

데이터 추출 과정을 적용한 Block-wise Adaptive Predictive PLS

Block-wise Adaptive Predictive PLS using Block-wise Data Extraction

김성영, 이범석*, 정창복, 최수형

(Sung Young Kim, Bomsock Lee, Chang Bock Chung, and Soo Hyoung Choi)

Abstract : Recursive Partial Least Squares(RPLS) method has been used for processing the on-line available multivariate chemical process data and modeling adaptive prediction model for process changes. However, RPLS method is unstable in PLS model updating because RPLS method updates PLS model by merging past PLS model and new data. In this study, Adaptive Predictive Partial Least Squares(APPLS) method is suggested for more sensitive adaptation to process changes. By expanding APPLS method, block-wise Adaptive Predictive Partial Least Squares(block-wise APPLS) method is suggested for a larger scale data of chemical processes. APPLS method has been applied to predict the reactor properties and the product quality of a direct esterification reactor for polyethylene terephthalate(PET), and block-wise APPLS method has been applied to predict the cetane number using NIR Diesel Spectra data. APPLS and block-wise APPLS methods show better prediction and updating performance than RPLS method.

Keywords : adaptive predictive PLS, APPLS, block-wise APPLS, PLS, RPLS, NIR diesel fuel spectra

I. 서론

최근의 화학공정은 공정의 자동화로 인하여 실시간으로 조업 상황을 모니터링하는 것이 가능해짐에 따라 화학공정 품질변수의 실시간 예측 성능을 향상시키는 문제에 관심이 집중되고 있다. 이를 위해서 실시간으로 얻어지는 데이터를 이용해 공정 변화를 반영하여 모델을 갱신하는 적응 예측 모델(adaptive predictive model)을 수립하는 것이 필요하다. 모델을 갱신하지 않으면 시간에 따른 공정의 변화에 대응하지 못하여 모델의 예측성능이 저하된다. 본 논문에서는 과거의 조업 데이터를 이용하여 PLS 모델을 세운 후, 블록 단위로 새롭게 얻어지는 데이터를 이용해 모델을 갱신하는 형태의 Block-wise Adaptive Predictive Partial Least Squares(Block-wise APPLS) 방법을 개발하였다.

가장 단순한 PLS 모델 갱신 방법은 공정의 진행에 따라 새롭게 얻어지는 데이터를 과거의 데이터와 모두 결합하여 새로운 PLS 모델을 계속해서 만들어 나가는 방법이다. 이 방법은 예측 성능은 우수하지만 시간이 지남에 따라 모델의 크기가 너무 커지고 공정의 변화에 대응하지 못하는 단점을 지니고 있다. 이런 이유로 인하여 실제 공정의 변화에 따른 모델 갱신 방법에 대한 몇몇 연구가 수행되어져 왔다.

Wold[4]는 exponentially weighted moving PCA/PLS 방법을 개발하여 모델 갱신에 이용하였다. 이 방법은 새롭게 얻어지는 데이터에 더 많은 가중치를 부여하여 시간에 따른 공정의 변화에 모델을 적응시키는 장점이 있는 반면에, 비정

상적인 공정의 거동에 대해 적절히 대응하지 못하는 단점이 있다[15]. Gallagher[5]등은 시간의 흐름에 따라 비정상적인 거동이 감지 될 때까지 데이터를 축적한 후에 공정 모델을 갱신하는 exponentially weighted moving average(EWMA)와 exponentially weighted moving covariance(EWMC)가 결합된 PCA 방법을 개발하였다. 그러나 이 방법은 오차 측정 방법에 따라 모델의 신뢰도가 달라지는 문제점을 지니고 있다[15]. 이러한 문제점들을 해결하기 위해 Qin[2]은 moving window approach를 Recursive Partial Least Squares(RPLS)에 결합한 방법을 제안하였다. RPLS 방법은 Hellend[1]등에 의해 처음 제안된 방법으로서 과거의 PLS 모델에 새로운 데이터를 추가하여 PLS 모델을 갱신하는 방법이다. 이는 많은 입력력 변수가 존재하는 촉매 개질기의 생산제품의 옥탄가 예측[2], Nitrogen Oxides의 배출농도 예측[9]등의 연구에 이용되어졌다. RPLS 방법은 새로운 데이터와 과거의 데이터를 모두 사용하여 새로운 PLS 모델을 만드는 방법에 비해 PLS 모델이 커지지 않고 PLS 모델의 갱신을 쉽게 해주는 장점이 있지만, 예측 성능은 떨어지는 단점이 있다.

본 논문에서는 이러한 단점을 개선하기 위하여 과거의 공정 데이터를 이용하여 만든 PLS 모델에 새로운 데이터가 들어오면 기존의 PLS 모델에서 오차가 큰 데이터는 제거하고 새롭게 들어온 데이터를 모델에 포함시키는 Adaptive Predictive PLS(APPLS) 방법을 개발하였다. APPLS 방법은 모델을 갱신할 때, 기존 모델에서 오차가 가장 큰 데이터가 제거되므로 PLS 모델의 크기는 일정하게 유지하면서, 동시에 RPLS 방법에 비해 예측 성능이 향상되는 효과를 가질 수 있다. 이러한 APPLS 방법을 확장하여, 데이터의 양이 많아질 경우 공정의 변화에 효과적으로 대응하기 위해 새롭게 들어오는 데이터를 블록 단위로 처리하는 block-wise APPLS 방법을 개발하였다.

본 연구에서는 PET 중합공정에 사용되는 직접 에스테르

* 책임저자(Corresponding Author)

논문접수 : 2006. 4. 14., 채택확정 : 2006. 5. 9.

김성영, 이범석 : 경희대학교 환경응용화학부

(alla79@hanmail.net/bslee@khu.ac.kr)

정창복 : 전남대학교 응용화학공학부(chungeb@chonnam.ac.kr)

최수형 : 전북대학교 화학공학부(soochoi@chonbuk.ac.kr)

※ 본 논문은 한국과학재단(과제번호 : R01-2003-000-10697-0)에서 지원하여 연구하였음.

화. 반응기에서 얻어지는 데이터[10]에 대해 APPLS 방법을 적용하여 반응기의 유출수의 유량과 유출수에 포함된 EG의 유량, 그리고 생산제품 내에 존재하는 acid end group의 농도를 예측해 보았고, NIR Diesel Spectra 데이터[17]를 이용하여 디젤유의 세탄가를 예측하는 문제와 같이 입출력 데이터가 많은 문제에 대해 block-wise APPLS 방법을 적용해 보았다. 또한 같은 문제에 대해 기존의 PLS 모델을 수정하지 않고 계속하여 사용하는 방법(PLS1), 새롭게 얻어지는 입출력 데이터를 기존의 PLS 모델에 계속 포함시켜 매번 새로운 PLS 모델을 만들어 수정하는 방법(PLS2)과 타 문헌에서 제시된 RPLS 방법[1]을 사용하여 본 논문에서 제시된 APPLS 방법과 block-wise APPLS 방법의 예측 성능을 비교해 보았다.

II. 이론

1. Partial Least Squares(PLS) 방법

PLS는 다변량 입출력 변수에 대한 데이터를 동시에 분석하는 통계적 기법으로 공정을 통해 측정된 여러 변수의 관계성이나 인과관계를 동시에 고려하여 공정을 정확하게 해석하는데 도움을 준다. PLS 방법은 새로운 잠재 변수(latent variable)들을 찾아내 모델에 사용하는 방법으로 PLS에 사용되는 잠재변수들은 서로 선형적으로 독립이며 입력 잠재변수와 출력 잠재변수 사이에는 매우 높은 상관관계를 가지게 된다. 입력 변수 데이터 블록(X)은 score 행렬 T와 loading 행렬 P의 곱으로 나타내어지고, 출력 변수 데이터 블록(Y)은 score 행렬 U와 loading 행렬 Q의 곱으로 표시되어진다.

$$X = TP^T + E \tag{1}$$

$$Y = UQ^T + F \tag{2}$$

$$U = TB + G \tag{3}$$

새로운 데이터가 들어왔을 때 PLS 모델을 수정하지 않고 기존의 PLS 모델을 사용하여 품질변수들을 계속 예측해 나가는 PLS1 방법은 모델 갱신을 하지 않으므로 계산 시간은 단축되지만 공정의 변화에 대한 대응을 전혀 하지 못하는 단점이 있다. 새롭게 얻어지는 입출력 데이터를 기존의 데이터에 계속 포함시켜 매번 새로운 PLS 모델을 만드는 PLS2 방법은 새로운 데이터가 들어올 때마다 모델을 갱신하기 때문에 공정이 진행될수록 모델의 예측성능은 향상되지만 기존 데이터와 새로운 데이터를 모두 사용하여 매번 PLS 모델을 만들기 때문에 시간이 지남에 따라 PLS 모델의 크기가 너무 커지는 단점이 있다.

2. Recursive Partial Least Squares(RPLS) 방법

RPLS 방법은 시간이 흐름에 따라 발생하는 공정의 변화들을 모델에 반영하기 위해 새로운 공정 데이터를 이용하여 PLS 모델을 갱신하는 방법이다[1]. 이 방법은 기존의 데이터를 이용하여 PLS 모델을 수립하고 새로운 데이터 X_1, Y_1 이 얻어지면 이를 이용해서 기존의 PLS 모델을 갱신하는 알고리즘이다. RPLS 알고리즘을 간단히 설명하면 다음과 같다.[2]

먼저 기존의 데이터 행렬 X, Y에 NIPALS 알고리즘[3]을 수행하여 PLS 모델을 수립한다.

$$[X, Y] \rightarrow NIPALS \rightarrow [T, W, P, B, Q] \tag{4}$$

기존의 데이터 블록으로 수립된 PLS 모델에 새로운 데이터 $\{X_1, Y_1\}$ 이 들어오면 다음과 같이 RPLS 모델을 갱신한다.

$$X_{new} = \begin{bmatrix} P^T \\ X_1 \end{bmatrix}, Y_{new} = \begin{bmatrix} BQ^T \\ Y_1 \end{bmatrix} \tag{5}$$

위와 같이 새롭게 들어온 데이터는 기존의 모델과 결합하여 새로운 데이터 블록을 만들게 되고, 이를 이용해 다시 PLS(NIPALS 알고리즘) 과정을 반복하는 방법으로 모델을 갱신하게 된다.

RPLS 방법은 기존의 PLS 모델과 새로운 데이터를 이용하여 모델을 갱신하므로 새롭게 구성된 데이터 블록의 크기가 작게 유지되어 PLS 모델의 크기가 커지지 않는다는 장점이 있다. 하지만 기존의 데이터와 새로운 데이터가 합쳐지는 PLS2 방법에 비해 모델의 예측 성능이 떨어지고 신뢰도가 저하되는 단점을 갖고 있다.

3. 데이터 추출 과정을 적용한 Adaptive Predictive PLS 방법

Adaptive Predictive PLS(APPLS) 방법은 실시간으로 진행되는 공정의 새로운 데이터가 들어왔을 때, 위의 PLS1, PLS2와 RPLS 모델 갱신 방법에서 나타나는 문제점들을 해결하기 위한 새로운 모델 갱신 방법이다. 이 방법은 새로운 데이터가 들어와서 모델을 갱신하기 전에 기존의 모델에서 오차가 큰 데이터를 찾아내 PLS 모델을 구성하는 입출력 데이터 행렬에서 제거하는 데이터 추출 과정을 미리 수행한다.

$$\begin{aligned} &\text{Eliminate } [X, Y]_i \text{ from} \\ &\text{for Max } |Y_i - Y_{m,i}| \quad (i=1, \dots, n) \end{aligned} \tag{6}$$

여기서 Y는 PLS 모델을 구성하는 실제 공정데이터이고 Y_m 은 모델 예측 값이다. 그리고 i는 입출력 데이터 행렬의 열 번호를 나타내고 n은 PLS 모델의 입출력 데이터의 개수를 나타낸다. $[X, Y]_i$ 는 예측 오차가 가장 큰(Max $|Y_i - Y_{m,i}|$) i번째 데이터를 의미한다. 위와 같은 데이터 추출 과정을 거쳐 실제 공정 값과 예측 값과의 오차가 가장 큰 i번째 실제 공정 데이터를 기존의 PLS 모델에서 제거하고, 새로운 데이터가 얻어지면 제거된 데이터의 자리를 채워가는 방식으로 모델 갱신이 진행된다. 자세한 APPLS 알고리즘은 표 1에 나타내었다.

다음과 같이 APPLS 모델 갱신 방법은 기존의 데이터 중에서 예측 오차가 가장 큰 데이터는 제거되고 새롭게 들어온 데이터가 모델에 포함되므로 PLS 모델을 구성하는 데이터의 개수가 항상 일정하게 유지된다. 따라서 PLS 모델의 크기가 일정하게 유지되고 모델을 계산하는 시간을 줄여줄 뿐만 아니라 RPLS 방법에 비해 모델의 예측성능도 향상되는 효과를 보인다.

표 1. APPLS 알고리즘.

Table 1. The APPLS algorithm.

①	$[X, Y] \rightarrow NIPALS \rightarrow [T, W, P, B, Q]$
②	$T = (P^T P)^{-1} X_1 P$: for new data X_1
③	$\hat{Y}_1 = TBQ^T$: prediction for new data for Y_1
④	$T = (P^T P)^{-1} X P$: old data block $[X, Y]$
⑤	$Y_m = TBQ^T$: prediction for old data for Y
⑥	Eliminate $[X, Y]_i$ from $[X, Y]$ for Max $ Y_i - Y_{m,i} $ ($i=1, \dots, n$) : data extraction
⑦	new data $\{X_1, Y_1\}$ is available, $X_{new} = \begin{bmatrix} X \\ X_1 \end{bmatrix}, Y_{new} = \begin{bmatrix} Y \\ Y_1 \end{bmatrix}$
⑧	Return to step ①.

4. Block-wise Adaptive Predictive PLS 방법

단위 시간당 많은 양의 입출력 데이터가 얻어지는 대규모 화학공정에서는 하나의 데이터를 이용해 PLS 모델을 갱신하는 APPLS 방법은 사용하기 어렵다. 본 논문에서는 이러한 문제점을 해결하기 위하여 입출력 데이터를 블록 단위로 축적하여 PLS 모델 갱신을 수행하는 block-wise APPLS 방법을 개발하였다.

Block-wise APPLS 방법은 초기 PLS 모델에 사용되는 데이터의 개수 만큼을 하나의 블록으로 처리한다. 즉, 기존 PLS 모델에 사용된 입출력 데이터의 개수만큼 블록으로 새로운 데이터를 받아들이고, 그 후 블록으로 데이터 추출 과정을 수행하여 PLS 모델을 갱신한다. block-wise APPLS 알고리즘은 위의 표 1에 나와 있는 APPLS 알고리즘을 블록의 개념으로 확장한 것으로 자세히 설명하면 다음과 같다.

먼저 기존의 데이터 행렬 $[X, Y]$ 에 NIPALS 알고리즘을 수행하여 PLS 모델을 수립할 때 사용된 X 데이터 행렬이 $(k \times l)$ 크기의 행렬이라면 모델 갱신에 사용되는 블록 데이터 X_1 의 크기는 $(k \times l)$ 이 된다. 즉, 초기 PLS 모델에 사용된 데이터의 개수 만큼을 블록의 단위로 지정하여 다음과 같은 과정을 반복 수행한다.

$$\begin{aligned}
 & [X, Y] \rightarrow NIPALS \rightarrow [T, W, P, B, Q] \\
 & T = (P^T P)^{-1} X_1 P \tag{7} \\
 & \hat{Y}_1 = TBQ^T \text{ : prediction for new data for } Y_1 \\
 & X_{new} = \begin{bmatrix} X \\ X_1 \end{bmatrix}, Y_{new} = \begin{bmatrix} Y \\ Y_1 \end{bmatrix}
 \end{aligned}$$

기존의 데이터로 PLS 모델을 수립하고, 지정된 블록의 크기만큼 새로운 데이터를 예측하고, 새로운 데이터를 받아들이는 과정이 끝나면 다음과 같이 새롭게 갱신된 X 데이터 행렬에서 예측 오차가 큰 데이터를 순서대로 블록의 크기만큼 제거하는 데이터 추출 과정을 수행하게 된다.

$$\begin{aligned}
 & T = (P^T P)^{-1} X P \\
 & Y_m = TBQ^T \text{ : prediction for old data for } Y \tag{8} \\
 & \text{Eliminate } [X, Y]_{j=1, \dots, k} \text{ from } [X, Y] \\
 & \text{for Max}_{(j=1, \dots, k)} |Y_i - Y_{m,i}|_j \text{ (} i=1, \dots, 2k \text{)}
 \end{aligned}$$

(8)과 같이 블록 크기만큼 데이터 추출 과정을 반복하고 난 후 처음으로 돌아가서 새로운 데이터를 받아들이고 예측하는 과정을 반복한다. 이처럼 block-wise APPLS 알고리즘은 블록 단위로 모델을 세우고, 갱신한 다음 모델로부터 블록 단위로 데이터 추출 과정을 반복하게 된다.

III. PET 중합공정 중 직접 에스테르화 반응기

PET 중합공정 중 직접 에스테르화 반응기의 데이터를 이용하여 APPLS 방법의 예측 성능을 확인해 보았다. 초기 PLS 모델에 사용된 데이터의 수는 5개이고, 12개의 새로운 데이터를 하나씩 예측하면서 모델을 갱신하였다. PET 중합공정은 긴 반응시간과 높은 반응 온도, 대용량의 다단계 공정시설을 필요로 하는 대표적인 에너지 소비 공정이다[10]. PET 직접 에스테르화 반응기를 간단히 나타내면 그림 1과 같다.

직접 에스테르화 반응기에는 TPA와 EG가 원료로 들어가며, TPA와 EG의 반응정도는 반응기에서 유출되는 water-EG 수증기의 양으로서 측정할 수 있다. 직접 에스테르화 반응에서 생성되는 부산물 중 디에틸렌글리콜(DEG)은 PET의 물리적 및 화학적 물성에 가장 큰 영향을 미치는 물질이므로 반응기 내부의 DEG 농도를 정확하게 예측하는 것이 매우 중요하다. 본 논문에서는, 반응기의 반응율에 큰 영향을 미치는 반응온도 T [°C], 압력 P [atm], 반응시간 τ [hr], 원료로 사용되는 TPA와 EG의 몰비율 α [-] 등을 입력 변수로 사용하였다. 출력 변수로는 반응기에서 유출되는

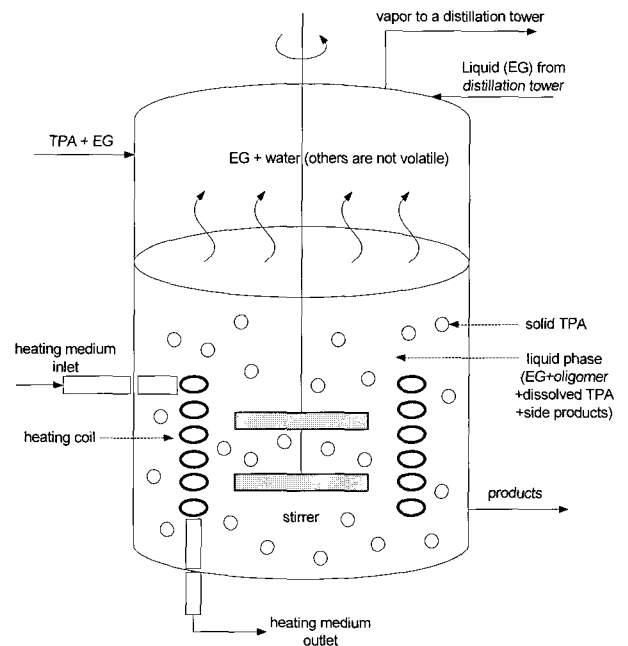


그림 1. PET 중합 공정 중 직접 에스테르화 반응기.
Fig. 1. Direct esterification reactor in PET synthesis.

표 2. 직접 에스테르화 반응기에 대한 PLS1, PLS2, RPLS, APPLS의 예측 성능 비교.

Table 2. Comparison of prediction by PLS1, PLS2, RPLS, APPLS for a direct esterification reactor.

	F^v_{water}				F^v_{EG}				C^0_{acid}			
	PLS1	PLS2	RPLS	APPLS	PLS1	PLS2	RPLS	APPLS	PLS1	PLS2	RPLS	APPLS
	0.01559	0.01559	0.01559	0.01559	0.00089	0.00089	0.00089	0.00089	0.00731	0.00731	0.00731	0.00731
	0.02967	0.03250	0.01414	0.03965	0.00066	0.00043	0.00154	0.00126	0.02848	0.02965	0.02119	0.03430
	0.08226	0.06630	0.05260	0.04781	0.00497	0.00606	0.00562	0.00458	0.07374	0.05371	0.04526	0.03532
	0.17523	0.07990	0.09297	0.06185	0.01100	0.00618	0.00603	0.00491	0.13672	0.04450	0.06298	0.03181
Root	0.26424	0.03335	0.08901	0.02556	0.01473	0.00062	0.00373	0.00049	0.21190	0.02316	0.07518	0.02016
Square	0.17523	0.01069	0.08900	0.00492	0.01100	0.00159	0.00373	0.00063	0.13672	0.00393	0.07518	0.00593
Error	0.13193	0.00285	0.04330	0.01947	0.01039	0.00260	0.00061	0.00175	0.10214	0.00564	0.03458	0.01716
	0.13643	0.00964	0.00449	0.02072	0.00359	0.00521	0.00680	0.00602	0.10461	0.01163	0.00247	0.01940
	0.12952	0.00661	0.00691	0.00543	0.00524	0.00171	0.00165	0.00079	0.10238	0.00464	0.00223	0.00374
	0.11557	0.00455	0.01395	0.02076	0.00178	0.00479	0.00346	0.00496	0.09742	0.00218	0.00496	0.01064
	0.17988	0.00624	0.06431	0.00192	0.01076	0.00157	0.00898	0.00227	0.14021	0.00177	0.04279	0.00687
	0.04311	0.01494	0.13676	0.03878	0.00384	0.00108	0.00691	0.00055	0.03555	0.01031	0.10466	0.02767
평균	0.12322	0.02360	0.05192	0.02521	0.00657	0.00273	0.00416	0.00242	0.09810	0.01654	0.03990	0.01836

수증기의 유량 F^v_{water} 과 EG vapor의 유량 F^v_{EG} , 그리고 반응기 내부에 존재하는 acid end group의 농도 C^0_{acid} 를 사용하였다.

PLS 모델을 수립할 때 잠재변수의 개수는 2개로 정하였다. 초기 5개의 데이터를 이용하여 PLS 모델을 수립하고 6번째 데이터부터 차례대로 12개의 데이터를 받아들이면서 PLS1, PLS2, RPLS와 APPLS 방법을 사용하여 출력변수를 예측하였다. 초기 PLS 모델을 수립하고 새롭게 얻어진 데이터의 입력변수를 이용하여 새로운 출력변수를 예측한 후에, 새로운 입력력 공정 변수들을 사용하여 모델을 갱신하였다. 12개의 데이터를 차례대로 위 4가지 방법에 적용시켜 예측 값과 실제 공정 값을 비교해 보았다.

표 2는 12개의 데이터를 예측한 값과 실제 공정 값과의 제곱근오차(root square error)와 이들의 평균을 나타낸 것이다. 그림 2는 표에 나타난 제곱근오차 값들을 누적시켜 비교한 그래프이다.

표 2에서 F^v_{water} , F^v_{EG} , C^0_{acid} 에 대한 PLS1, PLS2, RPLS, APPLS 방법의 제곱근오차 평균값을 비교해 보면 PLS1, RPLS 방법에 비해 PLS2와 APPLS 방법의 예측 성능이 우수함을 확인할 수 있다. PLS1 방법은 초기의 PLS 모델에서 새로운 데이터가 들어와도 모델을 갱신하지 않았으므로 공정이 진행될수록 모델의 변화에 대응하지 못하여 다른 방법들에 비해 가장 낮은 예측 성능을 나타내며, PLS2 방법은 새롭게 얻어지는 데이터를 기존의 데이터와 합쳐 PLS 모델을 수립하므로 PLS1 방법보다 예측 성능이 매우 뛰어나다. 하지만 공정이 계속 진행될수록 PLS 모델에 사용되는 데이터의 수가 계속 늘어나게 되므로 PLS 모델의 크기가 점점 커지는 단점이 있다.

RPLS 방법은 기존 PLS 모델과 새롭게 얻어지는 데이터를 합쳐서 모델을 갱신하므로 PLS1 방법보다는 좋은 예측 성능을 나타낼 수 있고, 항상 PLS 모델에 필요한 데이터의 수가 일정하게 유지되는 장점을 가진다. 하지만 표

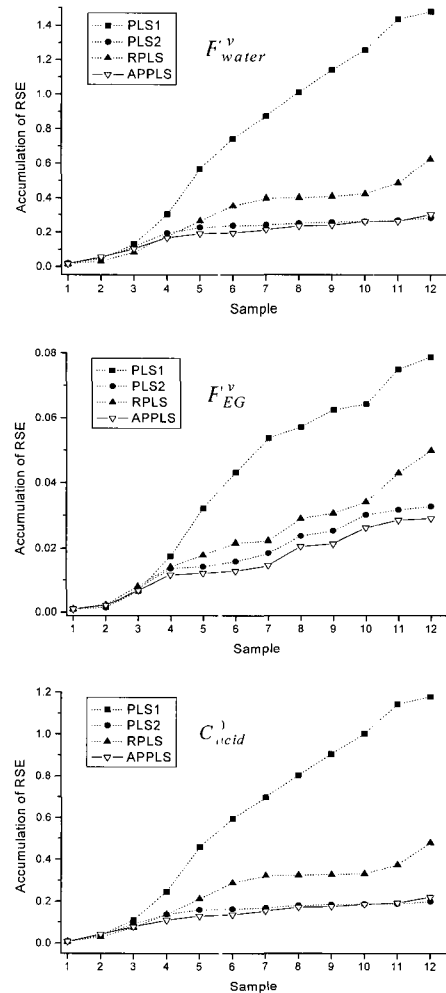


그림 2. PLS1, PLS2, RPLS, APPLS 방법의 제곱근오차의 누적 값 비교.

Fig. 2. Comparison with accumulated values of root square error by PLS1, PLS2, RPLS, APPLS.

2와 그림 2에서 볼 수 있듯이 PLS2 방법에 비해 예측 성능이 떨어진다.

APPLS 방법은 데이터의 추출 과정을 수행하면서 기존의 데이터와 새로운 데이터를 모두 이용해 모델을 갱신하므로 표 2와 그림 2에서 보듯이 PLS2 방법과 같이 뛰어난 예측 성능을 나타냄을 확인할 수 있다. 또한 모델에 사용되는 데이터의 개수도 RPLS 방법과 같이 항상 일정하게 유지되어 모델의 크기가 작게 유지된다는 장점을 가지며 예측 오차가 큰 데이터는 공정의 진행에 따라 제거되므로 RPLS 방법에 비해 갱신되는 모델의 안정성도 향상된다.

실제 화학공정에서는 단위 시간당 얻어지는 입출력 데이터의 개수가 많은 경우에 데이터 블록 단위로 PLS 모델을 갱신하는 방법을 사용한다[9]. 입출력 데이터의 개수가 많은 NIR Diesel Fuel Spectra data에 데이터 블록 단위로 PLS 모델을 갱신하는 block-wise APPLS 방법을 적용해 보았다.

IV. NIR Diesel Fuel Spectra data

최근 근적외선(NIR)을 이용한 실시간 모니터링 분야가 많은 주목을 받게 되어 실제 공정 데이터를 이용한 적응 예측 모델의 개발도 중요해지고 있다. 이러한 NIR Spectra 데이터를 이용한 연구는 참고문헌[6-8]에서 찾아볼 수 있다.

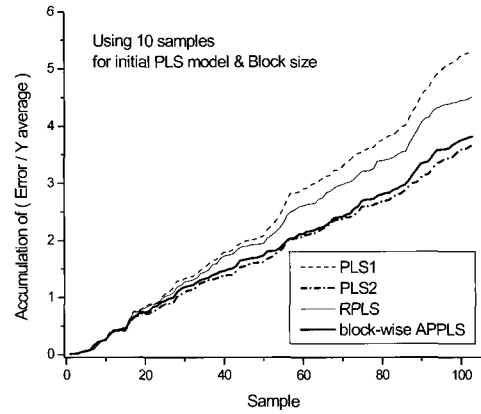
본 논문에서 제안한 block-wise APPLS 방법을 디젤유의 NIR Spectra 데이터에 적용해 보았다. 본 연구에서 사용된 NIR Diesel Fuel Spectra 데이터[17]는 Southwest Research Institute에서 측정되어진 것이고, 끓는점, 세탄가, 밀도, 어는점, total aromatics, 점도 등의 디젤유의 6가지 성질을 출력 변수로 가지고 있다. 본 논문에서는 6가지의 출력 변수 중 세탄가를 대상으로 block-wise APPLS 방법을 적용해 보았다. 세탄가는 디젤유의 규격으로서 가솔린의 앤티노크성(antiknock property)을 옥탄값으로 표시하는 것과 유사하게 시료와 동일한 착화성을 나타내는 표준연료에 혼합되어 있는 세탄의 부피 백분율을 가리킨다[11].

본 논문에서는 401 wavelength-wide NIR spectra에서 얻어지는 peak를 입력변수로 사용하였고, 세탄가를 출력변수로 하여 각 PLS 모델 갱신 방법을 적용해보았다. 사용한 샘플의 수는 총 113개이고, 이를 이용하여 첫 번째로 초기 PLS 모델에 사용되는 샘플의 수와 블록 단위를 샘플 10개로 지

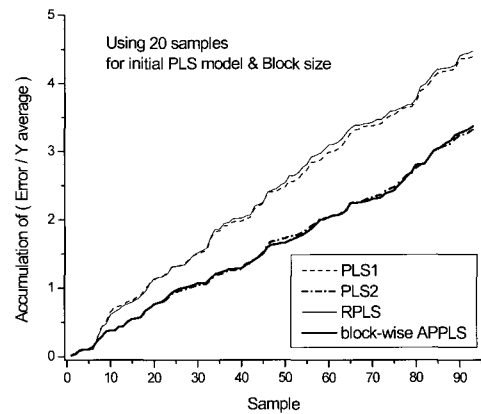
표 3. 모델과 블록에 사용된 데이터의 수의 변화에 따른 제곱근오차의 평균값 비교.

Table3. Comparison with root square error's average by changing # of samples for initial PLS model & block size.

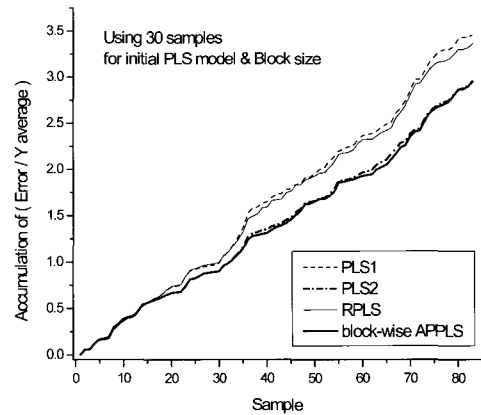
Average of Root Square Error	# of samples for Initial PLS model & block size		
	10 samples	20 samples	30 samples
PLS1	2.51559	2.32479	2.04060
PLS2	1.74824	1.75781	1.74457
RPLS	2.14762	2.37049	1.98714
block-wise APPLS	1.81892	1.78423	1.74265



(a)



(b)



(c)

그림 3. 모델과 블록에 사용된 샘플의 수의 변화에 따른 (root square error/Y average)의 누적 값 비교.

Fig. 3. Comparison with values of accumulated (root square error /Y average) by changing # of samples for initial PLS model & block size.

정하여 block-wise APPLS 방법을 적용하였고, 두 번째로 초기 PLS 모델에 사용되는 샘플의 수와 블록 단위를 20개의 샘플로, 마지막으로 30개의 샘플을 초기 PLS 모델과 블록 단위로 지정하고, 각각의 경우에 대해 나머지 샘플들을 블록 단위로 예측하고 모델을 갱신하였다. 초기 PLS 모델에 사용된 샘플의 수만큼 블록의 크기를 정하여 새로운 샘플이

들어왔을 때 데이터 추출 과정을 블록 단위로 실행하였다.

표 3에서는 초기 PLS 모델의 크기가 각각 샘플 10개, 20개, 30개일 때, PLS1, PLS2, RPLS 방법과 block-wise APPLS 방법으로 예측한 세탄가와 실제 세탄가와의 제공근오차의 평균값을 나타내었다. 표 3의 결과를 보면, 표 2의 결과와 같은 결과를 나타냄을 볼 수 있다. 즉, block-wise APPLS 방법이 RPLS 방법에 비해 평균 오차도 작게 나타나고 모델의 안정성도 뛰어나다.

그림 3은 각 방법에 의한 예측값 오차(root square error / y average)의 누적 값을 비교한 그래프이다. 초기 PLS 모델에 사용하는 샘플의 수와 모델 갱신에 사용되는 데이터 블록의 크기가 늘어날수록 block-wise APPLS 방법의 예측 성능이 점점 좋아지는 것을 알 수 있다. 그림 3(a), (b), (c)를 비교해보면 초기 PLS 모델의 크기와 데이터 블록의 크기가 커질수록 APPLS 방법이 PLS2 방법과 비슷한 예측 성능을 가지는 것을 볼 수 있다. 이는 모델 갱신에 사용되는 데이터 블록의 크기가 커질수록 APPLS 방법의 예측 성능이 더 좋아지는 것을 나타낸다.

V. 결론

본 논문에서는 화학공정의 품질 변수를 예측하는 PLS 모델로부터 예측이 나쁜 데이터를 제거하는 데이터 추출 과정을 도입하여 정상적인 혹은 비정상적인 화학공정의 변화에 적용하는 APPLS 방법을 제안하였고 PET 중합공정에 사용되는 직접 에스테르화 반응기의 운전데이터에 적용하여 예측 성능을 확인해 보았다. 제공근오차의 평균값과 누적 그래프를 이용해 PLS1, PLS2, RPLS 방법과 비교하여 APPLS 방법의 예측성능과 모델의 안정성을 확인해 보았다.

또한 대규모 화학공정과 같이 실시간으로 얻어지는 입출력 데이터의 양이 많고 새롭게 얻어지는 데이터의 양이 많아지는 경우에 대한 적응 예측 방법으로서 데이터 블록 개념을 도입한 block-wise APPLS 방법을 제안하였다. 이 방법을 NIR diesel fuel spectra 데이터에 적용하여 디젤유의 세탄가 예측에 사용해 보았다. APPLS 방법이 PLS1, PLS2, RPLS 등 다른 방법들보다 공정의 변화에 잘 적응한다는 것을 표 3, 그림 3의 결과를 통해 확인할 수 있었다. 초기 모델의 크기나 모델 갱신에 사용되는 데이터 블록의 크기를 샘플 10개, 20개, 30개로 증가시키기에 따라 block-wise APPLS 방법의 예측성능이 향상됨을 확인할 수 있었다. 만약 블록의 크기가 매우 커지게 되면 데이터 추출 과정에서 블록을 설정해주는 의미가 없어지게 된다. 즉, 블록의 크기가 매우 커지면 PLS2 방법과 같이 모델의 크기가 매우 커져 계산 시간이 늘어나게 되는 단점을 갖게 된다. 그러므로 대상이 되는 화학공정의 특성을 고려하여 적절한 블록의 크기를 설정해줘야 할 것이다.

참고문헌

[1] K. Helland, H. E. Bernsten, O. S. Borgen, and H. Marten, "Recursive algorithm for partial least squares regression," *Chemometrics and Intelligent Laboratory Systems*, vol. 14, pp. 129-137, 1991.

[2] S. Qin, "Recursive PLS algorithms for adaptive data modeling," *Computers Chem. Engng.*, vol. 22, no. 4/5, pp. 503-514, 1998.

[3] P. Geladi and B. R. Kowalski, "Partial least squares regression : A tutorial," *Analytica Chimica Acta*, 185, pp. 1-17, 1986.

[4] S. Wold, "Exponentially weighted moving principal component analysis and projection to latent structures," *Chemometrics and Intelligent Laboratory Systems*, vol. 23, pp. 149-161, 1994.

[5] N. B. Gallagher, B. M. Wise, S. W. Butler, D. D. White, and G. G. Barra, "Development and benchmarking of multivariate statistical process control tools for a semiconductor etch process: improving robustness through model updating," *Proceedings of the ADCHEM 97*, pp. 78-83, 1997.

[6] M. J. Kim, Y.-H. Lee, and C. h. Han, "Real-time classification of petroleum products using near-infrared spectra," *Computers Chem. Engng.*, vol. 24, pp. 513-517, 2000.

[7] M. V. Garcia-Mencia, J. M. Andrade, P. Lopez-Mahia, and D. Prada, "An empirical approach to update multivariate regression models intended for routine industrial use," *Fuel*, vol. 79, pp. 1823-1832, 2000.

[8] C. C. Felicio, L. P. Bras, J. A. Lopes, L. Cabrita, and J. C. Menezes, "Comparison of PLS algorithms in gasoline and gas oil parameter monitoring with MIR and NIR," *Chemometrics and Intelligent Laboratory Systems*, vol. 78, pp. 74-80, 2005.

[9] 윤경우, 이영학, 한중환, "블록들의 유사성을 고려한 Adaptive Block-wise PLS," *화학공학*, 제41권, 제5호, pp. 592-597, 2003.

[10] J-Y Kim, H-Y Kim, and Y-K Yeo, "Identification of kinetics of direct esterification reactions for PET synthesis based on a generic algorithm," *Korean J. of Chem. Eng.*, vol. 18, no. 4, pp. 432-441, 2001.

[11] 강익중, 정충섭, "국내 자동차용 경유의 성장과 세탄가," *화학공학*, 제38권, 제3호, pp. 384-354, 2000.

[12] 최주환, 정용진, 최웅수, 최영상, 권오관, "C-NMR에 의해 결정된 탄소 유흥별 구조적 조성으로부터 디젤 연료의 세탄가의 예측에 관한 연구," *J. of Korean Ind. & Eng. Chemistry* vol. 4, no. 4, pp. 709-714, 1993.

[13] W. Li, H. H. Yue, S. Valle-Cervantes, and S. J. Qin, "Recursive PCA for adaptive process monitoring," *Journal of Process Control*, vol. 10, pp. 471-486, 2000.

[14] B. S. Dayal and J. F. MacGregor, "Recursive exponentially PLS and its application to adaptive control and prediction," *Journal of Process Control*, vol. 7, no. 3, pp. 169-179, 1996.

[15] X. Wang, U. Kruger, and B. Lennox, "Recursive partial

least squares algorithms for monitoring complex industrial processes," *Control Engineering Practice*, vol. 11, pp. 613-632, 2003.

[16] 김성영, 이범석, 정창복, 최수형, "다변량 통계분석법

을 이용한 PET 중합공정 중 직접 에스테르화 반응기의 거동 및 생산제품 예측," 제어 · 자동화 · 시스템공학 논문지, 제11권, 제6호, pp. 550-557, 2005.

[17] <http://www.evriwire.com/Data/SWRI/index.html>.



김 성 영

1979년 6월 8일생. 2002년 경희대학교 환경응용화학부(공학사). 2004년(공학석사). 2004년~현재 경희대학교 화학공학과 박사과정. 관심분야는 공정시스템.



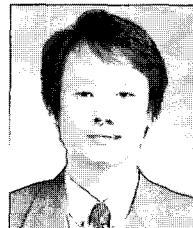
이 범 석

1961년 3월 22일생. 1984년 서울대학교 화학공학과 (공학사). 1986년(공학석사). 1992년 Purdue University (Ph.D.). 1993년~현재 경희대학교 환경응용화학대학 교수 및 산학협력기술연구원 상임연구원. 관심분야는 공정시스템.



정 창 복

1956년 3월 2일생. 1978년 서울대학교 화학공학과(공학사). 1982년 한국과학기술원 화학공학과(공학석사). 1988년 University of Michigan (Ph.D.). 1980년~현재 전남대학교 응용화학공학부 교수. 관심분야는 공정시스템.



최 수 형

1961년 1월 10일생. 1984년 서울대학교 화학공학과(공학사). 1986년(공학석사). 1990년 University of Missouri-Rolla(Ph.D.). 1993년~현재 전북대학교 화학공학부 교수. 관심분야는 공정시스템.