustering and decision trees, Decision Support Systems 42 (2006) 408–421.
[38] P.D. Wasserman, Neural Computing: Theory and Practice, Van Nostrand Reinhold, New York, USA, 1989.
[39] A.S. Weigend, D.E. Rumelhart, B.A. Huberman, Generalization by weight-elimination with application to forecasting, Advance Neural Information Processing System 3 (1990) 875–882.
[40] H. White, Artificial Neural Networks: Approximations and Learning Theory, Blackwell, Oxford, UK, 1992.

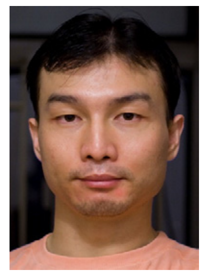
[41] T.J. Xiao, X.T. Qi, Price competition, cost and demand disruptions and coordination of a supply chain with one manufacturer and two competing retailers, Omega 36 (2008) 741–753.
[42] T.J. Xiao, X.T. Qi, G. Yu, Coordination of supply chain after demand disruptions retailers compete, International Journal of Production Economics 109 (2007) 162–179.
[43] T.J. Xiao, D.Q. Yang, Price and service competition of supply chains with risk-averse retailers under demand uncertainty, International Journal of Production Economics 114 (2008) 187–200.
[44] H. Yoo, R.L. Pimmel, Short-term load forecasting using a self-supervised adaptive neural network, IEEE transactions on Power Systems 14 (2) (1999) 779–784.
[45] Y. Yu, T.M. Choi and K.F. Au, Web-based Sales Forecasting System with Applications in Fashion Business, Forthcoming in Research Journal of Textile and Apparel. in press.
[46] G.Q. Zhang, B.E. Patuwo, M.Y. Hu, Forecasting with artificial neural networks: the state of the art, International Journal of Forecasting 14 (1998) 35–62.
[47] Q.Y. Zhu, A.K. Qin, P.N. Suganthan, G.B. Huang, Evolutionary extreme learning machine, Pattern Recognition 38 (2005) 1759–1763.



Zhan-Li Sun received his PhD degree from the University of Science & Technology of China in 2005. He worked as a Research Associate at The Hong Kong Polytechnic University in 2006–2007 and Research Fellow at Nanyang Technological University in 2007–2008. He is currently with the National University of Singapore. His research interests include machine learning, signal and image processing.

Tsan-Ming Choi (Jason) received his PhD in supply chain management from The Chinese University of Hong Kong. He is currently an assistant professor at The Hong Kong Polytechnic University. His current research interests mainly focus on supply chain management. Over the past few years, he has actively participated in a variety of research and consultancy projects in supply chain management. He has consulted companies such as Bossini, First Glory, and Sun-Hing Ltd. He has published in journals such as Computers and Operations Research, Decision Support Systems, European Journal of Operational Research, IEEE Transactions (various), International Journal of Production Economics, Journal of Industrial and Management Optimization, Journal of the Operational Research Society, Omega, etc. He is a member of IEEE and INFORMS.

Kin-Fan AU is an associate professor in the Institute of Textiles and Clothing of The Hong Kong Polytechnic University. His research interest is in the business aspects of fashion and textiles, particularly in the area of global trading of fashion and textile products. Dr. Au has published many papers in textiles and related journals on topics of world trading, offshore production and modelling of textiles and apparel trade.



Yong Yu is currently a research associate at The Hong Kong Polytechnic University. He received his PhD from the Hong Kong Polytechnic University. His research interests are computer simulation, cluster computing, and artificial intelligence in business forecasting.