Indian Journal of Fibre & Textile Research
Vol. 32, June 2007, pp. 173-178

Support vector machines for predicting worsted yarn properties

Zhi-Jun Lü[a], Jian-guo Yang, Qian Xiang & Xiao-ling Wang

College of Mechanical Engineering, Donghua University, Shanghai 201620, P R China

Received 18 November 2005; revised received and accepted 11 April 2006

Support vector machines (SVMs) models have been presented for predicting worsted yarn properties using SVM regression algorithms. Model selection which amounts to search in hyper-parameter space is performed to study the suitable parameter conditions. The predictive powers of the SVM models have been estimated and the results are compared with ANN models. It is observed that under the small population circumstances, SVM models are still capable of maintaining the stability of predictive accuracy, and more suitable for noisy and dynamic spinning process.

Keywords: Artificial neural networks, Kernel function, Structure risk minimization, Support vector machines, Worsted yarn
IPC Code: Int.Cl.[8] G06F

1 Introduction

Worsted yarn production is a complex multi-step technological process, and the yarn properties are influenced by many factors such as fibre properties and machine sets.[1,2] Spinners have always attempted to estimate yarn qualities from existing spinning data in advance to select appropriate spinning parameters accordingly. Therefore, many researchers[3-5] have adopted different techniques for predicting yarn qualities from numerous spinning data over the past years, such as mechanistic model, linear multiple regression and even artificial intelligent methods. Among them, the artificial neural network (ANN) is a kind of successful technique which has been widely used in many aspects of textile production[6], including prediction of yarn properties.[7] However, a shortcoming of artificial neural networks is that it is very difficult to intuitively know at what point the system will over-fit the data, which could result in ANN model instability.[8] The increasing quality demands from the textile manufacturer make clear the need to explore novel ways of quality prediction furthermore.

In recent years, support vector machines (SVMs), which is a statistical learning theory based machine leaning formalism, is gaining popularity due to its many attractive features and promising empirical performance.[9] Unlike ANN models, which rely on empirical risk minimization (ERM), SVM models are based on the principle of structure risk minimization (SRM), which equips the later with greater potential to generalize. Since the foundation of the SVMs paradigm was laid down by Vapnik[10] and Scholkőpf *et al.*[11] in mid 1990, the applications in many engineering fields have emerged, such as text categorization, computer vision, bioinformatics, and even fabric defect detection.[9,12] However, up to now, the studies related to predict yarn properties have not been reported. In this work, an attempt has been made to predict the worsted yarn properties from a real data set to assess the application of support vector machines as a mill-specific prediction tool. Results have been obtained from comparison with those of ANN models. The relative algorithm, model selection and experiments are presented in detail.

2 SVM Regression Algorithms

The main objective of regression is to approximate a function $f(x)$ from a given noisy set of samples $G = (x_i, y_i)_{i=1}^{N}$. The basic idea of support vector machines for regression is to map the data $\boldsymbol{x}$ into a high dimensional feature space via a nonlinear mapping and to perform a linear regression in this feature space, as shown below:

$$f(x) = \sum_{i=1}^{D} w_i \phi_i(x) + b \qquad \ldots(1)$$

where $\{\phi_i(x)\}_{i=1}^{D}$ are the features; and b and $\{w_i\}_{i=1}^{D}$, the coefficients that have to be estimated from the data. Thus, a nonlinear regression in the low dimensional input space is transferred to a linear

*To whom all the correspondence should be addressed.
E-mail: Lvzj@mail.dhu.edu.cn

regression in a high dimensional (feature) space. The coefficients $\{w_i\}_{i=1}^{D}$ can be determined from the data by minimizing the function, as shown below:

$$R_{reg} = R_{emp} + \lambda\|w\|^2 = \frac{1}{N}\sum_{i=1}^{N} |f(x_i) - y_i|_\varepsilon + \lambda\|w\|^2 \quad \ldots(2)$$

where λ is a regularization constant and the Vapnik's ε-insensitive loss function defined by the following relationship:

$$|f(x)-y|_\varepsilon = \begin{cases} \{|f(x)-y|-\varepsilon\}, & (|f(x)-y| \geq \varepsilon) \\ 0 & otherwise \end{cases} \quad \ldots(3)$$

where ε is a precision parameter representing the radius of the tube located around the regression function (Fig. 1); the region enclosed by the tube is known as ε-insensitive zone. The SVM regression algorithm attempts to position the tube around the data, as shown in Fig. 1. The optimization criterion in Eq. (3) penalizes those data points whose y values lie more than ε distance away from the fitted function $f(x)$. In Fig. 1, the size of the stated excess positive and negative deviations are depicted by ξ and ξ* respectively, which are termed as slack variables. Outside the [−ε, ε] region, the slack variables assume non-zero values. The SVM fits $f(x)$ to the data in a manner such that (i) the training error is minimized by minimizing ξ_i and ξ_i^*, and (ii) $\|w\|^2$is minimized to increase the flatness of $f(x)$ or penalize over complexity of the fitting function. Vapnik showed that the following function possessing finite number of parameters can minimize the regularized function in Eq (2):

$$f(x,\alpha,\alpha^*) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*)k(x_i,x) + b \quad \ldots(4)$$

where $\alpha_i\alpha_i^* = 0, \alpha_i, \alpha_i^* \geq 0$, $i = 1,\Lambda,N$ and the kernel function $k(x_i,x)$ describes the dot product in the D-dimensional feature space, as shown below:

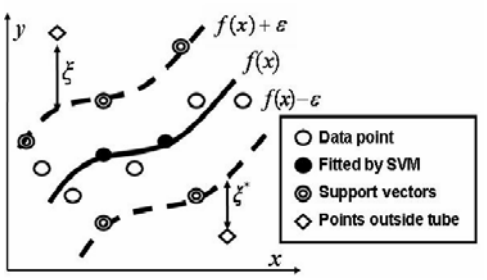

Fig. 1 — A schematic illustration of SVMs using ε-sensitive loss function

$$k(x_i, x) = [\phi(x_i).\phi(x)] \quad \ldots(5)$$

It is important to note that the features Φ_j need not be computed; rather kernel function that is very simple and has a known analytical form is needed. The only condition required is that the kernel function has to satisfy Mercer's condition. Some of the mostly used kernels include polynomial, radial basis function, and sigmoidal. Consider that for Vapnik's ε-insensitive loss function, the Lagrange multipliers α_i,α_i^* are sparse, i.e. they result in nonzero values after the optimization by Eq. (2) only if they are on the boundary, which means that they satisfy the Karush–Kuhn–Tucker conditions. The coefficients α_i,α_i^* are obtained by maximizing the following form:

$$R(\alpha^*,\alpha) = -\frac{1}{2}\sum_{i,j=1}^{N} (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)k(x_i,x_j) - \varepsilon\sum_{i=1}^{N} (\alpha_i^* + \alpha_i) + \sum_{i=1}^{N} y_i(\alpha_i^* - \alpha_i) \quad \ldots(6)$$

$$\text{Subject to} \sum_{i=1}^{N} (\alpha_i^* - \alpha_i) = 0,\ 0 \leq \alpha_i^*, \alpha_i \leq C \quad \ldots(7)$$

Only a number of coefficients α_i,α_i^* will be different from zero, and the data points associated to them are called support vectors. Parameters C and ε are free and have to be decided by the user. Computing b requires a more direct use of the Karush-Kuhn-Tucker conditions that lead to the quadratic programming problems. The key idea is to pick those values for a point x_k on the margin, i.e. α_k or α^*_k in the open interval (0, C). One x_k would be sufficient but for stability purposes it is recommended that one take the average over all points on the margin. More detailed description of SVM for regression can be found in literature.[10, 11]

3 Materials and Methods

3.1 Data Acquisition

Twenty-six different data samples from real worsted spinning were acquired for the study (Table 1). The bold digits (sample numbers w21-w26) indicate the data sets used for testing purpose. To simplify the problem, like most ANN models[8], some fibre properties and processing information have been selected as the SVM model's inputs, such as mean fibre diameter (MFD, μm), diameter distribution (CV_D, %), hauteur (HT, mm), fibre length distribution (CV_H, %), short fibre content (SFC, %), yarn count

Table 1—Sample details and yarn properties data

Sample No.	Material and process variable										Yarn properties			
	MFD	CV_D	HT	CV_H	SFC	DR	SS	TN	TW	CT	CV%	ED	EB	BF
W1	19.50	21.8	72.7	44.7	11.2	20.5	8800	34	769	12.2	20.5	80	15	120
W2	18.6	21.3	69.8	48.9	13.5	19	8800	33	680	13.2	20.3	89	14	80
W3	18.8	22.1	66.3	46.2	11	21.28	8600	32	751	13.2	21.2	90	12	65
W4	18.8	22.1	66.3	46.2	11	21.28	8600	32	741	13.2	19.5	65	10	80
W5	19.7	21.4	73	49.2	12.6	20.7	8800	32	730	14.5	20.6	89	14	85
W6	17.5	22	65	47	12.8	18.5	8800	32	730	13.2	20.1	70	12	72
W7	18.8	22.1	66.3	46.2	11	15.42	9000	29	900	20.7	15.32	30	17.3	168.5
W8	17.8	23	65.6	48	12.8	19	8800	32	750	12.8	20.2	65	11	68
W9	16.7	19.2	65	40.6	10.2	12.8	8900	29	990	17.2	15.92	35	17.6	123
W10	22.5	20.8	67.5	43	11.2	17.92	8900	29	640	17.5	20.01	80	10.9	75
W11	17.6	21.1	64.8	44.4	12	19.8	8500	36	833	11.1	23.2	72	11	63
W12	19.50	21.8	72.7	44.7	11.2	20.5	8800	34	769	12.2	20.3	75	15	125
W13	18.7	22.2	64.7	49.3	12	15.42	9000	29	901	20.7	16.11	30	18	130
W14	16.7	19.2	65	40.6	12	21.23	8500	36	833	10.4	23.2	72	11	63
W15	21.2	22.1	73.3	48.7	12.6	19.2	9000	29	660	16.7	15.92	35	17.6	123
W16	17.6	21.7	64.8	44.4	12	21.5	8800	33	788	11.6	20.3	80	19.9	80
W17	18.8	22.1	66.3	46.2	11	21.28	8800	32	751	13.2	21	85	14	70
W18	21.2	22.1	73.3	48.7	12.6	17.92	8900	29	640	17.5	20.01	80	10.9	75
W19	18.5	21.8	67.6	51	14.9	21.28	8600	32	741	13.2	19.5	65	10	80
W20	17.5	21.3	65.4	50.1	16.5	19.8	8800	34	806	10.9	20.05	66	18	72
W21	**19.7**	**21.4**	**73**	**49.2**	**12.6**	**19.8**	**9000**	**32**	**690**	**15.2**	**21**	**60**	**11.2**	**103**
W22	**17.5**	**21.7**	**65**	**47**	**14.2**	**22.8**	**8800**	**32**	**741**	**13.2**	**19.85**	**60**	**12**	**75**
W23	**19.8**	**21.3**	**74**	**45.4**	**9.4**	**14.78**	**9000**	**28**	**880**	**21.6**	**16.49**	**30**	**15**	**152**
W24	**17.5**	**21.7**	**65**	**47**	**14.2**	**22.8**	**8800**	**32**	**741**	**13.2**	**20.3**	**65**	**12**	**75**
W25	**16.7**	**20**	**58.6**	**46.4**	**17.4**	**21.12**	**8800**	**34**	**835**	**10.4**	**19.9**	**60**	**10**	**70**
W26	**18.8**	**22.1**	**66.3**	**46.2**	**11**	**21.28**	**8600**	**32**	**751**	**13.2**	**21.2**	**90**	**12**	**65**

MFD—mean fibre diameter (μm), CV_D—diameter distribute (%), HT—hauteur (mm),
CV_H—fibre length distribution (%), SFC—short fiber content (%), DR—draft ratio, SS—spinning speed (rpm),
TN—traveler number, TW—twist (tpm), CT—yarn count(*tex*), CV*%*—unevenness (%),
ED—end-down per 1000 spindle hour, EB—elongation-at-break (%), BF—break force (cN)

(CT, tex), twist (TW, tpm), draft ratio (DR), spinning speed (SS, rpm), and traveler number (TN). Four yarn properties, namely unevenness (CV %), elongation-at-break (EB, %), break force (BF, cN) and end-down per 1000 spindle hour (ED), served as the SVM model's outputs. Before training and modeling, the data in the process database had to be normalized so that they were bounded within the prescribed range of 1 and 0. Scaling of the input value (v_j) was carried out according to the following equation:

$$x_j = \frac{v_j - \min(v_{\mathrm{IK}\,n})}{\max(v_{\mathrm{IK}\,n}) - \min(v_{\mathrm{IK}\,n})} \quad \ldots(8)$$

where x_j is the scaled value; and $\min(v_{\mathrm{IK}\,n})$ and $\mathrm{man}(v_{\mathrm{IK}\,n})$, the respective maximum and minimum values within each input array.

3.2 Model Selection

In any predictive learning task, an appropriate representation of examples as well as the model and parameter estimation method should be selected to obtain a high level of performance of the learning machine. In fact, the task of learning amounts to selecting the model of optimal complexity and estimating parameters from training data. In this study, ν-support vector regression machines[10] were used and under the approach, the following parameters[13] were selected:

- The penalty term *C* which determines the tradeoff between the complexity of the decision function and the number of training examples misclassified.
- The sparsity parameter ν in accordance with the noise in the output values to get the highest generalization accuracy.
- The kernel function $k(x,y)$.

As reported earlier[14], the sparsely parameter v usually may be chosen in the interval [0.3, 0.6], here v=0.583. The following radial basis function (RBF) kernel is used:

$$K(x, y) = \exp(-\|x - y\|^2 / 2\sigma^2),\ \sigma > 0 \quad \ldots(9)$$

where the width of the RBF kernel parameter σ can be determined, in general, by an iterative process selecting an optimum value based on the training data set.

3.3 Cross-validation and Grid-search

There are two key parameters selected, namely C and σ, while using RBF kernels to model. Unfortunately, it is not known beforehand which C and σ are the best for one problem. The goal is just about to identify good (C, σ) so that the model can accurately predict unknown data, i.e. testing data. Therefore, a common approach is to separate training data into two parts, of which one part is considered unknown in training the model. Then the prediction accuracy on this set can more precisely reflect the performance on predicting unknown data. The procedure for improved model is called cross-validation. The cross-validation procedure can also prevent the over-fitting problem furthermore. In n-fold cross-validation, the training set was firstly divided into n subsets of equal size. Sequentially, one subset is tested using the model trained on the remaining $n - 1$ subsets. In this study, the regression function was built with a given set of parameters {C, σ}. The performance of the parameter set is measured by the computational risk, here mean squared error [MSE, Eq.(10)] on the last subset. The above procedure is repeated n times, so that each subset is used once for testing. Averaging the MSE over the n trials gives an estimate of the expected generalization error for training on sets of size $\frac{n-1}{n} \times l$, where l is the number of training data.

$$MSE = \frac{\sum_{i=1}^{m} (y_{ti} - y_{pi})^2}{m} \quad \ldots(10)$$

where m is the sample number of tested subset in the training set $m = \frac{l}{n}$; and y_{ti} and y_{pi}, the i^{th} observed and predicted values respectively. In order to capture the better pairs of C and σ, a grid-search on C and σ is employed. Firstly, in term of possible range of the two parameters, C and σ were divided into q pairs, then each pair of the parameters was tried using cross-validation and the one with the best cross-validation accuracy was picked up as optimal parameters of the model.

3.4 Artificial Neural Networks Model

Artificial neural network (ANN) is also a powerful data modeling tool simulating the behaviour of biological human neurons. The main difference between the ANN and the SVM is in the principle of risk minimization (RM).[10] In case of SVMs, structural risk minimization (SRM) principle is used minimizing an upper bound on the expected risk, whereas in ANNs the traditional empirical risk minimization (ERM) is used, minimizing the error on the training data. To demonstrate the difference in generalization performance between SVM model and ANN model under the given structure, ANN predictive models were also considered in the study. The most representative back propagation algorithm was employed in model training. The relationship between model inputs and outputs was completely same as the SVM model. The training function 'trainbr' has been transferred from Matlab ANN toolbox to realize the training of these models. The transformation function of one hidden neuron is logsig while purelin is the output layer function which can be also calculated by means of ANN toolbox of Matlab7.0 software.[15]

4 Results and Discussion

4.1 Optimizing SVM Models

One of the primary aspects of developing a SVM regression model is the selection of the penalty term C and the width of the RBF kernel parameter σ. To optimize the two parameters, the grid-search method was applied in the present work. In fact, optimizing the model parameters needs an iterative process which can continuously shrink the searching area and as a result, obtain a satisfying solution. Table 2 shows the final searching area and optimal values of the four SVM models respectively. The results of the grid-search experiments are shown in Figs 2-5 as along with the corresponding contour plots for the risk (mean squared error, MSE) versus σ and C.

4.2 Comparing Prediction Performance of SVM and ANN Models

After the completion of model development or training, all the four models based on SVM or ANN were subjected to the unseen testing data set. Statistical parameters, such as the correlation

Table 2—Final search area and optimal values of σ and *C*

Output parameter	Searching area		Optimal value
Unevenness (%)	σ ∈[0.5, 2.5]	*C* ∈[1500, 1700]	σ = 0.973, *C* = 1606
Elongation-at-break (%)	σ ∈[0.01, 0.1]	*C* ∈[2, 15]	σ = 0.016, *C*= 14.55
Breaking force (cN)	σ ∈[0.003, 0.03]	*C* ∈[50,150]	σ = 0.012, *C* = 101.19
Ends-down	σ ∈[0.01, 0.05]	*C* ∈[1, 30]	σ = 0.287, *C* = 2.975

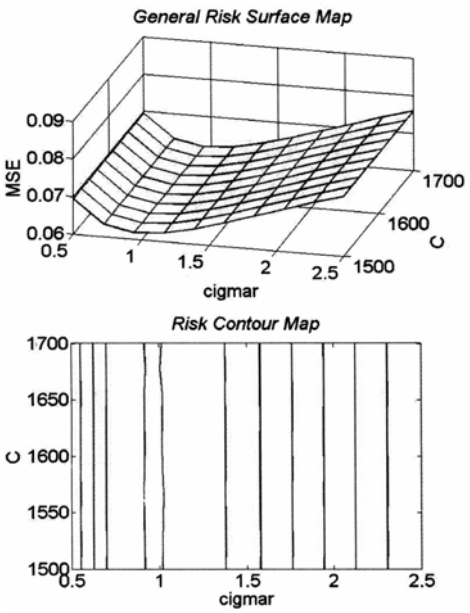

Fig.2 — Surface plot and contour for the predictive risk of yarn unevenness (CV%)

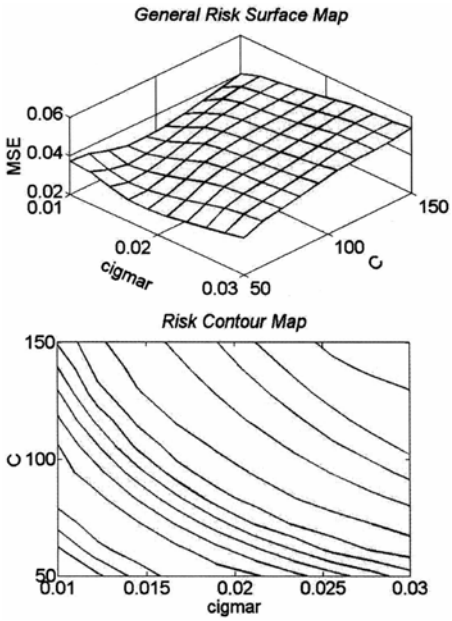

Fig.4 — Surface plot and contour for the predictive risk of breaking force (BF)

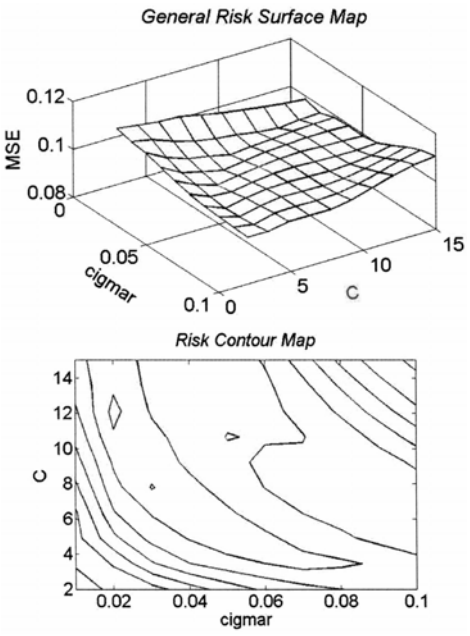

Fig.3 — Surface plot and contour for the predictive risk of elongation-at-break (EB)

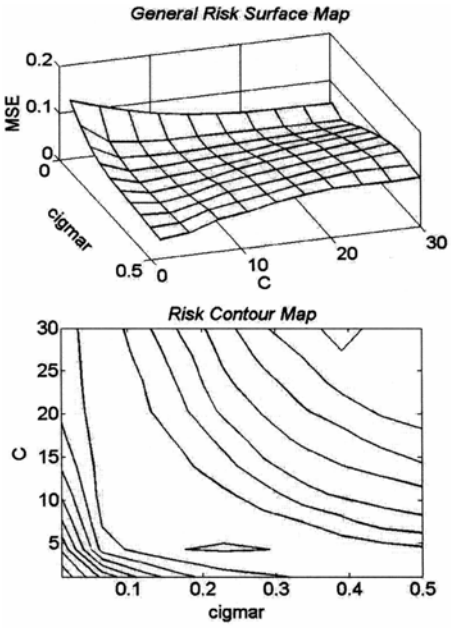

Fig.5 — Surface plot and contour for the predictive risk of ends-down (ED)

Table 3 — Comparison of predictive power of SVM-based and ANN-based models

Sample No.	ANN models				SVM models			
	CV%	EB	BF	ED	CV%	EB	BF	ED
W21	19.32	13.81	113.89	70.41	19.66	12.85	116.24	72.06
W22	20.52	16.55	61.91	75.78	20.88	12.25	76.87	72.40
W23	15.62	12.32	153.46	39.40	16.84	15.59	156.57	42.22
W24	20.66	16.55	61.91	75.78	20.75	12.25	76.87	72.40
W25	22.60	19.77	47.00	69.84	19.66	12.76	76.86	59.31
W26	20.70	11.87	66.76	79.22	21.20	12.59	66.62	81.27
Correlation coefficient (R)	0.76	0.67	0.96	0.88	0.88	0.80	0.99	0.91
Mean squared error	0.01	0.12	0.07	0.03	0.003	0.05	0.01	0.03
Mean error%	5.73	24.35	13.67	19.99	2.85	9.23	5.52	17.29
Cases with over 10% error	1	6	5	6	0	2	2	3

coefficient (R) between the actual and the predicted values, mean squared error, and mean error%, were used to compare the predictive power of the SVM and ANN models. Results are shown in Table 3. It has been observed that for ANN models, the mean error (%) of three models is more than 10% except for the CV% which remains about 5%, the correlation coefficients (R) of the CV% and EB models are very low, i.e. 0.76 and 0.67 respectively. However, for SVM models, the mean error is less than 10%, except for the ED which is still high, and the correlation coefficient (R) of all the models is improved to more than 0.80. On the other hand, the cases with over 10% error also decrease from 5 and 6 in ANN models to 2 and 3 in SVM models. In fact, among all the four yarn properties considered, end-down per 1000 spindle hours could be affected by different operators and observers to some extent[2,16], which data often result in undermining the prediction accuracy of various regression models. Even so, for ED almost all statistical parameters using SVM model seem to be much better than using ANN model.

5 Conclusions

The studies indicate that under the real data sets and small population circumstances, the predictive power of ANN models appears to decrease, but SVM models are still capable of maintaining the stability of predictive accuracy to some extent. Therefore, SVM models are more suitable for noisy and dynamic spinning process. Of course, like other emerging industrial techniques, applied issues on SVM reaffirm the due commitment to their further development and investigation, such as the problems how to design the kernel function and how to set the SVM hyper-parameters to make the industrial model development more easily. SVMs are becoming a very promising alternative for yarn producers to optimize the spinning process.

Acknowledgement

The authors are thankful to the National Science and Technology Department, People Republic of China, for funding the project under the contract number 2006BAF01A44.

References

1 Sette S, Wyns B & Boullart L, *Eng Appl Artif Intell,* 17(2004) 199.
2 Lord Peter R., *Handbook of Yarn Production* (*Technology, Science and Economics*) (Woodhead Publishing Limited, Abinhton England), 2003, 276.
3 Hearle J W S, Grosberg P & Backer S, *Structural Mechanics of fibres, Yarns and Fabrics*, Vol.1 (Wiley-Interscience, NY), 1969 .
4 Majumdar Abhijt, Majumdar Prabal Kumar & Sarkar Bijan, *Indian J Fibre Text Res*, 30(2005) 19.
5 Cheng Y S J & Cheng K P S, *Text Res J*, 74(8) (2004) 718.
6 Chattonpadhyay R & Guha A, *Text Prog* , 35(2004) 1.
7 Sette S, Boullart L, Van Langenhove L & Kiekens P, *Text Res J* , 67(1997) 84.
8 Beltran Rafael, Wang Lijing & Wang Xungai, *Text Res J*, 74(2004) 757.
9 V David Sanchez A, *Neurocompting*, 55(2003) 5.
10 Vapnik V N, *The Nature of Statistical Learning Theory* (Spring-Verlag, NY), 1995.
11 Scholkőpf B, Burges C & Smola A, *Advances in Kernel Methods—Support Vector Learning* (MA: MIT Press, USA), 1999.
12 Karras D A, *Proceedings, the International Joint Conference on Neural Networks*, Vol.3 (Portland, Oregon, US), 2003, 2322.
13 Cherkassky Vladimir & Ma Yunqian, *Neural Networks*, 17(2004) 113.
14 A Chalimourda thanassia, Scholkopt B & Smola A, *Neural Networks*, 17(2004) 127
15 Howard Demuth, Mark Beale & Martin Hagan, *Neural Network Toolbox User's Guide* (Ver. 4). http://www.mathworks.com/access/helpdesk/help/toolbox/nnet, 2005.

16 James Lappage, *Text Res J*, 75(2005) 507