NIRS Feature Extraction Based on Deep Auto-encoder Neural Network

Ting Liu, Zhongren Li, Chunxia Yu, Yuhua Qin

PII: S1350-4495(17)30226-8
DOI: http://dx.doi.org/10.1016/j.infrared.2017.07.015
Reference: INFPHY 2338

To appear in: *Infrared Physics & Technology*

Received Date: 26 April 2017
Revised Date: 12 July 2017
Accepted Date: 12 July 2017

Please cite this article as: T. Liu, Z. Li, C. Yu, Y. Qin, NIRS Feature Extraction Based on Deep Auto-encoder Neural Network, *Infrared Physics & Technology* (2017), doi: http://dx.doi.org/10.1016/j.infrared.2017.07.015

NIRS Feature Extraction Based on Deep Auto-encoder Neural Network

Ting Liu[a], Zhongren Li[b*], Chunxia Yu[b], and Yuhua Qin[c]

[a]College of Information Science and Engineering, China Ocean University, No.238 Songling Road, Qingdao 266100, China

[b]R&D Center, China Tobacco Yunnan Industrial Co., Ltd, No.367 Hongjin Road, Kunming 650231, China. Email: aidewen@hongta.com

[c]College of Information Science and Technology, Qingdao University of Science and Technology, No.99 Songling Road, Qingdao 266061, China

Abstract

As a secondary analysis method, Near Infrared Spectroscopy (NIRS) needs an effective feature extraction method to improve the model performance. Deep auto-encoder (DAE) can build up an adaptive multilayer encoder network to transform the high-dimensional data in to a low-dimensional code with both linear and nonlinear feature combinations. To evaluate its capability, we experimented on the spectra data obtained from different categories of cigarette with the method of DAE, and compared with the principal component analysis (PCA). The results showed that the DAE can extract more nonlinear features to characterize cigarette quality. In addition, the DAE also got the linear distribution of cigarette quality by its nonlinear transformation of features. Finally, we employed k-Nearest Neighbor (kNN) to classify different categories of cigarette with the features extracted by the linear transformation methods as PCA and wavelet transform-principal component analysis (WT-PCA), and the nonlinear transformation methods as DAE and isometric mapping (ISOMAP). The results showed that the pattern recognition mode built on features extracted by DAE was provided with more validity.

Keywords: Feature extraction; Near infrared spectroscopy; Deep auto encoder; Cigarette pattern recognition

1 Introduction

Due to advantages of near-infrared spectroscopy(NIRS), it has become a widely used technique in many industries. However, as a secondary analytical method, NIRS needs to build a mathematical model to interpret the relation between spectra data and reference value by conventional analytical method. Multivariate regression methods are needed to make the model from sample data spectra & reference data[1]. With regard to the sampling cost in industry applications, the number of samples(n) is much smaller than the number of features(d) in most situations. The modeling methods cannot get good prediction results without well conditioned sample matrices like $n >> d$[2]. In other words, an effective method for the dimension reduction is necessary to NIRS applications. Meanwhile, the extraction of spectrum features is conductive to remove noise and redundant feature in the spectra. Principal component analysis (PCA) [3-5] is a simple and widely used method, which converts a great number of relevant variables