

휴대용 NIR분석장치용 software개발

이기홍, 이문용, 조문신*, 유순유*, 이정표**, 김석준**
 영남대학교, 카사테크(주)*, SK주식회사**

Development of software for portable NIR spectroscopy

Kihong Lee, Moonyoung Lee, Moonshin Zo*, Soonyoo Yoo*, Jeongpyo Lee**,
 Sukjoon Kim**

School of Chemical Engineering & Technology, Yeungnam University
 CASA Tech.*
 SK Corp.**

서론

NIR spectroscopy는 NIR(Near InfraRed)분광분석법을 이용하여 정유 및 석유화학 공정에서 품질 관리 및 최적화를 수행하는 기법을 수행하는 하나의 방법이다. 그 성능은 실제공정에서 오랜 기간의 현장 테스트를 통하여 검증되어 왔으며, 정유 및 석유화학공정에서 많은 문제들에 대한 해법을 제시해왔다. 본 연구에서 개발한 NIR Plus software는 대형의 장치가 필요한 on-line spectroscopy가 아닌 전문지식이 없는 사람들도 쉽게 사용할 수 있는 소형의 장치를 개발하고, 그 장치의 제어 및 모델링을 수행하기 위한 software로서 개발되었다.

이론

Spectrum으로 구성된 입력행렬 \mathbf{X} 는 K개의 파장값을 갖는 N개의 sample로 이루어져 있다(N by K). 측정되는 물리량(ROn, ARO, Viscosity,...)으로 이루어진 행렬 \mathbf{Y} 는 각 sample당 M개의 물리량이 나열된 행렬이다(N by M). 여기에서는 두 행렬 \mathbf{X} 와 \mathbf{Y} 를 맺어주는($\mathbf{Y}=\mathbf{XA}$) K by M행렬 \mathbf{A} (Coefficient or Estimator)를 구하여 이를 통하여 미지의 sample에 대한 spectrum으로부터 각각의 물리량을 구하고자 하는 것이다. 이 작업 중 outlier detection을 위하여 PCA 및 Mahalanobis distance를 사용하였고, modeling을 위하여 PLS algorithm을 사용하였다.

1. PCA(Principal Component Analysis) : 임의의 행렬 \mathbf{X} 를 Singular Value Decomposition을 통하여 표현하면 아래와 같이 표현할 수 있다.

$$\mathbf{X} = \lambda_1 \begin{pmatrix} \hat{t}_1 \\ \vdots \\ \hat{t}_N \end{pmatrix} \begin{pmatrix} \hat{p}_1 & \dots & \hat{p}_K \end{pmatrix} + \lambda_2 \begin{pmatrix} \hat{t}_2 \\ \vdots \\ \hat{t}_N \end{pmatrix} \begin{pmatrix} \hat{p}_2 & \dots & \hat{p}_K \end{pmatrix} + \dots + \lambda_k \begin{pmatrix} \hat{t}_k \\ \vdots \\ \hat{t}_N \end{pmatrix} \begin{pmatrix} \hat{p}_k & \dots & \hat{p}_K \end{pmatrix}$$

여기서 \hat{t}_i 는 N차원 단위 vector로 score vector라 하고, \hat{p}_k 는 K차원 단위 vector로 loading vector라 한다. 구해진 score vector를 좌표로 표시한 것을 score chart라 하며, 이는 원래의 좌표 점들을 새로운 좌표로 표시하는 것이라 할 수 있는데 loading vector는 이 새로운 좌표의 각 축을 나타내는 단위 vector이다. 새롭게 구해진 좌표의 의미는 각각의 축에서 새롭게 구성한 좌표의 eigenvalue의 크기 순으로 직교 좌표가 구성되었다는 것이다. 이와 같이 자료행렬을 분해하여 유용한 자료를 뽑는 것을 PCA라 한다.

2. Mahalanobis Distance : Mahalanobis distance는 PCA로부터 구해진 score chart로부터 새로운 PC축간의 scale을 같게 만들어 각 좌표간의 variance를 같게 놓고 구한 거리를 Mahalanobis distance라 한다.

$$D_{M,i}^2 = (n-1)^2 \sum_{j=1}^k \left[s_{ij}^2 \sum_{i=1}^n s_{ij}^2 \right]$$

여기서 index j는 모델에 사용된 모든 k개의 Principal Component를 나타내고, index i는 training set의 n개의 sample을 나타낸다.

3. MLR(Multilinear Regression) : 입력과 출력 행렬을 각각 X, Y라 할 때, 두 행렬의 관계를 선형으로 가정하면 다음과 같이 나타낼 수 있다.

$$\mathbf{Y} = \mathbf{X} \cdot \mathbf{A} + \mathbf{e}, \quad \mathbf{A} = (\mathbf{X}^t \mathbf{X})^{-1} (\mathbf{X}^t \mathbf{Y})$$

이 방법이 성립하기 위해서는 $\mathbf{X}^t \mathbf{X}$ 행렬에 singularity가 없어야 하는데 이런 sample data는 종속성을 보이는 경우가 많아 추정행렬 \mathbf{A} 를 구하기가 불가능하다.

4. PCR(Principal Component Regression) : Singular value decomposition을 통하여 얻어진 score vector들과 Y행렬을 이용하여 추정행렬 A를 구한다.

$$\mathbf{A} = \mathbf{P}(\mathbf{T}^t \mathbf{T})^{-1} \mathbf{T}^t \mathbf{Y}$$

이 경우 \mathbf{X} 를 principal component들의 합으로 구성하여 \mathbf{X} 의 몇 가지 주성분만을 고려한 경우 noise나 outlier를 제거하여 자료행렬을 단순화시킬 수 있고, singularity가 일어나지 않는다는 장점이 있지만, score vector들이 \mathbf{Y} 와 무관하게 \mathbf{X} 공간에서의 크기 순으로 정의된 것이므로 \mathbf{Y} 를 잘 예측하리라는 보장이 없다.

5. PLS(Partial Least Square) : PCR의 경우 X의 주성분들이 Y의 주성분을 잘 예측한다는 보장이 없어, 이를 보완할 수 있는 방법이 PLS algorithm이다.

PLS algorithm의 추정행렬 \mathbf{A} 는 다음과 같이 구할 수 있다.

$$\begin{aligned} \mathbf{U} \mathbf{P}^t \mathbf{W} &= \mathbf{T} \mathbf{P}^t \mathbf{W} = \mathbf{X} \mathbf{W} \\ \therefore \mathbf{U} &= \mathbf{X} \mathbf{W} (\mathbf{P}^t \mathbf{W})^{-1} \\ \therefore \mathbf{Y} &= \mathbf{U} \mathbf{Q}^t = \mathbf{X} \mathbf{W} (\mathbf{P}^t \mathbf{W})^{-1} \mathbf{Q}^t = \mathbf{X} \mathbf{A} \\ \therefore \mathbf{A} &= \mathbf{W} (\mathbf{P}^t \mathbf{W})^{-1} \mathbf{Q}^t \end{aligned}$$

6. Selecting the Number of PLS Factors

일반적으로 PLS factor의 수를 결정하는 것은 PLS calibration의 중요한 부분이다. factor의 수가 너무 적을 경우, 너무 적은 정보로 인해 큰 prediction error를 얻을 수 있다. 반대로 factor의 수가 너무 많을 경우, 모델링에 포함될 수 있는 noise나 outlier에 의해 모델의 결과가 너무 불안정하게 될 수 있다. 최적의 factor의 수를 결정하기 위하여 PRESS(Prediction Residual Error-Sum of Square)를 계산하여, 사용자가 시각적으로 보고 최적의 factor의 개수를 결정할 수 있도록 하였다.

Software구성 및 주요기능

NIR Plus software는 Figure 1,2와 같이 NIR Plus main program과 NIR Plus modeling software의 두 부분으로 나누어져 있다.

NIR Plus main program에서는 modeling을 위한 data들의 import 및 export를 수행한다. 또한 modeling software의 호출, hardware제어 software를 호출하고, modeling된 결과를 export한다.

NIR Plus modeling software에서는 주로 main program으로부터 넘겨받은 data를 이용하여 PLS modeling을 수행하고, 결과를 분석할 수 있도록 한다.



Figure 1. NIR Plus main screen

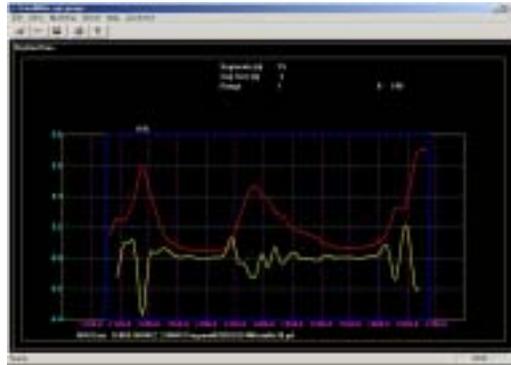


Figure 2. Modeling & result screen

NIR Plus main program에서는 Microsoft Excel에 정해진 양식으로 입력된 sample의 NIR spectrum data들인 wavelength를 project 및 group으로 관리하고, 간단한 data editing 기능을 제공한다. 또한 modeling을 수행한 후, 결과파일들을 graph 및 sheet로 보여주고, Excel sheet를 자동으로 생성한다.

Modeling software에서는 PLS modeling을 수행하고, 결과 분석 및 factor의 수를 결정하기 위하여 PLS factor의 개수에 따라 각각의 PRESS, Mahalanobis distance, calibration 및 prediction 결과를 graph로 보여준다. 각각의 결과에 대한 전환은 사용자가 편하게 사용할 수 있도록 하기 위하여 간단한 키 조작으로 가능하게 하였다.

개발된 소프트웨어의 주요 기능은 다음과 같다.

1. 소형화된 분석 장치의 구동 드라이버로서 하드웨어 제어기능
2. PLS modeling을 위한 project 및 group file관리
3. 쉬운 modeling(PLS & PCA)
4. Case study기능
5. Microsoft Excel과 연계하여 유저 사용에의 편의성 제공
6. 쉬운 report기능
7. 하드웨어에 대한 Correction 기능
8. 각 기능들의 module화 - 기능추가 및 update의 용이성
9. 다양한 hardware제어용 engine과의 연결가능
10. 다양한 형태의 결과 보기 지원 (아래의 Figure 3,4,5,6,7,8과 같이 modeling에 따른 Mahalanobis distance, PRESS, Trend, Error, 45° Plot, 등을 보여주고, 같은 형식으로 hardware에 따른 correction결과를 graph로 보여준다. 또한 모든 결과들은 main software로 넘겨져 Microsoft Excel data로 자동 저장된다.)



Figure 3. Mahalanobis distance

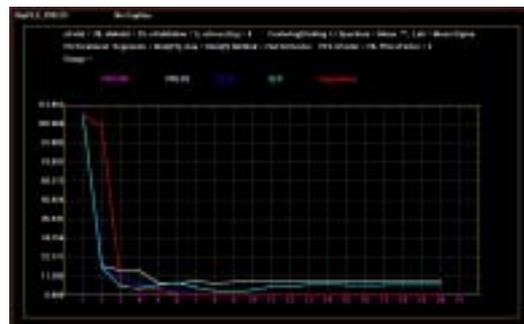


Figure 4. PRESS Plot

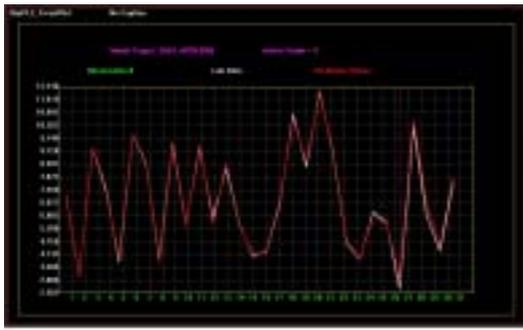


Figure 5. Trend Plot



Figure 6. Error Plot

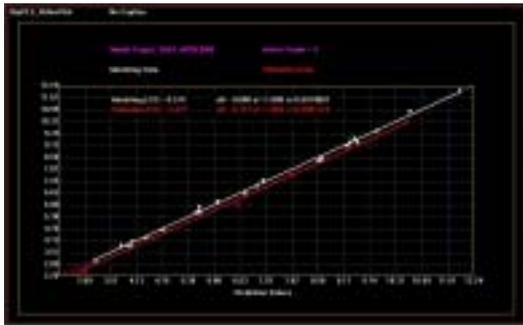


Figure 7. 45° Plot

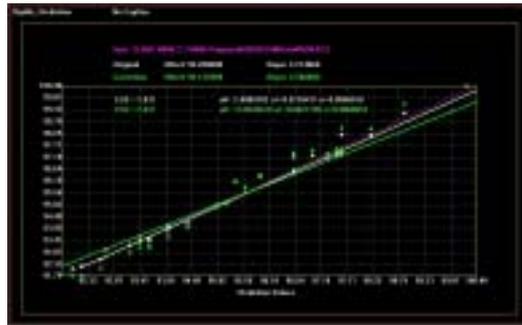


Figure 8. Correction Plot



Figure 9. Correction dialog box

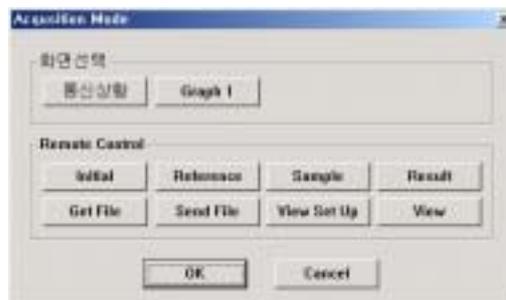


Figure 10. hardware 제어 dialog box

결론

이번에 개발된 software는 SK주식회사에서 개발된 소형의 NIR spectroscopy장치를 위하여 개발되었으며, 개발 tool은 Visual C++ 및 Visual Basic을 사용하였다. 사용된 Algorithm은 modeling을 위하여 PLS를 사용하였으며, 결과분석을 위하여 PCA 및 Mahanobis distance를 사용하였다. 또한 각 기능들을 module화 하여 프로그램의 update를 용이하게 할 수 있도록 하였다.

감사의 글

본 연구는 SK주식회사의 지원에 의해 수행되었습니다.

참고문헌

1. L.Eriksson, E.Johansson, N.Kettaneh-Wold and S.Wold., "Introduction to Multi- and Megavariate Data Analysis using Projection Methods(PCA & PLS)", UMETRICS, 1999
2. "VISION Documentation", FOSS NIRSystems, 1999