PROCEEDINGS OF SCIENCE

Application of Factor Analysis in Computer Color Matching for Textile Dyeing

Mengmeng Yin[1]
College of Information Engineering, Qingdao University ,Qingdao, 266071, China

Bingsen Zhang[2]
College of Information Engineering, Qingdao University ,Qingdao, 266071, China
E-mail: `bszhang@qdu.edu.cn`

Xiaodan Deng
College of Information Engineering, Qingdao University ,Qingdao, 266071, China

Firstly, we analyze relevant data which we've achieved from dyeing and printing enterprise by using the method of factor analysis in order to acquire the correlation coefficients between tristimulus values and the concentrations of the three dyes. For the sake of convenience, we call tristimulus values as CMY and define the concentrations of the three dyes as d_1, d_2 and d_3 ; thus we can speculate necessary mathematical model in our research. Secondly, we can obtain the relationship between CMY and d_1, d_2, d_3 by using regression analysis. Finally, we get the predicting values of d_1, d_2 and d_3 by Newton iteration method on the premise of our knowing in respect of the values of C, M and Y. Subsequently, we can obtain the experimental errors. Experimental results show that the errors of the mathematical model established by these methods are very small, which prove to us that the model can be used in actual production.

CENet2015
12-13 September 2015
Shanghai, China

[1]Speaker
[2]The project is supported by the National Natural Science Foundation of China (60973158).

1. Introduction

Nowadays, computer color matching technology has solved lots of problems such as time-consuming, laborious and poor reproducibility in the process of artificial color matching[1]. The theory of computer color matching technology is based on Kubelka-Munk theory (K-M theory) [2]. K-M theory is premised on the opaque medium, and the ink used in printing is transparent or translucent; therefore, K-M theory is insufficient. Yuli Pan et al. have made a deep research on computer color matching. They applied the principal component analysis in practical issues and have made great achievements [3]. In this article, we choose to make use of factor analysis to predicate the correlation between the concentrations of three dyes and tristimulus values. Factor analysis features many advantages in terms of analysis because the factor rotation is used to help explain the factor.

2. Processing Flow

Obtain the correlation coefficients between CMY and d_1, d_2, d_3 by using factor analysis method in SPSS.

Speculate the necessary mathematical model.

Get the relationship between CMY and d_1, d_2, d_3 by using regression analysis in the software of SPSS.

Figure out the predicting values of d_1, d_2 and d_3 by Newton iteration method, then work out the errors of the concentrations of three dyes.

3. Factor Analysis

3.1 Basic Idea of Factor Analysis

The basic idea of factor analysis is to divide the original data into groups according to the relationships among all the variables. In this process, we should keep the correlation of variables within the same group higher and the correlation of variables in different groups lower. Each group of variables represents a basic structure, which is called common factor[4-7].

3.2 Mathematical Model of Factor Analysis

Suppose there be n samples, each of which has p indicators. There is strong correlation among the p indicators, which are represented by $X_1, X_2, \ldots, X_p$. $\vec{x}_{(i)} = (x_{i1}, x_{i2}, \ldots, x_{ip})'$ is the i^{th} element, $i = 1,2,\ldots,n$.

Each sample is influenced by m common factors, which are recorded as $F_1, F_2, \ldots\ldots, F_m$. So each sample can be expressed as an expression:

$$X_i = a_{i1}F_1 + a_{i2}F_2 + \ldots\ldots + a_{im}F_m + \varepsilon_i \qquad (i=1,2,\ldots\ldots,p) \qquad (3.1)$$

We call Formula (3.1) the factor model.

In Formula (3.1), $F_1, F_2, \ldots\ldots, F_m$ are common factors of $\vec{X}$, and $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p$ are noises, which are called special factors. Generally speaking, common factors impact all the measures of $\vec{X}$, while ε_i influences X_i only. In addition, every special factor is irrelevant; at meanwhile, all the special factors are irrelevant to all the common factors. $\underset{\sim}{A} = (a_{ij})_{p*m}$ in the model is the coefficients matrix to be estimated, which is called the factor loading matrix.

3.3 Factor Rotation

The purpose of establishing the mathematical model is not only to find the common factors depending on the factor loading matrix, but to understand the actual meaning of every factor. Because the factor loading matrix is not unique, we rotate the matrix by multiplying orthogonal matrix and factor loading matrix to make the rotated factor loading matrix simpler, which is called the factor rotation. As a result, it can explain the common factors reasonably.

As to a variable, the factor rotation makes a big load on a common factor, while the loads on the other common factors are small. More precisely, only one of the correlation coefficients between the variable and the common factors is big. So factor analysis can explain the actual issue better than principal component analysis.

The common rotation methods of the factor loading matrix are varimax orthogonal rotation, oblique rotation, etc., among which the most common one is varimax orthogonal rotation.

4. Regression Analysis

In data set, the value of dependent variable Y relies on the independent variables $X_1, X_2, \ldots, X_p$, and the relationship is linear. Now given a set of values of independent variables $X_{1i}, X_{2i}, \ldots, X_{pi}$, the dependent variable Y will be valued in accordance with the following equation:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_p X_{pi} + \varepsilon_i \tag{4.1}$$

Formula (4.1) is called the multiple linear regression model[8-10].

In Formula (4.1), $\beta_0, \beta_1, \ldots, \beta_p$ are model parameters; and ε_i is the unknown factor, which will affect Y_i in a certain way other than the P independent variables selected in the model.

Numerous planes can represent the distribution of the sample points. In order to ensure the best fit between the regression equation and the sample points, we adopt the least square method to obtain the needed plane.

5. Establishment and Solution of the Mathematical Model

5.1 Establishment of the Mathematical Model

We analyze the experimental data by the method of factor analysis in SPSS, and we can get the correlation coefficients between CMY and d_1, d_2, d_3. The results are shown in Table 1.

	d_1	d_2	d_3
C	0.957	-0.034	-0.052
M	0.717	0.556	0.080
Y	0.549	0.385	0.698

Table 1: Correlation Coefficients

We find from Table 1 that the correlation coefficient between C and d_1 is close to 1.0, and the correlation coefficient between C and d_2 or d_3 is very small; thus we draw the conclusion that C is mainly determined by d_1. Similarly, M is determined by d_1 and d_2, and Y depends on the values of d_1, d_2 and d_3. More precisely, the relational model between CMY and d_1, d_2, d_3 is:

$$\begin{cases} C = f_1(d_1) \\ M = f_2(d_1, d_2) \\ Y = f_3(d_1, d_2, d_3) \end{cases} \tag{5.1}$$

5.2 Solution of The Mathematical Model

5.2.1 C—d1 Relationship

We've known that the value of C is determined by d_1, so the relationship may like $C = ad_1^2 + bd_1 + c$ or $C = ad_1^3 + bd_1^2 + cd_1 + e$. We analyze the data by using quadratic curve estimation and cubic curve estimation in SPSS simultaneously. Their trajectories are depicted in Fig. 1. The figure tells us that the cubic curve is better than the quadratic one in fitting degree.

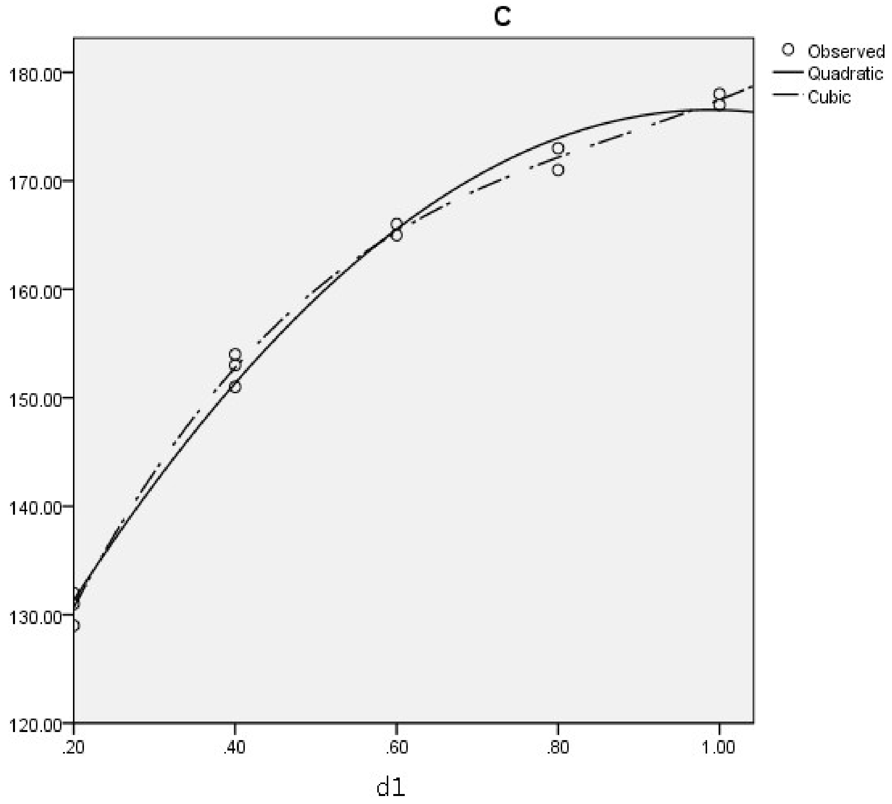

Figure 1: C—d_1 graph

We use cubic polynomial to fit the expression $C = f_1(d_1)$, and the relevant parameters are shown in Table 2.

Model summary			Coefficients			
R^2	Residual	F	constant	d_1	d_1^2	d_1^3
0.997	12.528	759.950	94.853	219.791	-221.038	83.855

Table 2: Correlation Coefficients of Cubic Polynomial about C

So, the relationship between C and d_1 is：

$$C = 94.853 + 219.791d_1 - 221.038d_1^2 + 83.855d_1^3 \tag{5.2}$$

From Table 2, we can know that the determination coefficient is close to 1.0 and the residual error is small, which proves to us that the regression effect of this model is remarkable; therefore, this cubic curve model is very close to the actual data value, and it can track the results well.

5.2.2 M—d1，d2 Relationship

We have known that the value of M is determined by the independent variables d_1 and d_2. As a result, we use the linear regression method to predict the expression $M = f_2(d_1, d_2)$ in SPSS software. Relevant parameters are shown in Table 3.

Model summary			Coefficients					
R^2	Residual	F	constant	d_1	d_2	d_1*d_2	d_1^2	d_2^2
0.997	7.026	408.496	110.571	98.882	102.448	-52.593	-32.341	-42.082

Table 3: Correlation Coefficients of Regression Equation about M

So the regression equation is:

$$M = 110.571 + 98.882d_1 + 102.448d_2 - 52.593d_1d_2 - 32.341d_1^2 - 42.082d_2^2 \tag{5.3}$$

From Table 3, we can know that the determination coefficient is 0.997 and the residual error is 7.026 which is very small. At the same time, the value of F test is 408.496, which is far above the threshold. All of these data prove the significance of the model, which can meet the actual needs.

5.2.3 Y—d1，d2，d3 Relationship

We analyze the data making use of linear regression method and try our best to make the value of F test large and residual as small as possible. Through analysis, we can obtain the expression between Y and d_1,d_2,d_3:

$$Y = 127.889 + 43.357d_1 + 33.109d_2 - 39.642d_1d_3 - 19.058d_2d_3 - 10.582d_2^2 + 54.055d_3^2 \tag{5.4}$$

The relevant parameters of the regression model are shown in Table 4.

Model summary			Coefficients						
R^2	Residual	F	constant	d_1	d_2	d_1*d_3	d_2*d_3	d_2^2	d_3^2
0.998	1.905	484.045	127.889	43.357	33.109	-39.642	-19.058	-10.582	54.055

Table 4: Correlation Coefficients of Regression Equation about Y

The data in Table 4 tell us that the the determination coefficient is close to 1.0, the the residual error is very small, and the value of F test is far from the threshold. All of these data prove the accuracy of the fitting polynomial.

6. Error Analysis

We solve the established mathematical model by using Newton iteration method to get the predicting values of d_1, d_2 and d_3 on the basis of our knowing about the values of CMY, and then the experimental errors can be obtained [11]. The experimental data are shown in Table 5.

PoS(CENet2015)041

Tristimulus values			Sample concentration			Predicting concentration			Error		
C	M	Y	d1	d2	d3	d1	d2	d3	d1	d2	d3
129	145	147	0.2	0.2	0.4	0.1886	0.1900	0.4090	0.0114	0.0100	-0.0090
131	159	150	0.2	0.4	0.4	0.2025	0.3984	0.3797	-0.0025	0.0016	0.0203
132	179	172	0.2	1	0.8	0.2097	1.0138	0.7984	-0.0097	-0.0138	0.0016
151	161	170	0.4	0.2	0.8	0.3794	0.2127	0.7956	0.0206	-0.0127	0.0044
153	171	173	0.4	0.4	0.8	0.4028	0.3810	0.8064	-0.0028	0.0190	-0.0064
154	183	161	0.4	0.8	0.4	0.4151	0.7841	0.3864	-0.0151	0.0159	0.0136
165	187	177	0.6	0.8	0.8	0.5941	0.7796	0.8003	0.0059	0.0204	-0.0003
171	178	164	0.8	0.2	0.4	0.7589	0.1694	0.4389	0.0411	0.0306	-0.0389
173	190	170	0.8	0.6	0.4	0.8318	0.5948	0.3865	-0.0318	0.0052	0.0135
178	186	168	1	0.2	0.4	1.0181	0.2186	0.4052	-0.0181	-0.0186	-0.0052
177	192	181	1	0.6	0.8	0.9840	0.5953	0.7990	0.0160	0.0047	0.0010

Table 5: Predicting Values and Errors of Dye Concentrations

We find from Table 5 that the predicating values of d_1,d_2 and d_3 are very close to the sample values. In other words, the experimental errors of the model are very small, which are within the permissible range. In short, this model can solve the problem in computer color matching well.

7. Conclusion

In this paper, we take factor analysis method to analyze the correlation among all the experimental data. The mathematical model can be applied to the textile dyeing and color matching to meet the demand of practical production. On the one hand, the factor analysis can explain the factors well because of factor rotating technique, but at the same time, the computation becomes more complicated. On the other hand, we need to determine the number of the factors in advance. In this sense, it is possible for us to do experiments many times in order to achieve the best results.

In our model, we can solve the issue when it has limited data; but when the initial data about the sample have large size, the computation will become complicated. As for this problem, we will carry out thorough researches.

References

[1] J. J. Liang. *Application of Computer in Staining Technique*[J]. Knitting Industries, 2003(3):53-56(2003) (In Chinese).

[2] W. J. Hao, X. P. Zhao. *Kubelka-Munk Single Constant Corlor Theory and Practice*[J]. China Printing and Packaging, 2009(3): 43-47(2009) (In Chinese).

[3] Y. L. Pan, B. S. Zhang, Q. Xu. *Research on Application of Principal Component Analysis in Computer Color Matching for Textile Dyeing*[J]. Journal of Qingdao University (Engineering and Technology Edition), 26(4): 19-22(2011) (In Chinese).

[4] D. Y. Fu. *Applied Multivariate Statistical Analysis*[M]. Higher Education Press, Bei Jing, 167-196(2013) (In Chinese).

[5] X. Q. He. *Multivariate Statistical Analysis (the third edition)*[M]. China Renmin University Press, Bei Jing, 143-172 (2012) (In Chinese).

[6] X. Qu, T. Guo, W. Wang, et al. *Measuring speed consistency for freeway diverge areas using factor analysis*[J]. Journal of Central South University, 20: 267-273(2013).

[7] S. Jung , S. Lee, *Exploratory factor analysis for small samples*[J], Behav Res Methods. 43(3):701-9(2011 Sep).

[8] L. Feng, *Principle of Regression Analysis Method and Actual Operation in SPSS*[M]. China Finance Publishing House, Bei Jing, 1-232 (2004) (In Chinese).

[9] M. Yan, B. S. Zhang, D. M. Wang. *Research on Application of Polynomial Regression Analysis for Computer Color Matching in Textile Dyeing*[C]//Applied Mechanics and Materials.Switzerland, 602: 719-722(2014).

[10] Y. Y. Wu, B. S. Zhang and M. Li, *Research on Application of Curve Regression Analysis in Computer Color Matching for Textile Dyeing*[J]. Journal of Qingdao University (Engineering and technology edition), 28 (1) p.22-26 (2013) (In Chinese).

[11] H. Liu, *Newton Iteration Method for Nonlinear Equation Solutions and Its Application*[J]. Journal of Chongqing Institute of Technology (Natural Science Edition),8: 026(2007).