Khalid A. A. Abakar, Chongwen Yu

College of Textiles, Donghua University,
Shanghai 201620, People's Republic of China
E-mail: yucw@dhu.edu.cn

Application of Genetic Algorithm for Feature Selection in Optimisation of SVMR Model for Prediction of Yarn Tenacity

Abstract
A proposed hybrid genetic algorithm (GA) approach for feature selection combined with support vector machines for regression (SVMR) was applied in this paper to optimise a data set of fibre properties and predict the yarn tenacity property. This hybrid approach was compared with a noisy model of SVMR that used all the data set of fibre properties as input in the prediction. The GA for feature selection was used as the preprocessing stage that aimed to find and select the best attributes or variables that most effect or are related to the prediction of yarn tenacity. The hybrid approach showed better predictive performance than the noisy model. However, the results indicated the suitability of GA for feature selection in the choice of the best fibre property attributes that give the preferred performance and high accuracy in the prediction of yarn tenacity.

Key words: *genetic algorithm, feature selection, support vector machines for regression, yarn properties.*

Introduction

The most important stage after the analysis of the dataset prepared for modelling is the selection of variables (or features) to use as predictors. This process of feature selection is a very important strategy to follow in preparing data for data mining [1]. A major problem of data mining in large data sets with many potential predictor variables is that it is not easy to ensure if the additional variables are added to the model, and it may or may not be able to predict a number better than in regression models or discriminate better between classes in a classification model. But in most cases, using more features can increase the system complexity, which, being a bad feature, may greatly degrade the performance of the system. Thus selecting a subset of the best features is important [1, 2]. On the other hand, the greatest danger in data mining is the overfit problems - where you fit the noise data so a complex model may overfit the training data.

Overfit models look good while training but fall apart on evaluation when used on new data [1]. However, statisticians' investigation experiments researched the relationship between complexity and accuracy and found that the process of feature selection is one way of avoiding the overfit by regulating the complexity of the model.

Feature selection aims to reduce the number of variables in the model, thus lessening the effect of the curse of dimensionality by removing irrelevant or redundant variables, or noisy data, which has the following immediate positive effects for analysis: speeds up processing of the algorithm, enhances data quality, increases the predictive power of the algorithm, and makes the results more understandable and accurate [1]. In recent years some optimisation methods have been used such as the genetic algorithm (GA) with an artificial neural network (ANN) to optimise input parameters for obtaining yarn properties [3 - 5].

In this paper we use evolutionary feature selection, that is the genetic algorithm (GA), to choose a good subset of features of the fibre property data set used for predicting the target detection yarn tenacity property. Thus a hybrid approach based on the combination of GA based feature selection with a support vector machine algorithm for regression (SVMR) was applied to optimise the data set of fibre properties and then predict the yarn tenacity.

Data set

Cotton fibre and corresponding yarn data were collected from a published review [6]. A total of twenty-five different cotton samples were collected and measured by an Uster high volume instrument (HVI). The fibre properties used as an input data set were fibre strength in cN/tex, fibre length in cm, elongation in %, trash content in Cnt, length uniformity in %), yellowness in +b, micronaire in M, and reflectance in Rd, as shown in ***Table 1***. The tenacity (cN/tex) of corresponding yarn was used as the output data and predicted. The samples were divided into two data sets: the training set, containing the first twenty samples used for the development of the models used in this paper, a statistical description of which is shown in ***Table 1***; and the testing set, containing the last five samples used to validate the predictivity of the models which were not used in this paper.

Review of feature selection approach

Feature selection

Feature selection is the study of algorithms for reducing the dimensionality of data to improve machine learning performance. For a data set with N features and M dimensions (or features, attributes), feature selection aims to reduce M to M', and $M' \leq M$. It is an important and widely used approach for dimensionality reduc-

Table 1. *Attribute and statistical description of training set data.*

Attribute number	Attribute name	Range	Mean	Standard deviation	CV%
1	Fibre strength, cN/tex	17.80 - 32.00	22.58	3.652	16.2
2	Fibre length, mm	27.18 - 35.05	29.61	2.225	7.52
3	Trash content, Cnt	0.000 - 9.000	4.150	1.981	45.74
4	Length uniformity, %	0.420 - 0.525	0.454	0.027	5.95
5	Micronaire, M	3.600 - 4.800	4.325	0.361	8.35
6	Elongation, %	0.056 - 0.062	0.058	0.002	3.45
7	Yellowness (+b)	8.90 - 13.50	10.595	1.398	13.19
8	Reflectance, Rd	0.648 - 0.794	0.750	0.043	5.73

tion and is commonly used in applications where original features need to be retained [7].

The structure of a feature selection system consists of four basic components: input, search, evaluation, and output. The *output* of any feature selection system can be either a *ranked list* or a *subset* of features.

In the context of learning, the *input* to a feature selection system is the data which can be 1). supervised – all instances are associated with class labels, as in supervised learning; 2). unsupervised – no class labels are available, as in unsupervised learning, and 3). some instances have class labels and the rest do not, as in semi-supervised learning.

To rank the features or select a feature, the subset can be phrased as a *search* problem in which various search strategies can be employed. Depending on how a feature selection system works together with a learning system, we can apply different models of feature selection such as wrapper, filter, or embedded. A filter model relies on measures about intrinsic data properties. A wrapper model involves a learning algorithm (e.g., a classifier, or a clustering algorithm) in determining the feature quality. An embedded model embeds feature selection in the learning of a classifier.

This necessitates the *evaluation* of feature selection to help us understand how the benefit of the removal of features can help machine learning [7].

The search for relevant features can be realised in two ways: 1). feature ranking – features are ranked according to the intrinsic properties of the data so that the top *k* features can be chosen according to the need or a given threshold, and 2). subset selection – a subset of the feature is selected from the full set of features, and there is no relevant difference between those in the subset selected. Subset selection can be carried out in various ways: forward selection, backward elimination, and random [7].

GA feature selection

Evolutionary feature selection (EFS) is a bio-inspired methodology for explicit modification of input data of a learning system. EFS uses an evolutionary algorithm such as the genetic algorithm (GA). The genetic algorithm (GA) is used to find a mapping from the original data representation space onto a secondary representation space; the mapping consists in dropping off some of the features (attributes) from the original representation so that the dimensionality of the resulting representation space is not greater than that of the original space [7].

A multitude of attributes are often involved in real-world machine learning problems. Those attributes often individually have low informative content and cannot provide a satisfactory performance of the learning system. Particularly this applies to the problem of fibre and yarn property relationships because in most cases these are nonlinear and complex. Thus in prediction with many low-quality attributes, the algorithms might tend to build classifiers that perform poorly in terms of classification or regression accuracy. This problem may be alleviated by removing some features from the original representation space (feature selection). Unfortunately many learning algorithms lack the ability of discovering intricate dependencies between attributes, which is a necessary precondition for successful feature selection. This gap is filled out by EFS, which uses GA to get rid of superfluous attributes and to construct new features.

The benefits of EFS expected include reduced dimensionality of the input space, better predictive accuracy of the learning system, faster training and querying, and better readability of the knowledge acquired [7].

Typically the genetic algorithm plays a role in maintaining a population of solutions (individuals), each of which encoding a particular subset of features. Solutions undergo mutations, crossing-over, and selective pressure that promote the well-performing ones. Selective pressure is exerted by the fitness function, which estimates the solution's quality by measuring some properties of the secondary representation space. ***Figure 1*** illustrates how evolutionary feature selection can work. This usually involves three steps:

1. Decoding of solution (retrieving mapping from the encoded solution).
2. Transforming the training set into a secondary representation space according to the mapping.
3. Estimating the quality of the secondary representation space, which after appropriate conversion (e.g., scaling) becomes the solution's fitness.

SVMR algorithm

Support vector machines (SVM) were originally developed for the classification problem by Vapnik and co-workers [8]. This technique was built on the structural risk minimisation principle. Now, with the introduction of the ε-insensitive loss function, SVM has been extended to solve nonlinear regression estimation. By using the kernel function, SVM plays a role in mapping the data to a high dimensional feature space and then finds a linear separating hyperplane with the maximal margin in that high dimensional space. Several kernel functions are available for nonlinear transformation of the input space, such as linear, polynomial and Gaussian radial basis function (RBF) kernels [9].

The technique has demonstrated much success in prediction studies in textile engineering exactly in the fibre and yarn relationship area, giving powerful accuracy with good performance in many studies[10-12] However, a detailed description of SVM theory was discussed in review[8]. In this paper we used a support vector machine for regression (SVMR) with the common kernel function (radial basis function (RBF)) to optimise the operation of the prediction of the yarn tenacity. The kernel function is defined by

$$K(x_i, x_j) = \exp(-\left\|x_i, x_j\right\| / 2\gamma)^2 \quad (1)$$

where $\gamma > 0$ is the parameter that controls the width of the Gaussian and plays a role in controlling the flexibility of the resulting classifier.

Experimental results and comparisons

Genetic algorithm (GA) for feature selection

The genetic algorithm is a powerful tool for the optimisation procedure that has a large field of application. A genetic algorithm for feature selection was applied to select a good subset of features used for predicting target detection from fibre properties, as well as in the training data set.

The genetic algorithm selection scheme used in this paper was Boltzmann and the genetic algorithm parameters selected were population size, maximum number of generations, p initialise, p mutation, p crossover, and crossover type. The meaning of those parameters was expressed as follows**:**

- **population size**: number of individuals per generation, (*Range:* integer; 1 - +∞),
- **maximum number of generations**: number of generations after which the algorithm was terminated (*Range:* integer; 1 - +∞),
- **p initialise**: initial probability for an attribute to be switched on (*Range:* real; 0.0 - 1.0),
- **p mutation**: probability for an attribute to be changed (*Range:* real; 0.0 - 1.0),
- **p crossover**: probability for an individual to be selected for crossover (*Range:* real; 0.0 -1.0),
- **crossover type**: type of crossover. (*Range:* one_point, uniform, shuffle) The type used in this paper was uniform.

A genetic algorithm works as follows:
Generate an initial population consisting (population_size) of individuals. Each attribute is switched on with probability (p_initialize)

1. For all individuals in the population,
2. Perform mutation, i.e. set used attributes for unused with probability (p_mutation) and vice versa,
3. Choose two individuals from the population and perform crossover with probability (p_crossover). The type of crossover can be selected by (crossover_type),
4. Perform selection, map all individuals to sections on a roulette wheel whose size is proportional to the individual's fitness and draw (population_size) individuals at random according to their probability,
5. As long as the fitness improves, go to 2.

The output of the genetic algorithm is represented as the weights of the attributes and the performance of their best attributes was selected.

Optimisation of parameters and implementation of SVMR

The parameters of the support vector machine for regression (SVMR) such as the complexity parameter C, the value of ε-insensitive loss function, and the width of RBF kernel function γ, were optimised by using the grid search approach, in which after inserting the ranges of parameters and trials depending on the RMSR error, optimal parameters of the yarn tenacity property were chosen.

To implement the support vector machine for regression, we used the sequential minimal optimization (SMO) algorithm to solve the quadratic programming (QP) optimisation problem.

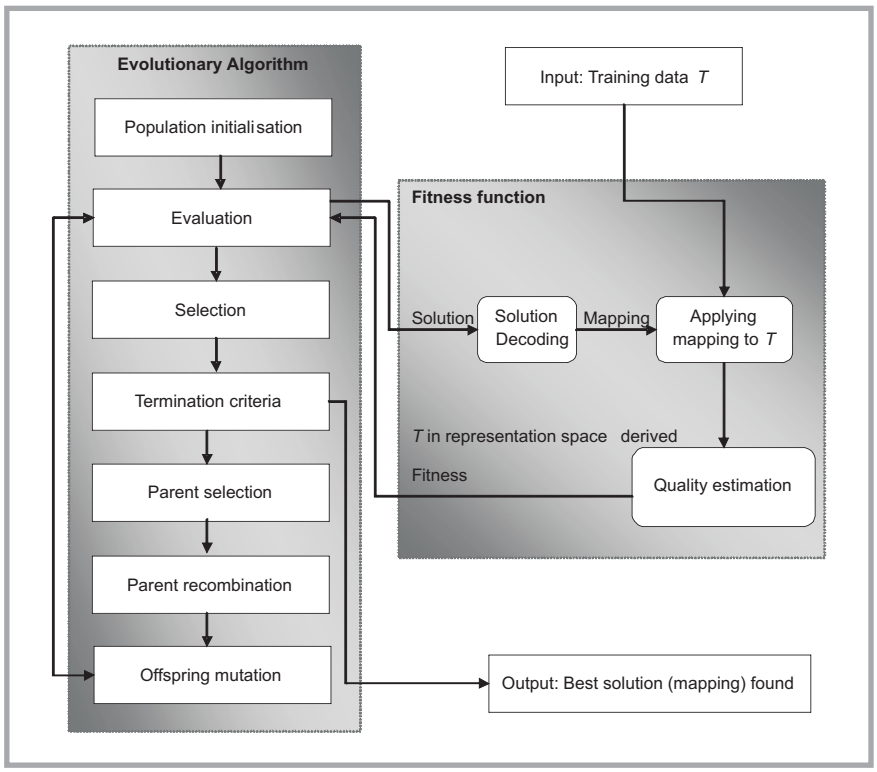

Figure 1. *Evolutionary feature selection and construction.*

During training, the support vector machine requires the solution of a very large quadratic programming (QP) optimisation problem. SMO breaks the large QP problem into a series of smallest possible QP problems, which are solved analytically, avoiding using time-consuming numerical QP optimisation.

The most popular algorithm is the improved SMO (RegSMOImproved) by Shevade et al [13]. This algorithm was proposed to overcome the limitation of the other algorithm called the original SMO algorithm (RegSMO), described by Smola and Schoelkopf [14] and proposed by Platt [15].

To evaluate the prediction performance of each algorithm, we used the cross-validation techniques, which randomly divides the data set into 10 folds or groups, creates a model using 9 of the sets and tests it on the remaining group. This procedure is repeated until each of the 10 groups has served as a test group. Error estimates are calculated and then averaged. Here the whole training data set was randomly divided into 10 groups and the model trained on 9 groups, with the one remaining group used for testing each time.

The errors used as an indicator of the predictive performance of the models were the root mean-squared error (RMSE), mean absolute error (MAE), and relative error (RE %). The definitions of all these functions are given by the following equations

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y_p)^2} \qquad (2)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - y_p\right| \qquad (3)$$

$$RE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - y_p\right| / \left|y_i\right| \times 100\% \qquad (4)$$

where n represents the total number of data points in the data set, and y_i, y_p are the actual and prediction values, respectively. The squared error (SE) and correlation coefficient (R) were also considered.

Results and comparisons

To carry out our experiments, the hybrid GA/SVMR algorithm was run using the A RM software program. The optimum genetic algorithm parameters to optimise yarn tenacity were the population size = 5, maximum number of generations = 30, p initialise = 0.5, p mutation = -1.0, and p crossover = 0.9, as shown in ***Table 2***.

Optimised parameters of the support vector machine for regression (SVR) were selected using the grid search approach, and depending on the smaller RMSE error the parameters were as follows:

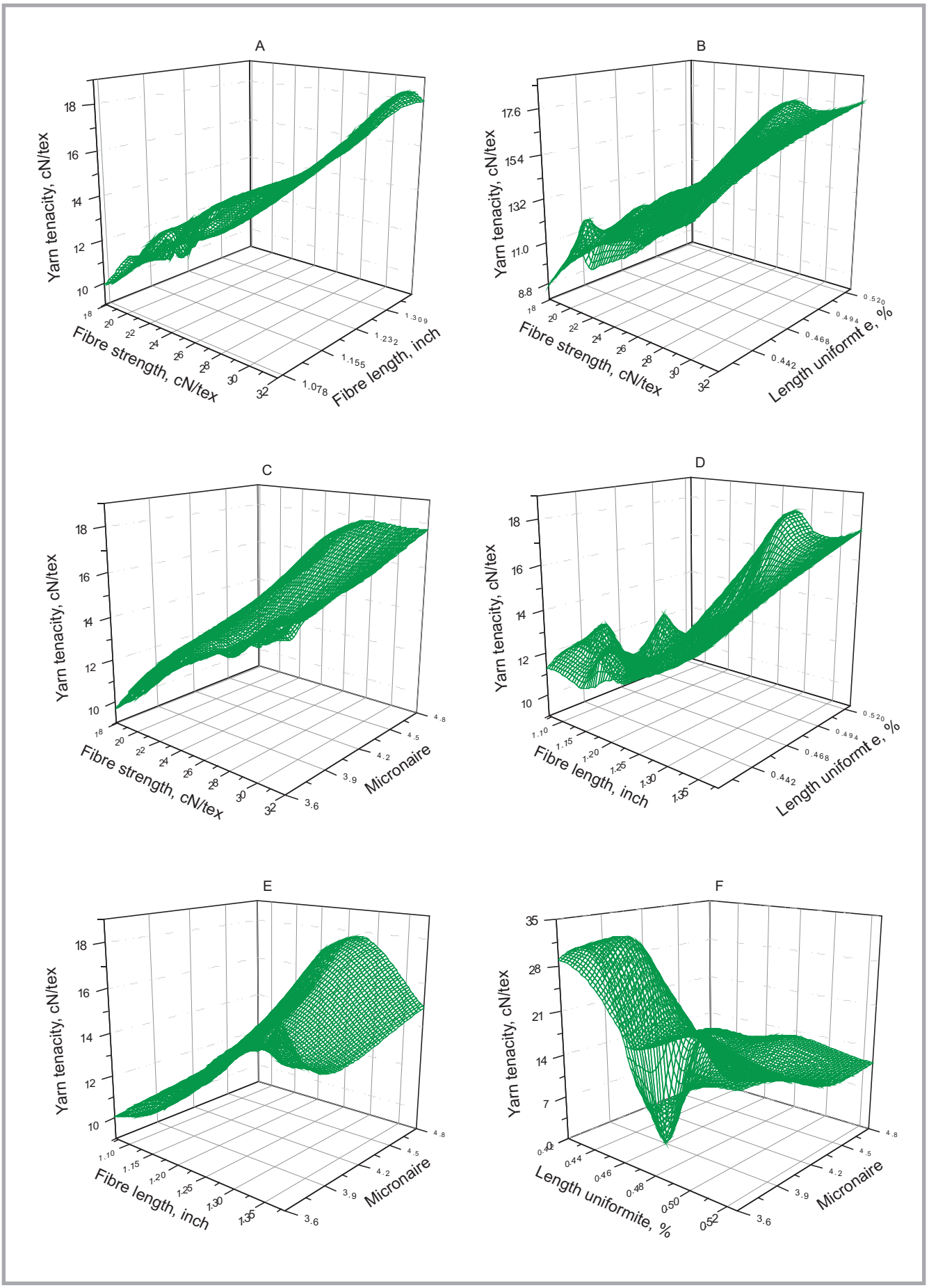

Figure 2. *Influence of (A) fibre strength and fibre length, (B) fibre strength and length uniformity, (C) fibre strength and micronaire, (D) fibre length and length uniformity, (E) fibre length and micronaire, and (F) length uniformity and micronaire on yarn tenacity.*

complexity parameter C = 150, value of ε-insensitive loss function = 0.04, and width of RBF kernel function γ = 0.01, as shown in ***Table 3***.

Table 4 summarises our results of the two models of SVMR; the noisy model of SVMR used the whole input data set or all attributes, and the optimised model of the hybrid GA-SVMR used the attributes selected.

First the preprocessing stage of the genetic algorithm for feature selection showed that the better attributes selected were the fibre strength, fibre length, length uniformity, and micronaire. As can be observed from these results, all those attributes selected means that it has a strong relation with the prediction of yarn tenacity. ***Figure 2*** showed the influence of the fibre property attributes selected on the yarn tenacity property. Thereafter the (GA) for feature selection removed the attributes that are less related to the yarn tenacity predicted, such as trash, elongation, yellowness (+b), and reflectance (Rd).

The errors used as an indicator of the predictive performance are utilised to compare the results. The results in ***Table 4*** illustrate that the optimised model of hybrid GA/SVMR with selected attributes gives smaller RMSE, MAE, RE% and SR errors than the original noisy model with all attributes; however, correlation coefficient R looks same in both models.

As can be observed from these results, there is a model with selected features with a lower dimension and higher performance value when compared with the noisy model consisting of all the features. This means that the prediction performance can be significantly improved using a small feature subset, while the usage of all the features does not guarantee better prediction results. Therefore by applying a prediction-driven dimensionality reduction mechanism based on GA for the feature selection scheme, only the most essential features are kept.

These results indicate that the suitability of the genetic algorithm (GA) for the selection of better attributes that give the preferred performance and high accuracy in the prediction of yarn tenacity.

Conclusions

In this paper, we presented a hybrid approach of GA/SVMR. The GA for feature selection was used as the preprocessing stage to find and select the best attributes or variables that most affect or relate to the prediction of yarn tenacity. This hybrid approach was compared with a noisy model of SVMR that used all attributes of fibre properties as input or predictors.

In the hybrid approach, the genetic algorithm for feature selection gave us better selected attributes for fibre strength, fibre length, length uniformity and micronaire, which were kept and used as predictors. Other attributes such as trash, elongation, yellowness (+b), and reflectance (Rd) were selected as less related attributes, removed from the second representation and not used.

It can be seen that from the results that the genetic algorithm for feature selection is capable of selecting a good set of features to discriminate the target from the data set.

However, the results of comparing the hybrid approach of GA/SVMR with the noisy model of SVMR show that the hybrid approach of GA/SVMR is able to obtain very high prediction accuracy and better performance than the noisy model with all attributes or all input data set variables.

Table 2. *Optimal parameters of GA parameters.*

population size	maximum number of generations	p initialize	p mutation	p crossover
5	30	0.5	-1.0	0.9

Table 3. *Optimal parameters of SVM model based on RBF kernel for predicting yarn tenacity.*

Property	Optimal parameters of SVM model based on RBF Kernel		
	γ	ε	C
Yarn tencaity, cN/dtex	0.01	0.04	150.0

Table 4. *Comparison of GA feature selection with the noisy model.*

Errors	Noisy model	Genetic algorithm
Feature subset	All features	1,2,4,5
Errors		
RMSE	0.527	0.449
MAE	0.474	0.420
RE, %	3.73	3.45
SR	0.329	0.226
R	0.800	0.799

Acknowledgement

The authors gratefully acknowledge the support of Natural Science Foundation of China (No. 51173023) and "Shu Guang" project (Shanghai Municipal Education Commission and Shanghai Education Development Foundation (No. 10GG17)).

References

1. Robert N, John E, Gary M. Handbook *of Statistical Analysis and Data Mining Applications*. Elsevier Inc, 2009, UK.
2. Bir B, Yingqiang L, Krzysztof K. *Evolutionary Synthesis of Pattern Recognition Systems. Monographs in Computer Science*. Springer, 2005: p. 142.
3. Sette S, Boullart L, Van Langenhove L, Kiekens P. *Textile Res. J*. 1997; 67, 2: 84.
4. Sette S, Boullart L, Van Langenhove L. *Textile Res. J*. 2000, 70, 5: 375.
5. Van Langenhove L, Sette S. Ed. W Krug & A Verbraeck. *Simulation in Industry* 2002, 469.
6. Cheng YSJ, Cheng KPS. *Textile Res. J*. 2004. 74(8), 718
7. Claude S, Geoffrey I. W. *Encyclopedia of Machine Learning*. Springer Science+Business Media, LLC , New York, USA, 2010.
8. Vapnik VN. *The nature of statistical learning theory*. Springer-Verlag New York, 1995.
9. Campbell C, Ying Y. *Learning with Support Vector Machines, in Synthesis Lectures on Artificial Intelligence and Machine Learning*. Morgan & Claypool, 2011: 1-95.
10. Jian GY, Zhi JL, Qian X. Fuzzy Systems and Knowledge Discovery, 2007. FSKD 2007. In: *4th International Conference,* 2007, 1, 696.
11. Anindya G, Pritam C. *Fibers and Polymers* 2010; 11, 1: 84.
12. Deogratias N, How WX. *Fibers and Polymers* 2011; 12, 4: 546.
13. Shevade SK, Keerthi SS, Bhattacharyya C, Murthy KRK. Neural Networks. *IEEE Transactions* 2000; 11, 5: 1188.
14. Smola AJ, Schölkopf B. *Statistics and Computing* 2004; 14, 3: 199.
15. Platt JC. *Advances in Kernel Methods-Support Vector Learning*. Cambridge, MA, MIT Press, 1999.

Received 27.07.2012 Reviewed 23.05.2013