Available online at www.sciencedirect.com
ScienceDirect
Analytica Chimica Acta 595 (2007) 114–119
ANALYTICA CHIMICA ACTA
www.elsevier.com/locate/aca

Non-destructive method for determination of hydroxyl value of soybean polyol by LS-SVM using HATR/FT-IR

Marco F. Ferrão [a,*], Simone C. Godoy [b], Annelise E. Gerbase [b], Cesar Mello [c], João Carlos Furtado [d], Cesar L. Petzhold [b], Ronei Jesus Poppi [e]

[a] *Departament of Chemistry and Physics, Universidade de Santa Cruz do Sul, C.P. 188, CEP: 96815-900, Santa Cruz do Sul, RS, Brazil*
[b] *School of Chemistry, Universidade Federal do Rio Grande do Sul, C.P. 15003, CEP: 91501-970, Porto Alegre, RS, Brazil*
[c] *School of Chemistry, Universidade de Franca, C.P. 32, CEP: 14404-600, Franca, SP, Brazil*
[d] *Departament of Computer Science, Universidade de Santa Cruz do Sul, C.P. 188, CEP: 96815-900, Santa Cruz do Sul, RS, Brazil*
[e] *School of Chemistry, Universidade Estadual de Campinas, C.P. 6154, CEP: 13083-970, Campinas, SP, Brazil*

Received 3 October 2006; received in revised form 22 February 2007; accepted 26 February 2007
Available online 4 March 2007

Abstract

This paper presents the use of least-squares support vector machine (LS-SVM) for quantitative determination of hydroxyl value (OHV) of hydroxylated soybean oils by horizontal attenuated total reflection Fourier transform infrared (HATR/FT-IR) spectroscopy. A least-squares support vector machine (LS-SVM) calibration model for the prediction of hydroxyl value (OHV) was developed using the range 1805.1–649.9 cm^{-1}. Validation of the method was carried out by comparing the OHV of a series of hydroxylated soybean oil predicted by the LS-SVM model to the values obtained by the AOCS standard method. A correlation coefficient equal to 0.989 and RMSEP = 4.96 mg of KOH/g was obtained. This study demonstrates a better prediction ability of the LS-SVM technique to determine OHV in hydroxylated soybean oil samples by HATR/FT-IR spectra.

Keywords: Least-squares support vector machine; Hydroxyl value; Hydroxylated soybean oil; Horizontal attenuated total reflectance; Chemometrics

1. Introduction

The environmental and sustainability aspects of using oleochemical polyols are of great importance to the polyurethane industry considering the broad range of applications of these materials. Oleochemical polyols can be used in the production of VOC-free, two-component polyurethane coatings and floorings, adhesives, and thermoplastic polyurethanes [1–3].

For polyurethane preparation, it is important to know the final hydroxyl value (OHV) of the soybean polyol. The hydroxyl value (OHV) is usually determined by titration methods such as the American Oil Chemists' Society (AOCS) hydroxyl value (OHV) determination (AOCS Cd 13–60) used in this work [4]. The hydroxyl value is expressed in mg of KOH/g of oil. This method is reliable and reproducible if carried out under standardized conditions, but it is time-consuming, labor-intensive, reasonably sensitive, largely dependent on the skills of the analyst, uses large amounts of sample and reagents, and some of them (pyridine, acetic anhydride) are hazardous and difficult to dispose off.

Similar problems were also observed in other chemical analyses of fats and oils based on titration methods. Therefore, spectroscopic methods are being increasingly used to replace wet chemical procedures. Infrared spectroscopy is one that has found increasing use due to its low cost, shorter time of analysis, non-destructiveness, small quantities of sample, in addition to accuracy and reliability when associated with chemometric methods [5–7]. Moreover, FTIR coupled with horizontal attenuated total reflectance (HATR) accessory simplifies many of the sample handling problems commonly associated with infrared analysis and is readily amenable to routine quality control applications [8].

There are many papers in the literature devoted to measurements of different samples using accessories for attenuated total reflection (ATR) and many of them using chemometric procedures [9–19]. HATR is a well-known technique that produces

* Corresponding author. Tel.: +55 51 37177390; fax: +55 51 37177632.
E-mail address: ferrao@unisc.br (M.F. Ferrão).

0003-2670/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.aca.2007.02.066

good quality and highly reproducible spectra, if good contact can be established between the sample and the internal reflectance element (IRE) [16].

Neural networks as multilayer perceptrons and radial basis functions networks have been used in a wide range of fields, including control theory, signal processing and linear or non-linear modeling [20,21]. A promising methodology called support vector machines (SVM) [20–23] has been recently used for classification, non-linear function and density estimation leading to excellent results when compared to standard procedures employed in such problems. In chemometrics, recent works have used SVMs for the classification [24–28] and quantification [29–36] problems.

2. Least-squares support vector machines (LS-SVM)

A promising methodology called support vector machines (SVM) [21] has been introduced into the chemical community to perform non-linear classification, multivariate function estimation or non-linear regression. Zomer et al. [24] applied SVM for the discrimination of analytical chemical data applied from the pharmaceutical industry. Brudzewski et al. [25] proposed an application in the calibration of an electronic nose for milk recognition. Also, SVM has been used for polymer classification [26], to identify bovine spongiform encephalopathy in serum [27] and, by Pierna et al. [28], to compare the performance of different supervised discrimination methods for the classification of starches according to type of chemical modification.

Later, computational calculations of SVM were simplified by Suykens et al. with the implementation of a least squares version for support vector machines (LS-SVM) [37]. This requires solving a set of linear equations, instead of the quadratic programming used in classical SVM. Only a few applications have been published using LS-SVM in chemometrics, such as the paper of Thissen et al. [30] where the determination of monomer masses during a copolymerization reaction using Raman spectra and the quantification of ethanol, water and 2-propanol in ternary mixtures using NIR spectra were proposed. Also, Chauchard et al. [31] compared classical linear regression techniques and LS-SVM regression for the prediction of total acidity in fresh grapes and Cogdill et al. [32] estimated physical properties of wood from NIR spectroscopy using LS-SVM. Borin et al. [33] applied LS-SVM to determine common adulterants (starch, whey or sucrose) found in powdered milk samples, using near infrared spectroscopy with direct measurements by diffuse reflectance and Gaus et al. [34] using UV-resonance Raman spectroscopy and SVM to develop the method for identification of lactic acid bacteria from yogurt. Luan et al. [35] developed a quantitative structure activity relationship (QSAR) non linear model for polychlorinated dibenzofurans (PCDFs), dibenzodioxins (PCDDs), and biphenyls (PCBs) binding to the aryl hydrocarbon receptor (AhR) based on SVM. Shi and Liu [36] proposed the predict melt index (MI) of polypropylene using LS-SVM soft-sensor model of propylene polymerization process from other process variables.

In LS-SVM, like most linear regression models, a linear estimation is carried out between the regressors (x) and the dependent variable (y): $y = w^{\mathrm{T}}x + b$, where w is the regression coefficient. The regression is calculated by minimizing a cost function (C) containing a penalized regression error, as follows:

$$C = \frac{1}{2}w^{\mathrm{T}}w + \frac{1}{2}\gamma\sum_{i=1}^{N}e_i^2 \tag{1}$$

such that:

$$y_i = w^{\mathrm{T}}x_i + b + e_i \tag{2}$$

for all $i = 1$ to N.

The first part of this cost function is a weight decay that is used to regularize weight sizes and penalize large weights. Due to this regularization, the weights converge to smaller values. Large weights deteriorate the generalization ability of the LS-SVM because they can cause excessive variance. The second part of Eq. (1) is the regression error for all training data. The parameter γ, which has to be optimized by the user, gives the relative weight of this part as compared to the first part.

Analyzing Eq. (1) and its restriction given by Eq. (2), a typical problem of convex optimization [37] is formulated, which can be solved by using the Lagrange multipliers method [38].

$$\mathrm{La} = \frac{1}{2}\|w\|^2 + \gamma\sum_{i=1}^{N}e_i^2 - \sum_{i=1}^{N}\alpha_i\{w^{\mathrm{T}}x_i + b + e_i - y_i\} \tag{3}$$

An important task in this approach considers that the weight coefficients (w) can be written as an expansion of the Lagrange multipliers with the corresponding x_{i}:

$$w = \sum_{i=1}^{N}\alpha_i x_i \tag{4}$$

with $\alpha_{\mathrm{i}} = 2\gamma e_{\mathrm{i}}$.

Putting the result of Eq. (4) into the original regression line ($y = w^{\mathrm{T}}x + b$), the following result is obtained:

$$y = \sum_{i=1}^{N}\alpha_i x_i^{\mathrm{T}}x + b = \sum_{i=1}^{N}\alpha_i\left\langle x_i^{\mathrm{T}}, x\right\rangle + b \tag{5}$$

In applications involving non-linear regression it is enough to replace the inner product $\langle x_{\mathrm{i}}, x\rangle$ of Eq. (5) by a kernel function $K = (x_i^{\mathrm{T}}x)$. This leads to the following non-linear regression function

$$y = \sum_{i=1}^{N}\alpha_i K(x_i, x) + b \tag{6}$$

The kernel functions more widely used are: the radial basis function (RBF) given by $\exp(-\|x_{\mathrm{i}} - x_{\mathrm{j}}\|^2/2\sigma^2)$ and the polynomial $\langle x_{\mathrm{i}}, x_{\mathrm{j}}\rangle^{\mathrm{d}}$, where σ^2 is the width of the RBF function and d is the polynomial degree, which should be optimized by the user.

In this work, a new methodology was proposed using least-squares support vector machines (LS-SVM) to determine OHV of hydroxylated soybean oil by HATR FT-IR spectra.