Journal of Food and Drug Analysis

Volume 22 | Issue 2 Article 7

Modeling research on wheat protein content measurement using near-infrared reflectance spectroscopy and optimized radial basis function neural network

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.jfda-online.com

Original Article

Modeling research on wheat protein content measurement using near-infrared reflectance spectroscopy and optimized radial basis function neural network

Xiaodong Mao, Laijun Sun, Guangyan Hui, Lulu Xu*

Key Laboratory of Electronics Engineering, College of Heilongjiang Province, Heilongjiang University, Harbin, China

ARTICLE INFO

Article history:
Received 10 June 2013
Received in revised form
14 September 2013
Accepted 30 September 2013
Available online 18 February 2014

Keywords:
Protein
PSO algorithm
RBF neural network
Wheat

ABSTRACT

In this study, near-infrared reflectance spectroscopy and radial basis function (RBF) neural network algorithm were used to measure the protein content of wheat owing to their nondestructiveness and quick speed as well as better performance compared to the traditional measuring method (semimicro-Kjeldahl) in actual practice. To simplify the complex structure of the RBF network caused by the excessive wave points of samples obtained by near-infrared reflectance spectroscopy, we proposed the particle swarm optimization (PSO) algorithm to optimize the cluster center in the hidden layers of the RBF neural network. In addition, a series of improvements for the PSO algorithm was also made to deal with its drawbacks in premature convergence and mechanical inertia weight setting. The experimental analysis demonstrated that the improved PSO algorithm greatly reduced the complexity of the network structure and improved the training speed of the RBF network. Meanwhile, the research result also proved the high performance of the model with its root-mean-square error of prediction (RMSEP) and prediction correlation coefficient (R) at 0.26576 and 0.975, respectively, thereby fulfilling the modern agricultural testing requirements featuring nondestructiveness, real-timing, and abundance in the number of samples.

1. Introduction

China is a large agricultural country, and wheat is one of the most favorite and widely distributed foods among the Chinese. The country's annual production of wheat is up to 110 million tons, accounting for up to 20% of the global wheat production [1]. Owing to the increasing pursuit of healthy diet among the local populace, the requirements to ensure the quality of wheat flour have also improved. The protein content and quality of wheat grain are the most important factors to evaluate the quality of wheat. Under normal

* *Corresponding author*. Key Laboratory of Electronics Engineering, College of Heilongjiang Province, Heilongjiang University, Harbin, China.
E-mail address: slaijun@126.com (L. Sun).

http://dx.doi.org/10.1016/j.jfda.2014.01.023

circumstances, the traditional method (semimicro-Kjeldahl) is used in the measurement of wheat protein content. Nevertheless, the drawback of this method is that the process is relatively tedious and slow, so it is not applicable to the rapid measurement of a large number of samples. So it is of huge practical significance to work out a fast, easy, and accurate method for the measurement.

In recent years, many studies have tended to focus on the combination of near-infrared reflectance spectroscopy (NIRS) with intelligent modeling techniques to achieve rapidness in the measurement of wheat protein content. The former has been widely used in the process of analysis, industrial control, nondestructive measurement, and other related fields [2–4]. As for the common methods of modeling, multiple linear regression, partial least squares, and artificial neural network are the most frequently applied ones, among which, the artificial neural network algorithm, radial basis function (RBF), in particular, is favored by many researchers because of its ease of use, high fitting, and high nonlinear approximation ability. The RBF neural network is a kind of feed forward network that has been widely applied because of its strong global optimization ability and good generalization ability [5,6]. Therefore, in this study we combined NIRS with RBF neural network to establish the prediction model of wheat protein content measurement.

In actual applications, the number of clustering centers in the hidden layers of the RBF neural network and the output weight value have a large impact on RBF neural network performance, so finding out the exact number of clustering centers in the hidden layers is of critical importance, and inappropriate choices of the number can easily lead to the "dimensionality curse" [7]. However, up to now, there is no effective method to calculate the optimal value of the number in theory, so the value can only be obtained by a large number of experiments. To a certain extent, it increases the difficulty of the application of RBF neural network and restricts the wide use of the RBF neural network in practice. To solve this problem, this paper proposes a particle swarm optimization (PSO) algorithm to optimize the number of cluster centers in the hidden layers of the RBF neural network and put forward relative improvements for the premature convergence and mechanical inertia weight setting of the PSO algorithm. In other words, the method applied in this study is the improved PSO algorithm, which is used to optimize the number of cluster centers in the hidden layers of the RBF neural network. The results demonstrated that the optimization ability of the improved PSO algorithm has been greatly improved. In addition, the improved PSO algorithm is effective in optimizing the number of cluster centers in the hidden layers of the RBF neural network and the output weight value—and thus the accuracy of the network output also increased.

2. RBF neural network

The RBF neural network is a typical three-layer feed forward network that consists of an input layer, a hidden layer, and an output layer. The basic principle is that RBF is used as the "base" of hidden units to constitute a hidden layer space, so the input vector could be directly mapped to the hidden space (no need for weight connection). When the number of clustering centers in the hidden layers of the RBF neural network is determined, the mapping relation is also established. The mapping from hidden layer space to output space is linear, and the network output is the sum of the linear weight of implicit unit output. The weight here refers to the adjustable parameter in the network.

Suppose the number of the input nodes in RBF neural network is n, the number of the hidden layer nodes is m, and the number of the output node is 1; then the output can be presented as:

$$y_j = \sum_{i=1}^{h} \omega_{ij} \exp\left(-\frac{\left\| x_p - c_i \right\|^2}{2\sigma_i^2} \right) \quad j = 1, 2, \ldots, m, \tag{1}$$

where ω_{ij} is the output weight, c_i is the ith center of the RBF hidden node, and σ_i is the width of the RBF hidden node. These three parameters have a great influence on the performance of the RBF neural network prediction model, especially on the determination of the number of hidden layer centers, because the complexity of the approximation function and the input space dimension of the actual sample are in exponential growth relation. Therefore, when the number of input nodes is large, the RBF neural network model can easily lead to "dimensionality curse".

3. PSO algorithm and its improvement

3.1. *Introduction of PSO algorithm*

The origin of the PSO algorithm is the artificial life and evolutionary computation theory, whose prototype is an imitation of the predation behavior of birds. The algorithm initializes a group of particles in the solution space, and each particle represents a potential optimal solution of the extremal optimization. The characteristics of the particles are represented by three indicators: position, speed, and fitness value. The fitness value is computed via the fitness function, and the quality of its value represents the quality of the particles [8]. The particles move in the solution space, and the position of the particles is updated by tracking the individual extremum Pbest value and the group extremum Gbest value. Every time the position of particles changes, the fitness value would be computed at once. And by comparing the fitness value with the individual extremum Pbest value and the group extremum Gbest value, the individual extremum Pbest value and the group extremum Gbest value could be updated [9].

3.2. *The improvement of PSO algorithm*

The PSO algorithm has a high converging speed and wide applicability, but it converges slowly and diverges easily in its later evolutionary stage. So, setting of the inertia weight ω is of vital importance, because it makes the particles maintain their motion inertia and gives a trend of expanding searching space. When ω takes a greater value, global search stands in advantage and the converging speed would be faster, but it would be difficult to obtain accurate solutions. When ω takes a smaller value, local search is more of an advantage and more

accurate solutions can be obtained, but the converging speed is relatively slower [10,11]. Currently, all the common strategies attempt to decrease linear weight coefficient ω with the increase in the number of iterations. However, the mere decreasing of ω cannot meet the requirement of a faster searching speed in the earlier stage and a slower searching speed in the later stage. In the actual searching process, the weighting transformation of PSO cannot be reflected simply by the linear decrease in ω. So, in this study, we adjusted the weighting coefficient ω in a dynamic manner—that is to say, we made adjustments to ω with the different characteristics of the particles.

If we assume that the size of the current particle swarm is N, then $f_{avg} = \frac{1}{N}\sum_{i=1}^{N} f_i$ is the average of the fitness value of all particles. f_i is the ith particle fitness value in the current iterative process. The fitness value of the global optimum particle is denoted as f_{max} in the current round of iteration, f_{havg} is the average of the fitness value of those particles with a higher value than f_{avg}, and Δ is defined as $\Delta = f_{max} - f_{havg}$.

(1) When $f_i \le f_{avg}$

The fitness value of particles would be smaller than the average fitness value in the current round of iteration, which is to say these particles have a low fitness value in the group. So, to improve the global searching capability, we need to set a higher inertia weight. This is expressed as follows:

$$\omega = 1.2 - \frac{1-\exp(-\Delta)}{1+\exp(-\Delta)} \tag{2}$$

(2) When $f_{avg} < f_i \le f_{havg}$

The fitness value of the particle is higher, and the global and local optimization capabilities of the particle are more prominent. The formula for weight can be expressed as:

$$\omega = \begin{cases} \omega - \left(\omega - \omega_{min}\right)\left[1 - \exp\left(-\frac{t}{t_{max}}\right)\right] & 0 \le t < t_{max} \\ \omega_{min} & t = t_{max} \end{cases} \tag{3}$$

As is shown in Formula 3, at the starting phase, higher values ought to be given to ω to take advantage of its higher global searching ability of the particles. Afterward, there should be a gradual decrease in ω in the latter stage to enhance the local searching ability of the particles.

(3) When $f_{havg} < f_i \le f_{max}$

The fitness value of the particles reaches the top. Therefore, it would be the time to reduce the weight value to improve the local searching capability of the particles. And this can be expressed as:

$$\omega = \omega - \omega_{min}\left(\frac{f_i - f_{havg}}{f_{max} - f_{havg}}\right) \tag{4}$$

As shown in Formula (4), the values of ω gradually decreased with the decrease in f_i. When $f_i = f_{max}$, the value of ω reaches the minimum, ω_{min}.

In Formulas (2–4), the ω on the right side still follows a linear descending trend, that is,

$$\omega = \omega_{max} - \frac{\omega_{max} - \omega_{min}}{k_{max}}k, \tag{5}$$

where k is the current time of iteration and k_{max} is the predetermined maximum number of iteration. In addition to the trouble in setting the weight ω, premature convergence is also a noticeable defect in the PSO algorithm. In order to solve this problem, we borrowed some ideas from the mutation operator of genetic algorithm, by introducing the mutation operation into the PSO algorithm. To be more exact, we reinitialized some variables at a certain probability. So it expanded the searching space of the shrinking population in iteration, allowing particles to escape from their optimal value positions in the previous search, thus enabling searching in a bigger space and also maintaining the population diversity. However, the mutation operator used in this paper is not ordinary or simple; it is a variation of the Sigmoid from the literature [12], and its specific form appears as follows:

$$P_m = 0.5 - \frac{1}{1+\exp(-k\Delta)} \tag{6}$$

Here, $k>0$; the definition of Δ is the same as described above.

4. Using PSO algorithm to optimize RBF neural network

In this paper, the operating process of the RBF neural network that was optimized by the improved PSO algorithm could be described as follows: first, initialize a few hidden cluster centers, and then optimize the connection weight of the RBF neural network by the improved PSO algorithm. After a certain number of rounds of iteration, if the network output error does not meet the requirement, we increased the number of cluster centers in the hidden layers, and used the improved PSO algorithm to optimize the network weight value. This process was repeated until the network error was within the required limits. The specific steps of this algorithm are explained below.

Step 1. Initialize the number of cluster centers in the hidden layers of the RBF neural network (usually input 1) and increase the step size l (usually input 1), the width of RBF function, and the target network error.

Step 2. Start training the network, and calculate the output error of each input vector. The one with the highest error value will be used as the weight vector to produce new cluster centers in the hidden layers. When the number increases to l, the network training will be stopped.

Step 3. Optimize the connection weights of the RBF neural network by the improved PSO algorithm, using $f = \sqrt{\sum_{i=1}^{n}|y_i - y'_i|}$ (where y_i is the actual output and y'_i is the target output) as the evaluation function for the improved PSO algorithm. The iteration will be stopped after n times. And if the network output still does not match the expected target, return to Step 2; otherwise, the algorithm will be stopped.

The wheat sample used in the experiment was given and tested by the Institute of Agricultural Quality and Safety under the Heilongjiang Agriculture Academy, China. The wheat samples were all produced in 2010, collected from a dozen of main wheat-producing areas in order to make the samples better reflect the actual condition of wheat in China. The protein content of wheat samples ranged from 10.8542 to 18.3139, demonstrating strong heterogeneity and high typicality. The model of the near-infrared spectroscopy

Table 1 – Predicted results of the model.

Sample number	Measured value (%)	Predicted value (%)	Relative error	Sample number	Measured value (%)	Predicted value (%)	Relative error
1	13.491	13.149	0.0253	21	14.36	13.694	0.0463
2	14.768	14.465	0.0205	22	12.86	12.781	0.0061
3	15.864	15.945	0.0051	23	12.78	12.534	0.0192
4	16.485	16.81	0.0197	24	14.69	14.631	0.0040
5	16.214	15.79	0.0261	25	13.63	13.232	0.0292
6	14.371	14.053	0.0221	26	13.43	13.241	0.0140
7	15.803	15.551	0.0159	27	14.34	14.047	0.0204
8	13.648	13.605	0.0031	28	13.73	13.794	0.0046
9	14.602	14.526	0.0052	29	13.85	13.672	0.0128
10	13.92	13.885	0.0025	30	14.37	14.289	0.0056
11	15.508	15.193	0.0203	31	13.98	13.909	0.0050
12	15.146	14.721	0.0280	32	14.68	14.374	0.0208
13	15.299	15.253	0.0030	33	15.74	15.092	0.0411
14	15.749	15.881	0.0083	34	13.57	13.44	0.0095
15	16.276	16.331	0.0033	35	15.27	15.049	0.0144
16	14.24	13.802	0.0307	36	15.66	15.52	0.0089
17	14.01	13.644	0.0261	37	16.38	15.932	0.0273
18	14.81	14.317	0.0332	38	14.28	14.081	0.0139
19	14.46	13.659	0.0553	39	14.41	14.438	0.0019
20	13.89	13.664	0.0162	40	16.88	16.792	0.0052

acquisition instruments used in the experiment was Infratec 1241 grain quality analyzer produced by FOSS. When collecting the near-infrared diffuse reflection spectra, the parameter settings of the instrument were as follows: spectral range, about 850–1050 nm; scanning step, 2 nm; sample number, 10; frequency of scans, 10 times. At room temperature, after warming up the instrument for 20 minutes, we filled the sample cup with decontaminated wheat samples and compressed them to collect spectra. After scanning all the samples, 14,000 absorbance points were collected altogether. Because each sample was scanned 10 times repeatedly, we took the average spectrum as the sample spectrum. Therefore, 140 groups of spectral data were used to establish the prediction model. According to the GB 2905-82 cereal, legume seed crude protein measurement method, the protein content in wheat was measured with the semimicro-Kjeldahl method, through which we obtained the data of the protein content of these 140 samples. These data were used as the desired output for the RBF neural network model.

First of all, 140 samples were divided into two groups: 100 samples were used to establish the model, and the remaining 40 samples were used to calibrate and inspect the model. The wave point of each sample obtained by NIRS analysis technology was 100; if all were denoted as the input of RBF neural network, the structure of the network would be very complicated. Hence, we used the Successive Projections Algorithm [13] to select the 11 most sensitive wave points as network output. Thus, the number of input layer nodes of RBF neural network was set to 11, and the number of output layer nodes was set to 1. Then, we optimized the number of hidden layer nodes and the connection weights between the hidden layers and the output layers using the improved PSO algorithm. The number of population particle was set to 20, the dimension of each particle was set to 11, and the iteration number of the evolutionary algorithm was set to 100.

We used 40 samples to test the optimized RBF neural network. The value of root-mean-square error of prediction (RMSEP) reached 0.26378, and the value of R reached 0.997. The predictive value and the relative errors are shown in Table 1. The predicted results suggested that the predictive values obtained by the RBF neural network that was improved by the optimized PSO algorithm were very close to the value calculated through chemical measurement. The minimum relative error between the predicted values and the measured values was 0.0019 and the maximum relative error was 0.0553, which could ensure the accuracy of the prediction model in terms of wheat protein content measurement to a large extent. Fig. 1 shows the difference between the predicted values and the measured values of those

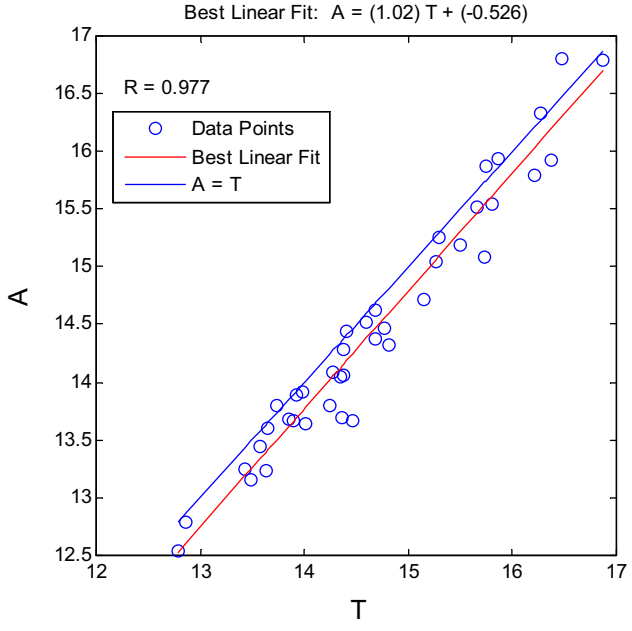

Fig. 1 – The diagram of fitting effect.

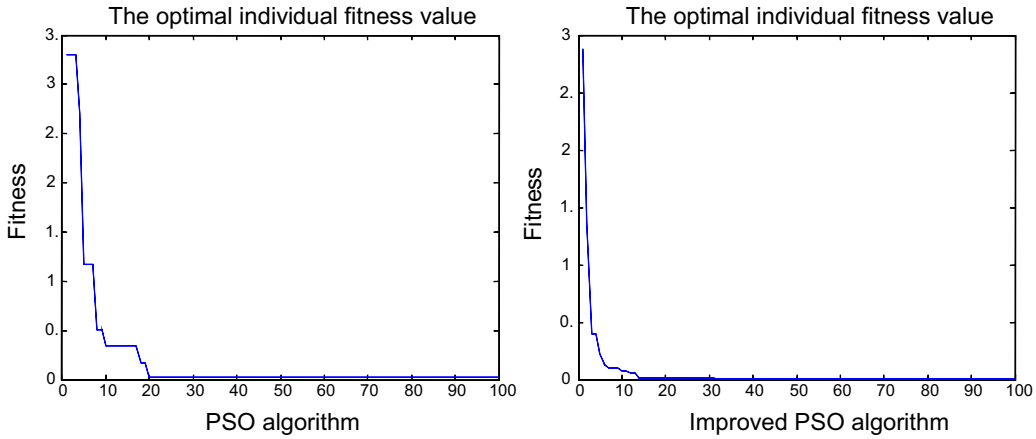

Fig. 2 – The evolution speed of the optimal individual fitness value.

40 samples. It can be seen in Fig. 1 that: (1) the high correlativity between the predicted values and the measured values indicated the model's ability of measuring wheat with varying qualities; (2) although some groups manifested lower correlativity, it was most probably caused by the changing environment or operator error in the actual experiment, and it did not affect the result of the model; (3) the redundancy in the number of samples with protein content between 13% and 17% might cause excessive fitting in this range, and thus cause inaccuracy in the data, so further improvements are still needed in this respect.

In addition, the results also showed that, when the RBF neural network has not been optimized, 27 cluster centers in hidden layers were required to keep the prediction error in check, with training time at 7 minutes and 24 seconds. However, the optimized RBF neural network required only nine cluster centers in the hidden layer, which greatly reduced the complexity of the network structure. Training time was only 2 minutes and 41 seconds, which also increased the efficiency of the process. In comparison to the classic PSO algorithm, the improved PSO algorithm used in this study significantly accelerated the evolution speed of the optimal individual fitness value in the PSO algorithm. The comparison is illustrated in Fig. 2.

5. Conclusion

This paper fully discussed the pros and cons of the RBF neural network, the PSO algorithm principle, and their improvement methods. The RBF neural network was optimized by the improved PSO algorithm, and 140 wheat samples from the major wheat-producing areas all over China in 2010 were used as the focus of this research. This study also demonstrated the feasibility and validity of the protein content measurement model, which is an integration of NIRS and optimized RBF neural network. As proven in the experimental analysis, the RBF neural network optimized by the improved PSO algorithm can realize high predicting ability in the measurement of the protein content in wheat samples, and the model's RMSEP and R values reached 0.26378 and 0.977, respectively. These figures indicate that the model satisfied the accuracy and speed requirements of the research. Meanwhile, the RBF neural network optimized by the improved PSO algorithm greatly reduced the complexity of network structure and also improved the training speed of the network. Moreover, the optimal individual fitness evolution speed of the improved PSO algorithm was greatly improved, which effectively avoided premature convergence.

Conflicts of interest

None.

Acknowledgments

The authors gratefully acknowledge the financial support from the 2013 graduate innovation project of Heilongjiang University, Harbin, China.

REFERENCES

[1] Wang L. 2003 wheat distribution and quality conditions. Int J Grain Process 2004;4:21–9.
[2] Liang X, Ji H. Applications of near infrared spectroscopy technology in analyzing the quality of crops. Zhong Guo Zhong Yao Za Zhi (Chin Agric Sci Bull) 2006;22:366–71 [In Chinese, English abstract].
[3] Yang C-W, Chen S, Fu O, et al. A robust identification model for herbal medicine using near infrared spectroscopy and artificial neural network. J Food Drug Anal 2011;19:9–17.
[4] Yao X, Zhang R, Li X, et al. Study on analysis of soybean quality by near infrared transmittance spectroscopy. Int J Soybean Sci 2006;4:417–20.
[5] Lu J-H, Zhang Y-B, Zhang Z-Y, et al. Application of wavelet transform-radial basis function neural network in NIRS for determination of rifampicin and isoniazide tablets. Zhong Guo Zhong Yao Za Zhi (Spectrosc Spectral Anal) 2008;28:1264–7 [In Chinese, English abstract].
[6] Huang G, Han L. Near infrared reflectance spectroscopy analysis of compost products using nonlinear support vector

machine with RBF nucleus. Zhong Guo Zhong Yao Za Zhi (Acta Opt Sin) 2009;29:3556–60 [In Chinese, English abstract].
[7] Zhong L, Rao W-B, Zou C-M, editors. Artificial neural networks and fusion applications technology. Beijing: Science Press; 2007.
[8] Zhen J, Hui-Lian L. Particle swarm optimization and application. Beijing: Science Press; 2009.
[9] Li Y, Tang X, Liu J. Application of particle swarm optimization algorithm in improving the stability of sensor. Zhong Guo Zhong Yao Za Zhi (Chin J Sci Instrum) 2010;31:1756–61 [In Chinese, English abstract].
[10] Wang S, Li A. Training fuzzy neural networks using particle swarm optimizer. Zhong Guo Zhong Yao Za Zhi (Chin J Sci Instrum) 2004;25:938–9 [In Chinese, English abstract].
[11] Yang C-H, Gu L-S, Gui W-H. Particle swarm optimization algorithm with adaptive mutation. Zhong Guo Zhong Yao Za Zhi (Comput Eng) 2008;34:188–90 [In Chinese, English abstract].
[12] Wu H, Zhu C, Chang B, et al. Adaptive genetic algorithm to improve group premature convergence. Zhong Guo Zhong Yao Za Zhi (J Xi'an Jiaotong Univ) 1999;33:27–30 [In Chinese, English abstract].
[13] Gao H-Z, Lu Q-P, Ding H-Q, et al. Choice of characteristic near-infrared wavelengths for soil total nitrogen based on successive projection Algorithm. Zhong Guo Zhong Yao Za Zhi (Spectrosc Spectral Anal) 2009;20:2951–4 [In Chinese, English abstract].