

Application of Asymmetric Support Vector Regression Considering Predictive Propensity

Dongju Lee[†]

Department of Industrial & Systems Engineering, Kongju National University

예측성향을 고려한 비대칭 서포트벡터 회귀의 적용

이 동 주[†]

공주대학교 산업시스템공학과

Most of the predictions using machine learning are neutral predictions considering the symmetrical situation where the predicted value is not smaller or larger than the actual value. However, in some situations, asymmetric prediction such as over-prediction or under-prediction may be better than neutral prediction, and it can induce better judgment by providing various predictions to decision makers. A method called Asymmetric Twin Support Vector Regression (ATSVR) using TSVR(Twin Support Vector Regression), which has a fast calculation time, was proposed by controlling the asymmetry of the upper and lower widths of the ϵ -tube and the asymmetry of the penalty with two parameters.

In addition, by applying the existing GSVQR and the proposed ATSVR, prediction using the prediction propensities of over-prediction, under-prediction, and neutral prediction was performed. When two parameters were used for both GSVQR and ATSVR, it was possible to predict according to the prediction propensity, and ATSVR was found to be more than twice as fast in terms of calculation time. On the other hand, in terms of accuracy, there was no significant difference between ATSVR and GSVQR, but it was found that GSVQR reflected the prediction propensity better than ATSVR when checking the figures. The accuracy of under-prediction or over-prediction was lower than that of neutral prediction. It seems that using both parameters rather than using one of the two parameters (p_1, p_2) increases the change in the prediction tendency. However, depending on the situation, it may be better to use only one of the two parameters.

Keywords : Support Vector Regression, Machine Learning, Predictive Propensity

1. 서 론

대부분의 머신러닝을 이용한 예측은 예측값이 실제값보다 작거나 크지 않은 대칭적인 상황을 고려한 중립예측이다. 하지만, 일부 상황에서는 과대예측이나 과소예측과 같은 비대칭적인 예측이 중립예측보다 나올 수도 있으며, 의사결정자에게 다양한 예측들을 제공하므로 써 좀 더 나

은 판단을 유도할 수 있다.

Wu et al.[10]은 SVR(Support Vector Regression)[6]에서 회귀선의 아래와 위의 폭(ϵ -tube)을 벗어나는 경우에 부과되는 벌점의 기울기의 비대칭으로만 비대칭성을 조절하는 비대칭적 SVR을 제안하였다. 또한, 이 방법을 적용하여 호주의 New South Wales 지역의 2019년 2월 1일부터 3월 30일까지의 전력사용관련 실데이터를 활용하여 과소예측(underprediction)하였을 때 약 42.19%~57.39% 정도의 비용절감이 발생한다고 하였다. 예측성향을 이용한 비대칭적 SVR의 활용에 대한 연구는 현재 연구가 진행되고 있는

Received 24 February 2022; Finally Revised 12 March 2022;
Accepted 14 March 2022

[†] Corresponding Author : djlee@kongju.ac.kr

분야라 활용사례가 거의 없다.

SVR의 정확성은 해치지 않으면서 계산시간을 향상한 TSVR(Twin Support Vector Regression)에 대한 연구들이 행해지고 있다.

TSVR(Twin Support Vector Regression)은 Peng[2]이 제안하였는데 2개의 제약식을 가진 1개의 큰 2차계획법문제를 푸는 기존의 SVR과 달리 1개의 제약식을 각각 가진 2개의 작은 2차계획법문제로 푸는 방법이다. 1개의 큰 2차계획법문제보다 2개의 작은 2차계획법문제로 나누어 풀기에 TSVR이 일반적인 SVR보다 약 4배 빠르게 해를 구한다고 하였다.

Singh et al.[5]은 TSVR에서 훈련데이터 전부를 사용하는 것이 아니라 그 일부를 무작위로 선택하는 Rectangular Kernel이라는 개념을 이용하는 RTSVR(Reduced Twin Support Vector Regression)을 제안하였다. RTSVR은 TSVR보다 빠른 시간 내에 TSVR과 비슷한 수준의 해를 제공하는 것을 실험을 통해 입증하였다. Xu and Wang[12]은 과적합(overfitting)을 어느 정도 해결하기 위해 데이터의 위치에 따른 벌점(penalty)을 기존보다 더 세분화하는 WTSVR(Weighted Twin Support Vector Regression)을 제안하였다.

Wang and Xu[8]은 TSVR의 쌍대문제에서 0과 1을 가지는 항들을 찾아내고 0의 항을 가지는 샘플들을 제거하는 방법(safe screen rule)을 적용한 후에 데이터를 훈련시켜, TSVR의 계산시간을 더욱 단축시키고 TSVR과 동일한 해를 제공하는 SSR-TSVR을 제안하였다.

한편, 비대칭적인 SVR에 대한 연구도 진행되고 있다. Xu et al.[11]은 비대칭적 ν -SVR을 제안하였다. SVR에서는 ϵ 은 사용자가 미리 정하는 매개변수이지만, ν -SVR에서는 벌점이 없는 데이터의 비율(ν)을 사용자가 미리 정하면 SVR모형을 풀 때 구해지는 값이 ϵ 이다. 1개의 매개변수를 이용하여 ϵ -tube의 상하폭의 비대칭성과 벌점의 비대칭성을 모두 조절하였다.

Lee and Choi[4]는 ϵ -tube의 상하폭의 비대칭성과 벌점의 비대칭성을 각각 조절하는 2개의 매개변수를 고려하는 손실함수를 제안하고 이를 이용한 GSVQR(Generalized Support Vector Quantile Regression)을 제안하였다. 조절하는 2개의 매개변수를 도입하므로 써 예측성향에 따른 좀 더 미세한 조절이 가능하였다. 하지만, 실험을 통한 GSVQR의 효용성을 입증하는 것이 부족하였다.

본 연구에서는 GSVQR의 손실함수를 이용하여 2개의 매개변수로 ϵ -tube의 상하폭의 비대칭성과 벌점의 비대칭성을 각각 조절하고 계산시간이 빠른 TSVR을 이용한 ATSVR(Asymmetric Twin Support Vector Regression)이라는 기법을 제안하였다. 또한, 기존의 GSVQR과 제안한 ATSVR을 적용하여 과대예측, 과소예측, 중립예측의 예측

성향을 이용한 예측을 실시하였다. 테스트함수와 실제사례인 공구마모량 예측사례에 매개변수를 이용한 예측성향별 예측을 실시하고, 이들의 유효성, 정확성, 계산시간 등을 파악하였다.

이어지는 제2장에서는 기존 SVR기법들에 대해 설명하고, 제3장에서는 ATSVR을 제안하였다. 제4장에서는 실험을 통해 GSVQR과 ATSVR을 적용하여 여러 가지 특징을 살펴보았다. 마지막으로 제5장에서는 결론을 도출하고 추후 연구과제들을 제시하였다.

2. 기존 SVR

제2장에서는 소프트마진 SVR, GSVQR, 그리고 TSVR을 설명한다.

2.1 소프트마진(Soft margin) SVR

소프트마진 SVR에서 사용되는 손실함수는 식 (1)과 같다.

$$L_{\epsilon}^1(u) = \begin{cases} 0, & -\epsilon \leq u \leq \epsilon \\ u - \epsilon, & u > \epsilon \\ -u - \epsilon, & u < -\epsilon \end{cases} \quad (1)$$

이러한 ϵ -둔감함수를 바탕으로 한 Soft Margin의 수학모형을 나타내면 식 (2)~식 (5)와 같다.

$$\text{Min } \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l (\xi_i + \eta_i) \quad (2)$$

$$\text{s.t. } y_i - \omega^T x_i - b \leq \epsilon + \xi_i \quad (3)$$

$$\omega^T x_i + b - y_i \leq \epsilon + \eta_i \quad (4)$$

$$\xi_i, \eta_i \geq 0 \quad (5)$$

위의 원문제에서 ξ_i, ξ_i^* 를 구하기 어려우므로 원문제를 라그랑주 쌍대 문제(Dual Lagrangian)로 나타내면 식 (6)~식 (8)과 같다.

$$\text{Min } \frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)(x_i, x_j) \quad (6)$$

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

$$\text{s.t. } \sum_{i=1}^l (\alpha_i^* - \alpha_i) = 0 \quad (7)$$

$$0 \leq \alpha_i, \alpha_i^* \leq C \quad (8)$$

여기서 α_i, α_i^* 는 위 수학모형의 제약식들의 라그랑주(Lagrangian) 변수이다.

2.2 GSQVR(General Support Vector Quantile Regression)

Lee and Choi[5]가 제안한 GSQVR에서의 ϵ -둔감함수 ($L_\epsilon^2(u)$)는 식 (9)와 같다.

$$L_\epsilon^2(u) = \begin{cases} 0, & -(1-p_1)\epsilon \leq u \leq p_1\epsilon \\ p_2(u-p_1\epsilon), & u > p_1\epsilon \\ -(1-p_2)(u+(1-p_1)\epsilon), & u < -(1-p_1)\epsilon \end{cases} \quad (9)$$

$L(u)_\epsilon^2$ 를 제약식으로 나타내기 위해 다시 정리하면

$$p_2(u-p_1\epsilon) \leq \xi_i$$

$$p_2(y_i - \omega^T x_i - b - p_1\epsilon) \leq \xi_i$$

$$y_i - \omega^T x_i - b \leq p_1\epsilon + \frac{1}{p_2}\xi_i$$

동일한 방법으로

$$-(1-p_2)(u+(1-p_1)\epsilon) \leq \eta_i$$

$$(1-p_2)(\omega^T x_i + b - y_i - (1-p_1)\epsilon) \leq \eta_i$$

$$\omega^T x_i + b - y_i \leq (1-p_1)\epsilon + \frac{1}{1-p_2}\eta_i$$

그러므로 수학모형은 식 (10)~식 (13)과 같다.

$$\text{Min} \quad \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l (\xi_i + \eta_i) \quad (10)$$

$$\text{s.t.} \quad y_i - \omega^T x_i - b \leq p_1\epsilon + \frac{1}{p_2}\xi_i \quad (11)$$

$$\omega^T x_i + b - y_i \leq (1-p_1)\epsilon + \frac{1}{1-p_2}\eta_i \quad (12)$$

$$\xi_i, \eta_i \geq 0 \quad (13)$$

식 (10)은 Soft Margin의 목적식처럼 첫 번째 항은 실제값 (y_i)과 추정값 $y = \omega_1^T x + b \pm \epsilon$ 의 차이의 제곱을 나타내며 두 번째 항은 오차변수(ξ_i, η_i)의 합을 나타내는데 이들을 최소화하고자 한다. 식 (11)은 실제값(y_i)이 ϵ -tube의 상한(upper bound)인 $\omega^T x_i + b + p_1\epsilon$ 보다 클지라도 $\frac{1}{p_2}\xi_i$ 보다는 같거나 작게 하는 제약식이다. 식 (12)는 ϵ -tube의 하한(lower bound)을 나타내는데 실제값(y_i)이 $\omega^T x_i + b - (1-p_1)\epsilon$ 보다 작을지라도 최소한 $-\frac{1}{1-p_2}\eta_i$ 보다는 같거나 크게 하는 제약식이다. 마지막 제약식은 별점 ξ_i, η_i 이 비음(non-negativity)임을 나타낸다.

p_1 을 이용하여 ϵ -tube의 폭의 좌우 비대칭 정도를 조절할 수 있으며, 허용한계인 $\pm 0.5\epsilon$ 을 벗어나는 ξ_i, η_i 의 값에 따른 벌금(penalty)의 기울기는 p_2 에 의해 비대칭 정도를

조절할 수 있다.

목적식을 Max에서 Min문제로 변환하여 라그랑주 쌍대 문제로 나타내면 식 (14)~식 (17)과 같다.

$$\text{Min}_{\alpha_i, \alpha_i^*} = \frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*)(x_i, x_j)(\alpha_i - \alpha_i^*) \quad (14)$$

$$+ \sum_{i=1}^n (\alpha_i p_1 \epsilon + \alpha_i^* (1-p_1)\epsilon) - \sum_{i=1}^n (\alpha_i - \alpha_i^*) y_i$$

$$\text{s.t.} \quad \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \quad (15)$$

$$0 \leq \alpha_i \leq p_2 C \quad \forall i=1, \dots, n \quad (16)$$

$$0 \leq \alpha_i^* \leq (1-p_2)C; \quad \forall i=1, \dots, n \quad (17)$$

예측값이 실제값보다 큰 과대예측, 예측값이 실제값보다 작은 과소예측, 중립예측을 위한 p_1, p_2 의 범위를 다음과 같이 조절할 수 있다.

- 과대예측 : $0 \leq p_1 \leq 0.5, 0.5 \leq p_2 \leq 1$
- 과소예측 : $0.5 \leq p_1 \leq 1$ 이고 $0 \leq p_2 \leq 0.5$
- 중립예측 : $p_1 = p_2 = 0.5$

2.3 TSVR(Twin Support Vector Regression)

Peng[2]이 제안한 TSVR은 식 (18)~식 (20), 식 (21)~식 (23)의 2개의 수학모형으로 이뤄져 있다.

$$\text{Min} \quad \frac{1}{2} \|Y - e\epsilon_1 - (Aw_1 + eb_1)\|^2 + C_1 e^T \xi \quad (18)$$

$$\text{s.t.} \quad Y - (Aw_1 + eb_1) \geq e\epsilon_1 - \xi \quad (19)$$

$$\xi \geq 0 \quad (20)$$

$$\text{Min} \quad \frac{1}{2} \|Y - e\epsilon_2 - (Aw_2 + eb_2)\|^2 + C_2 e^T \eta \quad (21)$$

$$\text{s.t.} \quad (Aw_2 + eb_2) - Y \geq e\epsilon_2 - \eta \quad (22)$$

$$\eta \geq 0 \quad (23)$$

여기서 e 는 모든 요소가 1인 벡터이며, $y_i \in R$ 일 때 Y 는 반응벡터로 $Y = (y_1; y_2; \dots; y_l)$ 이다. A_i 는 n 차원의 실수로 구성된 행벡터로 l 개가 있고 $A = (A_1; A_2; \dots; A_l)$ 이다. 그러므로 학습데이터는 (A, Y) 이다. $C_1, C_2 > 0, \epsilon_1, \epsilon_2 \geq 0$ 은 사용자가 미리 값을 정하는 매개변수(parameter)이며 ξ, η 은 여유변수(slack variable)이다.

TSVR은 ϵ 둔감 하한 회귀식 $f_1(x) = w_1^T x + b_1$ 과 ϵ 둔감 상한 회귀식 $f_2(x) = w_2^T x + b_2$ 을 구한다.

식 (18)과 식 (21)의 첫 번째 항은 함수 $y = w_1^T x + b_1 + \epsilon_1$ 혹은 $y = w_2^T x + b_2 - \epsilon_2$ 과 학습자료의 실제값(y_i)의 차이의 제곱이다. 두 번째 항은 오차변수(ξ, η)의 합인데 학습자료

를 과적합(overfitting)하는 것을 예방하기 위해 필요하다.

식 (19)와 식 (22)는 추정함수 $f_1(x)$, $f_2(x)$ 이 학습자료의 실제값(y_i)과의 거리가 ϵ_1 혹은 ϵ_2 보다는 같거나 크게 한다. 즉, 실제값은 ϵ 둔감 하한 회귀식보다 최소한 ϵ_1 보다 커야 하며, ϵ 둔감 상한 회귀식보다 최소한 ϵ_2 이상 작아야 한다. 여유변수 ξ , η 은 학습자료의 실제값과 추정함수의 거리가 ϵ_1 혹은 ϵ_2 보다 가까운 경우에 대한 오차의 측정값이다[2]. 식(20)과 식 (23)은 ξ , η 이 각각 비음임을 나타낸다.

TSVR은 2개의 2차계획법 문제로 표현되며, 각 2차계획법은 일반적인 SVR 모형에 속한 2종류의 제약식 중에서 1종류의 제약식만을 사용하여 하한 회귀식 혹은 상한 회귀식을 구하는데 사용된다. 이로 인해 Peng[2]은 TSVR이 일반적인 SVR보다 약 4배 빠르게 해를 구한다고 하였다.

$G=[Ae]$, $f=Y-e\epsilon_1$, $u=[w_1 \ b_1]^T$ 라고 할 때 식 (18)~식 (20)의 원문제를 라그랑주 쌍대 문제로 나타내면 식(24), 식 (25)와 같다.

$$\text{Max} -\frac{1}{2}\alpha^T G(G^T G)^{-1} G^T \alpha \quad (24)$$

$$+ f^T G(G^T G)^{-1} G^T \alpha - f^T \alpha$$

$$\text{s.t } 0 \leq \alpha \leq C_1 e \quad (25)$$

$h=Y+e\epsilon_2$ 라 할 때, 식 (21)~식 (23)의 원문제를 라그랑주 쌍대 문제로 나타내면 식 (26), 식 (27)과 같다.

$$\text{Max} -\frac{1}{2}\gamma^T G(G^T G)^{-1} G^T \gamma \quad (26)$$

$$- h^T G(G^T G)^{-1} G^T \gamma + h^T \gamma$$

$$\text{s.t } 0 \leq \gamma \leq C_2 e \quad (27)$$

3. ATSVR(Asymmetric Twin Support Vector Regression)

본 논문에서 제안하는 ATSVR의 수학모형은 식 (28)~식 (30), 식 (31)~식 (33)과 같다.

$$\text{Min } \frac{1}{2} \|Y - e\epsilon_1 - (Aw_1 + eb_1)\|^2 + C_1 e^T \xi \quad (28)$$

$$\text{s.t. } Y - (Aw_1 + eb_1) \geq p_1 e\epsilon_1 - \frac{1}{1-p_2} \xi \quad (29)$$

$$\xi \geq 0 \quad (30)$$

$$\text{Min } \frac{1}{2} \|Y - e\epsilon_1 - (Aw_1 + eb_1)\|^2 + C_2 e^T \eta \quad (31)$$

$$\text{s.t. } (Aw_2 + eb_2) - Y \geq (1-p_1)e\epsilon_2 - \frac{1}{p_2} \eta \quad (32)$$

$$\eta \geq 0 \quad (33)$$

식 (28)과 식 (31)은 목적식인데 첫 번째 항은 실제값(Y)과 추정값의 차이의 제곱이고 두 번째 항은 오차변수의 합인데 이들을 최소화하고자 한다. 식 (29)는 실제값(Y)과 하한(lower bound)인 $\omega^T x + b_1 + p_1 \epsilon_1$ 과의 차이가 최소한 $-\frac{1}{1-p_2} \xi$ 보다는 크거나 같아야 한다. 또한, ξ 은 목적식에서 최소화해야 하는 일종의 별점이다. 식 (30)은 별점(ξ)이 비음(non-negativity)임을 나타낸다.

식 (32)는 실제값(Y)는 상한(upper bound)인 $\omega^T x + b_2 + (1-p_1)\epsilon_2$ 의 차이가 최소한 $\frac{1}{p_2} \eta$ 보다는 작거나 같아야 한다. 또한, η 은 목적식에서 최소화해야 하는 일종의 별점이다. 식 (33)은 별점(η)이 비음(non-negativity)임을 나타낸다.

GSVQR과 동일하게 과대예측, 과소예측, 중립예측을 위한 p_1, p_2 의 범위를 다음과 같이 조절할 수 있다.

- 과대예측 : $0 \leq p_1 \leq 0.5, 0.5 \leq p_2 \leq 1$
- 과소예측 : $0.5 \leq p_1 \leq 1$ 이고 $0 \leq p_2 \leq 0.5$
- 중립예측: $p_1 = p_2 = 0.5$

식 (31)~식 (33)의 원문제를 라그랑주쌍대 문제로 나타내면 식 (34), 식 (35)와 같다. 자세한 유도과정은 Appendix에 주어져 있다.

$$\text{Max} -\frac{1}{2}\gamma^T G(G^T G)^{-1} G^T \gamma \quad (34)$$

$$- h^T G(G^T G)^{-1} G^T \gamma + (h^{T-p_1} \epsilon_2 e^T) \gamma$$

$$\text{s.t } 0 \leq \gamma \leq p_2 C_2 e \quad (35)$$

같은 방법으로 식 (28)~식 (30)에 대한 라그랑주쌍대문제를 구하면 식 (36), 식 (37)과 같다.

$$\text{Max} -\frac{1}{2}\alpha^T G(G^T G)^{-1} G^T \alpha \quad (36)$$

$$+ f^T G(G^T G)^{-1} G^T \alpha - (f^T + (1-p_1)\epsilon_1 e^T) \alpha$$

$$\text{s.t. } 0 \leq \alpha \leq (1-p_2) C_1 e \quad (37)$$

위 라그랑주 쌍대문제들을 각각 풀어 우리가 구하고자 하는 회귀식을 구하면 식 (38)과 같다.

$$f(x) = \frac{1}{2} (f_1(x) + f_2(x)) \quad (38)$$

$$= \frac{1}{2} (w_1 + w_2)^T x + \frac{1}{2} (b_1 + b_2) \quad (38)$$

비선형 추정을 위해서, 비선형 지도(nonlinear map) $\phi: R^n \rightarrow H$ 가 사용되는데, 여기서 H 는 특징공간(feature space)이다.

H 에서 변환 함수끼리의 스칼라곱인 $\phi(u)^T \phi(v) = K(u^T, v)$ 들을 모아놓은 집합을 커널(kernel)이라고 한다. 이때, 추정하고자 하는 회귀식은 $f_1(x) = K(x, A^T)w_1 + b_1$, $f_2(x) = K(x, A^T)w_2 + b_2$ 이다. 비선형인 경우의 ATSVR의 수학적모형은, 식 (39)~식 (41), 식 (42)~식 (44)와 같다.

$$\text{Min } \frac{1}{2} \|Y - e\epsilon_1 - (Aw_1 + eb_1)\|^2 + C_1 e^T \xi \quad (39)$$

$$\text{s.t. } Y - (K(A, A^T)w_1 + eb_1) \geq p_1 e\epsilon_1 - \frac{1}{1-p_2} \xi \quad (40)$$

$$\xi \geq 0 \quad (41)$$

$$\text{Min } \frac{1}{2} \|Y - e\epsilon_2 - (Aw_2 + eb_2)\|^2 + C_2 e^T \eta \quad (42)$$

$$\text{s.t. } K(A, A^T)w_2 + eb_2 - Y \geq (1-p_1)e\epsilon_2 - \frac{1}{p_2} \eta \quad (43)$$

$$\eta \geq 0 \quad (44)$$

$G = [K(A, A^T)e]$, $f = Y + e\epsilon_1$, $u_1 = [w_1^T b_1]^T$ 로 두고, 식 (39)~식 (41)의 라그랑주 쌍대문제를 구하면 식 (34), 식 (35)로 나타낼 수 있고, $h = Y + e\epsilon_2$, $u_2 = [w_2^T b_2]^T$ 로 두고, 식 (42)~식 (44)의 라그랑주 쌍대문제를 구하면 식 (36), 식 (37)로 나타낼 수 있다.

라그랑주 쌍대문제를 각각 풀어 우리가 구하고자 하는 회귀식을 구하면 식 (45)와 같다.

$$f(x) = \frac{1}{2} (f_1(x) + f_2(x)) \quad (45)$$

$$= \frac{1}{2} (w_1 + w_2)^T K(A, x) + \frac{1}{2} (b_1 + b_2)$$

커널을 사용하는 경우에 대한 ATSVR을 구하는 알고리즘을 소개하고자 한다. 선형의 경우에는 커널 사용을 하는 것만 제외하면 소개하는 알고리즘과 동일하다. 먼저, 사용되는 행렬의 차원을 위해 기호들을 나타내면 다음과 같다.

l_1 : 학습용 데이터 셋의 관측치 수

l_2 : 테스트용 데이터 셋의 관측치 수

l_{sv} : 서포트벡터에 해당하는 관측치 수

n : 변수(feature)의 수

Step 1. α 와 γ 를 구한다.

$H = [K(A, A^T) e]$, $f = Y - e\epsilon_1$ 라 할 때, 식(36), (37)을 풀어 α 를 구한다. $h = Y + e\epsilon_2$ 라 할 때, 식(34), (35)를 풀어 γ 를 구한다. 여기서 α 와 γ 는 각각 $l_1 \times 1$ 크기의 행렬이다.

Step 2. u_1 , u_2 를 구한다. $u_1 = [w_1 b_1]^T$, $u_2 = [w_2 b_2]^T$

$$u_1 = (H^T H)^{-1} H^T (f - \alpha)$$

$$u_2 = (H^T H)^{-1} H^T (h + \gamma)$$

여기서 u_1 , u_2 는 각각 $(l_1 + 1) \times 1$ 크기의 행렬이다.

Step 3. $\frac{1}{2}(u_1 + u_2)$ 를 구한다.

Step 4. 테스트용 데이터 셋을 이용하여 예측한다.

$$f(x) = \frac{1}{2} (f_1(x) + f_2(x)) = \frac{1}{2} (w_1 + w_2)^T K(A, x^T)$$

$$+ \frac{1}{2} (b_1 + b_2)$$

여기서 $K(A, x)$ 의 A는 테스트용 데이터셋의 A 행렬 ($l_2 \times n$)이고 x 는 학습용 데이터 셋의 A 행렬(x^T 는 $n \times l_1$)이므로 $K(A, x)$ 는 커널연산 후 $l_2 \times l_1$ 행렬이 된다.

4. 실험

소개하는 ATSVR과 기존의 GSVQR의 예측성향 (과대 예측, 과소예측, 중립예측)에 따른 정확도와 계산시간을 살펴보기 위해 3개의 테스트 함수와 1개의 실제 사례를 활용하였다. GSVQR의 예측성향에 따른 정확도와 계산시간에 대한 실험이 기존의 연구에서는 충분치 않았으므로 이번 실험을 통해 살펴보고자 한다. 3개의 테스트 함수들은 Wang et al.[7]과 Huang et al.[11]의 논문에서 사용된 함수들로 다음과 같다.

$$\text{Function 1: } f_1(x_1, x_2) = \sin(x_1 \cdot x_2), D = [-2, 2]^2$$

$$\text{Function 2: } f_2(x_1, x_2) = e^{x_1 \sin(\pi \cdot x_2)}, D = [-1, 1]^2$$

$$\text{Function 3: } f_3(x_1, x_2, x_3, x_4) = e^{2x_1 \sin(\pi \cdot x_4)} + \sin(x_2 \cdot x_3), D = [-0.25, 0.25]^4$$

노이즈(noise)를 추가하기 위해 정규분포 $N(0, 0.1^2)$ 를 이용하여 난수를 발생시키고 이 값들을 함수 $f(x)$ 에 더하였다. 각 함수들에서 총 500개의 데이터를 무작위로 생성하였고, 이들 중 80%에 해당하는 400개는 10-fold 교차검증(cross validation)을 적용하여 초매개변수(hyper-parameter)를 구하기 위해 사용되었다. 나머지 100개의 데이터는 테스트(test)용으로 활용되어 정확도를 구하는데 사용되었다.

R로 코딩되었으며, Intel(R) Core(TM) i3-610의 CPU가 탑재된 컴퓨터에서 실험이 진행되었다. p_1 , p_2 를 이용하여 예측성향을 조절하였으며, 예측성향에 따른 실제예측의 편향을 확인하기 위해 ME(Mean of Errors)가 사용되었고, 예측의 정확도를 측정하기 위해 RMSE(Root Mean of Squared Errors)와 MAE(Mean Absolute Errors) 또는 MAPE(Mean of Absolute Percentage Errors)가 사용되었다.

또한, 계산시간이 측정되었다. Y_i 는 i 번째 실제값, \hat{Y}_i 는 i 번째 예측값, n 은 데이터 수라고 할 때 ME, RMSE, MAE, MAPE는 식 (46)~식 (49)와 같다.

$$ME = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)}{n} \quad (46)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}} \quad (47)$$

$$MAE = \frac{\sum_{i=1}^n |Y_i - \hat{Y}_i|}{n} \quad (48)$$

$$MAPE = \frac{\sum_{i=1}^n \frac{|Y_i - \hat{Y}_i|}{Y_i} \times 100}{n} \quad (49)$$

ME의 값이 음수이면 예측값이 실제값보다 큰 과대예측, 양수이면 예측값이 실제값보다 작은 과소예측, 0에 가까울수록 중립예측을 나타낸다. 또한, ME의 크기는 예측의 편향 정도를 나타낸다. 예측의 정확도를 나타내는 RMSE, MAE, MAPE는 작을수록 예측이 정확함을 의미한다.

이미 언급하였듯이, p_1 을 이용하여 ϵ -tube의 폭의 좌우 비대칭 정도를 조절할 수 있으며, p_2 에 의해 벌금의 비대칭 정도를 조절할 수 있다. 그러므로, 중립예측을 위해서는 $p_1 = p_2 = 0.5$ 를 사용하였고, 과소예측을 위해서는 벌금의 비대칭만 고려한 $p_1 = 0.5, p_2 = 0.01$ 와 ϵ -tube의 폭의 좌우 비대칭 정도만 고려한 $p_1 = 0.99, p_2 = 0.5$, 폭과 벌금의 비대칭을 모두 고려한 $p_1 = 0.99, p_2 = 0.01$ 이 사용되었다. 과대예측을 위해서는 벌금의 비대칭만 고려한 $p_1 = 0.5, p_2 = 0.99$ 와 ϵ -tube의 폭의 좌우 비대칭 정도만 고려한 $p_1 = 0.01, p_2 = 0.5$, 폭과 벌금의 비대칭을 모두 고려한 $p_1 = 0.01, p_2 = 0.99$ 가 사용되었다.

테스트 함수들은 모두 비선형이므로 비선형 커널 중 하나인 R.B.F.(Radial Basis Function) 커널이 사용되었다. $\epsilon = 0.1$ 이 사용되었으며, $C = 10^1 \sim 10^6, \gamma = 10^{-2} \sim 10^1$ 의 범위를 고려하였다. 중립예측인 $p_1 = p_2 = 0.5$ 로 하였을 때의 최적 초매개변수이며, 다른 예측성향에도 동일한 초매개변수가 사용되었다. 함수별 최적 초매개변수는 <Table 1>와 같다. GSVQR과 ATSVR에 무관하게 함수별 최적 초매개변수는 동일하게 나타났다.

<Table 1> Optimal Hyperparameters for Functions f_1, f_2, f_3

Function	GSVQR		ATSVR	
	C	γ	C	γ
f_1	10^2	10^0	10^2	10^0
f_2	10^2	10^0	10^2	10^0
f_3	10^2	10^{-2}	10^2	10^{-2}

f_1 의 경우 반응변수의 값이 -1에서 1사이의 값들이 대부분이라 0에 가까운 값들이 다수 존재하기에 오차의 절대값과 실제값의 비율로 계산하는 MAPE는 큰 값이 나올 수 있어 적절치 않으므로, 오차의 절대값만을 이용하는 MAE를 사용하였다. f_2, f_3 의 반응변수는 모두 양수이므로 MAE대신 MAPE를 사용하였다.

GSVQR과 ATSVR을 적용하였을 때의 각 함수별 결과가 <Table 2>, <Table 3>, <Table 4>에 주어졌다. 중립예측인 $p_1 = p_2 = 0.5$ 을 기준으로 ME값을 상대적으로 판단해 볼 때 중립예측보다 양수이면 과소예측, 음수이면 과대예측이라고 할 수 있다. 모든 경우에 있어 p_1, p_2 의 값의 변화에 따른 과소예측, 과대예측이 제대로 이루어졌다. 또한, 과소예측과 과대예측의 크기는 p_1, p_2 중 하나만 고려한 경우보다 모두 고려한 경우가 크다고 할 수 있다. p_1, p_2 중 어느 값이 예측성향에 더 큰 영향을 미치는지는 함수에 따라 달라질 수 있으며, 특히 p_1 은 ϵ 의 설정된 값에 영향을 받는다.

정확도는 RMSE, MAE 혹은 MAPE값을 고려해 볼 때 중립예측이 가장 작은 값을 보여 정확도가 높으며, p_1, p_2 모두를 이용한 과대예측과 과소예측의 경우가 정확도가 낮게 나타났다.

계산시간을 보면, GSVQR은 6초에서 8초 사이이며, ATSVR은 3.5초대로 ATSVR이 2배 이상 빠른 것으로 나타났다.

<Figure 1>과 <Figure 2>에는 GSVQR과 ATSVR에 함수별 반응변수값과 p_1, p_2 에 따른 예측값들이 주어졌다. 반응변수 값은 점으로 표현되었으며, 예측값들은 실선으로 표현되었다. GSVQR이 ATSVR에 비해 p_1, p_2 의 변화에 따라 예측값들이 더 잘 변화되었다. 즉, GSVQR이 예측성향에 따른 예측이 더 잘 이뤄지는 것으로 보인다.

실제사례에 대한 실험을 위해 Lee[3], Wu et al.[9]의 연구에서 사용되었던 3축 고속 CNC 기계가공에서 밀링커터의 공구마모량을 예측하기 위해 사용되었던 센서데이터들과 공구마모량 데이터를 사용하였다. 선형예측과 비선형예측이 사용되었는데, 비선형예측을 위해서는 R.B.F. 커널이 사용되었다.

데이터의 80%는 10-fold 교차검증(cross validation)을 적용하여 초매개변수(hyper-parameter)를 구하기 위해 사용되었다. 나머지 20%의 데이터는 테스트(test)용으로 활용되어 정확도를 구하는데 사용되었다. 나머지는 테스트함수들에 대한 실험조건과 동일하다.

$\epsilon = 0.1$ 이 사용되었으며, 선형커널의 경우에는 $C = 10^0 \sim 10^5$ 의 범위를, R.B.F. 커널의 경우에는 $C = 10^1 \sim 10^6, \gamma = 10^{-5} \sim 10^1$ 의 범위를 고려하였다. 중립예측인 $p_1 = p_2 = 0.5$ 로 하였을 때의 최적 초매개변수이며, 다른 예측성향에도 동일한 초매개변수가 사용되었다. 함수별 최적 초매개변수는 <Table 5>와 같다.

<Table 2> Results for f_1 by Method

Predictive Propensity	p_1, p_2	Method	ME	RMSE	MAE	Time(sec)
Neutral Predciton	0.50, 0.50	GSVQR	0.002	0.228	0.179	6.94
		ATSVR	-0.007	0.170	0.137	3.58
Underprediction	0.50, 0.01	GSVQR	0.112	0.265	0.207	8.00
		ATSVR	0.010	0.167	0.137	3.60
	0.99, 0.50	GSVQR	0.051	0.234	0.182	6.86
		ATSVR	0.038	0.172	0.140	3.45
	0.99, 0.01	GSVQR	0.161	0.289	0.225	7.77
		ATSVR	0.048	0.171	0.139	3.49
Overprediction	0.50, 0.99	GSVQR	-0.102	0.270	0.219	7.87
		ATSVR	-0.020	0.176	0.141	3.51
	0.01, 0.50	GSVQR	-0.047	0.233	0.185	6.73
		ATSVR	-0.054	0.179	0.143	3.56
	0.01, 0.99	GSVQR	-0.151	0.292	0.240	8.22
		ATSVR	-0.060	0.185	0.147	3.67

<Table 3> Results for f_2 by Method

Predictive Propensity	p_1, p_2	Method	ME	RMSE	MAPE	Time(sec)
Neutral Predciton	0.50, 0.50	GSVQR	-0.044	0.177	13.646	7.03
		ATSVR	-0.049	0.194	14.254	3.52
Underprediction	0.50, 0.01	GSVQR	0.080	0.201	21.071	8.71
		ATSVR	-0.033	0.184	14.108	3.35
	0.99, 0.50	GSVQR	0.005	0.172	14.372	7.23
		ATSVR	-0.005	0.189	15.020	3.46
	0.99, 0.01	GSVQR	0.128	0.225	28.535	8.72
		ATSVR	0.003	0.187	15.132	3.57
Overprediction	0.50, 0.99	GSVQR	-0.173	0.247	16.079	8.70
		ATSVR	-0.061	0.196	14.447	3.38
	0.01, 0.50	GSVQR	-0.093	0.195	13.859	7.80
		ATSVR	-0.095	0.210	14.313	3.60
	0.01, 0.99	GSVQR	-0.222	0.284	18.98	8.05
		ATSVR	-0.101	0.213	14.544	3.51

<Table 4> Results for f_3 by Method

Predictive Propensity	p_1, p_2	Method	ME	RMSE	MAPE	Time(sec)
Neutral Predciton	0.50, 0.50	GSVQR	0.001	0.096	8.160	6.65
		ATSVR	-0.002	0.105	8.937	3.48
Underprediction	0.50, 0.01	GSVQR	0.167	0.212	21.908	8.01
		ATSVR	0.031	0.109	9.559	3.84
	0.99, 0.50	GSVQR	0.050	0.108	9.447	6.46
		ATSVR	0.041	0.110	9.980	3.75
	0.99, 0.01	GSVQR	0.216	0.252	28.642	8.05
		ATSVR	0.063	0.126	11.726	3.51
Overprediction	0.50, 0.99	GSVQR	-0.176	0.204	15.660	7.62
		ATSVR	-0.026	0.123	10.749	3.51
	0.01, 0.50	GSVQR	-0.048	0.107	8.643	6.12
		ATSVR	-0.039	0.119	9.475	3.55
	0.01, 0.99	GSVQR	-0.225	0.248	18.795	7.66
		ATSVR	-0.063	0.144	12.260	3.52

<Table 5> Optimal Hyperparameters for Tool Wear Data

Kernel	GSVQR		ATSVR	
	C	γ	C	γ
Linear	10^2	-	10^3	-
R.B.F.	10^3	10^{-2}	10^3	10^{-2}

선형 커널과 R.B.F. 커널을 사용했을 때의 GSVQR과 ATSVR의 결과가 <Table 6>에 주어져 있다. 중립예측인 $p_1 = p_2 = 0.5$ 을 기준으로 ME값을 상대적으로 판단해 볼 때 모든 경우에 있어 p_1, p_2 의 값의 변화에 따른 과소예측, 과대예측이 제대로 이루어졌다. 또한, 과소예측과 과대예측의 크기는 p_1, p_2 중 하나만 고려한 경우보다 모두 고려한 경우가 크다고 할 수 있다. p_1, p_2 중 어느 값이 예측성향에 더 큰 영향을 미치는지는 함수에 따라 달라질 수 있다. 정확도는 RMSE, MAPE값을 고려해 볼 때 중립예측이 가장 작은 값을 보여 정확도가 높으며, p_1, p_2 모두를 이용

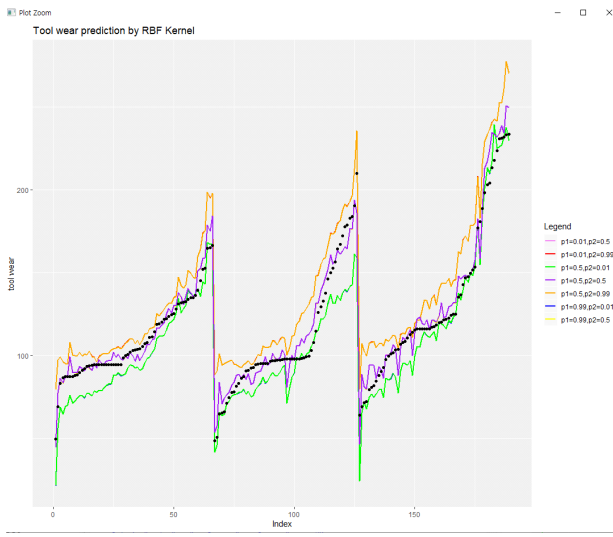
한 과대예측과 과소예측의 경우가 정확도가 낮게 나타났다. R.B.F. 커널의 경우가 선형커널보다 RMSE, MAPE 값이 작으므로 정확도가 높게 나타났다.

계산시간을 보면, GSVQR은 42초~58초이며, ATSVR은 18초~25초로 ATSVR이 2배 이상 빠른 것으로 나타났다. 또한, 선형커널을 사용한 경우가 R.B.F. 커널을 사용한 경우보다 계산시간이 약간 빠른 것으로 나타났다.

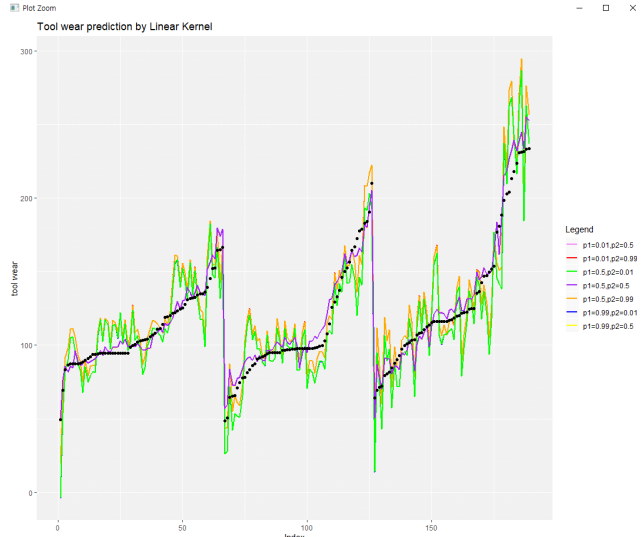
<Figure 3>과 <Figure 4>에는 공구마모량과 p_1, p_2 에 따른 GSVQR과 ATSVR의 예측값들이 주어져 있다. 공구마모량 측정값은 점으로 표현되었으며, 예측값들은 실선으로 표현되었다. 선형커널의 경우에는 GSVQR의 예측성향에 따른 예측값들은 공구마모량을 따라 잘 변화되고 있으나 ATSVR은 변화의 폭이 심한 것으로 보인다. R.B.F. 커널의 경우에도 GSVQR의 예측성향에 따른 예측값들은 공구마모량에 따라 잘 변화되고 있으나, ATSVR은 공구마모량에 따라 변화는 잘 되고 있으나 예측성향에 따른 변화는 제대로 보여주지 못하고 있다.

<Table 6> Results for tool Wear by Method and Kernel

Predictive Propensity	p_1, p_2	Method	Kernel	ME	RMSE	MAPE	Time(sec)
Neutral Prediction	0.50, 0.50	GSVQR	Linear	-2.655	7.803	5.048	42.29
			RBF	-3.045	5.724	4.075	48.54
		ATSVR	Linear	-3.031	7.977	5.313	20.78
			RBF	-2.710	6.936	4.559	23.74
Underprediction	0.50, 0.01	GSVQR	Linear	9.253	13.680	11.741	44.54
			RBF	6.284	10.440	9.140	57.26
		ATSVR	Linear	1.023	21.046	24.202	20.03
			RBF	-2.197	12.118	9.339	23.94
	0.99, 0.50	GSVQR	Linear	-2.606	7.787	5.037	52.72
			RBF	-2.996	5.700	4.051	53.00
		ATSVR	Linear	-2.985	7.960	5.306	18.20
			RBF	-2.680	6.927	4.553	22.74
0.99, 0.01	GSVQR	Linear	9.302	13.714	11.791	45.18	
		RBF	6.333	10.469	9.170	58.86	
	ATSVR	Linear	1.042	21.038	24.117	19.86	
		RBF	-2.180	12.134	9.359	24.20	
Overprediction	0.50, 0.99	GSVQR	Linear	-15.675	18.482	11.880	54.95
			RBF	-11.257	14.240	9.062	56.79
		ATSVR	Linear	-6.662	20.600	13.392	20.96
			RBF	-3.798	12.429	8.142	24.78
	0.01, 0.50	GSVQR	Linear	-2.704	7.820	5.060	41.86
			RBF	-3.094	5.750	4.100	49.77
		ATSVR	Linear	-3.077	7.995	5.320	18.34
			RBF	-2.739	6.946	4.565	23.35
	0.01, 0.99	GSVQR	Linear	-15.724	18.524	11.914	54.12
			RBF	-11.306	14.279	9.097	55.88
		ATSVR	Linear	-6.687	20.624	13.400	21.49
			RBF	-3.818	12.448	8.154	25.03



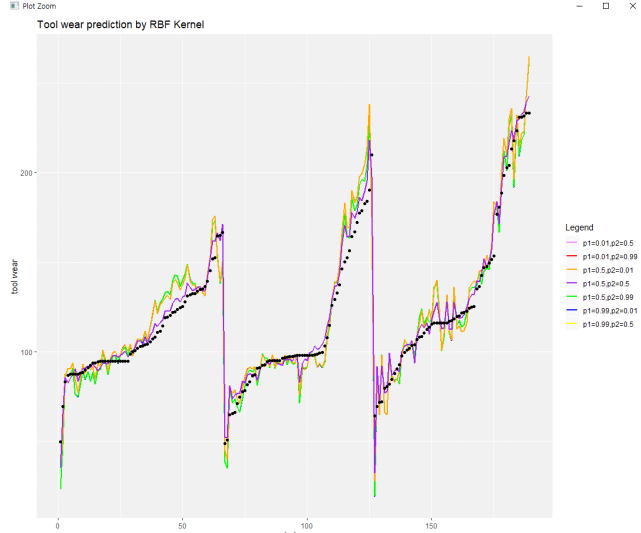
a) Linear Kernel



a) Linear Kernel



b) R.B.F. Kernel



b) R.B.F. Kernel

<Figure 3> Prediction for tool wear by GSVQR

<Figure 4> Prediction for tool wear by ATSVR

5. 결론 및 미래연구과제

본 연구에서는 중립예측, 과소예측, 과대예측의 예측성향을 고려하여 SVR을 이용한 예측을 실시하고 그 차이를 비교하였다. 특히, 2개의 매개변수(p_1, p_2)를 이용한 ATSVR이라는 기법을 제안하고, 기존의 GSVQR 기법의 결과와 비교하였다. GSVQR기법의 예측성향에 따른 실험이 충분치 않고 장단점에 대한 연구가 충분치 않아 이번 연구를 통해 다양한 테스트 함수와 공구마모데이터를 활용하여 그 특징을 살펴보았다.

GSVQR과 ATSVR 모두 2개의 매개변수를 활용하는 경우 예측성향에 따른 예측이 가능하였으며, ATSVR이 계산시간 측면에서는 2배 이상 빠른 것으로 나타났다. 한편, 정확도

측면에서는 ATSVR과 GSVQR은 큰 차이가 없었으나, 그림으로 확인해 볼 때 GSVQR이 ATSVR보다 예측성향을 더 잘 반영하는 것으로 나타났다. 2개의 매개변수(p_1, p_2)중 1개의 매개변수를 활용하는 것보다 2개의 매개변수 모두 활용하는 것이 예측성향의 따른 변화를 크게 하는 것으로 보인다. 하지만, 상황에 따라 2개 중 1개의 매개변수만 활용하는 것이 나올 수도 있다. 또한, 중립예측보다 과소예측이나 과대예측의 정확도가 낮게 나타났다.

미래의 연구과제로는 ϵ 폭의 비대칭성을 조절하는 p_1 은 ϵ 의 값에 영향을 많이 받는다. 다양한 ϵ 값에 따른 변화에 대한 추가 연구가 필요하다. 또한, 예측성향에 따른 실제 문제로의 적용에 대한 연구가 현재로서는 많지 않으므로 많은 실제사례를 발굴하고 이들 사례에 대한 ATSVR과

GSVQR 등의 적용을 통한 연구가 필요하다.

References

- [1] Huang, X., Shi, L., Pelckmans, K., and Suykens, J., Asymmetric ϵ -tube support vector regression, *Computational Statistics and Data Analysis*, 2014, Vol. 77, pp. 371-382.
- [2] Peng X., TSVR: An efficient Twin Support Vector Machine for regression, *Neural Networks*, 2010, Vol. 23, pp. 365-372.
- [3] Lee, D.J., Tool Lifecycle Optimization using ν - Asymmetric Support Vector Regression, *Journal of the Society of Korea Industrial and Systems Engineering*, 2020, Vol. 43, No. 4, pp. 208-216.
- [4] Lee, D.J., and Choi, S.J., Generalized Support Vector Quantile Regression, *Journal of the Society of Korea Industrial and Systems Engineering*, 2020, Vol. 43, No. 4, pp. 107-115.
- [5] Singh, M. Chadha, J., Ahuja, P., and Jayadeva, Chandra, S., Reduced Twin Support Vector Regression, *Neurocomputing*, 2011, Vol. 74, pp. 1474-1477.
- [6] Vapnik, V., *Statistical Learning Theory*, New York, NY: Wiley, 1998.
- [7] Wang, S., Huang, X., and Yan, Y., A neural network of smooth hinge functions, *IEEE Trans. Neural Netw.*, 2010, Vol. 21, No. 9, pp. 1381-1395.
- [8] Wang, H. and Xu, Y., Scaling up twin support vector regression with safe screening rule, *Information Science*, 2018, Vol. 465, pp. 174-190.
- [9] Wu, D., Jennins, C., Terpenney, J., Gao, R.X., and Kumara, S., A Comparative Study on Machine Learning Algorithms for Smart Manufacturing: Tool Wear Prediction Using Random Forests, *Journal of Manu. Sci. and Engineering*, 2017, Vol. 139, No. 7, pp. 1-10.
- [10] Wu, J., Wang, Y.-G., Tian, Y.-C., Burrage, K., and Cao, T., Support Vector Regression with Asymmetric Loss for Optimal Electric Load Forecasting, *Energy*, 2021, Vol. 223, 119969.
- [11] Xu, Y., Li, X., Pan, X., and Yang, Z., Asymmetric ν -twin support vector regression, *Neural Comput. & Appl.*, 2018, Vol. 30, pp. 3799-3814.
- [12] Xu, Y. and Wang, L., A weighted twin support vector regression, *Knowledge-Based Systems*, 2012, Vol. 33, pp. 92-101.

ORCID

Dongju Lee | <http://orcid.org/0000-0001-6650-9270>

<Appendix> ATSVR에서 원문제의 라그랑주 쌍대문제로의 변환과정

식 (31)~ 식(33)의 원문제를 이용하여 라그랑주 쌍대문제를 구하는 과정은 다음과 같다. 먼저 라그랑주 함수(L_1)를 구하면 식 (3)과 같다.

$$L_1 = \frac{1}{2} \|Y - e\epsilon_1 - (Aw_2 + eb_2)\|^2 + C_2 e^T \eta - \gamma^T \left((Aw_2 + eb_2) - Y - (1-p_1)e\epsilon_2 + \frac{1}{p_2}\eta \right) - \delta^T \eta \quad (50)$$

그러므로, KKT 조건을 구하기 위해 L_1 에 대하여 변수들의 편미분을 계산하고 0으로 두면 식 (51)~식 (53)과 같다.

$$\frac{\partial L_1}{\partial w_2} = -A^T(Y + e\epsilon_2 - (Aw_2 + eb_2)) - A^T \gamma = 0 \quad (51)$$

$$\frac{\partial L_1}{\partial b_2} = -e^T(Y + e\epsilon_2 - (Aw_2 + eb_2)) - e^T \gamma = 0 \quad (52)$$

$$\frac{\partial L_1}{\partial \eta} = C_2 e - \frac{1}{p_2} \gamma - \delta = 0 \quad (53)$$

P6의 제약식을 추가하면 식(54), (55)과 같다.

$$(Aw_2 + eb_2) - Y \geq (1-p_1)e\epsilon_2 - \frac{1}{p_2}\eta \quad (54)$$

$$\eta \geq 0 \quad (55)$$

그러므로, 상보여유정리는 식 (56), 식 (57)과 같다.

$$\gamma^T \left((Aw_2 + eb_2) - Y - (1-p_1)e\epsilon_2 + \frac{1}{p_2}\eta \right) = 0, \gamma \geq 0 \quad (56)$$

$$\delta^T \eta = 0, \delta \geq 0 \quad (57)$$

$C_2 e - \frac{1}{p_2} \gamma - \delta = 0$ 이고 $\delta \geq 0$ 이므로 $\frac{1}{p_2} \gamma \leq C_2 e$ 이다. 그러므로 $0 \leq \gamma \leq p_2 C_2 e$ 이다.

식 (51)와 식 (52)는 식 (58)과 같다.

$$-\begin{bmatrix} A^T \\ e^T \end{bmatrix} \left((Y + e\epsilon_2) - [A \ e] \begin{bmatrix} w_2 \\ b_2 \end{bmatrix} \right) - \begin{bmatrix} A^T \\ e^T \end{bmatrix} \gamma = 0 \quad (58)$$

$G = [A \ e]$, $h = Y + e\epsilon_2$, $u_2 = [w_2^T \ b_2^T]^T$ 라고 하자. 그러면, 식 (58)은 식 (59)로 나타낼 수 있다.

$$-G^T h + G^T G u_2 - G^T \gamma = 0 \quad (59)$$

그러므로, u_2 , u_2^T 에 대한 식으로 나타내면 식 (60), 식 (61)과 같다.

$$u_2 = (G^T G)^{-1} (G^T h + G^T \gamma) \quad (60)$$

$$u_2^T = (h^T G + \gamma^T G) (G^T G)^{-1} \quad (61)$$

$G^T G$ 가 반양정치행렬(positive semidefinite matrix) 이기는 하지만 역행렬을 구하기 어려울 수 있기에 아주 작은 임의의 값인 σ 를 더해서 계산할 수 있는데, 식 (60)은 식 (62)와 같다.

$$u_2 = (G^T G + \sigma I)^{-1} G^T (h + \gamma) \quad (62)$$

이다. 여기서 I 는 identity matrix이다.

식 (50)을 G, h, u_2 을 이용하여 다시 정리하면

$$\begin{aligned} G &= [A \ e], \quad h = Y + e\epsilon_2, \quad u_2 = [w_2^T \ b_2]^T \\ L_2 &= \frac{1}{2}(Y + e\epsilon_2 - (Aw_2 + eb_2))^T(Y + e\epsilon_2 - (Aw_2 + eb_2)) + C_2e^T\eta \\ &\quad - \gamma^T((Aw_2 + eb_2) - Y - (1-p_1)e\epsilon_2 + \frac{1}{p_2}\eta) - \delta^T\eta \\ L_2 &= \frac{1}{2}(h - Gu_2)^T(h - Gu_2) - \gamma^T(Gu_2 - Y - (1-p_1)e\epsilon_2) + (C_2e - \frac{1}{p_2}\gamma - \delta)\eta \end{aligned}$$

식 (53)에 의해 $(C_2e - \frac{1}{p_2}\gamma - \delta)\eta = 0$ 이므로

$$\begin{aligned} &= \frac{1}{2}(h - Gu_2)^T(h - Gu_2) - \gamma^T(Gu_2 - Y - (1-p_1)e\epsilon_2) \\ &= \frac{1}{2}(h - Gu_2)^T(h - Gu_2) - \gamma^T(Gu_2 - Y - (1-p_1)e\epsilon_2) + \gamma^T e \epsilon_2^T \gamma \\ &= \frac{1}{2}(h - Gu_2)^T(h - Gu_2) - \gamma^T(Gu_2 - Y - e\epsilon_2) + \gamma^T(1-p_1)e\epsilon_2 - \gamma^T e \epsilon_2 \\ &= \frac{1}{2}(h - Gu_2)^T(h - Gu_2) + \gamma^T(h - Gu_2) - \gamma^T p_1 e \epsilon_2 \end{aligned}$$

이다. $\gamma^T(h - Gu_2)$ 는 1×1 행렬이므로 $(\gamma^T(h - Gu_2))^T = (h - Gu_2)^T \gamma$ 로 표현할 수 있다. 또한, $\gamma^T p_1 e \epsilon_2$ 도 1×1 행렬이므로 $(\gamma^T p_1 e \epsilon_2)^T = p_1 \epsilon_2^T \gamma$ 로 표현할 수 있다. 그러므로,

$$\begin{aligned} L_2 &= \frac{1}{2}(h - Gu_2)^T(h - Gu_2) + (h - Gu_2)^T \gamma - p_1 \epsilon_2^T \gamma \\ &= \frac{1}{2}h^T h - \frac{1}{2}h^T Gu_2 - \frac{1}{2}u_2^T G^T h + \frac{1}{2}u_2^T G^T Gu_2 + h^T \gamma - u_2^T G^T \gamma - p_1 \epsilon_2^T \gamma \\ &= \frac{1}{2}h^T h - \frac{1}{2}h^T Gu_2 + \frac{1}{2}u_2^T (-G^T h + G^T Gu_2 - G^T \gamma) + h^T \gamma - \frac{1}{2}u_2^T G^T \gamma - p_1 \epsilon_2^T \gamma \end{aligned}$$

$h^T h$ 는 상수이므로 생략가능하다. 또한, 식 (58) $-G^T h + G^T Gu_2 - G^T \gamma = 0$ 이므로

$$L_2 = -\frac{1}{2}h^T Gu_2 + h^T \gamma - \frac{1}{2}u_2^T G^T \gamma - p_1 \epsilon_2^T \gamma$$

여기에 식 (60), 식 (61)를 적용하면

$$\begin{aligned} L_2 &= -\frac{1}{2}h^T G(G^T G)^{-1}(G^T h + G^T \gamma) + h^T \gamma - \frac{1}{2}(h^T G + \gamma^T G)(G^T G)^{-1}G^T \gamma - p_1 \epsilon_2^T \gamma \\ &\quad - \frac{1}{2}h^T G(G^T G)^{-1}G^T h - \frac{1}{2}h^T G(G^T G)^{-1}G^T \gamma + h^T \gamma - \frac{1}{2}h^T G(G^T G)^{-1}G^T \gamma \\ &\quad - \frac{1}{2}\gamma^T G(G^T G)^{-1}G^T \gamma - p_1 \epsilon_2^T \gamma \end{aligned}$$

여기서 $-\frac{1}{2}h^T G(G^T G)^{-1}G^T h$ 는 상수이므로 생략 가능하다. 그러므로, 다음과 같이 표현할 수 있다.

$$-\frac{1}{2}\gamma^T G(G^T G)^{-1}G^T \gamma - h^T G(G^T G)^{-1}G^T \gamma + (h^T - p_1 \epsilon_2^T) \gamma$$