Chemometrics and Intelligent Laboratory Systems 158 (2016) 21–30

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

A generalized fuzzy linguistic model for predicting component concentrations in an optical gas sensing system

Yanxia Wang[a,c], Hui Cao[a], Xingyu Yan[a], Yan Zhou[b,*], Xueqin Liu[c], Seán McLoone[c]

[a]State Key Laboratory of Electrical Insulation and Power Equipment, School of Electrical Engineering, Xi'an Jiaotong University, Xi'an, 710049, China
[b]School of Energy and Power Engineering, Xi'an Jiaotong University, Xi'an, 710049, China
[c]School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Belfast, BT9 5AH, UK

ARTICLE INFO

Article history:
Received 22 April 2016
Received in revised form 26 July 2016
Accepted 27 July 2016
Available online 7 August 2016

Keywords:
Optical gas sensing systems
Generalized fuzzy linguistic model
Parameter optimization

ABSTRACT

Motivated by environmental protection concerns, monitoring the flue gas of thermal power plant is now often mandatory due to the need to ensure that emission levels stay within safe limits. Optical based gas sensing systems are increasingly employed for this purpose, with regression techniques used to relate gas optical absorption spectra to the concentrations of specific gas components of interest (NO_x, SO_2 etc.). Accurately predicting gas concentrations from absorption spectra remains a challenging problem due to the presence of nonlinearities in the relationships and the high-dimensional and correlated nature of the spectral data. This article proposes a generalized fuzzy linguistic model (GFLM) to address this challenge. The GFLM is made up of a series of "If-Then" fuzzy rules. The absorption spectra are input variables in the rule antecedent. The rule consequent is a general nonlinear polynomial function of the absorption spectra. Model parameters are estimated using least squares and gradient descent optimization algorithms. The performance of GFLM is compared with other traditional prediction models, such as partial least squares, support vector machines, multilayer perceptron neural networks and radial basis function networks, for two real flue gas spectral datasets: one from a coal-fired power plant and one from a gas-fired power plant. The experimental results show that the generalized fuzzy linguistic model has good predictive ability, and is competitive with alternative approaches, while having the added advantage of providing an interpretable model.

1. Introduction

In order to demonstrate compliance with regularity requirements on thermal power plant emissions, monitoring the concentrations of pollutants such as nitrogen oxides, sulfur oxides and oxycarbides in flue gas emissions is now mandatory in many countries [1,2]. Gas sensing methods are diverse due to the chemical and physical effects that can reflect gas characteristics [3]. One common sensing principle is the electrochemical variations that occur between target gases and different sensor materials, such as metal oxide semiconductors, polymers, and carbon nanotubes [4,5]. In recent decades, optical spectroscopy based methods have become increasingly popular for gas sensing [6–10], due to their high sensitivity, selectivity and stability. These methods measure the chemical composition dependent absorption of light that occurs at different wavelengths when light passes through the flue gas [11]. By analyzing the measured absorption spectra, the concentration of specific components of the gas can be predicted by regression models [12].

Many regression methods for spectral data have been reported [13]. As a well-known multivariate regression algorithm, the classical partial least squares (PLS) can only establish linear relationships between absorption spectra and component concentrations [14–16]. In experiments, however, there are many conditions that can lead to nonlinearity such as instrument variation and analyte characteristics [17]. Nonlinear modeling methods such as multilayer perceptron (MLP) neural networks [18], radial basis function (RBF) networks [19], and support vector machines (SVM) [20,21] can be used to learn the nonlinear relationships. However, these methods typically require substantial computational effort to train, and by virtue of their black-box structure, cannot provide understandable heuristic knowledge [22].

Linguistic models are built up by fuzzy rules that express human-readable descriptions in a format suitable for regression analysis [23,24]. A fuzzy rule is a logical linguistic "If-Then" statement [25], where the "If" expression is referred to as the antecedent and the

* Corresponding author.
E-mail address: yan.zhou@mail.xjtu.edu.cn (Y. Zhou).

http://dx.doi.org/10.1016/j.chemolab.2016.07.012
0169-7439/© 2016 Elsevier B.V. All rights reserved.

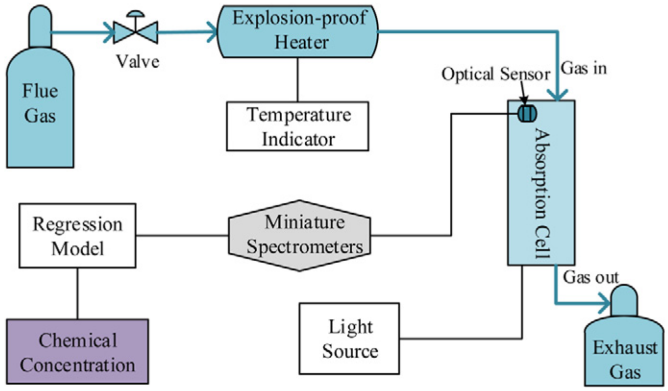

Fig. 1. Optical gas sensing system.

"Then" expression as the consequent. The antecedent expresses input conditions in terms of fuzzy linguistic labels. Two forms of consequent are normally employed in fuzzy models; the first expresses the output directly as linguistic labels and is referred to as the Mamdani fuzzy rule [26], while the second defines the output as a linear function of the inputs and is called the Takagi–Sugeno formulation. The latter is preferred for modeling applications because it produces a crisp output without defuzzification, and yields reduced complexity regression models [27,28]. In our previous work, we discussed a technology with a series of Takagi–Sugeno fuzzy rules for quantitative analysis [29]. Considering the nonlinearity of spectral data, we proposed a quadratic polynomial equation as the rule consequent [30]. Nevertheless, the predefined form of the rule consequent employed may limit the approach's power to handle variation in nonlinear complexity for different chemical concentration estimation tasks.

In this article, a generalized fuzzy linguistic model (GFLM) suited to the optical gas sensing system modeling problem is presented. The model consists of a sequence of "If-Then" fuzzy rules. In the rule antecedent, the input variables are absorption spectra. The rule consequent is a general nonlinear polynomial function expressed as a function of the absorption spectra. Least squares and gradient descent are both adopted to optimize the model. To demonstrate the performance of GFLM, it is compared with PLS, SVM, MLP and RBF models for flue gas spectral datasets from coal-fired and gas-fired power plants.

The reminder of the paper is organized as follows. The optical gas sensing system and GFLM are described in Section 2. In Section 3, the experimental setup (datasets and procedure) is described in detail, while Section 4 presents and discusses the experimental results. Finally, Section 5 concludes the paper.

2. Gas sensing system with a generalized fuzzy linguistic model

2.1. Optical gas sensing system

The schematic diagram of an optical gas sensing system is shown in Fig. 1. Flue gas is drawn into an explosion-proof tubular heater and heated to a predefined temperature. It is then transferred to an absorption cell where lights from a known light source is shone through the gas onto miniature spectrometers which measure the absorption spectrum.

2.2. Generalized fuzzy linguistic model (GFLM)

GFLM is functionally equivalent to a series of logical "If-Then" fuzzy rules. The antecedent "If" presents the conditions, using fuzzy linguistic labels instead of crisp numbers. The consequent "Then" is a general nonlinear polynomial function expressed in terms of the input variables. The "If-Then" fuzzy rules thus assume the form:

If u_1 is A_1 and u_2 is A_2 and $\cdots$ and u_n is A_n

$$\text{Then } f_i = \sum_{j_1=1}^{n} k_{j_1}^{1} u_{j_1} + \sum_{j_1=1}^{n}\sum_{j_2=j_1}^{n} k_{j_1 j_2}^{2} u_{j_1} u_{j_2} + \cdots + \sum_{j_1=1}^{n}\sum_{j_2=j_1}^{n}\cdots\sum_{j_m=j_{m-1}}^{n} k_{j_1 j_2 \cdots j_m}^{m} u_{j_1} u_{j_2} \cdots u_{j_m} + b$$

where $\{u_1, u_2, \ldots u_n\}$ is the input vector, $\{A_1, A_2, \ldots A_n\}$ are the linguistic labels, $\{k_{j_1}^1, k_{j_1 j_2}^2, \ldots, k_{j_1 j_2 \cdots j_m}^m, b\}$ is the vector of consequent parameters, and m is the highest degree considered. When $m = 1$, this fuzzy rule reduces to the classical Takagi–Sugeno type rule. A given GFLM consists of a series of these rules, each one having the same highest polynomial degree, m.

To initialize the model, a clustering technique is first used to determine the initial locations of the linguistic labels. The number of clusters R can be predefined by users or automatically determined as part of the clustering process.

Gaussian functions are employed to generate the membership degree of each linguistic fuzzy set. The firing strength of the i-th rule is then calculated as:

$$\omega_i = \prod_{j=1}^{n} \mu(u_j) = \prod_{j=1}^{n} e^{-\frac{\left(u_j - c_{i,j}\right)^2}{2\sigma_{i,j}^2}} \tag{1}$$

where $\prod$ performs fuzzy AND, and $\{c_{i,j}, \sigma_{i,j}\}$ is the antecedent parameter set. The Gaussian function varies when these parameters change, thus exhibiting various firing strengths. The output is computed as:

$$y = \frac{\sum_{i=1}^{R} \omega_i f_i}{\sum_{i=1}^{R} \omega_i} = \sum_{i=1}^{R} \bar{\omega}_i f_i \tag{2}$$

where f_i is the output of i-th rule.

Thus, we have constructed the generalized fuzzy linguistic model. Next, a learning procedure needs to be developed. For simplicity, we assume that the parameters can be decomposed into a nonlinear set $S_N = \{c_{i,j}, \sigma_{i,j}\}$ and linear set $S_L = \{k_{j_1}^1, k_{j_1 j_2}^2, \ldots, k_{j_1 j_2 \cdots j_m}^m, b\}$. Now given values of the elements of S_N, we can determine estimates for S_L by solving:

$$XS_L = Y \tag{3}$$

where X is a regressor matrix whose elements are a function of S_N and the model inputs, i.e. $X = [x_{kj}] = g_j(S_N, u_1(k), u_2(k), \cdots, u_n(k))$. The linear least squares solutions to Eq. (3), which minimizes $\|XS_L - Y\|^2$, is given by:

$$\hat{S}_L = \left(X^T X\right)^{-1} X^T Y \tag{4}$$

While Eq. (4) is concise in notation, $X^T X$ can often be ill-conditioned or singular leading to numerical issues if computed directly; singular value decomposition (SVD) provides a stable approach to address this [31]. However, determining the solution in this manner is computationally expensive which can be an issue if X is large. Alternatively, S_L can be computed using the recursive least squares estimator [32,33], defined as:

$$\begin{cases} S_L(n+1) = S_L(n) + P_{n+1} x_{n+1} \left(y_{n+1} - x_{n+1}^T S_L(n)\right) \\ P_{n+1} = P_n - \frac{P_n x_{n+1} x_{n+1}^T P_n}{1 + x_{n+1}^T P_n x_{n+1}} \end{cases} \tag{5}$$