Selecting optimal spectral regions for sensor analysis

V. Galyanin[1], A. Bogomolov[1,2]

[1] *Samara State Technical University, Samara, Russia*
[2] *J&M Analytik AG, Essingen, Germany*

Replacement of the wide-range general-purpose spectroscopy by inexpensive sensor systems customized for a specific application is a distinct trend in industrial process analysis today. One of the modern approaches applies light emitting diodes (LEDs) for the sample illumination with subsequent monochromatic detection of diffusely reflected or transmitted light.

The accuracy of such sensor systems greatly depends on the number of LEDs and their arrangement along the spectral region involved into the analysis. The system optimization can be done using full spectral data, and resulting multivariate model can be taken as a benchmark. Specific character of the problem makes existing variable selection techniques (including interval methods) inefficient.

New algorithmic approaches suggested in this work have been specifically intended for the development of sensor analyzers. The optimization has been tested on a previously published dataset [1] including 96 spectra of raw milk in visible and short-wave near infrared region used to predict fat and protein content by means of PLS regression. Optimization of the number and wavelength ranges of the light sources in the developed milk sensor was performed using two different approaches: exhaustive search for a reduced spectral resolution with a predefined grid and genetic variable selection algorithm specially adjusted for the problem at hand. It has been shown that the sensor system can be realized with five LEDs without any loss in accuracy (compared with full-spectral PLS prediction). Increasing LED number to seven noticeably improves the accuracy of fat content prediction and, moreover, reduces the model complexity compared to the spectral data.

Suggested approach can be adapted to a wide range of practical applications of optical spectral analysis.

References

1. A. Bogomolov, A.Melenteva, Scatter-based quantitative spectroscopic analysis of milk fat and total protein in the region 400–1100 nm in the presence of fat globule size variability, Chemometrics and Intelligent Laboratory Systems 126 (2013) 129–139, http://dx.doi.org/10.1016/j.chemolab.2013.02.006