Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 154 (2019) 154–160

Procedia Computer Science

www.elsevier.com/locate/procedia

8th International Congress of Information and Communication Technology, ICICT 2019

An Algorithm for Establishing A Model of Optimal Cotton Blending

Qing E Wu, Yuan Gang Gao, Hu Chen*, Ying Hui Guo, Chang Sheng Fan, De Ming Yang

QingE Wu：School of Electrical and Information Engineering, Zhengzhou University of Light Industry, Zhengzhou, 450002, China.

Abstract:

Due to the small amount of sample data and poor quality in cotton mills, this leads to large errors in the prediction of yarn quality to make the model difficult to train. This paper uses the appropriate data stretching algorithm after data preprocessing. Then, based on the features extracted by the main factors, the mapping relationship model of cotton blending is established. Finally, the evaluation and the correction of model are established. Actual tests have shown that the cost of cotton mills is decreased about 15% by the system and is increased about 25% for the profit by the system.

Keywords: feature extraction, optimal cotton blending, model evaluation, model correction;

1. Introduction

In order to reduce time and costs, it is necessary to properly mix cotton according to the quality parameters of yarn provided by customers in the production and practice of cotton spinning mills. Experts and scholars in the industry had been trying to establish a streamlined predictive model, mainly including regression statistics [1], competitive neural networks [2]and fuzzy mathematics [3], etc.

Xueyue Li [4] used WEKA (data mining open source software) feature selection method to analyze the influence of performance index of raw cotton on yarn quality, and combined with competitive neural network to

* Hu Chen. Tel.:15936274956

*E-mail address:*1593700132@qq.com; chenhu@zzuli.edu.cn.

10.1016/j.procs.2019.06.023

predict dryness and strength of yarn. The main factor analysis method [5] was widely used. In recent years, it had attracted much attention in the field of unbalanced data, such as image classification and recognition, information retrieval, medical diagnosis, network intrusion detection. Chuang Gao [6] studied the prediction model of yarn quality based on the combination of genetic algorithm, principal component analysis and competitive neural network. Yang S [7] established a physical or theoretical model to describe the relationship between yarn structure and yarn properties. And they provided the data needed for the model. Kothari N [8] showed that the degree of variation in plant fiber length and length uniformity were related to genotype. Chattopadhyay R [9] discussed how to use artificial neural networks to predict yarn properties from fiber parameters. Muhlstadt M [10] studied the fiber volume fraction gradient (G) in the laminated plain woven fabric (PWF), and solved the shortening problem by appropriately measuring the G in the laminated PWFs, and proposed the corresponding model. Ji R [11] proposed a cotton heterogeneous fiber classifier based on support vector machine (decision tree support vector machine, DTSVM). Experiments show that the recognition rate of different heterogeneous fibers is greater than 92%. Dhawan S [12] used the binary classification concept of support vector machine and the multi-class classifier concept of neural network to identify the spam and data categories received online and achieved good results.

This paper screens the sample data. Then the paper extracts the data features and builds a relational model. Finally, the model is tested and applied in production practice after multiple model evaluating and calibrating model. The results show that the system model can provide theoretical guidance for practical production.

2. Data Preprocessing

The sample data is filtered prior to modeling. There are two ways to implement this: one is manual filter, which is to manually eliminate the data that is obviously wrong in the database; the other is software filter, that use writing a filtering algorithm in the program to achieve culling of the wrong data. The filtered input data is normalized, and the normalization is written as formula (1):

$$x_i = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

Where x_i represents the normalized data; x represents the unprocessed data; $x_{\min}$ represents the minimum value of unprocessed data; $x_{\max}$ represents the maximum value of unprocessed data.

After expanding the sample data set, the sample data is extended here because the sample data is difficult to train due to the small sample data. According to related experience, the main factors affecting the strength are the original cotton nep, micronaire value, etc. Here, the raw cotton nep is considered to be x_i, and the strength is considered to be y_i, then it is assumed that there are $n+1$ data points: $(x_0, y_0), (x_1, y_1), (x_2, y_2), \cdots$).

The algorithm is as follows.

a. Calculation step: $h_i = x_{i+1} - x_i \quad (i = 0,1,2,\cdots$;

b. Bring the data node into the matrix equation to obtain the second derivative value m_i;

c. Calculate the coefficient of the fitted curve a_i, b_i, c_i, d_i.

d. In each subinterval $x_i \le x \le x_{i+1}$ create a formula (2):

$$g_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3 \tag{2}$$

The extension of the data is carried according to this equation.

3. Data Feature Extraction

Extracting data features in this paper refer to find the main factors from a number of factors. The purpose is to build an accurate model for the next step. For multivariate problems, there is a certain correlation among different indicators. The data feature extraction is to find the best classification feature subset, and sort the samples according to the intensity of the classified information. The algorithm is shown in Table 1.

A set of parameter $F_i(i=1,2,3,\cdots$, where i represents the number of subsets, $F_j^{\max}(j=1,2,3,\cdots$ represents the optimal feature subset in the parameter set, j represents the number of optimal feature subsets.

Table 1. Algorithm steps for extracting data features.

Algorithm steps for extracting data features
For i = 1:n-1 A, B, C arrays of correctness, sensitivity, and specificity, respectively. for j = 1:m-1 P = Posi; N = Neg; Using classification model to classify and calculate evaluation indicators and put the results into the array; end; q = Q (A, B, C); Output q that contributes the least to the classification. Posi (:, q); Neg (:, q); Remove the non-conforming q and leave F_{n-i} into next cycle. end;

4. Model Establishment

4.1 Relational model

In this paper, we establish a fuzzy relational model, that is, the mapping relationship between input and output factors is established through a relational table, as shown in formula (3).

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & & \cdots & & \cdots \\ x_{21} & x_{22} & \cdots & & \cdots & & \cdots \\ x_{31} & x_{32} & \cdots & \vdots \quad \vdots & \cdots & \vdots & \cdots \\ x_{41} & x_{42} & \cdots & & \cdots & & \cdots \end{pmatrix} \tag{3}$$

$x_{i1}, x_{i2}, \cdots$ represents an m-dimensional vector, which represents the input factors of strength, nep, strips, hairiness and single strength; $y_{i1}, y_{i2}, \cdots$ represents an n-dimensional vector representing the original cotton nep, the original cotton clone, the original cotton strength and the raw cotton impurity rate. The relationship matrix *R* is the established input-output relationship matrix.

4.2 Output-based optimization rule modeling

For multiple output results, as shown in formula (4), how to choose the optimal result output is particularly important. There are two analysis methods.

(1) Principal factor analysis

The input and output factors can be expressed as (x_i, y_i), where x_i represents an m-dimensional vector that represents all attributed parameter values in the data set. There are two types of decision-making attributes, namely the value y_i=1, -1, respectively, which indicate the main factor and the secondary factor. After analyzing the above problems, the resulting model is shown in formula (4).

$$f(x)=\begin{cases} +1 & y_i\alpha_i^*(x_i,x)+b^*>0 \\ -1 & y_i\alpha_i^*(x_i,x)+b^*\le 0 \end{cases} \tag{4}$$

Since α_i corresponding to the non-target object is 0, whose value is not zero corresponds to the target object. So the above formula is only performed on the target object.

In order to distinguish which are the main factors and the secondary factors. Kernel technology is used here, as shown in formula (5).

$$f(x)=\begin{cases} x & x>0 \\ \lambda x & x\le 0 \end{cases} \tag{5}$$

Linear indivisibility means that for each instance x, it is difficult to be distinguished, as shown in Fig. 1. (a).

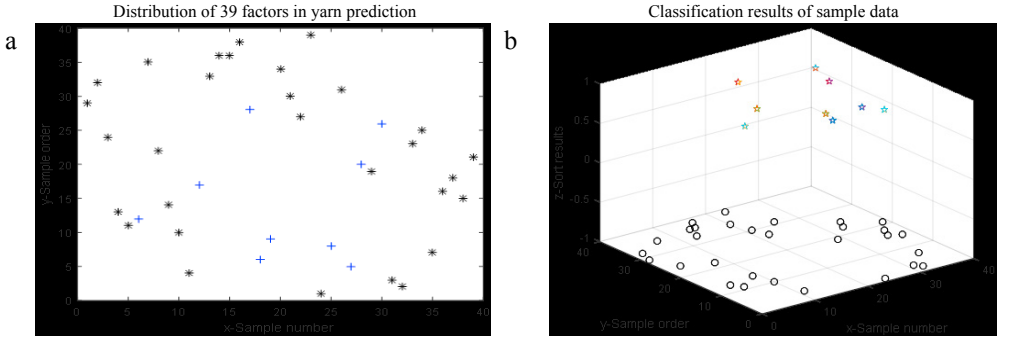

Fig. 1. (a) Distribution of 39 factors in yarn prediction; (b) Classification results of sample data.

There are 39 factors considered. Because the units vary from factors, and the values vary greatly. This brings great inconvenience to draw. In this paper, the number of the factor is represented by the x-axis, and the position is sorted in ascending order by the y-axis.

After the main element analysis algorithm, the classification result is shown in Fig. 1. (b) . The x-axis represents the number of the sample element; the y-axis represents the sort number of the sample; and the z-axis represents the sample element marker value (+1 or -1). The sample parameters are marked by principal factor analysis: the primary influence on the output factor is +1, and the secondary impact is -1. This divides the sample into two layers (z=+1 and z=-1), which is convenient for assigning weights when building the model.

(2) Core factor analysis

The core factor analysis is an algorithm after modeling the mapping relationship between input and output factors, output vector is composed $[y_{i1}, y_{i2}, \cdots$ of multiple output factors. Select the mean of each column to form a new vector $[\bar{y}_{i1}, \bar{y}_{i2}, \cdots$, and treat the vector as a core point, as shown in formula (6).

$$\delta_{ij} = \left| y_{ij}^2 - \bar{y}_{ij}^2 \right| \tag{6}$$

The output sets of vectors are variance with the corresponding elements of the core point vector respectively, as shown in formula (6). Then, the variance is compared with the core point data (or threshold) to vote, and the vector with more votes can be regarded as the optimal result satisfying the condition.

5. Model Evaluation

The classification model evaluation indicators in this paper have precision rate, recall rate and F_1 value. For the two classification problems, the classification can be divided into true positive cases (T_t), false positive cases (T_f), true negative (F_t) and false negative (F_f) according to the categories of real categories and classifier predictions. The precision, recall rate and F_1 value are shown in formula (7), (8) and (9), respectively.

$$P = \frac{T_t}{T_t + T_f} \tag{7}$$

$$R = \frac{T_t}{T_t + T_f} \tag{8}$$

$$F_1 = \frac{2 * P * R}{P + R} \tag{9}$$

6. Model Calibration

(1) Correction mapping w

The sum of w in each row is 1. The following methods are used for correction as formula (10).

$$w_n = w_{n-1} + \mu\left(w_n - w_{n-1}\right) \tag{10}$$

Where w_n represents the current weight; w_{n-1} represents the weight calculated from the previous calculation; w_0 represents the initial weight, and the empirical value needs to be obtained according to the sample training; μ represents the learning rate. Through machine learning, the mapping R is gradually adjusted.

(3) Output correction

The calibration model as shown in formula (11).

$$y' = y_0 + \eta\left(y' - y_0\right) \tag{11}$$

Where y' represents the current output value, y_0 represents the sample output value, and η represents the learning rate. Using machine learning, the system can achieve the desired output after a limited number of iterations.

7. Experiment and Simulation

7.1 Analysis of results

Here, the predicted value of the micronaire value of raw cotton process parameter is compared with the real value, as shown in Fig. 2. (a).

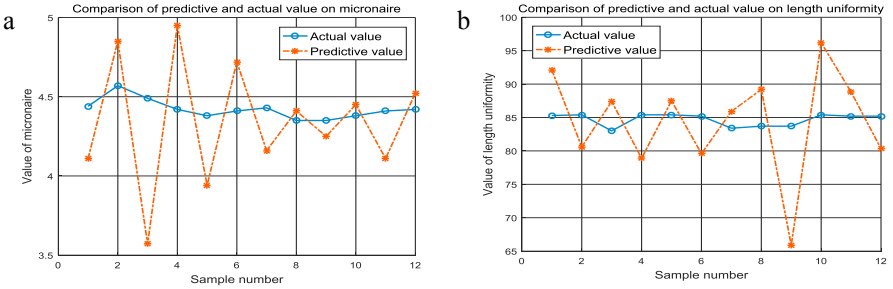

Fig. 2. (a) Comparison of prediction with real values of micronaire; (b) Comparison of prediction with real value of length uniformity

In Fig. 2. (a), the full line indicates the actual value. The imaginary line indicates the predicted value. From the figure of the actual value of the micronaire value it fluctuates little. The predicted value fluctuates around the real value. It can be seen that the prediction result of the micronaire value is ideal.

The predicted value of the length uniformity of raw cotton process parameters is compared with the real value, as shown in Fig. 2. (b). The full line indicates the true value of the length uniformity, and the imaginary line indicates the predicted value of the length uniformity. From the figure of a part of the predicted value, it falls outside the range $[x_i(1-0.15), x_i(1+0.15)]$, where x_i represents the number i true value of the length

uniformity, but most of the predicted values can meet requirement.

7.2 Compared with existing methods

The value of micronaire and length uniformity of raw cotton process parameters has an impact on strength, nep, strips, hairiness and single strength factors. In this paper, the two factors of micronaire value and length uniformity are selected to compare the prediction results of the proposed algorithm and the competitive neural network algorithm. The relative error curve of the micronaire value predicted by the algorithm and the competitive neural network algorithm as shown in Fig. 3. (a).

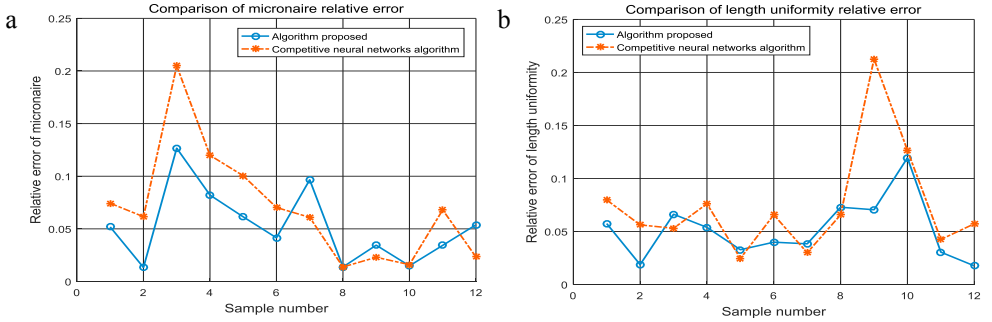

Fig. 3. (a) Comparison of micronaire relative error; (b) Comparison of length uniformity and relative error.

The relative error of about 80% example predicted by the method in this paper for a micronaire is kept below 8%. However, about 60% of the micronaire value predicted by the competitive neural network method keep the relative error below 8%, and the maximum relative error exceeds 20%. Therefore, the algorithm in this paper is better to predict the micronaire value.

The uniformity of length will also have different effects on factors such as strength, levelness, hairiness and single strength. The results of the algorithm and competitive neural network predicted are shown in Fig. 3. (b). From Fig. 3. (b) and experiment, that the max relative error of the prediction length uniformity of the competitive neural network algorithm is more than 20%, and only about 60% relative error of the competitive neural network algorithm can be controlled below 8%. However, the maximum relative error of the length uniformity predicted by the algorithm in this paper is about 10%. More than 80% of the sample data can control the relative error below 8%. Overall, the results predicted by this algorithm are better than those predicted by competitive neural networks.

8. Conclusion

Compared with the existing competitive neural network technology, the efficiency and accuracy of prediction improved by the algorithm model, and the evaluation and correction model in this paper. The test results show that the system can meet the basic requirements of practical production and provide theoretical guidance for yarn quality prediction.

Acknowledgement

This work is supported by Science and Technology Innovation Team Project of Henan Province University (19IRTSTHN013); Center Plain Science and Technology Innovation talents (ZYQR201810020); Key Science and Technology Program of Henan Province (172102410063) respectively.

References

[1] Liu M Z, Shao Y H, Wang Z, et al. Minimum deviation distribution machine for large scale regression[J]. Knowledge-Based Systems, 2018, 146.

[2] Araujo E, Bentes G E F, Zangaro R. Body Sway and Global Equilibrium Condition of the Elderly in Quiet Standing Posture by using Competitive Neural Networks[J]. Applied Soft Computing, 2018, 69:625-633.
[3] Putti F F, Lrag F, Cpc G, et al. A Fuzzy mathematical model to estimate the effects of global warming on the vitality of Laelia purpurata orchids [J]. Mathematical Biosciences, 2017, 288:124-129.
[4] Xueyue Li, Xiaoyan Gao, Zhigang Wu, Chongwen Yu. WEKA-Artificial Neural Network Prediction of Raw Cotton Performance and Yarn Quality[J]. Shanghai Textile Technology,2014,42(06):7-9+38.
[5] Lei R, Xiong Y, Li Y, et al. Main factors influencing the formation of thermogenic solid bitumen [J]. Organic Geochemistry, 2018.
[6] Chuang Gao, Chongwen Yu, Jun Wang, Yu Weidong. Study on the Prediction Model of Cotton Quality Index for Yarn Quality[J]. Cotton textile technology,2016,44(06):18-22.
[7] Yang S, Gordon S. 4–Fiber-to-yarn predictions [J]. Engineering of High-Performance Textiles, 2018:81-106.
[8] Kothari N, Hague S, Hinze L, et al. Boll sampling protocols and their impact on measurements of cotton fiber quality[J]. Industrial Crops & Products, 2017, 109:248-254.
[9] Chattopadhyay R. 4 – Artificial neural networks in yarn property modeling [J]. Soft Computing in Textile Engineering, 2011, 3(2):105-125.
[10] Mühlstdt M, Seifert W, Arras M.L, et al. 3D Model of Intra-Yarn Fiber Volume Fraction Gradients of Woven Fabrics[J]. Composite Structures, 2017, 180:944.
[11] Ji R, Li D, Chen L, et al. Classification and identification of foreign fibers in cotton on the basis of a support vector machine [J]. Mathematical & Computer Modelling, 2010, 51(11):1433-1437.
[12] Dhawan S, Simran. An enhanced mechanism of spam and category detection using Neuro-SVM [J]. Procedia Computer Science, 2018, 132:429-436.