ondalys
Donnez du sens à vos données

BRUKER

Support Vector Regression (SVR) on PLS Scores applied to Near Infrared Data Sets

➢ Context

In NIR spectroscopy, the most common calibration method is the PLS regression. It is a linear multivariate calibration method, efficient on spectral data and easy to implement.

However, this method reaches its limits when the data to predict is complex, for instance when many products or recipes are analyzed or when non-linear correlations are present.

In this study, conducted by Bruker Optics and Ondalys, two spectroscopic data sets were acquired with Bruker FT-NIR instruments MPA or TANGO:

- 1500 spectra of various intermediate liquid products across the raw sugar production process, with reference values for two parameters of interest (Brix and POL %)
- 9656 spectra of various finished feeds (poultry, swine and ruminant), with reference values for five parameters of interest (protein, fat, moisture, ash and fibre %)

In each dataset, the heterogeneity of the products and the non-linearity between the parameters of interest and the spectra can hinder the performances of the PLS model and would require a huge workload to create and maintain calibrations for each subset of products. In this case, Machine Learning (ML) methods such as Artificial Neural Networks (ANN) or Support Vector Machines (SVM) are a good alternative to explore.

➢ Ondalys solution

The use of SVM for quantitative analysis, called Support Vector Regression (SVR) presents some advantages over other ML methods:

- Complex data sets with a higher variance in the spectra can be handled in a single model
- Linear and non-linear relationships can be modeled
- An smaller number of samples is required compared to ANN
- SVR is relatively easy to implement, although the optimization of model parameters (in this case the 3 parameters epsilon, cost and gamma) is necessary

⇨ The application of SVR on PLS score values is proposed and compared with the results of the optimized PLS regression on the two data sets. This approach benefits from the data compression and outlier detection provided by the PLS scores, as opposed to using SVR directly on the spectra.

➢ Results

Each data set (sugar and feed) was divided into a calibration set and a tuning set used for optimization, using the same samples for PLS and SVR. For PLS, the optimization on the tuning set was performed by the function in Quant2 (Bruker Optics GmbH & Co. KG) in which selected spectral ranges and pre-processing methods are combined. For SVR, a Grid Search was used to select the best parameters based on the RMSEP of the tuning set.

For all the predicted parameters, the SVR models on the PLS scores resulted in lower errors than the PLS models built with the optimal pretreatment and variable range selection.

Moreover, SVR models were also built on the PLS scores without variable selection in order to reduce the optimization time. In this case, the performances of the PLS models were strongly degraded whereas those of the SVR models remained satisfactory.

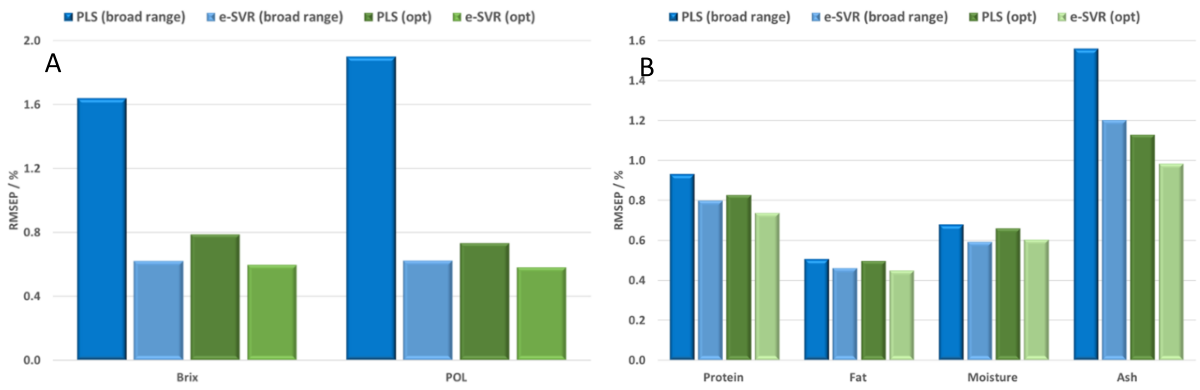

Figure 1: RMSEP values on tuning samples for models with broad range (9080-5140 cm-1) in blue, and optimized range in green (A: sugar data set, B: feed data set)

This study demonstrates the advantage of SVR on PLS scores, providing a simple approach with very good results, especially for complex data sets with high spectral variances and non-linearities.

Contact- us **Ondalys**

contact@ondalys.fr

www.ondalys.fr

☎ 04 67 67 97 87

ONDALYS S.A.S.
S.A.S. au capital de 7 500 €
RCS Montpellier 447 624 024

Siège Social
4, rue Georges Besse
34830 CLAPIERS - France

Agence Auvergne Rhône Alpes
Rond-point de l'échangeur - Les Levées
69360 SOLAIZE - France

+33 (0)4 67 67 97 87
contact@ondalys.fr
www.ondalys.fr