Sensors and Actuators B 218 (2015) 97–104

Contents lists available at ScienceDirect

Sensors and Actuators B: Chemical

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

Selecting optimal wavelength intervals for an optical sensor: A case study of milk fat and total protein analysis in the region 400–1100 nm

Vladislav Galyanin[a,*], Anastasiia Melenteva[a], Andrey Bogomolov[a,b]

[a] Samara State Technical University, Molodogvardeyskaya Street 244, 443100 Samara, Russia
[b] Global Modelling, Rembrandtstraße 1, 73433 Aalen, Germany

ARTICLE INFO

Article history:
Received 23 December 2014
Received in revised form 24 March 2015
Accepted 28 March 2015
Available online 15 April 2015

Keywords:
Optical sensor
Light-emitting diode
Variable selection
Genetic algorithm
Milk analysis

ABSTRACT

A broad-band optical sensor analyzer, based on a set of light-emitting diodes (LED), for milk fat and protein analysis has been simulated and optimized using full-spectrum data in the wavelength range 400–1100 nm obtained in a designed experiment. Genetic Algorithm (GA) has been adapted to find an optimal set of wavelength intervals to be used for analysis in order to get the best prediction accuracy. Weighting and averaging of the spectral variables within the chosen intervals has been applied to take the LED emission spectra and integrating diode detection into account. Partial least-squares (PLS) regression models built on seven and six selected intervals for fat and protein, respectively, exhibit no performance loss compared to the corresponding full-spectrum models.

Suggested approach is universal and can be used to customize any LED-based or similar optical sensor system for a specific analytical problem prior to the construction. The GA-based algorithm of searching optimal de-resolved spectral intervals can be used as a general variable selection method for multivariate calibration.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

The number, frequency and diversity of analyses routinely performed in the modern industry are constantly increasing to meet growing quality and safety requirements. The leading role in the instrumental analysis belongs to the optical spectroscopy that is based on the measured interaction effects between the matter and electromagnetic radiation in different wavelength regions [1]. Visible (Vis) and near infrared (NIR) spectroscopies are the most favored techniques for industrial quality monitoring today [2,3].

Replacement of the wide-range general-purpose spectroscopic analysis by multivariate sensor systems customized for a specific application is a distinct need in modern industrial analysis [2,4–6]. Two main advantages of customized optical sensor analyzers are their relative technical simplicity and compactness. These features extend the applicability range of such sensors toward in-line, field and remote analysis, also enabling the measurement under hard conditions. Due to the price reduction, they can be installed in multiple sites, thus, increasing the level of object or process control. The absence of complex signal processing electronics, as in the full-range spectrometers, creates the potential for reducing measurement times.

One of the modern approaches used in optical sensor systems applies light emitting diodes (LEDs) or color filters for the sequential illumination of a sample followed by the simple photometric detection of diffusely reflected or transmitted light; see [4–6] and references in [6].

Regardless of the technical embodiment, the accuracy of an optical sensor system greatly depends on the number of LEDs (or other monochromatic elements) and their arrangement along the spectral region chosen for the analysis. The task of system optimization can be solved mathematically, using a representative full-spectrum data. The respective full-range multivariate regression model can be then taken as a benchmark for the evaluation of sensor system performance.

Milk fat and protein analysis in Vis and short-wave (SW) NIR range elaborated in [7–9] is based on subtle differences in low-selectivity scattering patterns of differently sized colloidal particles, and therefore, is an excellent pilot system for testing the low-resolution sensor system as an alternative to full-featured spectroscopy. Kucheryavskiy et al. [10] experimentally investigated the replacement of Vis/SW-NIR spectroscopy with a sensor system including a set of blue, green and red LEDs. Milk samples placed into a Petri dish were illuminated by the tree LEDs, one after another; and the resulting light spot was captured by a digital

* Corresponding author. Tel.: +7 927 294 45 12.
E-mail address: v.galyanin@gmail.com (V. Galyanin).

http://dx.doi.org/10.1016/j.snb.2015.03.101
0925-4005/© 2015 Elsevier B.V. All rights reserved.

Table 1
Composition of raw milk standards.

Sample	Fat (%, w/w)	Protein (%, w/w)
F1	2.22	3.58
F2	3.26	3.60
F3	4.08	3.53
F4	5.26	3.48
P1	3.55	2.98
P2	4.00	3.36
P3	4.65	3.76
P4	4.19	4.14

camera set on the opposite side. The experiment has shown that milk fat and protein contents can be predicted from the resulting images with a practically meaningful accuracy. In that case, the LED selection for the analysis was arbitrary. It is expected that the prediction performance can be significantly improved by selecting the optimal number and wavelengths of the illumination LEDs.

A wealth of published works has been devoted to the development of LED-based optical sensors for quantitative analysis on one hand, and to the selection of spectral variables on another hand. However, the problem of finding an optimal sensor configuration, i.e. choosing a set of LEDs providing minimal prediction error in a specific application, has not been systematically studied and solved before. Optical sensor optimization task described in this paper is similar to the general variable selection problem, but it also has some clear distinctions. Firstly, the solution should include a limited number of spectral channels, typically less than ten bands, as the inclusion of a larger number of respective elements (e.g. LEDs or filters) may be technically challenging or economically prohibitive. Secondly, the spectral channels in the sensor system are significantly broader than in spectroscopy. The latter feature requires that the selection of optimal wavelength during the spectrum-based sensor simulation is accompanied by a de-resolution, and consequently by averaging and weighting of spectral variables within chosen intervals. This requirement makes existing variable selection techniques inappropriate for solving the problem at hand [11,12].

In this study, genetic algorithm (GA) has been adapted as an engine of our interval selection routine for solving the problem of optical sensor optimization. GA is a general optimization approach widely used for spectral variable selection, specifically, in calibration [4,13–15]. Novel problem of finding an optimal configuration (i.e. the LED number and their spectral characteristics) of a sensor analyzer for milk fat and protein determination in the Vis/SW-NIR range (400–1100 nm), as a cost-effective and portable alternative to the spectroscopic technique described in [8], was chosen as a case study here.

2. Materials and methods

2.1. Experimental data

The experiment was performed using two sets of raw milk standards (QSE GmbH, Wolnzach, Germany) with predominantly varying fat (F) or protein content (P) (Table 1). Four standard samples in F- and P-sets were mixed pair-wise in proportions 1:2 and 2:1 in all possible combinations to produce (together with the initial pure standards) two series of 16 samples each (respectively F- and P-series). In total, 32 samples with exactly known fat and total protein content were analyzed. Due to the applied experimental design, the correlation between fat and protein contents in the samples was low, as required by the robust calibration design of a multi-component system [16,17]. Every sample was then measured three times: in its initial state and after two subsequent ultrasound homogenizations for 10 s each. The homogenization was gentle and aimed at the simulation of the natural variability of the fat globule size distribution in milk. Therefore, the resulting dataset consisted of 96 samples and respective spectra (Table S2, supplementary materials).

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.snb.2015.03.101

The milk spectra were acquired by a diode-array Vis/SW-NIR spectrophotometer (400–1100 nm) in diffuse-transmission mode using a glass cavity with inner path lengths of 4 mm.

Further details can be found in [8], where similar experimental setup and the same equipment were used.

2.2. Data analysis

Prior to the modelling the spectra were mean-centered. No further pretreatment of initial spectra was performed. Partial least-squares (PLS) regression [18] was used to perform the calibration of fat and total protein contents.

The model performance was characterized by the root mean-square errors (RMSE, Eq. (1)) of calibration ($RMSE_C$), prediction ($RMSE_P$, for the validation set) and cross-validation ($RMSE_{CV}$) as well as by the respective coefficients of determination R^2 (Eq. (2)). $RMSE_{CV1}$ and $RMSE_{CV2}$ respectively correspond to the classical leave-one-out (full cross-validation) and a segmented cross-validation formed by groups of individual samples at all homogenization degrees (three measurements each, technically – 32-fold cross-validation). The validation strategy is further described in [8].

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{k}\left(\hat{y}_i - y_i\right)^2}{k}} \quad (1)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{k}\left(\hat{y}_i - y_i\right)^2}{\sum_{i=1}^{k}(y_i - \bar{y})^2}; \quad \text{where} \quad \bar{y} = \frac{\sum_{i=1}^{k} y_i}{k} \quad (2)$$

where y_i and $\hat{y}_i$ are known and predicted fat or protein content values respectively, k – is the number of samples in the validation set (for CV k is the number of segments).

2.3. Interval selection with resolution reduction

The suggested approach makes use of full-range spectroscopy data to find a set of spectral intervals forming new variables, the sensor optical channels, optimized for a certain analytical task. Spectral variables falling into the chosen intervals are averaged and can optionally be weighted to simulate the response functions of respective monochromatic elements (LEDs, filters or detectors) used in the sensor analyzer.

Considering the problem combinatorial intensiveness and objective function complexity an appropriate optimization routine should combine the computational speed with the effectiveness in finding a viable (ideally, the global) optimum. In general, the interval optimization may be based on multiple criteria, for instance, on a combination of different regression modelling and validation statistics, such as standard errors or root mean-square errors of calibration (SE_C, $RMSE_C$), cross-validation (SE_{CV}, $RMSE_{CV}$) and prediction (SE_P, $RMSE_P$), or the respective coefficients of determination R^2. In this case, a desirability function accumulating the chosen criteria, possibly weighted, should be used as an optimization target [19].

Generally, there are no intrinsic restrictions on the interval properties: their widths, positions and weighing functions. However, the restrictions can be optionally introduced for two reasons: firstly, for a better approximation of real-life sensor elements, and

Download English Version:

https://daneshyari.com/en/article/739814

Download Persian Version:

https://daneshyari.com/article/739814

Daneshyari.com