

SVM Load Forecasting using Cross-Validation

趙南熏[†]

(Nam-Hoon Jo)

Abstract - In this paper, we study the problem of model selection for Support Vector Machine(SVM) predictor for short-term load forecasting. The model selection amounts to tuning SVM parameters, such as the cost coefficient C and kernel parameters and so on, in order to maximize the prediction performance of SVM. We propose that Cross-Validation method can be used as a model selection algorithm for SVM-based load forecasting technique. Through the various experiments on several data sets, we found that the difference between the prediction error of SVM using Cross-Validation and that of ideal SVM is less than 5%. This shows that SVM parameters for load forecasting can be efficiently tuned by using Cross-Validation.

Key Words : Load Forecasting, Support Vector Machine, Model Selection, Cross-Validation

1. 서 론

전력 수요 예측은 매우 활발히 진행되고 있는 연구 분야이며, 전력시장이 경쟁체제로 바뀌어 가면서 그 중요성이 급속히 증가되고 있다. 전력수요예측 연구는 크게, 시계열 모델(time series model)을 이용한 방법 [1][2], 인과 모델(causal model)을 이용한 방법[3], 인공지능기법을 이용한 방법 [4][5][6][7]등으로 나눌 수 있는데, 최근에는 인공지능기법을 적용한 전력 수요 예측기법 연구가 매우 활발히 진행되고 있다. 이러한 인공지능기법 중 가장 많은 관심을 받아 온 기법은 1990년대 초부터 연구되어 온 신경회로망(Neural Networks; NN)이다 [6][7]. 그러나, 신경회로망에는 몇 가지 단점이 지적되고 있다. 첫째로 학습에 필요한 시간이 오래 소요되고, 둘째로 과대적합(overfitting)시 예측성능이 저하되며, 셋째로 최적의 설계변수를 결정할 때 국소적 최소점(local minimum)으로 수렴하는 것 등이 취약한 점으로 알려져 있다. 이러한 단점을 보완할 수 있는 인공지능기법으로 최근에 각광받고 있는 것이 지원벡터머신(Support Vector Machine; SVM)이다 [8][9]. SVM은 기존의 통계적 학습 방법에서 이용되는 '경험적 위험 최소화' (Empirical Risk Minimization; ERM)와는 다른 '구조적 위험 최소화' (Structural Risk Minimization; SRM)을 이용하여 일반화 오차를 감소시키는 기법이다. SVM은 초기에 패턴인식과 같은 분류(classification) 문제에 적용되어 왔으나, 최근에 전

력수요예측과 같은 비선형 회귀(nonlinear regression)에도 적용되기 시작하였다. 최근에, SVM을 이용한 전력수요예측 기법이 기존의 신경회로망기법보다 우수한 예측성능을 보일 수 있다는 가능성이 참고문헌 [10]에서 제시되었다. [10]에서는 기존의 신경회로망을 이용한 국내 전력수요 예측연구결과인 [11]와 예측성능을 비교하여 SVM 파라미터가 적절히 선정될 경우, SVM 예측성능이 신경회로망의 예측성능보다 우수하다는 것을 보였다. 하지만, 참고문헌 [10]에서는 우수한 예측성능을 제공하는 SVM 파라미터의 선정에 대한 방법은 제시되지 않아, 실제적인 응용에는 한계가 있었다.

본 논문에서는 SVM 전력수요예측에서 예측성능에 적지 않은 영향을 주는 SVM 파라미터의 선택기법에 대한 연구를 수행하고자 한다. SVM 파라미터를 체계적으로 선택하는 방법으로는 VC 한계 최소화기법, 베이시안 학습기법, 교차검증기법 등이 알려져 있다 [8][9]. 이 중에서 상대적으로 간단한 교차검증을 실제의 전력수요예측에 적용할 경우 우수한 파라미터를 찾을 수 있는지에 대한 연구가 필요하다. 만약, 교차검증에 의해 선택된 SVM 파라미터가 우수한 예측성능을 주지 않는다면 보다 복잡한 기법을 필요로 할 것이지만, 우수한 성능을 제공한다면 교차검증을 사용하는 것으로도 충분할 것이다. 교차검증기법은 기본적으로 학습데이터를 두개의 집합으로 나누어, 학습시에는 이 중 하나의 집합(순수학습 데이터)만 사용하고 또 다른 집합(검증 데이터)에 대해서는 학습된 SVM의 예측성능을 평가하여 최상의 예측성능을 주는 파라미터를 선정하는 것이다. 이와 같은 방법을 이용하면 체계적으로 SVM 파라미터를 선정할 수 있지만, 교차검증의 특성상 어느 정도 학습데이터의 특성에 의존하는 파라미터가 선정될 가능성이 있다. 따라서, 교차검증으로 선정된 파라미터가 실제의 예측시에 우수한 예측성능을 줄 수 있는지를 실험적으로 검증할 필요가 있다.

[†] 교신저자, 正會員: 崇賓大 工大 電氣工學部 助教授 · 工博

E-mail : nhjo@ssu.ac.kr

接受日字 : 2006年 8月 3日

最終完了 : 2006年 9月 26日

본 논문에서는 교차검증을 SVM 전력수요예측에 적용하여, 실제로 우수한 SVM 파라미터 선정이 가능함을 실험적으로 입증할 것이다.

본 논문의 구성은 다음과 같다. 2장에서는 지원벡터머신(SVM)의 기본적인 개념 및 용어에 대해서 알아본다. 3장에서는 SVM을 이용한 전력수요예측 기법을 소개하고, 효과적인 SVM 파라미터 선정에 사용될 수 있는 교차검증 기법을 소개한다. 4장에서는 모의실험을 통하여 교차검증이 SVM 파라미터를 효과적으로 선택할 수 있음을 보이며, 마지막으로 5장에서는 본 논문의 결론을 제시한다.

2. 지원벡터머신(Support Vector Machine; SVM)

2.1 선형 지원벡터회귀(Support Vector Regression)

선형 SV regression의 목적은 선형 함수

$$f(x) = \langle w, x \rangle + b \quad (1)$$

를 찾아서, $f(x)$ 와 모든 학습 데이터(training data)의 최대 편차가 ϵ 이내가 되면서 동시에 w 의 노름(norm)을 최소화 하는 것으로 다음과 같은 문제로 표현할 수 있다 :

$$\begin{aligned} \min_{w, b, \xi, \xi_i^*} & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{\ell} (\xi_i + \xi_i^*) \\ \text{subject to} & \begin{cases} y_i - \langle w, x_i \rangle - b \leq \epsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, i = 1, \dots, \ell \end{cases} \end{aligned} \quad (2)$$

여기서, 상수 C 는 ϵ 보다 큰 오차를 얼마만큼 허용할 것인가와 함수 $f(x)$ 의 편평함(flatness) 사이의 타협점을 결정하는 것으로, **비용계수(Cost Coefficient)**라고 부른다. 비용계수 C 가 상대적으로 큰 경우에는 학습데이터의 오차를 가능한 작게 만드는 것이고, 반대로 C 가 상대적으로 작은 경우에는 w 의 노름을 작게 만들기 때문에 함수 $f(x)$ 를 가능한 편평하게 만드는 것이다. 최적화문제 (2)는 쌍대 최적화(Dual optimization) 문제로 표현할 수 있다 :

$$\begin{aligned} \max & \begin{cases} -\frac{1}{2} \sum_{i,j=1}^{\ell} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle x_i, x_j \rangle \\ -\epsilon \sum_{i=1}^{\ell} (\alpha_i + \alpha_i^*) + \sum_{i=1}^{\ell} y_i (\alpha_i - \alpha_i^*) \end{cases} \\ \text{subject to} & \begin{cases} \sum_{i=1}^{\ell} (\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C] \end{cases} \end{aligned} \quad (3)$$

최적화 문제 (3)을 풀어서 해 α_i, α_i^* 를 구하면, 선형 SVM은

$$f(x) = \sum_{i=1}^{\ell} (\alpha_i - \alpha_i^*) \langle x_i, x \rangle + b \quad (4)$$

로 표현할 수 있으며, 상수항 b 는

$$b = \begin{cases} y_i - \langle w, x_i \rangle - \epsilon & \text{for } \alpha_i \in (0, C) \\ y_i - \langle w, x_i \rangle + \epsilon & \text{for } \alpha_i^* \in (0, C) \end{cases}$$

를 이용하여 계산한다. 지원벡터회귀에 대한 보다 자세한 내용은 참고문헌 [8],[9],[10]을 참조하기 바란다.

2.2 비선형 지원벡터회귀

앞서 살펴본 선형 SVM을 비선형으로 확장하기 위하여 커널 기교(Kernel trick)를 사용한다. 커널 기교란 어떤 사상(mapping) $\Phi(x)$ 를 통하여 원래의 데이터 공간을 고차원의

새로운 공간으로 사상시킨 후, 이 새로운 공간에서 선형 SVM regression 이론을 적용하는 것이다. 커널 함수를

$$k(x_1, x_2) = \langle \Phi(x_1), \Phi(x_2) \rangle$$

로 정의하면 (3)으로부터 비선형 SVM 최적화문제는 다음과 같이 표현된다 :

$$\begin{aligned} \max & \begin{cases} -\frac{1}{2} \sum_{i,j=1}^{\ell} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) k(x_i, x_j) \\ -\epsilon \sum_{i=1}^{\ell} (\alpha_i + \alpha_i^*) + \sum_{i=1}^{\ell} y_i (\alpha_i - \alpha_i^*) \end{cases} \\ \text{subject to} & \begin{cases} \sum_{i=1}^{\ell} (\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C] \end{cases} \end{aligned} \quad (5)$$

최적화 문제 (5)를 풀어서 해 α_i, α_i^* 를 구하면, (4)로부터 비선형 비선형 SVM식은

$$f(x) = \sum_{i=1}^{\ell} (\alpha_i - \alpha_i^*) k(x_i, x) + b$$

로 표현된다.

3. SVM 수요예측과 교차검증

3.1 SVM 수요예측

SVM 수요예측을 위해서는 SVM의 입력으로 사용되는 특징벡터를 선정하여야 한다. 이를 위해 본 논문에서는 참고문헌 [10]에서와 같이 요일, 온도, 과거 전력수요를 특징으로 선택하였다. 요일은 요일정보와 특수일 여부를 나타내기 위한 7자리 이진 코드인데, 앞의 6자리는 요일을 나타내며 마지막 7번째 이진 코드는 특수일 여부를 나타낸다. 또한, 과거 수요는 예측일 하루 전일의 최대 수요이며, 온도는 네 가지 온도, 즉 예측일의 최고온도, 예측일의 최저온도, 예측일 하루 전일의 최고온도, 예측일 하루 전일의 최저온도를 나타낸다. 따라서, SVM의 입력벡터는 다음과 같은 형태를 갖는다.

[과거 수요, 요일, 온도]

한편, ϵ 은 일반적으로 작을수록 우수한 일반화성능을 나타내기 때문에 $\epsilon = 0.5$ 를 선택하였다. 또한, 커널기교에서 사용되는 커널함수는 다음과 같은 가우시안 방사기저함수(RBF) 커널을 선택하였는데,

$$k(x_1, x_2) = \exp\left(-\frac{\|x_1 - x_2\|^2}{2\sigma^2}\right)$$

여기서 σ 는 RBF함수의 대역폭(bandwidth)이라고 부른다. SVM 학습시 비용계수 C 와 RBF 커널함수의 대역폭 σ 를 선택하여야 하는데, 기존의 SVM 연구결과인 [10]에서는 $C = 10^7$, $\sigma = 5, 10, 20, 30$ 를 선택하였다. 하지만, 이를 선택하는 체계적인 방법이 제시되지 않아서, 실제 시스템에의 응용에는 어려움이 있었다. 본 논문에서는 이러한 비용계수와 RBF함수의 대역폭을 체계적으로 선택할 수 있는 방법을 3.3에서 제시하도록 한다.

마지막으로, 예측오차 계산에는 절대값 최대오차(Maximum Absolute Error)와 퍼센트 절대값 평균오차(Mean Absolute Percentage Error; MAPE)가 많이 사용되는데 본 논문에서는 MAPE를 사용하였다:

$$MAPE = 100 \times \frac{1}{n} \times \sum_{i=1}^n \left| \frac{L_i - \hat{L}_i}{L_i} \right|$$

여기서, L_i 와 \hat{L}_i 는 i 번째 날의 실제 최대 전력수요와 예측된 최대 전력수요이며, n 은 예측기간내의 일수이다.

3.2 SVM파라미터가 예측성능에 미치는 영향 분석

최근에 신경회로망을 이용한 국내의 전력 수요 예측 연구 결과가 [11]에서 발표되었다. [11]에서는 입력값 선정과 학습 케이스 선정 등과 같이 예측정확성에 영향을 주는 요인에 대해 과거 논문을 조사하고 평가하였다. 이 결과를 토대로 최소구조 신경회로망을 제안하였는데, 기존의 연구결과 [12]에 비해서 정확도가 상당히 향상되었다. 따라서, [11]에서 제시된 수요예측을 위한 신경회로망은 그 파라미터 (예를 들면, 은닉층의 뉴런 개수) 튜닝이 매우 우수하게 수행되었다고 판단할 수 있다. 한편, SVM을 이용한 전력수요예측기법이 기존의 신경회로망기법보다 우수한 예측성능을 보일 수 있다는 가능성이 참고문헌 [10]에서 제시되었다. 참고문헌 [10]에서는 2002년 여름과 2003년 여름의 하루 최대 전력수요량을 예측하는데 참고문헌 [11]의 신경회로망과 SVM의 예측성능을 비교하였다. 그 결과 SVM 파라미터가 적절히 선정될 경우, SVM 예측성능이 신경회로망의 예측성능보다 우수함을 알 수 있었다. 하지만, 참고문헌 [10]에서는 우수한 예측성능을 제공하는 SVM 파라미터의 선정에 대한 방법은 제시되지 않고 있다. 실제로, SVM 전력수요예측은 SVM 파라미터를 어떤 값으로 선정하느냐가 예측성능에 매우 지대한 영향을 준다. 이를 확인하기 위하여, 여러 가지 C 와 σ 에 대해서, 2000년 여름(6, 7, 8, 9월)과 2001년 여름(6, 7, 8, 9월)을 학습데이터로 이용하여 SVM을 학습시킨 후, 2002년 8월의 하루 최대 전력수요량을 예측해 보았다. 그 결과를 그림 1에서 볼 수 있는데, 두 번째 경우, 즉 $C=2^{24}$, $\sigma=2^1$ 일 때 가장 우수한 성능을 보여주며, 다른 경우에는 예측성능이 상당히 악화됨을 알 수 있다. 따라서, 우수한 예측성능을 주는 적절한 SVM파라미터의 선정은 매우 중요함을 알 수 있다.

3.3 교차검증을 이용한 SVM파라미터의 선택

SVM의 예측성능에 상당한 영향을 미치는 비용계수 C 와 RBF 커널함수의 대역폭 σ 를 적절히 선택하는 가장 간단한 방법은 학습시 사용한 학습데이터의 예측오차를 최소화 하는 C 와 σ 를 사용하는 것이라고 생각할 수 있다. 하지만, 이러한 SVM 파라미터를 사용하는 경우, 과대적합(over-fitting) 문제가 발생하여 학습되지 않은 데이터에 대한 예측성능이 오히려 저하되는 현상이 발생한다는 것이 알려져 있다 [13]. 이와 같은 현상을 방지하고 일반화성능을 개선시키기 위하여 여러 가지 기법이 제시되고 있는데, 대표적인 방법으로 VC 한계(VC bounds) 최소화기법, 베이시안 학습기법, 교차 검증기법 등이 있다[8][9]. VC 한계 최소화기법이란, SVM의 일반화오차의 한계를 수식으로 정확히 유도할 수 있는데, 이 한계가 VC 차원으로 표현되며, 다시 VC 차원의 상한이 추가적인 2차 최적화문제를 풀어서 계산된다는 것을 이용하는 방법이다. 베이시안 학습기법은 SVM 파라미터의 확률분포와 조정화 항(regularization term)를 고려하는 방법으로 신경회로망에서도 유용하게 사용된 방법이다. 마지막으로,

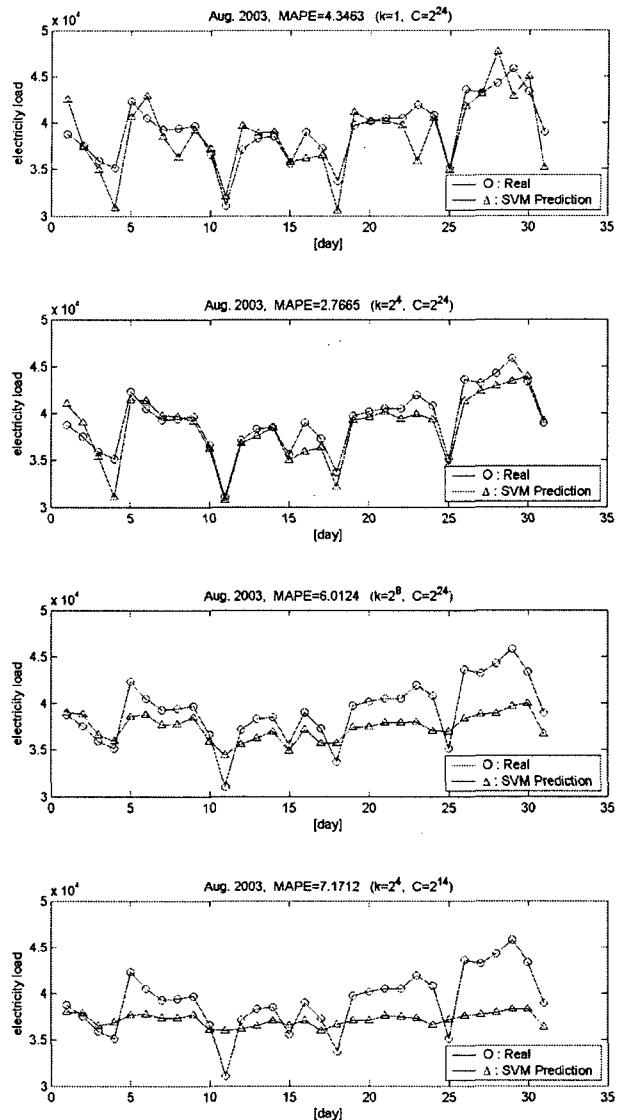


그림 1 여러 가지 C 와 σ 에 대한 SVM 예측결과
Fig. 1 SVM prediction results for various C , σ

교차검증기법은 학습시 주어진 데이터를 학습데이터와 검증 데이터로 나누어 검증데이터에 대한 최적의 모델을 선정하는 방법이다. 본 논문에서는 이들 중에서 상대적으로 간단하면서도 우수한 성능을 나타내는 교차검증(cross validation)을 사용하여 SVM전력수요예측의 최적 파라미터를 선정하고자 한다. 수요예측에 사용되는 교차검증은 다음과 같은 과정으로 진행되며, 그림 2는 교차검증의 수행방법 및 절차를 순서대로 표현한 것이다.

- 1) 학습 데이터 집합을 순수학습(training) 데이터와 검증(validation) 데이터로 나눈다.
- 2) 여러 가지 (σ , C) 쌍에 대해서 순수학습 데이터만을 사용하여 SVM을 학습시킨다 여기서, 검증 데이터는 SVM을 학습시킬때 전혀 사용되지 않는다..
- 3) 2)에서 학습된 SVM에 대해서 검증 데이터에 대한 예측오차를 계산한다. 이를 통하여 알려지지 않은 데이터에 대한 예측성능을 좀 더 정확히 평가할 수 있다.

- 4) 3)에서 계산된 검증 데이터에 대한 예측오차를 최소화 하는 (σ , C)를 최적의 SVM 파라미터 (σ^*, C^*)로 결정한다.
- 5) SVM 파라미터 (σ^*, C^*)를 이용하여 전체학습데이터, 즉 순수학습 데이터와 검증 데이터를 합한 데이터에 대해서 SVM을 다시 한번 학습시킨다.
- 6) 학습된 SVM을 이용하여 미지의 데이터에 대한 예측을 수행한다.

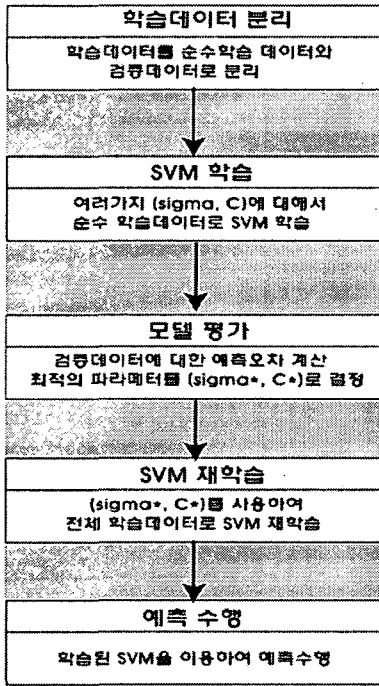


그림 2 교차검증 순서도
Fig. 2 Flowchart of Cross-validation

다음 장에서는 위에서 제시된 교차검증을 사용하여 학습된 SVM의 예측성과 실제로 최적의 SVM 파라미터로 학습된 SVM과의 예측성을 비교할 것이다. 이를 통하여 교차검증을 이용한 SVM 전력수요예측기법이 충분히 만족할 만한 효과를 낼 수 있음을 실험적으로 입증한다.

4. 모의실험

본 장에서는 모의실험을 통하여 3장에서 제시한 교차검증의 유효성을 검증하고자 한다. 이를 위하여 먼저, 비용계수와 대역폭이 취할 수 있는 값을 결정해야 하는데, 여러 번의 모의실험 결과 다음 값만을 고려해도 충분하였다: $\sigma = 2^1, 2^2, \dots, 2^9$, $C = 2^1, 2^2, \dots, 2^{29}$. 따라서, SVM 학습은 261(=9×29)개의 (σ , C) 쌍에 대해서 수행되어야 하며, 이 중에서 검증데이터에 대한 예측오차를 최소화 하는 (σ , C)를 선택하여 실제의 예측시스템에 사용하는 것이다. 참고로, 펜티엄4 2.8G, Matlab 6.6를 사용하여 교차 검증을 수행하는데 약 3분정도가 소요되었으며, 계산시간은 큰 문제가 되지 않는 것을 알 수 있다

또한, SVM학습시 입력 전처리 과정으로서 입력벡터의 정규화를 수행한다. 입력벡터 정규화를 수행하지 않으면 알고리즘이 수렴하지 않고 발산하거나, 수렴한다고 하더라도 성능이 별로 좋지 않은 경우가 대부분이다. 본 논문에서는 입력벡터의 모든 성분 변수가 0에서 1사이의 값이 되도록 입력벡터 정규화를 수행한 후 SVM을 학습시킨다. 여기서 주의할 점은, 전력수요를 예측할 때 사용하는 입력벡터도 반드시 위와 같은 정규화를 거친 후 사용하여야 한다는 것이다.

4.1 사례연구 1: 2002년 7월 전력수요 예측

첫 번째로, 2002년 7월의 하루 최대 전력수요량을 예측하는데 SVM에 교차검증을 적용하여 보았다. 학습시 사용가능한 데이터는 2000년 6, 7, 8, 9월, 2001년 6, 7, 8, 9월, 2002년 6월 데이터라고 가정하였다. 우선, 모든 학습 데이터를 이용하여 SVM을 학습시킨 후, 2002년 7월 데이터의 MAPE를 최소화 하는 가장 우수한 3가지 SVM 파라미터를 구해보면, 그 결과는 표 1과 같다.

표 1 최적의 SVM파라미터가 이용되었을 때의 2002년 7월 예측성능

Table 1 Prediction performance of July, 2002 when optimal SVM parameters were used

σ	C	MAPE (2002년 7월)
2^4	2^{23}	3.2586
2^5	2^{27}	3.2651
2^5	2^{26}	3.2690

위 결과는 SVM예측기법을 사용할 때 얻을 수 있는 최상의 예측성능이지만, 2002년 7월 데이터를 알고 있다는 비현실적인 가정을 토대로 (σ , C)를 결정하기 때문에 실제의 경우에는 사용할 수가 없다.

이제, 2002년 7월 데이터를 알 수 없을 때, (σ , C)를 결정하기 위하여 교차검증을 적용해 보자. 이를 위해서 2001년 7월 자료를 검증 데이터로 정하고, 나머지를 순수학습 데이터로 정하도록 한다. 순수학습 데이터로 SVM을 학습시킨 후, 검증 데이터의 MAPE를 최소화 하는 가장 우수한 3가지 SVM 파라미터는 표 2와 같다.

표 2 검증데이터인 2001년 7월의 예측성능

Table 2 Prediction performance of validation set June, 2002

σ	C	검증데이터(2001년 7월)에 대한 MAPE
2^5	2^{28}	1.9771
2^3	2^{21}	1.9819
2^4	2^{22}	1.9825

표 2로부터 최적의 SVM 파라미터는 (σ, C)=($2^5, 2^{28}$)이라고 판단할 수 있다. 마지막으로 이 파라미터를 이용하여 모든 학습데이터에 대해서 다시 SVM을 학습시킨 후, 2002년 7월 데이터의 MAPE를 계산하면 표 3과 같은 결과를 얻을 수 있다.

표 3 교차검증에 의해 SVM파라미터를 선택할 때의 2002년 7월 예측성능

Table 3 Prediction performance of July, 2002 when SVM parameters were chosen by cross validation

σ	C	MAPE (2002년 7월)
2^5	2^{28}	3.3130
2^3	2^{21}	3.3024
2^4	2^{22}	3.2825

따라서, 최상의 MAPE는 3.2586 (표 1)일 때, 교차검증에 의해 현실적으로 얻을 수 있는 최소 MAPE는 3.3130임을 알 수 있다. 이는 최상의 MAPE를 기준으로 할 때, 예측성능이

$$\frac{3.3130 - 3.2586}{3.2586} * 100 = 1.6\%$$

악화 되는 것을 의미하며, 교차검증으로 구한 최적 파라미터와 실제의 최적 파라미터의 예측성능이 거의 비슷함을 의미한다. 그림 3은 교차검증을 이용한 SVM 수요예측을 이용한 2002년 7월 전력수요 예측 결과를 보여준다.

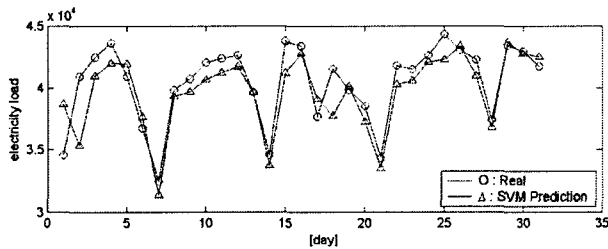


그림 3 2002년 7월의 SVM 예측결과
Fig. 3 SVM prediction results for July, 2002

4.2 사례연구 2: 2002년 8월 전력수요 예측

두 번째로, 학습시 사용가능한 데이터가 2000년 6, 7, 8, 9월, 2001년 6, 7, 8, 9월, 2002년 6, 7월 데이터일 때, 2002년 8월의 하루 최대 전력수요량을 예측하는데 교차검증을 사용하여 보았다. 우선, 학습 데이터를 모두 이용하여 SVM을 학습시킨 후, 2002년 8월 데이터의 MAPE를 최소로 하는 가장 우수한 SVM 파라미터를 구해보면

$$\sigma = 2^6, C = 2^{29}$$

이고

$$MAPE = 2.4706$$

이다 (사례연구 1에서 가장 우수한 3가지 SVM 파라미터의 결과가 비슷한 경향을 보임을 알 수 있으므로, 이제부터는 가장 우수한 한 가지 경우에 대해서만 분석하도록 한다).

사례연구 1과 마찬가지로, 위의 (σ, C)는 이론적으로 최상의 예측성능을 주는 SVM 파라미터이지만, 실제의 경우에는 사용할 수가 없다. 이제, 교차검증을 적용하기 위해서 2001년 8월 자료를 검증 데이터로 정하고, 나머지를 순수학습 데이터로 정하도록 하자. 순수학습 데이터로 SVM을 학습시킨 후, 검증 데이터의 MAPE를 최소화하는 가장 우수한 SVM 파라미터를 구하면

$$\sigma = 2^2, C = 2^{20}$$

이다. 이 파라미터를 이용하여 모든 학습데이터에 대해서 다시 SVM을 학습시킨 후, 2002년 8월 데이터의 MAPE를 계산하여 다음을 얻을 수 있다:

$$MAPE = 2.5871$$

따라서, 최상의 MAPE는 2.4706일 때, 교차검증에 의해 현실적으로 가능한 최상의 MAPE는 2.5871임을 알 수 있다. 이는 최상의 MAPE를 기준으로 할 때 예측성능이

$$\frac{2.5871 - 2.4706}{2.4706} * 100 = 4.7\%$$

악화됨을 의미한다. 2002년 8월 전력수요 예측 결과를 그림 4에서 볼 수 있다.

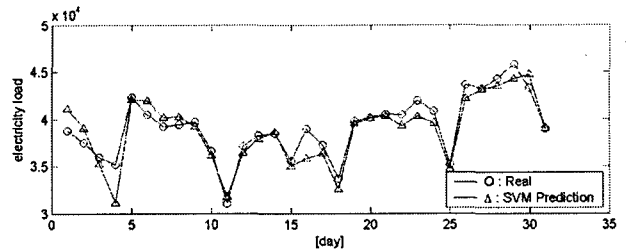


그림 4 2002년 8월의 SVM 예측결과
Fig. 4 SVM prediction results for August, 2002

4.3 사례연구 3: 2003년 7월 전력수요 예측

세 번째로, 학습시 사용가능한 데이터가 2001년 6, 7, 8, 9월, 2002년 6, 7, 8, 9월, 2003년 6월 데이터일 때, 2003년 7월의 하루 최대 전력수요량을 예측하는데 교차검증을 사용하여 보았다. 앞서와 마찬가지로 방법을 사용하면 2003년 7월 데이터의 MAPE를 가장 최소로 하는 SVM 파라미터는 $\sigma = 2^2, C = 2^{19}$ 이고 MAPE=1.7265이다. 이제, 2002년 7월 자료를 검증 데이터로 정하면, 검증 데이터의 MAPE를 최소화하는 SVM 파라미터는 $\sigma = 2^6, C = 2^{25}$ 이다. 이 파라미터를 이용하여 모든 학습데이터에 대해서 다시 SVM을 학습시킨 후, 2003년 7월 데이터의 MAPE를 계산하면 MAPE=1.7853이다. 따라서, 최상의 MAPE를 기준으로 할 때 예측성능이

$$\frac{1.7853 - 1.7265}{1.7265} * 100 = 3.4\%$$

악화됨을 의미한다. 2003년 7월 전력수요 예측 결과를 그림 5에서 볼 수 있다.

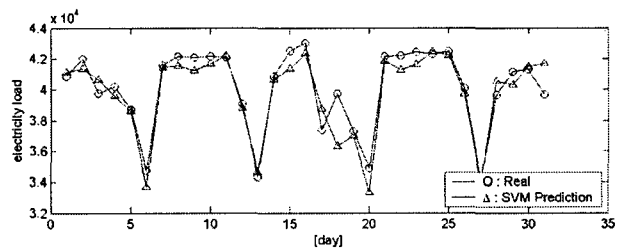


그림 5 2003년 7월의 SVM 예측결과
Fig. 5 SVM prediction results for July, 2003

4.4 사례연구 4: 2003년 8월 전력수요 예측

네 번째로, 학습시 사용가능한 데이터가 2001년 6, 7, 8, 9월, 2002년 6, 7, 8, 9월, 2003년 6, 7월 데이터일 때, 2003년 8월의 하루 최대 전력수요량을 예측하는데 교차검증을 사용하여 보았다. 2003년 8월 데이터의 MAPE를 가장 최소로 하는 SVM 파라미터는 $\sigma=2^6, C=2^{27}$ 이고 MAPE=2.6483이다. 2002년 8월 자료를 검증 데이터로 정하고, 나머지를 순수학습 데이터로 정하면, 검증 데이터의 MAPE를 최소화하는 SVM 파라미터는 $\sigma=2^4, C=2^{24}$ 이다. 이 파라미터를 이용하여 다시 2003년 8월 데이터의 MAPE를 계산하면 MAPE=2.7158 이다. 따라서, 최상의 MAPE를 기준으로 할 때 예측성능이

$$\frac{2.7158 - 2.6483}{2.6483} * 100 = 2.5\%$$

악화됨을 의미한다. 그림 6은 2003년 8월 예측 결과를 보여준다.

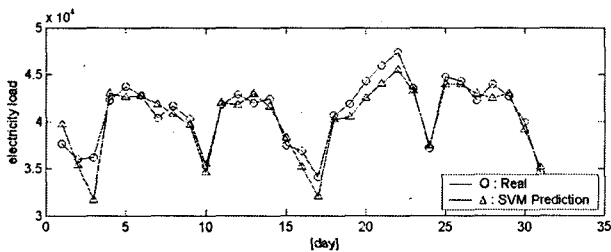


그림 6 2003년 8월의 SVM 예측결과

Fig. 6 SVM prediction results for August, 2003

4.5 사례연구 5: 2002년 여름과 2003년 여름의 전력수요 예측

마지막으로, 기존의 SVM 전력수요예측 연구결과인 [10]에서와 같이 다음 기간의 하루 최대 전력수요량을 예측하여 보자.

- ① 2002년 여름 (7월, 8월)
- ② 2003년 여름 (7월, 8월)

이때, 학습시 사용가능한 데이터는 다음과 같다고 하자.

- ① 2002년 여름 (7월, 8월) 예측시 사용한 학습데이터
 - 2000년 6월, 7월, 8월, 9월
 - 2001년 6월, 7월, 8월, 9월
 - 2002년 6월
- ② 2003년 여름 (7월, 8월) 예측시 사용한 학습데이터
 - 2000년 6월, 7월, 8월, 9월
 - 2001년 6월, 7월, 8월, 9월
 - 2002년 6월, 7월, 8월, 9월
 - 2003년 6월

우선, 학습 데이터를 모두 이용하여 SVM을 학습시킨 후, 예측기간의 MAPE를 최소로 하는 가장 우수한 SVM 파라미터와 예측성능은 표 4와 같다. 이제, 교차검증을 적용하여 실제 사용가능한 예측시스템의 예측성능을 계산해보자. 먼저, 2002년 여름 수요예측을 위해서는 2001년 7월과 8월 자료를, 2003년 여름 수요예측을 위해서는 2002년 7월과 8월 자료를 각각 검증 데이터로 정한다. 검증 데이터의 MAPE를 가장 최소로 하는 SVM 파라미터와 이를 사용한 SVM 예측성능은 표 5와 같다.

표 4 최적 SVM 파라미터를 선택할 때의 예측성능

Table 4 Prediction performance when optimal SVM parameters were chosen

	σ	C	MAPE
2002년	2^4	2^{23}	2.92
2003년	2^5	2^{29}	2.22

표 5 교차검증에 의해 SVM 파라미터를 선택할 때의 예측성능

Table 5 Prediction performance when SVM parameters were chosen by cross validation

	σ	C	MAPE
2002년	2^4	2^{26}	2.97
2003년	2^4	2^{23}	2.24

앞서와 마찬가지로 최상의 MAPE와 교차검증에 의한 MAPE의 차이는 매우 작음을 알 수 있고, 이로부터 교차검증이 전력수요예측을 위한 SVM 예측에서도 매우 효과적으로 사용될 수 있음을 알 수 있다. 한편, 기존 SVM 연구결과인 [10]에서는 $C=10^7, \sigma=5, 10, 20, 30$ 의 4가지 경우에 대해서 MAPE를 계산하고 $C=10^7, \sigma=20$ 인 경우에 2002년과 2003년 MAPE가 각각 2.93과 2.25이라는 것이 보고 되었다. 이를 표5의 결과와 비교하면, [10]의 결과가 약간 더 우수해 보이지만, [10]에서는 이러한 SVM 파라미터를 선택할 수 있는 방법이 명확히 제시되지 못하였다. 따라서, [10]의 결과만으로는 실제 상황에 사용하기에는 상당한 제약이 있다는 것을 알 수 있다.

5. 결 론

본 논문에서는 전력수요 예측에 사용되는 SVM의 효과적인 파라미터 선정방법에 대한 연구를 수행하였다. SVM은 기존의 통계적 학습 방법에서 이용되는 '경험적 위험 최소화'와는 다른 '구조적 위험 최소화'를 이용하여 일반화 오차를 감소시키는 기법이다. 하지만, SVM의 예측성능을 극대화하기 위해서는 SVM 파라미터를 효과적으로 선택하는 방법에 대한 연구가 필요하다. 기존의 SVM 전력수요예측기법에서는 SVM 파라미터를 선택하는 체계적인 방법이 제시되지 않아서 실제적인 응용에는 상당한 한계를 안고 있었다. 본 논문에서는 SVM 파라미터를 체계적으로 선택하기 위하여, 통계학에서 사용되는 교차검증을 SVM 전력수요예측에 사용할 것을 제안하였다. 4장의 5가지 사례연구 결과로부터 교차검증에 의해 선택된 SVM 파라미터를 사용하였을 때의 MAPE와 최적의 SVM 파라미터를 사용하였을 때의 MAPE 오차는 5%이내임을 알 수 있었다. 이는 교차검증에 의해서 SVM 파라미터를 효과적으로 선택할 수 있음을 실험적으로 입증한 것이다.

감사의 글

본 연구는 숭실대학교 교내연구비 지원으로 이루어 졌음

참 고 문 헌

- [1] D.G. Infield and D.C. Hill, "Optimal smoothing for trend removal in short term electricity demand forecasting," *IEEE Trans. Power Systems*, vol. 13, no. 3, pp. 1115-1120, 1998.
- [2] J.H. Park, Y.M. Park, and K.Y. Lee, "Composite modeling for adaptive short-term load forecasting," *IEEE Trans. Power Systems*, vol. 6, no. 2, pp. 450-457, 1991.
- [3] J.W. Taylor and S. Majithia, "Using combined forecasts with changing weights for electricity demand profiling," *J. Oper. Res. Soc.*, vol. 51, no. 1, pp. 72-82, 2000.
- [4] S.E. Papadakis, J.B. Theocharis, S.J. Kiartzis, and A.G. Bakirtzis, "A novel approach to short-term load forecasting using fuzzy neural networks," *IEEE Trans. Power Systems*, vol. 13, no. 2, pp. 480-492, 1998.
- [5] Kyung-Bin Song, Young-Sik Baek, Dug Hun Hong, and Jang, G., "Short-term load forecasting for the holidays using fuzzy linear regression method," *IEEE Trans. Power Systems*, vol. 20, no. 1, pp. 96-101, 2005.
- [6] H.S. Hippert, C.E. Pedriera, and R.C. Souza, "Neural Networks for Short-term Load Forecasting: A Review and Evaluation," *IEEE Trans. Power Systems*, vol. 16, no. 1, pp. 44-55, 2001.
- [7] Reis, A.J.R. and da Silva, A.P.A., "Feature extraction via multiresolution analysis for short-term load forecasting," *IEEE Trans. Power Systems*, vol. 20, no. 1, pp. 189-198, 2005.
- [8] C.H.C. Burges, "A tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, pp. 121-167, 1998.
- [9] N. Cristianini and J. Shawe-Taylor, "An introduction to Support Vector Machines," Cambridge University press, 2003.
- [10] 조남훈, 송경빈, 노영수, 강대승, "지원벡터머신을 이용한 단기전력 수요예측에 관한 연구," *대한전기학회 논문지*, Vol. 55A, No. 7, pp. 306-312, July, 2006
- [11] 하성관, 송경빈, 김홍래, "신경회로망과 하절기 온도 민감도를 이용한 단기 전력 수요 예측," *대한전기학회 논문지*, Vol. 54A, No. 6, pp. 259-266, Jun., 2005.
- [12] 지평식, 남상천, 임재운, 김정훈, "분류된 부하패턴을 근거로 한 단기 전력 수요 예측," *대한전기학회 논문지*, Vol. 47, No. 3, pp. 269-275, March, 1998
- [13] S. Haykin, *Neural Networks*, New Jersey: Prentice-Hall, 1999

저 자 소 개



조 남 훈 (趙南熏)

1992년 서울대 공대 전기공학과 졸업.
 2000년 서울대 대학원 전기공학부 졸업 (공학).
 2000년~2001년 서울대 자동화시스템공동연구소 연구원.
 2001년~2002년 삼성전자 DVS사업부 책임연구원.
 2002년~현재 숭실대학교 전기공학부 조교수.
 2006년~현재 대한전기학회 영문논문지 편집위원,
 2004년~현재 제어자동화시스템공학회 국문논문지 편집위원

Tel : 02-820-0643

E-mail : nhjo@ssu.ac.kr,

professor.jo@gmail.com

Homepage : <http://ee.ssu.ac.kr/~nhjo/>