CHINESE JOURNAL OF ANALYTICAL CHEMISTRY
Volume 45, Issue 8, August 2017
Online English edition of the Chinese language journal

Available online at www.sciencedirect.com
ScienceDirect

Cite this article as: Chin J Anal Chem, 2017, 45(8), 1137–1142.

RESEARCH PAPER

Qualitative Prediction of Yeast Growth Process Based on Near Infrared Spectroscopy

WANG Wei, JIANG Hui*, LIU Guo-Hai, MEI Cong-Li, JI Yi

School of Electrical and Information Engineering, Jiangsu University, Zhenjiang 212013, China

Abstract: To improve the yield of industrial fermentation, this study presented a method based on near infrared spectroscopy to predict the growth process of yeast. The spectral data were measured from fermentation sample by Fourier-transform near-infrared (FT-NIR) spectrometer in the process of yeast culture. Each spectrum was acquired over the range of 10000 to 4000 cm^{-1}. Meanwhile, the optical density (OD) values of fermentation sample were determined with photoelectric turbidity method. A method on the basis of competitive adaptive reweighted sampling (CARS) was used to select characteristic wavelength variables of NIR data, and then extreme learning machine (ELM) algorithm was employed to develop the categorization model about the four growth phases of yeast. The experimental results showed that only 30 characteristic wavelength variables of NIR data were selected by CRAS algorithms, and prediction accuracy of the training set and testing set of the CARS-ELM model was 98.68% and 97.37%, respectively. This study showed that near infrared spectral analysis technique was feasible to predict the growth process of yeast.

Key Words: Near-infrared spectroscopy; Growth of yeast; Competitive adaptive reweighted sampling; Extreme learning machine

1 Introduction

With the tensions of global fossil energy sources, the development of biomass energy gained significant attention. The determination of yeast growth process played an important guiding role in industrial fermentation[1–3]. At present, common methods such as cell count technique, plate colony-counting and weighing have been reported to predict the growth process of yeast cells. Although the detection process was intuitive and fast by using these methods, the detection results were greatly affected by operator factors, and the stability and uniformity of these methods were difficult to guarantee[4].

Near infrared (NIR) light is an electromagnetic wave between ultraviolet visible light and mid-infrared light, and the spectral information are from the absorption of frequency doubling and combination band of organic compounds containing hydrogen groups. Owing to the obvious difference of wavelength and intensity of near-infrared absorption of different groups or the same groups in different chemical environment, NIR spectroscopy uses these differences to indirectly measure physical and chemical parameters of organic compounds[5–7]. In yeast culture, the organic macromolecules in the substrate contain a large number of hydrogen-containing groups[8]. In recent years, some researchers employed NIR spectroscopy to detect the biomass, concentration of substrate and product in the process of microbial fermentation, and satisfactory results were achieved[9–11]. However, few studies were reported on the dynamic tracking of microbial growth using NIR spectroscopy. Therefore, a rapid description method of yeast growth process based on NIR spectroscopy analysis technique was proposed in this study. In addition, in recent years, a number of researches showed that NIR spectral information has complicated backgrounds with peak overlapping and weak signal[12]. Generally, NIR spectroscopy has hundreds of

Received 18 February 2017; accepted 7 June 2017
*Corresponding author. Email: h.v.jiang@ujs.edu.cn
This work was supported by the the Natural Science Foundation of Jiangsu Province, China (No. BK20140538), the China Postdoctoral Science Foundation (No. 2016M600381), the College Science Foundation of Jiangsu Province, China (No. 16KJB210003), and the Graduate Practical Innovation Foundation of Jiangsu Province, China (No.SJZZ16_0193).

DOI: 10.1016/S1872-2040(17)61030-2

variables and there exist some uninformative variables, redundant variables and serious collinearity among the wavelength variables[13]. Model calibration using complete spectral data would not only reduce the modeling speed, but also affect the accuracy and robustness of the model. Therefore, it is necessary to select spectral characteristic wavelength variables using an appropriate wavelength variable selection method prior to model calibration[14]. Given this, the competitive adaptive reweighted sampling (CARS) algorithm was adopted to select the characteristic wavelength variables of the preprocessed spectra, and then the qualitative description model, a method for identifying the four stages of yeast growth, was established in this study using extreme learning machine (ELM) to realize the high precision detection of yeast growth state by NIR spectrum analysis technique.

2 Experimental

2.1 Cultivation of yeast and acquisition of data

2.1.1 Expanding culture of yeast

Yeast strains (1 mL) were purchased from Shanghai Ruichu Biotech Co., Ltd. The yeast suspension was inoculated into a sterile malt medium to expand culture, and yeasts strains with good growth status were chosen as parent strains of the next extended culture until 40-mL yeast suspension was obtained.

2.1.2 Subpackage culture of yeast

After expanding culture of yeast, three 250-mL volumetric flasks were marked as I, II and III, and then 125 mL malt medium and 0.5 mL yeast suspension were loaded into each of the volumetric flasks respectively. These volumetric flasks were continuously cultured for 72 h in a constant temperature shock incubator, the temperature and rotation rate of the incubator was set to 28 °C and 110 rpm. According to the above scheme, six batches of yeast culture experiments were carried out in this study.

2.1.3 Sample collection

Sampling was carried out at 19 different time points with 4 hour interval during the yeast culture (0, 4, 8, 12, ⋯ , 72 h). In addition, to avoid contamination of sterile malt medium caused by multiple sampling, the 19 sampling time points were divided to three parts. The sampling was carried out in the volumetric flask I during the first 24 h, the sampling was implemented in the volumetric flask II during the next 24 h, and sampling of the last 24 h was carried out in flask III. Thus, 19 samples were obtained in each batch of yeast culture, and a total of 114 samples were obtained through six batches of yeast culture.

2.2 Acquisition of FT-NIR spectra

During the spectra collection, the temperature was kept around 25 °C and the humidity was kept at a steady level in the laboratory. The NIR spectra were collected in the transmittance mode using the Antaris II near-infrared spectrophotometer (Thermo Scientific Co., USA). Each spectrum was the average of 32 scanning spectra. The range of spectra was from 10000 cm^{-1} to 4000 cm^{-1}, and the original data were measured with a resolution of 8 cm^{-1}. To obtain more accurate spectral data, three different positions of each sample were collected, and then the mean of these three spectral data was calculated as raw spectral data of sample.

2.3 Measurement of OD value

The wavelength of the UV-2204PC spectrophotometer was firstly set at 600 nm and the light transmittance was adjusted to 100%. 3.5 mL of sterile malt extract medium was loaded into 1-cm cuvette as a control group. Then, yeast culture media were filtered by 0.45-μm microporous membrane, and the filtrate was added into cuvette as a testing group, which was used to measure the OD value by the spectrophotometer. Each sample was measured three times, and the mean of the three measured OD value was recorded as the OD value of the sample. During measurement, the bacterial suspension should be appropriately diluted to ensure the measured OD value between 0.1 and 0.65[14,15].

2.4 Analysis of data

2.4.1 Competitive adaptive reweighted sampling algorithm

The spectral data need to be selected to eliminate the redundancy and collinearity information between spectral variables. The competitive adaptive reweighted sampling (CARS) algorithm was a wavelength selection method, employing the effective principle survival of the fittest on the basis of Darwin's Evolution Theory[16]. The algorithm applied Monte Carlo approach for model sampling, 80% of the samples were randomly selected to establish the partial least square (PLS) model, and the wavelength point with a large absolute value of the regression coefficient was retained. Meanwhile, the wavelength points with small weight were removed. After repeated sampling, the subset with the lowest root mean square error of cross-validation (RMSECV) value was selected as the characteristic wavelength variables[17,18].

2.4.2 Extreme learning machine algorithm

The extreme learning machine (ELM) algorithm, proposed by Huang *et al*[19], was a learning algorithm for feed forward