REGULARIZATION ADAPTION PROCESSES FOR MULTIVARIATE CALIBRATION MAINTENANCE

Anit Gurung, John H. Kalivas, Erik Andries

Abstract
In the field of chemometrics, an important issue in multivariate calibration is model updating. Model updating is the adaption process in which a model obtained for a given set of samples and measurement conditions (primary) is updated to predict the analyte in new samples and measurement conditions (secondary). The calibration method partial least squares is applied with two new updating approaches. In one approach, only one updated model is obtained to predict the analyte amount in both primary and secondary conditions. The other approach forms two updated models in which one model is used to predict in primary conditions and the second model is used to predict in secondary conditions. Both approaches are evaluated with near-infrared spectral datasets. Datasets include spectra of sunflower and pharmaceutical tablets. Fusion process and single merits are used to select models. Model selection methods are evaluated based on prediction errors using selected models.

Objective
• Develop a new effective modal updating approach.

Model Updating Approach

Partial Least Squares (PLS)

1b-PLS: 1 updating model
\[ y = Xb + e \]

2b-PLS: 2 updating models
\[ y = Xb + e \]

Model Measures

• Bias:
  - Primary Calibration
  - Secondary Calibration
  - Median Bias (BM) of 0 saves models

• Variance:
  - Root Mean Squares Error (RMSE)

• U-Curves:
  - Bias-variance trade-off

Data Centering

Local mean centering
- Mean of M and y are used to center validation samples in secondary condition.

Experimental Design & Results

Corn Data
- 80 samples
- 5 different instruments

Set 1 (Mp6spec – Mp5spec):
- Primary: Mp6spec
- Secondary: Mp5spec
- 60% of samples ranging from 1000 to 4000

Set 2 (Mp5spec – Mp6spec):
- Primary: Mp5spec
- Secondary: Mp6spec
- 40% of samples ranging from 1000 to 6000

Results 1
- Model with minimum RMSEY plot sorted based on 1b-PLS

Results 2
- Model with minimum RMSEY plot sorted based on 2b-PLS

Pharmaceutical Tablet Data
- 900 calibration samples
- Type 1 (90 mg)
- Type 2 (25 mg)
- Type 3 (10 mg)
- Type 4 (270 mg)
- Each tablet type is produced in three batches

But, median fusion can not be disregarded as it can pick better models than sum fusion in some instances.

Future Work
- Apply model updating approaches with unlabeled data.
- Unlabeled data does not have analyte concentration values of samples.

Acknowledgement
This material is based upon work supported by the National Science Foundation under Grant No. CHE-1506417 (co-funded by CDS&E Program) and is gratefully acknowledged by the authors.

Observations

• Result 1 and 2-1:
  - Cross validation splits where the prediction error is large using 1b-PLS; generally, 2b-PLS lowers the prediction error significantly.

• Results 1-2, 1-5, 2-3, and 2-5:
  - Models have similar low prediction error for selected models

Conclusions

For model updating, it seems to be clear that:
- 2b-PLS works better than 1b-PLS on the dataset where the differences in spectra of samples in primary and secondary conditions are minimal.
- Both 1b-PLS and 2b-PLS give similar results when spectra are unique to each other like change in intensity.
- In this situation, it suggests to look for different method for the approach using two updating models.

Regarding the model selection method, based on the consistency of selecting the better model across most of the data, sum fusion seems to work best in picking the models.
- But, median fusion can be disregarded as it can pick better models than sum fusion in some instances.