Analytica Chimica Acta 1010 (2018) 20–28

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

Feature expansion by a continuous restricted Boltzmann machine for near-infrared spectrometric calibration

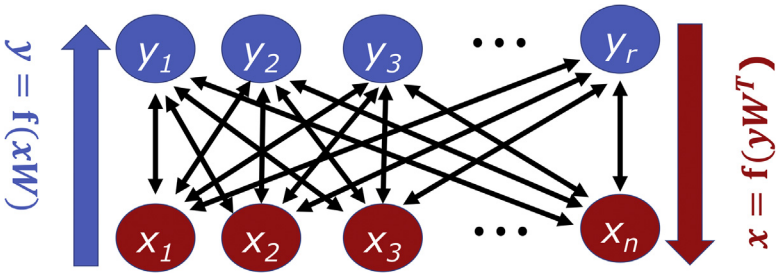
Peter de Boves Harrington

Ohio University Center for Intelligent Chemical Instrumentation, Department of Chemistry & Biochemistry, Clippinger Laboratories, Athens, OH, 45701-2979, USA

HIGHLIGHTS

- The Copiosity Principle is demonstrated.
- Calibrations by nonlinear mapping and feature expansion.
- The RBM algorithm was modified to work with NIR spectra.
- The results were statistically validated using Bootstrapped Latin Partitions.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 29 September 2017
Received in revised form
22 November 2017
Accepted 3 January 2018
Available online 31 January 2018

Keywords:
Restricted Boltzmann machines
Statistical learning
Contrastive divergence
Chemometrics
Calibration
Near infrared spectroscopy
Data expansion

ABSTRACT

A modified algorithm for training a restricted Boltzmann machine (RBM) has been devised and demonstrated for improving the results for partial least squares (PLS) calibration of wheat and meat by near-infrared (NIR) spectroscopy. In all cases, the PLS calibrations improved by using the abstract features generated from the RBM so long as the nonlinear mapping increased the dimensionality. The evaluations were validated using bootstrapped Latin partitions (BLPs) with 5 bootstraps and 3-Latin partitions which proved useful because of the statistical learning and random initial conditions of the RBM networks. By using a noise decay parameter, initial large amounts of noise could be used and the benefits of simulated annealing achieved as the noise level is slowly decreased. This paper demonstrates for the first time that using abstract features and enlarging the spectral data can improve the calibration results and exemplifies the Copiosity Principle. Two NIR reference datasets were evaluated. The first set of wheat spectra was calibrated for protein concentration and the second set of meat spectra was calibrated for moisture, fat, and protein concentration. The RBM feature extraction improved the linearity of the models and reduced embedded noise. The RBM also can help eliminate some difficult spectral preprocessing stages such as variable alignment and feature selection. RBMs benefit from derivative preprocessing of the NIR spectra or other preprocessing that enhances the differences among the spectra.

1. Introduction

The recent success of deep learning has renewed interest in artificial neural networks with many successful applications [1]. Deep learning refers to feed forward neural networks with many layers. Training multiple layers of a neural network is slow and difficult by backpropagation. One innovation to resolve slow training is to use layers of restricted Boltzmann machines (RBMs) that were invented by Hinton [2]. The advantage of these networks

E-mail addresses: peter.harrington@ohio.edu, harring10@msn.com.

https://doi.org/10.1016/j.aca.2018.01.026
0003-2670/

is that they are unsupervised and can select features in the data so that each layer can be trained independently. Hence, they are referred to as stackable. They also do not rely on backpropagation for machine learning, but instead, an algorithm devised by Hinton referred to as contrastive divergence [3]. This approach relies on stochastic learning which helps avoid local minima.

Deep learning has already proven useful for handwriting, voice, and image recognition in many applications that are used today [1]. It would make sense to exploit the feature selection aspects of the RBM as a preprocessing method for analytical measurements. Deep learning is just now making some advances in chemistry. It has been applied [4] and reviewed [5] with respect to drug discovery. Stackable RBMs were applied to a deep belief network (DBN) to improving biological activity prediction for a quantitative structure-activity relationship model [6]. Three-dimensional desorption electrospray ionization (DESI) mass spectrometry (MS) imaging of metabolic heterogeneity of cancerous tumors [4] is an application of deep learning to bioanalytical chemistry.

RBMs were designed for binary input data such as black and white images [7]. However, they have been used on data that were continuous in the interval of [0, 1] [8]. Chen and Murray developed and implemented a continuous RBM [9] which can accommodate any range of input data. One feature of their learning algorithm which is the scaling of the inputs for the sigmoid function was adapted for the algorithm presented here. Their algorithm was evaluated for spectral preprocessing, however, it failed to perform adequately.

The new algorithm presented in this paper resembles the Hinton algorithm with training by contrastive divergence using a single step. The scaling factors to control the sigmoid slopes were added which was devised by Chen and Murray [9] with a small modification to the learning step. One other departure from conventional RBM training is that random noise is added only during the forward propagation of the algorithm. The rationale for this modification is that the subspace defined by the spectra will be maintained. In other words, the weights $\boldsymbol{W}$ will be constructed from only linear combinations of the spectra.

The idea of stackable layers is intriguing with each layer capable of modeling higher order relationships in the data. It is worthwhile studying single hidden layer RBMs for their capability as feature selectors. If a layer of RBMs fails to transmit information, it can never be regained from a successive layer. RBMs can be combined with traditional chemometric classification and calibration algorithms. This paper will study RBMs as spectral conditioners for calibration and apply them to near-infrared (NIR) spectral measurements of wheat and meat. An enhanced algorithm will be presented that allows continuous data in the range between 0 and 1 to be modeled.

Although reconstruction of the spectra is interesting especially when measurements are missing from the data, it is not the subject of this study. This study presents the use of RBMs to expand the dimensionality of chemical data such as spectra to improve calibration. This approach is known as the Copiosity Principle [10]. The key idea is that by using a nonlinear mapping to a higher dimensional space, the calibration methods have more degrees of freedom to find the correlations in the dataset.

2. Theory

The data set $\boldsymbol{X}$ is represented as a matrix with m rows of spectra and n measurements (i.e., spectral resolution elements). The role of the RBM is to produce a nonlinear mapping from the data $\boldsymbol{X}$ to $\boldsymbol{Y}$ so that $\boldsymbol{Y}$ has m rows of r nonlinear features. In the computer science literature, the spectral data $\boldsymbol{X}$ are referred to as visible units and the features $\boldsymbol{Y}$ as hidden units, although in this case they are not hidden but will be output and used by PLS. The mapping occurs in two directions from $\boldsymbol{X}$ to $\boldsymbol{Y}$ and from $\boldsymbol{Y}$ to $\boldsymbol{X}$ using an $n \times r$ matrix of weights $\boldsymbol{W}$ that represents the connection strengths. The RBMs also include a nonlinear function that usually serves to restrict the range of the products with the weight matrix $\boldsymbol{W}$ and they allow the weights to grow to large values when the nonlinear function is saturated (i.e., have values at the limits of 0 and 1).

The network is a bipartite bidirectional graph, which means the weight matrix $\boldsymbol{W}$ is used in both the forward and reverse propagations. The transpose of the weight matrix is used in the reverse propagation so during training the weight matrix will converge to an approximately orthogonal form so that the transpose will be equal to the inverse. Fig. 1 is a schematic of a simple RBM. The nonlinear function can be any of several functions but for this work, the sigmoid function is used. The sigmoid is a very popular logistic function and is important with respect to fuzzy entropy calculations [11,12]. The sigmoid logistic function is given below for both forward (1) and backward (2) propagations.

$$\boldsymbol{y_i} = \left(1 + e^{-\boldsymbol{a_y} \odot (\boldsymbol{x_i W} + \boldsymbol{b_y} + \boldsymbol{N}(0,\sigma))}\right)^{-1} \tag{1}$$

$$\widehat{\boldsymbol{x}}_{\boldsymbol{i}} = \left(1 + e^{-\boldsymbol{a_x} \odot (\boldsymbol{y_i W}^T + \boldsymbol{b_x})}\right)^{-1} \tag{2}$$

The term $\boldsymbol{a_x}$ and $\boldsymbol{a_y}$ are scale parameter row vectors that control the slopes of the sigmoid functions in the forward and backward directions. They are multiplied element-wise $\odot$ with respect to the other terms in the exponent. Each spectrum $\boldsymbol{x_i}$ is multiplied by the weight matrix $\boldsymbol{W}$. The products are corrected with a bias $\boldsymbol{b_y}$ that is a row vector with r elements. Random deviates from the normal distribution $\boldsymbol{N}(0, \sigma)$ with mean 0 and standard deviation σ are added to the r bias corrected products. Then these values are transformed with the nonlinear logistic function to produce the outputs of the layer $\boldsymbol{y_i}$ that is row i of $\boldsymbol{Y}$.

Equation (2) has the same functional form except that it works in the reverse direction by use of the transpose of the weight matrix. The $\boldsymbol{a_x}$ and $\boldsymbol{b_x}$ are the respective scale and bias row vectors for the n spectral inputs just like $\boldsymbol{a_y}$ and $\boldsymbol{b_y}$ are the scale and bias corrections for the r weight outputs. The reverse direction will use $\boldsymbol{y_i}$ to reconstruct spectrum $\widehat{\boldsymbol{x}}_{\boldsymbol{i}}$. Another modification from literature is the random deviates are not used in the reverse direction. Later, $\widehat{\boldsymbol{x}}_{\boldsymbol{i}}$ will be used as the input for equation (1) to yield $\widehat{\boldsymbol{y}}_{\boldsymbol{i}}$.

Because the asymptotes of the sigmoid function are 0 and 1, it is important that the input data is properly scaled to fall within this range. The scaling is for the entire the calibration set is given below.

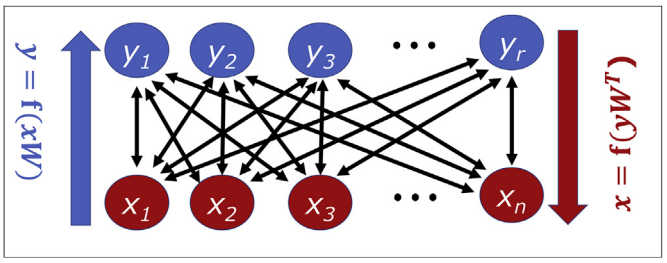


Fig. 1. Schematic of an RBM bipartite bidirectional network. The n input units are the red circles (i.e., spectra) and the r hidden units (i.e., features) are the blue circles. The connection strengths are stored in the weight matrix $\boldsymbol{W}$, which is used to propagate signals in the forward (up) and backward (down) directions. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)