Quantitative Regression Modeling of Cocoa Bean Content Based on Gated Dilated Convolution Network

Yayu Chen[1],Wenju Zhou[1], Minrui Fei*[1], Haikuan Wang[1] and Xiaofei Han[1]

[1] School of Mechanical Engineering and Automation, Shanghai University
Shanghai, China
{chenyayu, zhouwenju}@shu.edu.cn, mrfei@staff.shu.edu.cn,
hkwang@shu.edu.cn, hanxiaofei@i.shu.edu.cn.

Abstract. By analyzing the near-infrared spectrum, the quantitative relationship between the spectral data and the target component can be modeled. This paper proposes a predictive regression model based on one-dimensional convolutional neural network. Based on the traditional convolutional neural network, gating mechanisms and dilated convolutions are combined. The particle swarm optimization method is used to optimize the hyperparameters of one-dimensional convolution. The end-to-end near-infrared predictive regression model does not require wavelength selection. It is convenient to use and has a strong promotional value. Taking the public cocoa beans near-infrared data set as an example, the method can predict the moisture content and fat content in cocoa beans, and the effectiveness of the method is verified. Comparing the improved one-dimensional convolution with traditional one-dimensional convolution results and partial least squares regression, it shows better prediction accuracy and robustness.

Keywords: Convolutional neural network , Infrared spectroscopic data, Gating mechanisms, Dilated convolution.

1 Introduction

Near-infrared spectroscopy analysis technology has a wide range of applications due to its fast analysis speed, no need for chemical reagents, reduced environmental pollution, low cost, and easy online analysis. Near-infrared spectroscopy detection technology will not destroy the substance that predicts the information of a sample by establishing a correlation model between the near-infrared spectroscopy data of the sample and the true value of the attribute to be measured.

In the past few decades, due to the complexity of NIR spectrum and the diversity of analysis objects, NIR quantitative analysis generally uses multivariate information processing technology, such as partial least squares regression (PLSR)[1-3], multiple linear regression (MLR)[4], principal component regression. partial least squares regression is the most commonly used linear stoichiometric model. However, with the deepening of the application, these methods also expose many problems, such as: MLR will encounter collinearity problems and the limitation of the number of input

variables, PCR can not distinguish noise or valid information, and the running speed is slow, and PLS cannot be effective Handling nonlinear problems.

The study of machine learning method is a significant task to resolve the problem of accurate quantitative measurement in chemometrics. Especially in the industrial field, The machine learning algorithm of fast near-infrared detection has established several mathematical models to estimate the fidelity of measured data, such as artificial neural network (ANN)[5], support vector machine regression (SVR)[6] and so on. However, they all produce more parameters and complicated operation, which is not conducive to the promotion of NIR analysis technology.

Neural networks have powerful modeling capabilities and have been widely used in computer vision and speech recognition for multi-large-scale data and achieved great success. At the same time, due to its deep network structure and nonlinear activation ability, the application of neural network in modeling and analysis of near-infrared spectroscopy has been reported. Chao Ni et al. proposed a method to quantify the nitrogen content of pinus massoniae seedling leaves by using variable weight convolutional neural network [7]. Muhammad Bilal et al. proposed to reduce the eigenvector size by training L2 regularized sparse auto en-coder[8]. Salim Malek et al. used a one-dimensional convolutional neural network for feature extraction[9]

In view of the characteristics of deep learning's powerful expression ability, we improve the dilation gate convolution model and introduce it into the prediction regression of infrared spectrum. In order to optimize the convolution kernel and some other parameters, we used the PSO, which aims to study a non-destructive and effective method to identify the cocoa fat content and the moisture content and explore a new model applied to the learning of spectral characteristics of foods.

2 Neural Network

2.1 CNN

CNN model uses gradient descent to minimize the loss function, reverse adjustment of the weight parameters in the network layer by layer, and improve the accuracy of the network through frequent iterative training.

The convolution calculation layer is to extract feature on the original input. There can be multiple convolutional layers. We specify the size and weight of the convolution kernel. In the input matrix, select the data convolution calculation with the same size as the convolution kernel. Continuously extract and compress features, and then obtain relatively advanced features for convolution. The parameters of the convolution kernel of each convolution layer are the same, which is weight sharing, which significantly reduces the number of parameters.

In the excitation layer, the activation function is mainly used to smooth the results of the operation. Meanwhile, the activation function can remove the redundant data, and retain the main features of the data.

After the excitation layer, the pooling layer specifies the size of a window and replaces it with a value. It is commonly used to maximize the pool and average the pool. It can effectively reduce the size of the matrix and prevent overfitting.

The output layer outputs multiple neurons. Depending on the type of output, the output layer uses different types of activation. For regression models, we use linear activation.

CNN training can be seen as two stages of forward propagation and back propagation. Back propagation uses the BP algorithm. The difference between CNN and traditional multilayer neural networks is the addition of convolution and pooling operations, so CNN is in convolution The training process of the layer and the pooling layer is different from the neural network. Before the CNN network starts training, first, the network must be initialized, the connection weights are assigned random numbers in (-1, 1), and the calculation accuracy and maximum learning are given. Number of times. Second, randomly select input samples and corresponding expected outputs. Finally, start the training of the network.

2.2 Spectral Regression Model Based on CNN

The wavelength range of NIR is 780nm to 2526nm, and a complete nir spectrum sequence is relatively long. Therefore, dilated convolution and gating mechanisms are introduced .In order to identify larger features, all pooling layer uses maximum pooling.

Block Structure. The gating mechanism will control the flow of information to make it multi-path circulation. Aäron van den Oord [10]proved that it could model more complex interactions with LSTM-style gating mechanism.

$$y = \tanh(x * W_1 + b_1) \odot \sigma(x * W_2 + b_1) \tag{1}$$

W and b represent kernel and deviation, which represents the sigmoid function, $\odot$ is element-wise multiplication, and * is the convolution multiplier. The gradient of gating in LSTM style is

$$\begin{aligned} \nabla y = {} & \tanh'(x * W_1 + b_1)\nabla(x * W_1 + b_1)\Theta\sigma(x * W_2 + b_2) \\ & + \sigma'(x * W_2 + b_2)\nabla(x * W_2 + b_2)\Theta\tanh(x * W_1 + b_1) \end{aligned} \tag{2}$$

Among them, $\tanh'(x * W_1 + b_1)$, $\sigma'(x * W_2 + b_2) \in (0,1)$ .Under normal circumstances, the problem of disappearing gradients appears as the network depth increases.

We constructed a set of block structures (as shown in Fig 1), which is equivalent to two gating mechanisms in a set of block structures.

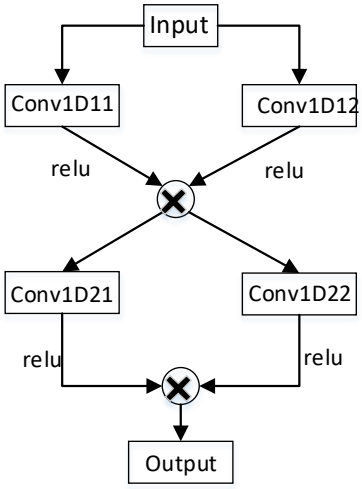

Fig. 1..block structure

Dilated Convolution. Because the complete sequence of a near-infrared spectrum is relatively long, dilated convolution is used to enable the CNN model to capture a longer distance without increasing model parameters. Dilated convolution was first developed from the wavelet decomposition algorithm[11] called " convolution with dilated filter ", and later applied to image and semantic segmentation. Dilated convolution modifies the convolution operator so that it can use filter parameters in a variety of ways. In a two-dimensional convolution,

$F: Z^2 \to R$ is a discrete function, $\Omega_r = [-r, r]^2 \cap Z^2$, $k: \Omega_r \to R$ is a discrete filter of size $(2r+1)^2$ [12].The two-dimensional discrete convolution operator * is defined as

$$(F*k)(p) = \sum_{s+t=p} F(s)k(t) \tag{3}$$

Then r is a dilation factor and $*_r$ can be defined as

$$(F*_r k)(p) = \sum_{s+rt=p} F(s)k(t) \tag{4}$$

So we can extend to one-dimensional expansion convolution,

$$(F*_r k)(p) = \sum_{s+rt=p} F(s)k(t), \quad t \in [-m, m] \cap Z \tag{5}$$

It can expand the vision of convolution kernel without increasing computation. Fig 2 shows two convolutional neural networks with three layers, and the size of the convolution kernel is 3. Traditional convolution obtains a feature through a three-layer convolutional layer. The first layer has 7 inputs. After three-layer expansion convolution, a feature is obtained. The first layer has 15 inputs. In the dilated one-dimensional convolution, if the dilated rate of the kernel increases exponentially, the size of the receiving domain is exponentially correlated with the depth of the layer.

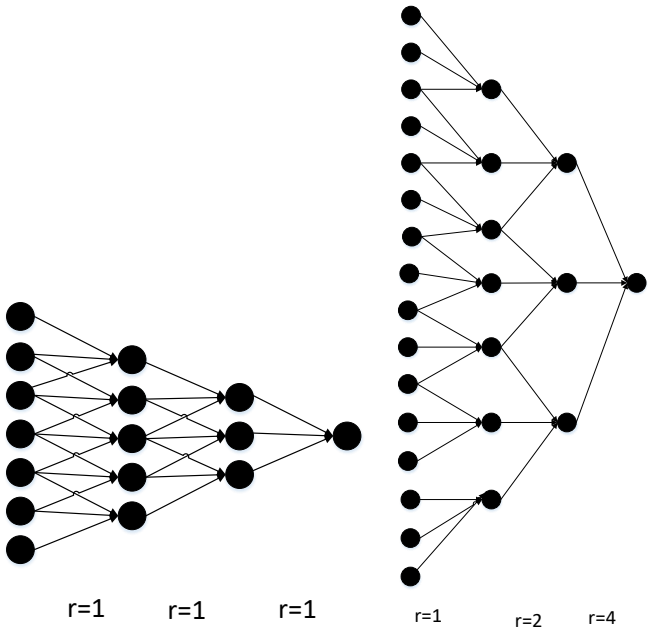

Fig. 2 . Left : a 1-D CNN with three conventional convolutional layers. Right: a 1-D CNN with three dilated convolutional layers

3 PSO Algorithm

Through the combination of local perception, weight sharing and pooling, CNN reduces the number of parameters in network training. However, CNN's hyperparameters, such as the size of convolution kernel and the number of convolution kernel ,all need to be set manually.

PSO was proposed by Kennedy and Eberhart in 1995[13][14].It is a global search algorithm, which has become a research hotspot due to its fast convergence and few parameters .The idea comes from simulating the behavior of birds flying for food in a cluster. Scattered birds then fly for food. There is an indirect mechanism to let them know the distance between food and the current position. Each particle representing

every bird, has its own velocity and position, can be determined by the problem of the definition of fitness function of particle fitness, and then iterate unceasingly, The particle's flight speed and the next position are calculated by the historical optimal solution of the particle and the global optimal solution until the global optimal solution is found. The PSO algorithm has the characteristics of fast search speed, easy algorithm implementation, and high precision.

Suppose that in a d-dimensional search space, m particles form a population, and the ith particle is represented as a d-dimensional vector $\mathrm{x}_i = (x_{i1}, x_{i2}, ..., x_{id}), i = 1, 2, ..., m$ The position of the ith particle in the d-dimensional search space is x_i , The velocity of the ith particle is also a d-dimensional vector $v_i = (v_{i1}, v_{i2}, ..., v_{id})$. At present, the optimal position searched for the ith particle is $p_i = (\mathrm{p}_{i1}, \mathrm{p}_{i2}, ..., \mathrm{p}_{id})$, and the optimal position searched for the entire population is $p_i = (\mathrm{p}_{g1}, \mathrm{p}_{g2}, ..., \mathrm{p}_{gd})$ .The standard particle swarm optimization algorithm uses the following formula to update the particle velocity and position:

$$\mathrm{v}_{\mathrm{id}}^{k+1} = w v_{\mathrm{id}}^{k} + c_1 r_1 (\mathrm{p}_{id}^{k} - x_{id}^{k}) + c_2 r_2 (p_{gd}^{k} - x_{id}^{k}) \tag{6}$$

$$x_{\mathrm{id}}^{k+1} = x_{\mathrm{id}}^{k} + v_{\mathrm{id}}^{k+1} \tag{7}$$

In the above formula, $\mathrm{v}_{\mathrm{id}}^{k}$ is the d-th component of the velocity of particle i in the k-th iteration.. x_{id}^{k} is the d-th component of the position of particle i in the k-th iteration. p_{id}^{k} is the d-th component of the local optimal value of particle i. $\mathrm{p}_{\mathrm{g}d}^{k}$ is the d-th component of the global optimal value. k represents the current iteration times, c_1, c_2 are the learning factor, r_1, r_2 are the random number between [0,1], and w is the inertia weight.

Assuming that the optimization goal is to find the minimum value, the update methods of p_{id}^{k} and $\mathrm{p}_{\mathrm{g}d}^{k}$ are:

$$\mathrm{p}_{id}^{k+1} = \begin{cases} x_i^{k+1} & f(x_i^{k+1}) < f(\mathrm{p}_{gd}^{k}) \\ \mathrm{p}_{gd}^{k} & f(x_i^{k+1}) \geq f(\mathrm{p}_{gd}^{k}) \end{cases} \tag{8}$$

$$\mathrm{p}_{gd}^{k} = \arg\min_{1 \leq i \leq N} (f(\mathrm{p}_{gd}^{k})) \tag{9}$$

Where $f(\bullet)$ is the objective function and N is the total number of particles in the particle swarm.

The steps of particle swarm optimization for CNN hyperparameters are:

(1) Determine the number and value range of CNN hyperparameters.

(2) Set the parameters of PSO , such as inertia factor, learning factor ,and maximum iteration number.
(3) Initialize the particle swarm, and each particle swarm represents a group.
(4) CNN was trained according to hyperparameters, the fitness function value of each particle was calculated, and the advantages and disadvantages of particles were sorted according to the fitness function value.
(5) Update the optimal position of the current particle and particle swarm according to the value of the fitness function.
(6) Update the position and speed of each particle to generate a new group of particles.
(7) Add 1 to the current number of iterations.
(8) If the maximum current iteration number is exceeded, the optimization is terminated; otherwise, step 4 is returned.
(9) According to the optimal position of PSO algorithm, the optimal value of CNN hyperparameters is obtained.

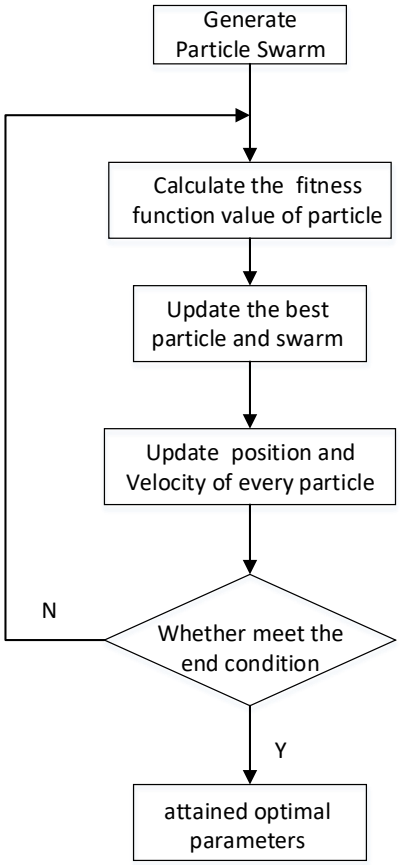

Fig.3 Particle swarm optimization algorithm (PSO) optimizes CNN hyperparameters flowchart

4 Experimental Methods

4.1 Dataset Description

The feasibility and effectiveness of the method proposed is based on the public cocoa bean near-infrared spectral dataset[15]. The cocoa bean near-infrared data set has 72 samples. Fat and water are the internal quality parameters of cocoa beans samples. They correspond to the chemical information in cocoa beans and consist of structured molecular bonds of C-H-O, C-H, O-H, N-H and R-O-H. Near-infrared spectroscopy is an electromagnetic wave with a wavelength in the range of 780 nm to 2 526 nm, which can reflect the combined frequency and frequency doubling of the vibrations of the near-infrared spectral region and the hydrogen-containing groups in organic molecules to determine the hydrogen-containing groups of organic molecules in the sample Characteristic information, specific wavelengths in the near infrared region can be used to describe the quality parameters of cocoa bean samples.

The wavelength range of near infrared spectrum in this data set is 1000nm~2500nm.Training sets are used to establish quantitative analysis model, the test sets are used to evaluate the generalization performance of the model. The spectra of each sample between 1000nm and 2000nm is more dispersed. The spectrum of each sample between 2000nm and 2200nm is relatively messy.

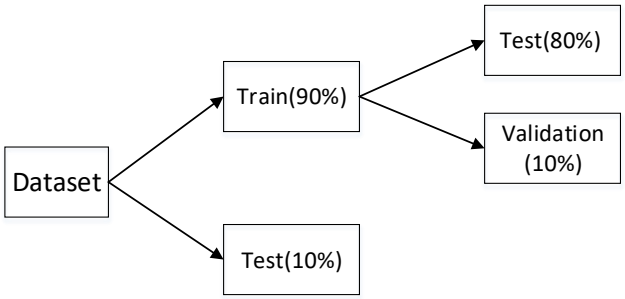

Fig . 4. Modeling the Dataset for Training and Testing the Model

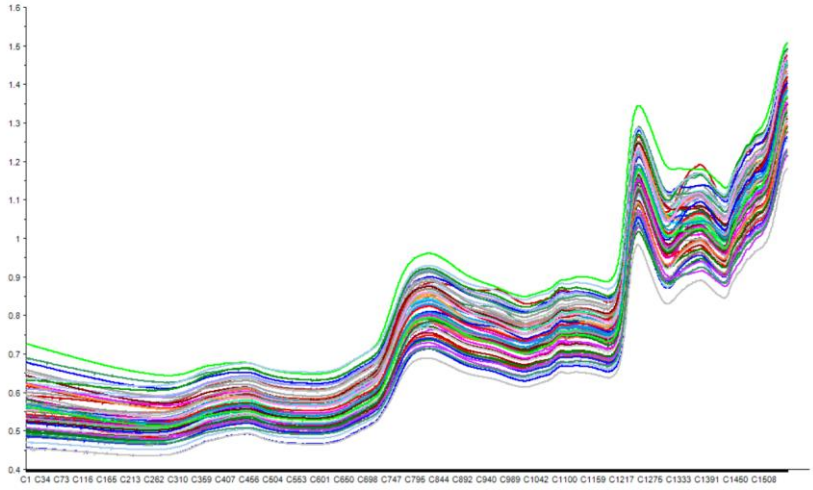

Fig. 5. Cocoa beans near infrared spectrum

4.2 Spectral Pretreatment

In order to eliminate the influence of noise on the model in the process of spectral measurement, the data preprocessing method is usually used to preprocess the spectral data before the model is established. Common pre-processing methods include derivative, multivariate scattering correction and vector normalization. Derivative preprocessing can reduce the impact of low-frequency noise and improve resolution. Although neural networks have powerful feature extraction, preprocessing the spectrum and selecting the wavelength of the spectrum can make the prediction results more accurate.

Experiments have shown that the cocoa beans have a high degree of spectral overlap after SNV, MSC, and derivatives. They can not clearly reflect the differences between the samples, which is not conducive to the regression prediction of the neural network. The spectral preprocessing in this paper first adopts Gaussian smoothing to eliminate the dimensional influence between the indicators and the influence of the size of the variable itself and the value of the variable, then standardizes the spectral data to speed up the training of neural networks.

4.3 Network Structure

The network structure we propose is as follows: we construct a group of block structures, which is equivalent to two gating mechanisms in a group of block structures, and the subsequent groups repeat the same pattern. Zero padding is applied to the one-dimensional convolution to make the input and output size the same. Use maximum pooling to take the largest feature point in the neighborhood. Add dropout to the pool-

ing layer to make some neurons in the pooling area randomly set to 0 with a probability of 0.1, which can reduce the probability of overfitting.

4.4 Model Evaluation Criteria

In order to evaluate the regression method and make a direct comparison with the latest results, we used the root mean square error (RMSE) and coefficient of determination (R2) . The RMSE is very sensitive to the large or small error in a group of measurements. Coefficient of determination (R2) represents that the percentage of variation in variable y can be explained by the controlled independent variable x. The higher the goodness of fit, the higher the degree of independent variable's interpretation of the dependent variable.

$$R2 = 1 - \frac{SSE}{SST} \tag{10}$$

SSE is sum of squares for error, SST is sum of squares fortotal.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{11}$$

N is the total number of test samples, y_i and $\hat{y}_i$ are the true and predicted values of x_{it} samples . Gain measured using accuracy can provide information about how much our method has improved [16], the formula is：

4.5 Parameter Setting

The improved one-dimensional convolutional neural network architecture includes some parameters: the number of convolutional layers L, and the number of samples m selected for one training and Particle swarm optimization parameters. As shown in table 1and table 2 are the set value of parameters. The parameter setting is greatly reduced, and it is easier to operate and use for users who lack professional knowledge. It can be seen from the set number of layers that for the regression model of the near infrared spectrum, the deeper the structure, the better the result. Since the network is not deep, there is no problem of the gradient disappearing.

Table 1. Best parameter values of the Improved PSO-1DCNN for each dataset

Dataset	L	m
moisture	1	4
fat	2	2

Table 2. Particle swarm optimization parameters

Parameter	Value
inertia factor	0.5
learning factor	0.5
Number of iteration	10
Number of particles	5

Fig 6 is a sample-by-sample comparison between the estimated value and the actual output value of the test set of cocoa bean moisture content data set.

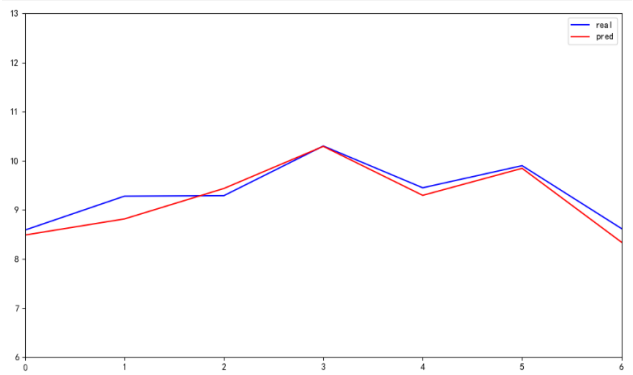

Fig. 6. Sample-by-sample comparison between the predictive value and the actual output value for the test set of cocoa moisture content data set

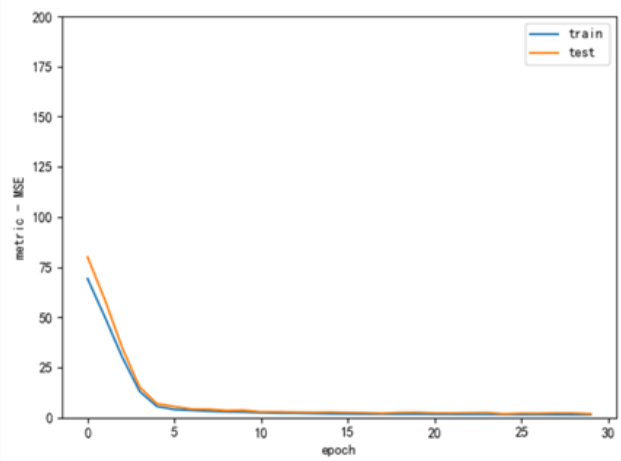

Fig. 7. Training error and verification error of water content

Fig 8 is a sample-by-sample comparison between the estimated value and the actual output value of the test set of cocoa bean fat content data set.

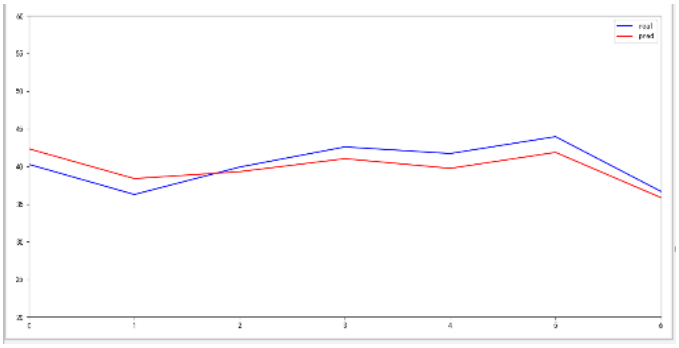

Fig.8. Sample-by-sample comparison between the predictive value and the actual output value for the test set of cocoa fat content dataset

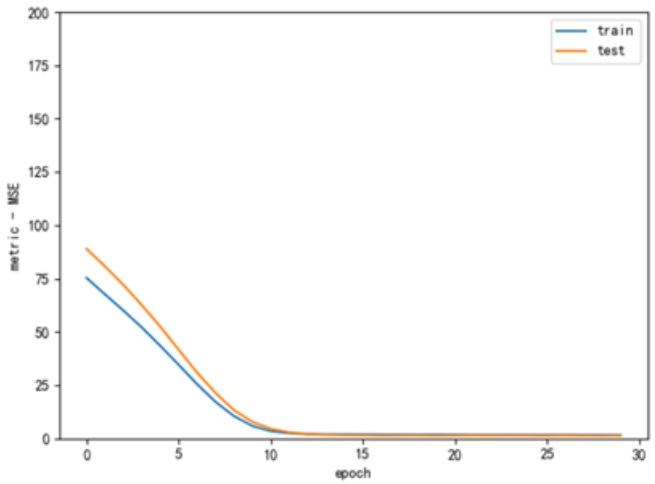

Fig.9. Training error and verification error of fat content

The figure above shows that the improved one-dimensional CNN can predict the moisture content and fat content of cocoa beans.

The improved CNN uses several blocks to superimpose, and its effect is not as good as blocks with few layers. Indicating that in the near infrared spectroscopy prediction regression model, the deeper the structure of the neural network does not mean the better the effect .

5 Discussion

In order to evaluate our method, we choose two regression methods, PLSR and traditional one-dimensional CNN. PLSR is the most commonly used linear stoichiometric model, and one-dimensional CNN is a common algorithm in machine learning. Table 3 and Table 4 are the moisture content and fat content of cocoa beans, and their R2 and RMSE are calculated by three methods. It can be seen that in predicting the moisture and fat content of cocoa beans, the improved PSO-1DCNN has the highest R2 and the lowest RMSE.

Table 3. Results for moisture content dataset

Method	R2	RMSE
Improved PSO-1DCNN	0.876	0.223
PSO-1DCNN	0.744	0.794
PLSR	0.861	0.472

Table 4. Results for fat content dataset

Method	R2	RMSE
Improved PSO-1DCNN	0.832	0.801
PSO-1DCNN	0.554	1.197
PLSR	0.825	0.846

6 Conclusion

We proposed a new method for chemometric data analysis based on CNN, improved the traditional one-dimensional convolutional neural network architecture, and added dilated convolution and gate mechanisms. Experimental results in cocoa beans near infrared spectroscopy show that the improved one-dimensional convolutional neural network can predict the moisture content and fat content of cocoa beans. Compared with the traditional one-dimensional convolutional neural network, the gate mechanism and expansion volume The integrated convolutional neural network has higher accuracy and better fault tolerance. Compared with traditional machine learning methods, simplifying the process of quantitative analysis of infrared spectroscopy is also more suitable for users who lack professional knowledge and is easy to promote.

Acknowledgments

This research is financially supported by Natural Science Foundation of China (61877065) and the National Key Research and Development Program of China （No. 2019YFB1405500).

References

1. Wold, Svante, Michael Sjöström, and Lennart Eriksson.:PLS-regression: a basic tool of chemometrics . Chemometrics and intelligent laboratory systems 58. 2 , 109-130 (2001)
2. Nicolai, Bart M., Karen I. Theron, and Jeroen Lammertyn.:Kernel PLS regression on wavelet transformed NIR spectra for prediction of sugar content of apple." Chemometrics and intelligent laboratory systems 85.2,243-252 (2007)
3. Arrobas, B. Gordillo, et al, F.:Raman spectroscopy for analyzing anthocyanins of lyophilized blueberries. 2015 IEEE SENSORS. IEEE, (2015)
4. Galvao, Roberto Kawakami Harrop, et al.:Multivariate analysis of the dielectric response of materials modeled using networks of resistors and capacitors. IEEE Transactions on Dielectrics and Electrical Insulation 20.3 ,995-1008 (2013)
5.Verikas, Antanas, and M. Bacauskiene.:Using artificial neural networks for process and system modelling. Chemometrics and Intelligent Laboratory Systems 67.2, 187-191 (2003)
6. Devos, Olivier, et al.:Support vector machines (SVM) in near infrared (NIR) spectroscopy: Focus on parameters optimization and model interpretation. Chemometrics and Intelligent Laboratory Systems 96.1 , 27-33 (2009)
7. Ni, Chao, Dong yi Wang, and Yang Tao.:Variable weighted convolutional neural network for the nitrogen content quantization of Masson pine seedling leaves with near-infrared spectroscopy. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 209 ,32-39 (2019)
8. Bilal, Muhammad, Mohib Ullah, and Habib Ullah.:CHEMOMETRIC DATA ANALYSIS WITH AUTOENCODER NEURAL NETWORK. Electronic Imaging ,679-1 (2019)
9. Malek, Salim, Farid Melgani, and Yakoub Bazi.:One-dimensional convolutional neural networks for spectroscopic signal regression. Journal of Chemometrics 32.5 ,e2977 (2018).
10. Van den Oord, Aaron, et al.:Conditional image generation with pixelcnn decoders. Advances in neural information processing systems. (2016)
11. Shensa, Mark J, F.:The discrete wavelet transform: wedding the a trous and Mallat algorithms. IEEE Transactions on signal processing 40.10,2464-2482 (1992)
12. Yu, Fisher, and Vladlen Koltun.:Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122 (2015)
13. Kennedy, James, and Russell Eberhart.:Particle swarm optimization. Proceedings of ICNN'95-International Conference on Neural Networks. Vol. 4. IEEE, (1995)
14. Shi, UYuhui, and Russell Eberhart.:A modified particle swarm optimizer. 1998 IEEE international conference on evolutionary computation proceedings. IEEE world congress on computational intelligence (1998)
15. Agussabti, Rahmaddiansyah, Purwana Satriyo, and Agus Arip Munawar.:Data analysis on near infrared spectroscopy as a part of technology adoption for cocoa farmer in Aceh Province, Indonesia. Data in Brief 29 (2020)

16 .AlHichri, Haikel, et al.:A novel fusion approach based on induced ordered weighted averaging operators for chemometric data analysis, Journal of Chemometrics 27.12 447-456.(2013)