

Particle Swarm Optimization Algorithm Establish the Model of Tobacco Ingredients in Near Infrared Spectroscopy Quantitative Analysis

Bibo Ma, Haiyan Ji

► **To cite this version:**

Bibo Ma, Haiyan Ji. Particle Swarm Optimization Algorithm Establish the Model of Tobacco Ingredients in Near Infrared Spectroscopy Quantitative Analysis. Daoliang Li; Yingyi Chen. 6th Computer and Computing Technologies in Agriculture (CCTA), Oct 2012, Zhangjiajie, China. Springer, IFIP Advances in Information and Communication Technology, AICT-393 (Part II), pp.92-98, 2013, Computer and Computing Technologies in Agriculture VI. <10.1007/978-3-642-36137-1_12>. <hal-01348219>

HAL Id: hal-01348219

<https://hal.inria.fr/hal-01348219>

Submitted on 22 Jul 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Particle swarm optimization algorithm establish the model of tobacco ingredients in near infrared spectroscopy quantitative analysis

Bibo Ma^{1,2} Haiyan Ji^{1,2*}

¹College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China

²Key Laboratory of Modern Precision Agriculture System Integration Research, Ministry of Education, China Agricultural University, Beijing, China 100083

Mabibo1986@163.com instru@cau.edu.cn

Abstract. 57 tobacco samples were chosen, the near-infrared diffuse reflectance spectra at five different parts of tobacco leaves were averaged as the original spectra, the range of 4000-9000 wavenumber of spectral information was selected after wavelength selection, first-order differential was used to eliminate the impact of offset, PCA was used to reduce the dimensions and select the most representative of the principal components of the sample information. After these, use PSO and the data of chemical composition of nicotine and total nitrogen, to establish the models of near-infrared spectra of the main ingredient in tobacco quantitative analysis. In which, nicotine and total nitrogen quantitative analysis of modeling set correlation coefficients were 0.8886,0.9290; the prediction set correlation coefficients were 0.8786,0.8487. RMSEC values were 0.4737,0.2215; RMSEP values were 0.6417,0.2677. The results show that: PSO can be used to establish the model of nicotine and total nitrogen by near infrared spectroscopy quantitative analysis in tobacco.

Keyword. Tobacco, Particle Swarm Optimization (PSO), Principal Component Analysis (PCA), Near infrared spectroscopy

¹ Bibo Ma: Master. Research field: Spectral analysis.

Haiyan Ji: Professor. Research field: Spectral analysis and sensing technique.

Introduction

Near infrared reflectance spectroscopy does not require any chemical pretreatment for the sample. Compared with traditional chemical analysis and other spectroscopy methods, it has many advantages, such as cheap, convenient, fast and non-destructive characteristics^[1].

Currently, there are some methods in the near-infrared spectroscopy quantitative analysis, such as partial least squares (PLS)、support vector machines(SVM) and artificial neural network(ANN). In 2007, Yan Qing Duan, Jia Jun Wang, and others use PLS method for quantitative analysis for the chemical composition of five kinds of tobacco sheet. The results showed that: near-infrared predictive value and the measured values have a good correlation, the model has better prediction accuracy^[2]; In 2007, Xu Liu, Hua Chen use the PCA-SVR and the ANN for a quantitative analysis of total sugar, reducing sugar, total nitrogen and nicotine content in tobacco. The results show that: the SVR model has a better prediction^[3]; In 2009, Yong Zhang, Qian Cong use the SVM for a quantitative analysis of total sugar, reducing sugar, nicotine and total nitrogen in tobacco. The results show that: the SVM model's classification accuracy rate can up to 100%, the quantitative analysis of four components of the prediction coefficient of determination (R^2), the prediction mean square error (RMSEP) and the average relative error (RME) show that: the model has a better prediction^[4].

In this paper, PSO method is used in near-infrared spectroscopic quantitative analysis of total nitrogen and nicotine content in tobacco, in order to prove that the PSO method can be used in the near-infrared quantitative analysis in tobacco.

The basic principles

1.1 The basic principle of PSO

PSO, the English name is Particle Swarm Optimization, is proposed by Kennedy and Eberhart in the study of birds and fish group's behavior in 1995. It's a swarm intelligence algorithm^[5]. In PSO, each optimization problem can be a bird in the search space (we can call it a particle). All the particles have an optimized to

determine the fitness value by a function and have a speed to determine their direction and speed for flying.

Firstly, the algorithm initializes a group of random particles, and then finds the optimal solution by iteration. In each iteration, the particles update their speed and position by individual extreme and global extreme. The algorithm assumes that all particles are moving in the personal best position and global best position, and mark

$x_i = (x_{i1}, x_{i2} \dots x_{in})$ to initialize the position of the particle, mark

$v_i = (v_{i1}, v_{i2} \dots v_{in})$ to initialize the velocity, mark $p_i = (p_{i1}, p_{i2} \dots p_{in})$ as the

best result for searching so far for the particle of i, mark $p_g = (p_{g1}, p_{g2} \dots p_{gn})$

as the best result for searching so far for the all particles. In each iteration, the particles update themselves by tracking the two optimal values, which is the value founded by particles themselves and the value founded by the entire population so far. After finding the two values, particles update their velocity and position according to the following formula^[6]:

$$v_{id}(new) = w \times v_{id}(old) + c_1 \times r_1 \times (p_{id} - x_{id}) + c_2 \times r_2 \times (p_{gd} - x_{id}) \quad (1)$$

$$x_{id}(new) = x_{id}(old) + v_{id}(new) \quad (2)$$

$$w = w_{\max} - t \times (w_{\max} - w_{\min}) / \max DT \quad (3)$$

In this formula, w is constant of non-negative, we call it the inertia factor, which is used to balance the global searching and the local searching. Shi and other people for the first time introduce to update formula for the inertia weight of the speed in PSO^[7]. In this paper, we propose $w_{\max} = 0.9$, $w_{\min} = 0.4$ ^[8]; $c_1 = 2.8$, $c_2 = 1.3$; r_1 and

r_2 are random number in $[0, 1]$, $\max DT$ is the maximum number of iterations,

in order to control the closing conditions of the algorithm; t is the current number of iterations. In this paper, w decrease gradually from $[0.9, 0.4]$ ^[9].

Firstly, calculate the new speed of the particles by the distance of the original particle velocity and position between the two optimal locations according to the formula (1). Then particles move to a new location according to the formula (2). So back and forth. If it reached the specified number of iterations, the program terminates.

1.2 The processes of PSO

(1) Initialize the initial position and velocity of a group of particles and set the number of iterations for 5000

(2) Set the mean square error as the fitness value and get the fitness value of each particle by calculating

(3) For every particle, compare it's fitness value with the best position (pbest,) if better than pbest, use it as the current optimal particle position (pbest);

(4) For every particle, compare it's pbest with the best position gbest of all particles, if better than gbest, use it as the current optimal particle position (gbest);

(5) According to the formula (1) and (2), adjust the speed and position of the particle;

(6) If it has not reach the end of the conditions, turn to step 2;

Here, the termination condition is that the required iterations can get the number of 5000.

2 Experimental part

2.1 The sources of material

In this paper, the tobacco data is K326 tobacco varieties, which is coming from the national 18 origins and is divided into three tobacco grades (B2F, X2F, C3F).

2.2 The collection of spectral information

In this paper, the spectra data is collected on five different parts of the tobacco leaf, and then take the average of the five parts as the spectrum. Tobacco original near infrared spectroscopy is showed in Figure 1.

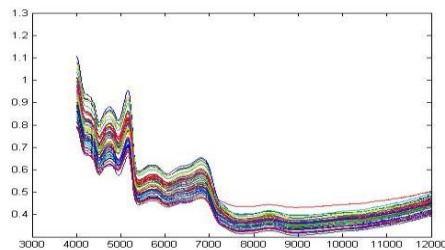


Fig. 1. Tobacco original near infrared spectroscopy

2.3 The preprocess of spectral

It can be seen from Figure 1, the main crest of tobacco information is concentrated in the 4000-9000 wavenumber. Therefore, in order to save modeling time, and eliminate the unwanted ingredients in the spectral information, in this paper, the wavenumber in the range of 4000-9000 cm^{-1} information are selected for the study.

The spectrum is showed in Figure 2 after wavelength selection. In this paper, use the first difference to eliminate the impact of the offset. The spectrum is showed in Figure 3 after the first difference. Then use the principal component analysis (PCA) to reduce the dimension and get the principal component of the spectrum. Use the PSO to create a system model. In the model, the main components of the spectral data is as the variable input of the model, the chemical values of nicotine and total nitrogen content is as the output of the model ^[10]. Use the data of mean square error between the calculated results and the real chemical value as the fitness function, in order to adjust the speed and position of the particle in the particle swarm and determine the global optimal solution. In the paper, the number of the principal components, the range of chemical value content of nicotine and the total nitrogen in modeling and in the prediction are showed in Table 1.

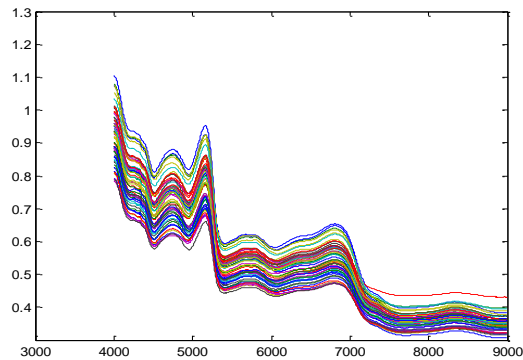


Fig. 2 spectra after the wavelength selection

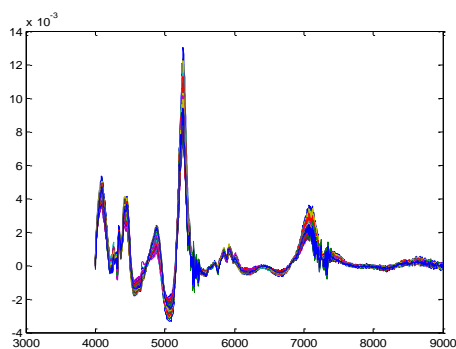


Fig. 3. spectra after first-order Differential

Table 1. Range of chemical content and the number of principal components

Sample Name	Samples number	The range of model (%)	The range of prediction (%)	Principal components
Nicotine	57	0.88-4.68	0.94-4.67	14
total nitrogen	57	1.45-3.83	1.58-3.55	15

2.4 The parameter settings of the model

In PSO, the parameter setting is very important. In order to improve the ability of modeling and prediction, in this paper, the algorithm parameter settings are showed in Table 2.

Table 2. The parameter settings of the model

The parameter settings of the model	
The location of the particles (X)	Random
The velocity of the particle (V)	Random, but between[-1,1]
Maximum number of iterations (MAXDT)	5000
Number of particles (N)	50
c1,c2	c1=2.8,c2=1.3

2.5 The establishment and evaluation of the model

In this paper, in all of 57 K326 tobacco data, use 40 samples as the set of modeling, 17 samples as the set of prediction. In the model, the total nitrogen modeling set of correlation coefficient is 0.9290, the prediction set of correlation coefficient is 0.8487; nicotine modeling set of correlation coefficient is 0.8886, the correlation coefficient of prediction set is 0.8786. Other related indicators of the rest of the model are showed in Table 3. Figure 4 is PSO regression of nicotine modeling set and prediction set. Figure 5 is PSO regression of total nitrogen modeling set and prediction set.

Table 3. The indicators of evaluating the model

Sample Name	Modeling R	Modeling RMSEC	Prediction R	Prediction RMSEP
Nicotine	0.8886	0.4737	0.8786	0.6417
total nitrogen	0.9290	0.2215	0.8487	0.2677

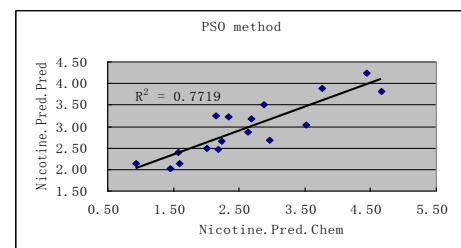
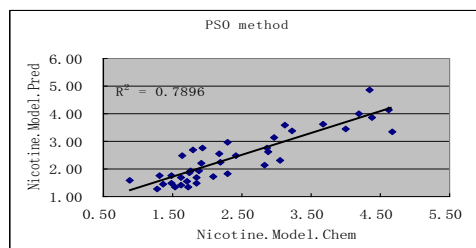


Fig. 4 PSO regression of nicotine modeling set and prediction set

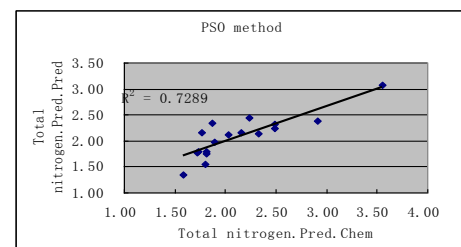
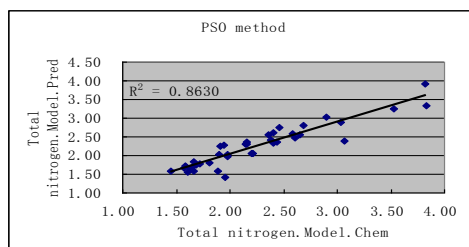


Fig. 5 PSO regression of total nitrogen modeling set and prediction set

3 Results and Discussion

57 tobacco samples were chosen, the near-infrared diffuse reflectance spectra at five different parts of tobacco leaves were averaged as the original spectra, the range of 4000-9000 wavenumber of spectral information was selected after wavelength selection, first-order differential was used to eliminate the impact of offset. PCA was used to reduce the dimensions and select the most representative of the principal components of the sample information. After these, use PSO and the data of chemical composition of nicotine and total nitrogen, to establish the models of near-infrared spectra of the main ingredient in tobacco quantitative analysis. In which, nicotine and total nitrogen quantitative analysis of modeling set correlation coefficients were 0.8886,0.9290; the prediction set correlation coefficients were 0.8786,0.8487. RMSEC values were 0.4737,0.2215; RMSEP values were 0.6417,0.2677. The results show that: PSO can be used to establish the model of nicotine and total nitrogen by near infrared spectroscopy quantitative analysis in tobacco.

References

- 1.Heng,L. Y., Long,L .Z., Dong,H. H.,et.al:China Light Industry Press,1-6 (2005)
- 2.Yan,Q.D., Jia,J. W., et.al : Laser and infrared. vol. 37, pp. 1058-1061(2007)
- 3.Xu,L. H., Cai,C.,et.al: Spectroscopy and Spectral Analysis. vol. 27, pp.2460-2463(2007)
- 4.Yong,Z., Qian, C., et.al : Journal of Chemical. vol. 30, pp. 697-700(2009)
- 5.Kennedy,J., Eberhart,R.C.: Particle swarm optimization. In Proceedings of the IEEE Conference on Neural Networks, pp. 1942-1948. IEEE Press, Washington DC (1995)
- 6.Shao,R. H.: Computer Engineering and Design. vol. 30, pp. 1977-1980(2009)
- 7.Shi,Y., Eberhart,R.C.:A modified particle swarm optimizer. In Proceedings of IEEE International Conference on Evolutionary Computation. pp. 69-73, IEEE Press ,Washington DC(1998)
- 8.Jun,W. W., Ding Wei Wang:Journal of Systems Engineering, vol. 20, pp. 194-198(2005)
- 9.Jiu,F. Xu.,Guo,L.M., Jin,R. N., Xiao,T. C.:He'nan Institute of Technology (Natural Science). vol. 22, pp. 68-72(2010)
- 10.Xin,J.W.,Yuan,L., Dan,Z., Chong,E. T.:Liaoning University of Technology. vol. 34, pp. 18-20(2007)