Rapid Determination of Hemicellulose Content in Corn Stalks by Near-infrared Spectroscopy Based on Dung Beetle Optimizer

Baihong TONG, Jinming LIU, Jianfei SHI*

College of Information and Electrical Engineering, Heilongjiang Bayi Agricultural University, Daqing 163000, China

Abstract Corn stalks are a kind of common organic fertilizer and feed material in agriculture in China, as well as an important source of modern biomass energy and new materials. Hemicellulose is an important component in corn stalks, and it is very important to determine its content in corn stalks. In this paper, the feasibility of near-infrared spectroscopy (NIRS) combined with chemometrics for rapid detection of hemicellulose content in corn stalks was studied. In order to improve the accuracy of NIRS detection, a new intelligent optimization algorithm, dung beetle optimizer (DBO), was applied to select characteristic wavelengths of NIRS. Its modeling performance was compared with that based on characteristic wavelength selection using genetic algorithm (GA) and binary particle swarm optimization (BPSO), and it was found that the characteristic wavelength selection performance of DBO was excellent, and the regression accuracy of hemicellulose quantitative detection model established by its preferred characteristic wavelengths was better than the above two intelligent optimization algorithms.

Key words Hemicellulose; Near-infrared spectrum; Characteristic wavelength selection; Intelligent optimization algorithm; Dung beetle algorithm

DOI: 10.19759/j.cnki.2164 – 4993.2024.05.018

Corn stalks are an important by-product in agricultural production, which contains hemicellulose, a kind of complex polysaccharide playing a connecting role between cellulose and lignin[1]. The content of hemicellulose in corn stalks is about 20% to 30%. Hemicellulose is composed of various sugar monomers (such as glucose, xylose, arabinose and mannose) connected by different glycosidic bonds. As corn stalks serves as a renewable resource, the content of hemicellulose in them has extensive research value and application potential. In order to solve the problems of complicated operation and long testing time in traditional chemical methods for the determination of hemicellulose content in corn stalks, this study combined near infrared spectroscopy (NIRS) and chemometrics to determine hemicellulose content[2].

NIRS is a spectral analysis technique based on near-infrared region (780 – 2 500 nm). This technique obtains the molecular information of samples by analyzing the absorption, reflection or transmission of near-infrared light by substances. It has been widely used in agriculture, food, medicine, chemical industry and other fields because of its non-destructive and rapid characteristics and no need for sample pretreatment[3]. When near-infrared light shines on samples, molecules in samples will absorb photons with a specific wavelength, which will lead to the stretching and bending vibration of molecular bonds. These vibrations correspond to specific absorption wavelengths, thus forming unique spectra. Through the analysis of these spectral data, we can infer molecular structures and chemical composition contents of samples.

Due to the high complexity of NIRS full-spectrum modeling, the redundant wavelength variables will affect the modeling accuracy and lead to poor model accuracy. In recent years, intelligent optimization algorithm has been widely used in NIRS field with its powerful global search ability. Many experts and scholars often use intelligent optimization algorithm to select NIRS characteristic wavelength variables[4-5], and the characteristic wavelength variable selection is used to select high-correlation wavelength variables from high-dimensional spectra to build models, thereby effectively reducing the redundant information of spectra and the dimension of spectral data and improving the generalization ability of models[6].

Therefore, a partial least squares (PLS) regression quantitative detection model was established based on dung beetle optimizer (DBO) to select the characteristic wavelength of hemicellulose, and it was compared with genetic algorithm (GA)[7] and binary particle swarm optimization (BPSO)[8] models to explore the feasibility of selecting the characteristic wavelengths of NIRS by new intelligent optimization algorithm DBO, so as to effectively improve the detection performance of NIRS regression models for hemicellulose content.

Data Set and Method

Data set

Firstly, 184 samples of corn stalks were put in a ventilated place for natural air drying, and then crushed by a crusher. Then, the crushed corn stalks were stored in sealed plastic bags for hemicellulose content determination and spectral data collection. The samples were subjected to integrating sphere diffuse reflection scanning using a TANGO Fourier near-infrared spectrometer (Bruker Optics, Ettlingen, Germany) in the spectrum collection range of 3 946 – 11 542 cm^{-1}, with a resolution of 8.0 cm^{-1}, and

Received: June 19, 2024 Accepted: August 31, 2024

Supported by San Heng San Zong Project of Heilongjiang Bayi Agricultural University (ZRCPY202314).

Baihong TONG (2000 –), male, P. R. China, master, devoted to research about application of intelligent optimization algorithm in near-infrared spectroscopy.

* Corresponding author. associate professor, E-mail: sjfjs1226@163.com.

the samples were scanned 32 times. Each sample was loaded three times to obtain the average of three scanning results as the original spectrum of the sample, and the wavelength variables of the original spectrum were 1 845.

Dung beetle optimizer

The principle of DBO is based on the living habits of dung beetles. Dung beetles feed on animal dung. They roll the dung into a ball and move it quickly and effectively to prevent it from being robbed by other dung beetles. Dung beetles use celestial clues (such as the sun, the moon and polarized light) to navigate and make dung balls roll in a straight line. If encountering obstacles, dung beetles will reposition themselves by "dancing" to find new paths. DBO searches for an optimal solution by simulating these behaviors. Specifically, dung beetles mainly include four roles in algorithm design, namely, dung-rolling beetles, reproductive dung beetles, small dung beetles and stealing dung beetles. The corresponding behaviors of dung beetles in these four roles are rolling ball, reproducing, foraging, and stealing and dancing. The ball-rolling behavior simulates dung beetles' process of rolling dung balls; the reproductive behavior simulates dung beetles' process of choosing spawning sites; the foraging behavior simulates the process of finding food; the stealing behavior simulates the competition between dung beetles; and the dancing behavior simulates the repositioning of dung beetles when they encounter obstacles. At present, DBO algorithm is mainly used in path planning, parameter optimization and other issues. In this study, it was applied to solve the NIRS wavelength selection problem, and then, corresponding modeling performance was analyzed.

In addition, the flexibility and diversity of DBO algorithm make it have obvious advantages in dealing with complex optimization problems. In the problem of NIRS wavelength selection, characteristic wavelength selection directly affects the prediction accuracy and generalization ability of the model. DBO can efficiently locate the optimal or near-optimal wavelength combination in a high-dimensional search space by simulating various behavior patterns of dung beetles, which avoids falling into a local optimal solution. Specifically, in the stage of rolling behavior, the algorithm explores the whole search space and finds potential high-quality wavelength sets; in the stage of reproductive behavior, excellent wavelength combinations are expanded and retained; and the algorithm can quickly get rid of obstacles and adjust the search direction by dancing, which further enhances the global search ability. The foraging behavior and stealing behavior of DBO further strengthen the exploration and development ability of the algorithm. The foraging behavior simulates searching in local area to capture the best wavelength combination more accurately, while stealing behavior introduces competition mechanism, which helps to improve the diversity of solutions and the robustness of the algorithm through information sharing and competition between different solutions. Finally, the DBO algorithm can fine-tune the local search while maintaining the global search ability, thus achieving better model performance in the wavelength selection process.

Algorithm evaluation parameters

The performance of the calibration model was evaluated by using adjusted coefficient of determination (R_c^2), coefficient of determination for prediction (R_p^2), root mean squared error of calibration (RMSEC), root mean squared error of prediction (RMSEP) and residual predictive deviation (RPD). Larger determination coefficient and RPD and smaller root mean square error indicate better model performance.

Results and Analysis

Data analysis and processing

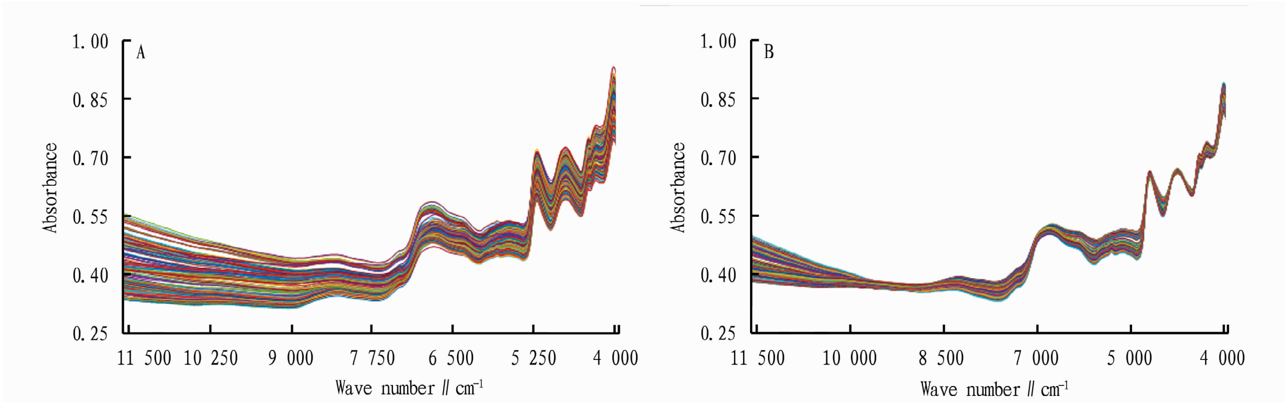
In order to improve spectral characteristics, correct baseline drift and spectral scattering, eliminate random noise generated by instruments and effectively improve the signal-to-noise ratio of spectral data, Savitzky-Golay (SG) smoothing and multivariate scattering correction were used for spectral preprocessing. The original spectra and preprocessed spectra of corn stalks are shown in Fig. 1. As can be seen from the figure, baseline drift of the spectra can be effectively eliminated after spectral preprocessing.

In the process of spectral scanning, spectral data may be affected by many factors such as instrument state, sample characteristics and measurement environment, which may lead to the condition that the data contains not only useful spectral information, but also a small number of abnormal samples. The existence of these abnormal samples will significantly weaken the prediction performance of NIRS correction models, and it is difficult to eliminate them by conventional preprocessing methods. After preprocessing, abnormal samples were removed from the corn stalk data set by using the MCCV residual mean variance distribution diagram method, and the results are shown in Fig. 2.

Four samples with high mean and high variance far from the origin position in the diagram were rejected as abnormal samples. The numbers of abnormal samples were 13, 14, 54 and 173. At this time, the number of experimental samples in the soil data set was 180. The KS method was adopted to divide the sample data into a calibration set and a verification set according to the ratio of 3 : 1. There were 135 samples in the calibration set and 45 samples in the verification set. The average hemicellulose contents in the calibration set and verification set were 26.28% and 20.04%, respectively. The hemicellulose contents in the calibration set was in the range of 9.48% – 38.59%, and the hemicellulose contents in the verification set were in the range of 9.51% – 38.54%. The standard deviations of the calibration set and verification set were 9.43% and 10.04%, respectively. The coefficients of variation of calibration set and verification set were 35.89% and 50.51%, respectively. Then, GA, BPSO and DBO were used to select characteristic wavelengths respectively, and PLS regression models were established by using the optimized characteristic wavelengths, respectively.

Algorithm performance evaluation

Taking the 10-fold RMSECV of PLS regression model as the objective function, the population size and evolutionary algebra were set to 100. Because of the randomness of intelligent optimization algorithms, the result of a single run is not enough to explain the modeling effect. Therefore, each intelligent optimization algorithm was run for 50 rounds to select characteristic wavelengths, and corresponding regression model was established. Comparing the modeling performance of different intelligent optimization algorithms under the same parameters, the best modeling results in 50 rounds are shown in Table 1.



A. Original spectra; B. preprocessed spectra.

Fig. 1 NIRS of corn stalk hemicellulose

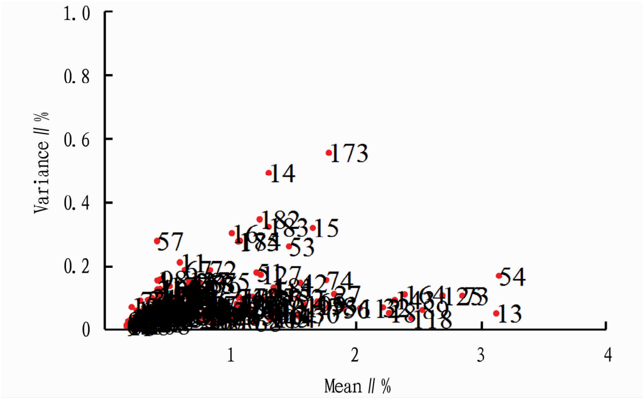


Fig. 2 Residual mean variance distribution diagram

Table 1 Performance evaluation indexes of algorithm PLS regression models

Indicators	Full-PLS	GA-PLS	BPSO-PLS	DBO-PLS
Wavelengths	1 845	952	933	924
R_c^2	0.994 6	0.995 0	0.994 7	0.995 0
R_p^2	0.992 6	0.994 1	0.994 3	0.994 7
RMSEC // %	0.686 5	0.668 5	0.688 0	0.672 7
RMSEP // %	0.836 1	0.754 4	0.735 2	0.714 7
RPD	11.989 5	13.144 7	13.486 8	13.873 7

It can be seen from Table 1 that the prediction performance of the models could be effectively improved by selecting characteristic wavelengths, and the *RPD* value was also increased from 11 of full spectra to 13. Compared with other two algorithms and full spectra, the DBO algorithm showed the largest *RPD* value and the best modeling effect, indicating that DBO is a feasible and efficient method for selecting characteristic wavelengths. Furthermore, it is not difficult to see from the table that the DBO algorithm selected the least wavelength points while obtaining high modeling performance, indicating that the algorithm has good global convergence performance.

Conclusions

From the evaluation indexes of PLS regression models, it could be seen that DBO algorithm had excellent model prediction performance. The R^2 and RMSEP of its verification set were 0.994 7 and 0.714 7%, respectively, and the detection accuracy error was very low, which could meet the actual detection requirements of hemicellulose. Compared with traditional intelligent optimization algorithms, DBO has obvious advantages in search efficiency and modeling accuracy. This intelligent optimization algorithm shows great potential in solving complex spectral data feature selection problems, and provides an effective tool for further improving the accuracy of NIRS model. The RPD value obtained was better than full spectra and other two traditional intelligent optimization algorithms while reducing modeling complexity. It strongly proves that DBO is an effective NIRS wavelength variable selection algorithm, which can be applied to the rapid detection of hemicellulose content in corn stalks by NIRS. This study provides a new way for the detection of hemicellulose content in corn stalks.

References

[1] WANG C, LIU JM, WANG CQ. Research progress on alkaline pretreatment of corn stalk[J]. Journal of Heilongjiang Bayi Agricultural University, 2022, 34(2): 23-31. (in Chinese).

[2] WANG C, WANG CQ, XU LL, *et al.* Research on rapid detection of corn moisture content based on near-infrared spectroscopy[J]. Journal of Heilongjiang Bayi Agricultural University, 2022, 34(4): 93-99. (in Chinese).

[3] LIU JM, CHENG QS, ZHEN F, *et al.* Rapid determination of C/N ratio for anaerobic fermentation feedstocks using near infrared spectroscopy based on GSA[J]. Transactions of the Chinese Society of Agricultural Machinery, 2019, 50(11): 323-330. (in Chinese).

[4] XU YH, WANG N, LIU JM. Research on rapid determination of lignocellulosic contents in corn stover using near infrared spectroscopy based on spectral intervals selection[J]. Chinese Journal of Analytical Chemistry, 2022, 50(10): 1587-1596. (in Chinese).

(Continued on page 92)

(Table 4)

Reflection after class	The contents of this section are mainly common types of poultry and livestock, which are closely related to life and have strong application. The teacher pay attention to the expansion of cases and case analysis when teaching this section, so that students can firmly grasp the knowledge of this section and gain ideological and political understanding at the same time. However, due to students' narrow knowledge and biased understanding ability, teachers should explain based on students' characteristics.
Blackboard writing	Overview of livestock and poultry raw materials (1) Quality characteristics and cooking application of common livestock meat (2) Quality characteristics and cooking application of common poultry meat (3) Common livestock viscera and miscellaneous materials

Conclusions and prospects

At present, China's catering industry is booming. In order to meet people's requirements for high quality in the catering industry, the catering industry needs a large number of culinary professionals with craftsmanship spirit and high professional ethics to undertake the task of revitalizing and developing the industry. Therefore, it is very necessary to carry out ideological and political construction in the teaching process of culinary majors. As far as this curriculum is concerned, the ideological and political construction of the curriculum should be considered from many aspects in the subsequent teaching process. Teachers should constantly improve the ability of integrating ideological and political education into cooking teaching, attach importance to the combination of theory and practice of ideological and political elements, stimulate students' enthusiasm, and better realize the goal of talent training.

References

[1] The Ministry of Education issued Guiding Outline of Curriculum Ideological and Political Construction in Colleges and Universities to comprehensively promote the ideological and political construction in colleges and universities[J]. New Education, 2020, 464(19): 32. (in Chinese).

[2] YUAN WJ. Study on the collaborative education of "curriculum ideological and political education" and ideological and political curriculum in colleges and universities[D]. Jishou: Jishou University, 2020. (in Chinese).

[3] CUI ZK, BI JC, ZHANG LW, *et al.* Discussion on the teaching of "culinary raw materials" under the background of application-oriented undergraduate education[J]. Agricultural products processing, 2018, 449(3): 72–74. (in Chinese).

[4] PENG H, YANG WN, CHEN WH, *et al.* Design and implementation of blended teaching of culinary raw materials course based on superstar learning platform[J]. Journal of Higher Education, 2021, 7(35): 5. (in Chinese).

[5] PENG H, QI XW, CHEN WH, *et al.* Study on the design and application of micro-course group of culinary raw materials[J]. The Processing of Agricultural Products, 2021(23): 4. (in Chinese).

[6] YANG ZS, QISI KANA, GONG W, *et al.* Teaching discussion on the integration of ideological and political education into "culinary raw materials"[J]. Yunnan Agriculture, 2022(9): 4. (in Chinese).

[7] YANG ZS, QISI KANA, SHI ZF, *et al.* On case teaching design of integrating moral education in "cooking ingredients": Taking "cereal raw materials" as an example[J]. Journal of Sichuan Tourismuniversity, 2023, 164(1): 85–89. (in Chinese).

[8] Overall framework of "Development of Chinese Students' Core Literacy" was officially released[J]. Shanghai Education, 2016, 992(27): 8–9. (in Chinese).

Editor: Yingzhi GUANG Proofreader: Xinxiu ZHU

(Continued from page 85)

[5] LIU DY, SUN XR, LIU CL, *et al.* Detection of ash content of wheat flour based on raman spectroscopy combined with simulated annealing[J]. Journal of the Chinese Cereals and Oils Association, 2019, 34(5): 128–133. (in Chinese).

[6] CHENG JH, CHEN ZG. Wavelength selection method for near infrared spectroscopy based on iteratively retains informative variables and successive projections algorithm[J]. Chinese Journal of Analytical Chemistry, 2021, 49(8): 1402–1409. (in Chinese).

[7] YAO Y, SHEN XM, QIU Q, *et al.* Predicting the biochemical methane potential of organic waste with near-infrared reflectance spectroscopy based on GA-SVM[J]. Spectroscopy and Spectral Analysis, 2020, 40(6): 1857–1861. (in Chinese).

[8] WANG ZQ, LIU JM, ZENG CH, *et al.* Rapid detection of protein content in rice based on Raman and near-infrared spectroscopy fusion strategy combined with characteristic wavelength selection[J]. Infrared Physics & Technology, 2023, 129: 104563.

Editor: Yingzhi GUANG Proofreader: Xinxiu ZHU