JOURNAL OF FOOD AND DRUG ANALYSIS 22 (2014) 230–235

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:

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

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