국소가중 서포트 벡터 머신을 이용한 품질 예측

조재규, 이동언, 송상옥, 윤인섭
서울대학교 응용화학부

Quality estimation Using Support Vector Machine based on locally weighted regression

Jae Kyu Cho, Dong Eon Lee, Sang Ok Song and En Sup Yoon
School of Chemical Engineering, Seoul National University

1. 서 론

가스 및 유류를 이용하는 화학공장의 경우, 비용이나 신뢰성 등의 요인으로 인하여 중요한 품질 변수를 실시간에 측정하기 어려운 경우가 많다. 이러한 문제점은 실시간으로 측정이 가능한, 우리가 관심을 갖는 영역내의 변수들의 인과관계에 바탕을 둔 추론 모델을 이용할 경우 해결될 수 있다. 추론 모델의 개발에 있어서, 이론적 모델과 경험적 모델이 사용될 수 있으나, 이론적 모델의 경우 특정 한계를 가지는 경우가 많아서 경험적 모델을 이용하여 개발하는 경우가 많다. 경험적 모델의 경우 주로 다변량 통계분석이나 인공신경망을 이용하여 개발된다.

다변량통계 방법으로 많이 이용되는 주성분분석(PCA)나 부분최소자승법(PLS)의 경우 많은 관심을 끌어 왔으나 비선형 데이터에 대해서 필요이상으로 데이터에 맞추는 경향이 있어 일반화(generalization)등의 여러 한계점을 드러내었다.
본 연구에서는 이러한 한계를 극복할 방법으로 Vapnik에 의해 개발된 서포트 벡터 머신(SVM)에 기초를 한 추론 모델을 이용하였다.

2. 이 론

2. 1. 서포트 벡터 머신

SVM(Support Vector Machine)은 통계적 학습 이론에서 유도한 학습 바이어스를 이용한 학습알고리즘으로 훈련되는 고차원의 feature space상의 선형식의 가상 공간을 이용한다. 이 학습 전략은 Vapnik에 의해 소개된 강력한 통계방법으로써 이미 폭 넓은 분야에서 다른 시스템보다 뛰어난 능력을 발휘하고 있다.

SVM은 분류(Classification)문제를 해결하기 위하여 고안된 것이지만, loss Function을 이용할 경우 회귀(Regression)문제에도 적용될 수 있다.

회귀문제는 주어진 입력 공간에서 출력공간으로 적합한 매핑(mapping) 한수를 찾아내는 것이다. 출력값과 실제값의 차이인 잔차(residual)이 SVM의 경우에는 loss function을 이용하여 결정된다.

모델의 학습 데이터로서 $\{(x_1, y_1), \ldots, (x_n, y_n)\} \subset X \times \vec{R}$ 을 고려해 보자.
이때 모델함수는 다음과 같다.

$$f(x) = \sum_{i=1}^{n} w_i \phi(x_i) + b \qquad \text{식 (1)}$$

여기서 $\phi_i(x)$는 입력공간에서 feature space로 비선형적으로 매핑된 점이다. 회귀의 목적은 위의 함수에 따른 risk function을 아래와 같이 최소화 하는 것이다.

$$\frac{1}{2}|w|^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*) \qquad \text{식 (2)}$$

제약조건은 아래와 같다.

$$\begin{cases} y_i - w\phi(x_i) - b_i \le \varepsilon + \xi_i \\ w\phi(x_i) + b_i - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \quad \xi_i^* \ge 0 \end{cases}$$

식 (3)

식 (2)에서 상수 C는 함수 f의 모델의 복잡도와 학습 데이터 적합도 사이의 절충정도(trade off)를 결정하는 것이다.

위의 식들은 loss function중에서 ε-insensitive loss function을 사용하여 기술된 것으로 아래와 같다.

$$|\xi|_\varepsilon = \begin{cases} 0 & \text{if } |\xi| \le \varepsilon \\ |\xi| - \varepsilon & \textit{otherwise} \end{cases} \qquad \text{식 (4)}$$

이러한 제한된 최적화는 아래와 같은 primal variables Lagrangian, $L_p(w, \xi, \xi^*)$을 풀면 얻어낼 수 있다.

$$L_P(w, b, \xi, \xi^*, \alpha_i, \alpha_i^*, \beta_i, \beta_i^*) \qquad \text{식(5)}$$

$$= \frac{1}{2}|w|^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*)$$

$$- \sum_{i=1}^{n} \alpha_i^* (y_i - w\phi(x_i) - b - \varepsilon + \xi_i^*)$$

$$- \sum_{i=1}^{n} \alpha_i (w\phi(x_i) + b - y_i + \varepsilon + \xi_i) - \sum_{i=1}^{n} (\beta_i^* \xi_i^* + \beta_i \xi_i)$$

Lagrangian $L_P(w, b, \xi, \xi^*, \alpha_i, \alpha_i^*, \beta_i, \beta_i^*)$는 primal variable w, b, ξ, ξ^*에 관하여 최소화되어야 하며, 동시에 non-negative Lagrangian multiplier $\alpha, \alpha^*, \beta, \beta^*$에 관하여 최소화되어야 한다. 식 (2)는 primal space뿐만 아니라 dual space에서도 해결할 수 있다.

dual space에서 해결하기 위하여 Karush-Khun-Tucker(KKT) 조건을 적용할 때 dual variables Lagrangian $L_d(\alpha, \alpha^*)$을 최대화시켜야 한다.

$$L_d(\alpha, \alpha^*) \qquad \text{식 (6)}$$

$$= -\varepsilon \sum_{i=1}^{n} (\alpha_i^* + \alpha_i) + \sum_{i=1}^{n} (\alpha_i^* - \alpha_i) y_i$$

$$-\frac{1}{2}\sum_{i,j=1}^{n}(a_i^*-a_i)(a_j^*-a_j)K(x_i,\ x_j)$$

여기서 제약조건은 아래와 같다.

$$\sum_{i=1}^{n}a_i^*=\sum_{i=1}^{n}a_i$$

$$\text{여기서}\quad 0\le a_i^*,\ a_i\le C, \qquad i=1,\ldots,n$$

식 (6)에서 $K(x_i, x_j)$는 feature space로 매핑된 점 $\phi(x_i)$, $\phi(x_j)$의 내적인 Kernel function이다. Kernel function의 사용은 데이터를 feature space로 매핑이 가능하게 해주며 그러한 공간에서 선형 학습을 가능하게 해 주며, 이때 모델함수는 식 (1)은 아래와 같이 바뀐다.

$$f(x,\ a_i,\ a_i^*)=\sum_{i=2}^{n}(a_i-a^*)K(x_i,\ \ x_j)+b \qquad \text{식 (7)}$$

2.2. 국소가중법

국소가중법은 모델을 근처의 데이터에 회귀하는 방법이다. 학습과정에서 각 데이터에 인접한 데이터에 가중치를 주어 추론모델을 형성한다.
각 점에서 이웃한 점들은 떨어진 거리에 따라 가중치를 부여하며, 이때 사용하는 함수로는 Gaussian kernel, exponential kernel, quadratic kernel, tricubic kernel 등이 있으며 이중 tricubic kernel을 이용하였다.

Tricubic kernel은 Cleveland에 의하여 처음 도입이 되었다. 유클리드 거리(Euclidian distance; ρ)를 이용여 국소 가중법을 실행하며, 다음과 같이 표현된다.

$$W(u)=\begin{cases}(1-u^3)^3 & \text{if}\quad |u|\le 1\\ 0 & otherwise\end{cases} \qquad \text{식 (8)}$$

이때 weight function은 아래와 같다.

$$w_i=W(\rho(x\ ,\ x_i)/d(x)) \qquad \text{식 (9)}$$

3. 국소 가중법을 이용한 서포트벡터 머신

본 연구에서는 표준 서포트 벡터 머신개념에 국소가중 회귀 개념을 추가하여 추론모델을 개발하였다. 식 (2)의 상수 C의 경우 모델의 복잡도와 학습데이터에 적합도의 절충 상수의 역할을 하지만, 모든 학습 데이터들이 동일하게 추론모델에 기여를 한다. 그러나 실제 데이터들은 산재되어 있고 예측값과 가까운 입력값일 수 록 추론모델에 더 많은 기여를 해야 하는 것이 타당하다. 따라서 미리 정해진 상수 C보다 입력데이터간의 거리 함수에 따른 함수값 C를 사용하여 학습을 진행하는 것이 제안된 모델이다. 제안된 추론모델에 따른 목적함수를 수식으로 표현하면 아래와 같다.

$$\frac{1}{2}|w|^2+\sum_{i=1}^{n}C_i(\xi_i+\xi_i^*), \qquad \text{식 (10)}$$

$$C_i=w_i(x_0)\times C$$

추론모델을 얻기 위해서는 변화된 제한조건에 따라 아래의 dual form function을 최소화 시키는 것이다.

$$L_d(\alpha,\ \alpha^*) \qquad\qquad \text{식 (11)}$$

$$= -\varepsilon \sum_{i=1}^{n} (\alpha_i^* + \alpha_i) + \sum_{i=1}^{n} (\alpha_i^* - \alpha_i) y_i$$

$$- \frac{1}{2} \sum_{i,j=1}^{n} (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) K(x_i,\ x_j)$$

여기서 제약조건은 아래와 같다.

$$\sum_{i=1}^{n} \alpha_i^* = \sum_{i=1}^{n} \alpha_i$$

$$\text{여기서}\ \ 0 \le \alpha_i^*,\ \alpha_i \le C_i, \qquad\qquad i = 1, \ldots, n$$

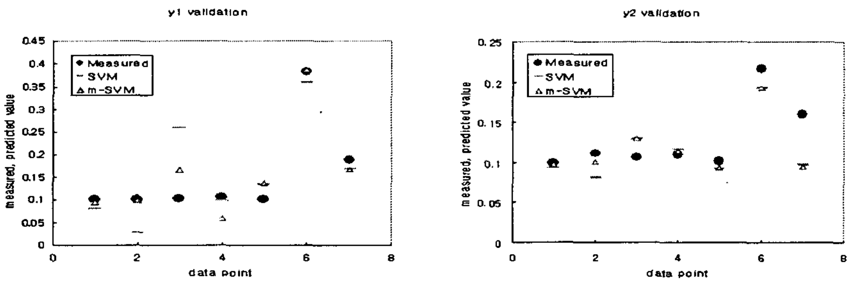
4. 사례연구

제안된 추론 모델을 고분자 테스트 공정의 불규칙적으로 산재된 데이터의 비선형 추론모델개발을 위하여 적용해 보았다. 학습데이터는 10개의 조작변수들과 4개의 출력변수들로 이루어져 있다. 국소가중법 적용을 위하여 유클리드 거리와 tricubic function을 사용하였다. 추론모델의 변수로서 C=5000, ε=0을 사용하였다.

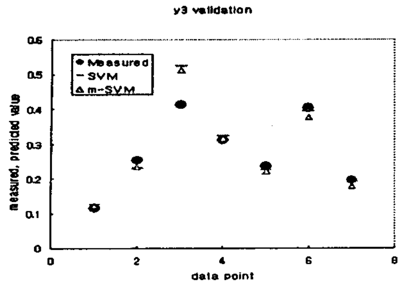
표 1. 국소가중 서포트 벡터 머신(m-SVM)과 표준 SVM의 평균제곱근 오차 (Root mean square error) 비교

	y1	y2	y3	y4
표준 SVM	0.0451	0.0194	0.0293	0.0260
m-SVM	0.0226	0.0188	0.0269	0.0224

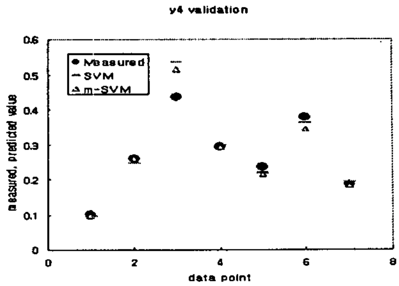
위의 표를 보면 전체적으로 국소가중 서포트 벡터 머신의 오차가 작은 것으로 나타나, 기존의 서포트 벡터 머신보다 성능이 뛰어남을 알 수 있다.



(a) 출력변수 1 (b) 출력변수 2



(c) 출력변수 3



(d) 출력변수 4

Fig 1. 제안된 추론모델을 이용한 출력값과 실제값의 비교

제안된 추론모델을 이용하여 출력값을 예측한 데이터와 실제값을 도시해본 결과 전반적으로 제안된 국소가중모델이 표준 서포트 벡터 머신보다 성능이 우월함을 나타내었다.

5. 결 론

본 연구에서 기존의 표준 서포트 벡터 머신에 국소가중법(locally weighted regression)의 개념을 도입하여 국소가중 서포트 벡터 머신을 제안하였다. 제안된 모델을 고분자 테스트 공정에 적용하여 본 결과 기존의 방법보다 우월한 성능을 나타내었다. 사례연구 결과, 예측값과 가까운 데이터일수록 가중치를 높게 부여한 국소가중법 적용의 타당성을 입증하였으며, 제안된 국소가중 서포트 벡터 머신의 향상된 예측력을 나타내었다.

6. 감사

본 연구는 서울대학교 화학공정신기술연구소와 자동화 시스템 공동연구소 및 교육인적자원부의 Brain Korea 21사업의 지원에 의한 것입니다.

7. 참고문헌

1. W. Cherkassky and F. Mulier, "Learning from data", *John Wiley & Sons*, 1998.

2. Cleveland, R. J. and McArthur, J. M. "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting", *Journal of American Statistical Association*, 83, 596-610.

3. N. Cristianini and J. Shawe-Tayler, *An introduction to Support Vector Machine*, Cambridge univ. press, 2001.

4. V. Kecman, *Learing and Soft Computation*, MIT press, 2001.

5. Vapnik, V., *The Nature of Statistical Learning Theory*, Springer, 1998.