

Support Vector Regression에 기반한 전력 수요 예측

이형로 · 신현정[†]

아주대학교 산업공학과

Electricity Demand Forecasting based on Support Vector Regression

HyoungRo Lee · Hyunjung Shin

Dept. of Industrial Engineering, Ajou University

Forecasting of electricity demand have difficulty in adapting to abrupt weather changes along with a radical shift in major regional and global climates. This has lead to increasing attention to research on the immediate and accurate forecasting model. Technically, this implies that a model requires only a few input variables all of which are easily obtainable, and its predictive performance is comparable with other competing models. To meet the ends, this paper presents an energy demand forecasting model that uses the variable selection or extraction methods of data mining to select only relevant input variables, and employs support vector regression method for accurate prediction. Also, it proposes a novel performance measure for time-series prediction, shift index, followed by description on preprocessing procedure. A comparative evaluation of the proposed method with other representative data mining models such as an auto-regression model, an artificial neural network model, an ordinary support vector regression model was carried out for obtaining the forecast of monthly electricity demand from 2000 to 2008 based on data provided by Korea Energy Economics Institute. Among the models tested, the proposed method was shown promising results than others.

Keyword: artificial neural network, energy, electricity demand forecasting, feature extraction, variable selection, support vector regression

1. 서론

전력 수요 예측은 전력거래소의 급전계획, 즉 가격결정 계획과 운영발전계획을 위하여 필수적인 요소이다. 전력수급 계획 측면에서 볼 때, 소비자에게 안정적으로 전력을 공급하기 위해서는 예비전력이 항상 준비되어야만 한다. 그러나 전기는 저장할 수 없기 때문에 전력 수요량을 미리 예측하고 이에 따라 전력을 생산하게 된다. 만약 전력 수요의 예측값이 정확하지 않아 전력을 실제 수요보다 대량 생산했을 경우 예비전력의 과다로 인한 비용의 낭비가 발생하며, 실제 수요보다 소

량 생산했을 경우 예비전력의 부족으로 인한 전력 수급 비상 또는 단전의 문제가 발생하게 된다(Han, 2009). 전력거래소의 급전계획과 관련해서는 시간대별 전력 수요의 예측이 중심이 되고 있다(Harvey *et al.*, 1993; Khotanzad *et al.*, 1996).

한편으로는 전력 수요에 대한 월별 수요 예측이 필요한 경우도 있다. 예를 들어 최근 지역난방공사의 열병합발전소 입지 선정의 경우를 들 수 있다. 열병합발전소는 대규모로 생산된 전기 및 열을 대단위 지역에 일괄 공급하는 시스템으로 이에 대한 지역난방 모형의 사업 허가신청 및 경제성 분석 시에는 중장기 전력 수요 예측 및 열 수요 예측을 필요로 한다. 이러

The authors would like to gratefully acknowledge support from Post Brain Korea 21 and the research grant from National Research Foundation of Korea (2010-0007804).

[†]연락처 : 신현정 교수, 443-749 경기 수원시 영통구 원천동 산 5번지 아주대학교 산업정보시스템공학부,

Fax : 031-219-1610, E-mail : shin@ajou.ac.kr

투고일(2011년 8월 31일), 심사일(1차 : 2011년 10월 11일, 2차 : 2011년 10월 31일), 게재확정일(2011년 11월 02일).

한 경우의 전력 수요 예측은 급전계획에서와 같은 시간대별 예측 보다는 월별 예측이 적절하다. 그러나 아직까지 이러한 목적으로 개발된 전력 수요 예측 모델에 대한 연구는 그 수가 드물다고 보여 진다.

전력 수요 예측 모델 개발과 관련하여 기술적으로는 다음과 같은 어려움이 있다. 최근 빈번하게 발생하는 이 이상 기후로 인한 폭서나 한파는 전력의 소비량을 급격하게 증가시킨다. 국내 전력시장의 경우에도 한여름의 폭염으로 인해 전력수급의 비상이 초래되는 상황이 비일비재하다. 기후에 영향을 많이 받는 전력 수요의 특성으로 미루어 볼 때, 이상기후 및 이상기온은 증장기 전력 수요 예측결과와의 불확실성을 높인다(Han *et al.*, 2009). 기상 관련된 원인 외에도, 급변하는 전력시장의 수요-공급 불균형, 불안정, 미래 상황에 대한 불확실성 등도 전력 수요 예측 모델의 성능의 안정성에 부정적 영향을 미치는 요인이라 할 수 있다. 따라서 이러한 어려움들을 우회하는 방안으로서 필요시 마다 신속하고 간편하게, 그리고 정확하게 수요예측을 수행할 수 있는 모델을 개발하는 것을 고려해 볼 수 있다. 이러한 조건을 만족시키기 위해서는 획득이 용이한 변수들로 모델의 입력변수를 구성하고 되도록이면 그 수를 최소화하는 것이 좋다. 또한 정확한 예측 값을 기대하기 위해서는 모델의 성능에 대한 우수성이 검증된 모델을 사용하는 것이 바람직하다.

따라서 본 연구에서는 앞에서 언급되었던 요구들을 충족시키는 월별 전력 수요 예측 모델을 제안하고자 한다. 제안하는 모델은 한국에너지경제연구원에서 제공한 지역난방공사의 월별 전기 에너지 데이터를 기반으로 전력 수요를 예측한다. 기술적인 측면에서는 다양한 데이터마이닝 기법들을 활용하여 모델을 구성한다. 우선, 특징추출(Feature Extraction) 기법과 변수선택(Variable Selection) 기법을 통하여, 사용 가능한 여러 종류의 변수들 중 예측에 중요한 요인이 되는 입력 변수들만을 선별한다. 예측 모델로는 예측 성능이 우수하다고 잘 알려진 Support Vector Regression(SVR) 모델을 사용한다. 마지막으로 모델의 시계열 예측성능에 대한 평가를 위하여 새로운 평가척도인 시프트인덱스(Shift Index)를 제안한다. 시프트인덱스는 기존의 시계열 예측성능 평가지표가 갖는 단점 즉, 실제에서는 사용할 수 없는 예측값을 산출하는 모델임에도 불구하고 좋다고 평가하는 오류를 보완할 수 있는 평가척도이다.

본 논문의 구성은 다음과 같다. 제 2장에서는 에너지 수요 예측에 관련한 기존 연구를 정리한다. 제 3장에서는 본 연구에서 제안하는 방법론의 기초 이론인 SVR과 이어 변수 선택 및 특징추출 방법에 대하여 소개한다. 이어, 본 연구에서 제안하는 전력 수요 예측 모델과 시프트 인덱스에 대하여 설명한다. 제 4장에서는 제안하는 모델의 객관적인 성능 검증을 위하여 전통적 시계열 예측기법 및 신경망과 비교 실험한 결과를 제시한다. 끝으로 제 5장에서는 제안한 방법의 개선점과 기대효과에 대하여 설명한다.

2. 기존 연구

기존 연구에서는 많은 논문들이 전력 수요 예측의 중요성에 대해 언급하고 있으며, 전력 수요를 예측하기 위한 다양한 입력변수와 방법론을 제시하고 있다. 많은 논문들에서 다양한 시계열분석방법을 제시하고 있으며 이 중 회귀모형(Regression)을 기반으로 한 예측 모델과 인공신경망(Artificial Neural Network, ANN)을 기반으로 한 예측 모델을 다수 제안하고 있다. 또한 이를 위해 활용한 입력변수가 매우 다양하게 나타나고 있는데 기존 연구에서 활용한 입력변수를 큰 카테고리 나눈다면 사회 · 경제적 요소, 시간 · 계절적 요소, 기후적 요소, 전력시장 요소로 나누어 볼 수 있다.

Kong *et al.*(2010)은 건구온도와 상대습도, 풍속, 일사량을 입력변수로 ANN 모델을 적용하고 이를 시뮬링으로 구현하여 도시계획을 위한 전력 수요 예측을 수행하였고 Nam *et al.*(2008)은 1인당 지역총생산을 입력변수로 다중회귀분석을 이용한 지역별 전력 수요량 예측을 수행하였다. Wi *et al.*(2007)은 기온과 전력 수요량을 입력변수로 하여 특수일 전력 수요를 예측하기 위해 온도효과를 고려한 데이터추출법을 적용한 다중회귀분석법을 제안하고 있다. Mirasgedis(2006)는 분야별 전력 수요를 예측하기 위해 분야별 전력 수요량을 입력변수로 사용하여 변수간의 상관관계를 제거한 자기회귀모델(Auto Regression, AR)을 제안하였으며, Romera *et al.*(2008)은 스페인 전력 수요 예측을 위해 스페인의 월별 전력 수요량을 입력변수로 사용하여 푸리에 계수를 활용한 ANN을 제안하였다. Choi *et al.*(2001)은 일별 전력 수요를 예측하기 위해 ANN 모델을 사용하였으며 기온을 입력변수로 사용하였다. Yang *et al.*(2005)은 기온 관련 변수와 냉각기의 상태지표를 입력변수로 하여 실시간으로 빌딩 내의 전력 수요를 예측할 수 있는 ANN 모델을 제안하고 있다. Pappas *et al.*(2010)은 일별 전력 수요량을 활용하여 Multi-Model Partitioning Theory를 적용한 AutoRegressive Moving Average (ARMA) 모델을 제안하였다. Srinivasan(2008)은 분야별 월별 전력 수요량을 입력변수로 하여 Group Method of Data Handling (GMDH) Network를 활용한 전력 수요 예측을 제안하였으며, Hong(2009)은 대만의 지역별 전력 수요를 입력변수로 하여 면역알고리즘에 기반한 SVR 모델을 제안하였다.

본 연구에서는 기존 연구에서 제안된 다양한 카테고리에 포함된 다양한 변수들을 입력변수로 활용하고, 방법론으로 활용되었던 회귀분석기반 모델인 AR과 ANN, SVR 모델을 비교·분석한다. 또한 다양한 변수 중 중요한 요인이 되는 변수를 선별하고자 변수선택 방법과 특징추출 방법을 적용한다.

3. 방법론

전력 수요를 예측하기 위한 예측 모델로 현재 데이터마이닝 분야에서 각광 받고 있는 SVR 모델을 활용하였다. SVR은 다양

한 분야에서 시계열 예측을 위해 활용되고 있다(Thissen *et al.*, 2003; Wenwu *et al.*, 2008). 하지만 전력 수요 예측에 활용된 사례는 많지 않기 때문에 본 연구에서는 SVR을 전력 수요 예측에 도입하여 예측 모델로 활용하였다.

또한 예측 모델의 입력변수로 활용되는 변수들을 선택 및 특징 추출하기 위하여 변수선택 및 특징추출 기법을 적용하였다. 입력변수들은 각각 전처리 과정을 거쳐 정제되며, 최종적으로 얻어진 예측 값은 MAPE와 본 연구에서 제안하는 시프트 인덱스를 통해 평가하였다.

3.1 Support Vector Regression

Support Vector Machine(SVM)은 분류(Classification)와 예측(Regression)에 응용할 수 있는 데이터마이닝 알고리즘이다. 크게 분류를 위한 Support Vector Classification(SVC)과 예측을 위한 Support Vector Regression(SVR)으로 나뉜다(Sholkopf *et al.*, 2002). 기본적으로 SVC는 다차원으로 구성된 데이터의 입력공간에 마진을 최대화 하는 초평면(Hyperplane)을 구하고, 이 초평면을 결정함수로 사용하여 미지의 데이터가 어떤 그룹에 속할지를 예측한다(Park, 2006). l 개의 학습데이터 $(x_1, y_1), \dots, (x_n, y_n)$ 가 있고 $x^i \in R^n$, $y_i \in 1, -1$ 일 때, SVC는 학습데이터를 레이블에 따라 두 개의 클래스로 분할하는 초평면을 찾는다. 이때, 초평면 $f(x) = w^T x + b$ 가 SVC의 결정함수가 된다. 따라서 SVC의 결정함수의 w 와 b 를 구하는 문제를 수리계획법으로 모형화하면 다음과 같이 이차계획법 문제로 모델링된다.

$$\begin{aligned} \max_{w,b,\xi,\xi^*} \quad & \frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i \\ \text{subject to} \quad & y_i (w^T x_i + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0, \quad i = 1, \dots, n \end{aligned} \quad (1)$$

회귀를 위한 SVR은 임의의 실수값을 예측 할 수 있도록 SVC를 일반화한 방법이다. SVC의 목적함수에 있는 $\sum_i \xi_i$ 대신에 식 (2)의 ϵ -무감도 손실함수(ϵ -intensive loss function) $L_\epsilon(x, y, f)$ 를 사용한다.

$$\begin{aligned} L_\epsilon(x, y, f) &= \max(0, |y - f(x)| - \epsilon) \\ &= \max(0, |y - (w^T x + b)| - \epsilon) \end{aligned} \quad (2)$$

ϵ -무감도 손실함수는 실제값 y 와 예측된 값 $f(x) = w^T x + b$ 간의 오차가 ϵ 이하면 $L_\epsilon(x, y, f) = 0$ 이고 ϵ 보다 크면 오차의 절대값에서 ϵ 만큼 차감한 값인 $L_\epsilon(x, y, f) = \max(0, |y - f(x)| - \epsilon)$ 를 갖게 된다. 이를 그림으로 나타내면 다음 <그림 1>과 같다.

즉, SVR에서는 실제 값과 예측 값을 가능한 한 ϵ 이내로 유지하면서 마진을 최대화하게 된다. 이를 수리모형으로 나타내

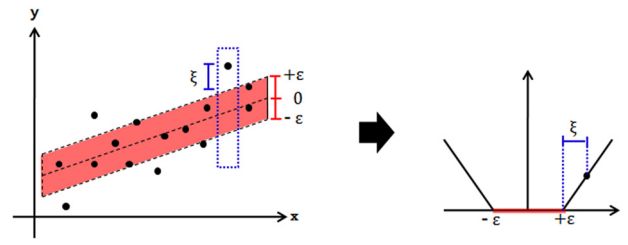


그림 1. ϵ -인텐시브 손실함수

면 $x_i \in R^m$ 이 입력변수이고 $y_i \in R^1$ 이 출력변수인 주어진 데이터 집합 $(x_1, y_1), \dots, (x_n, y_n)$ 에서 SVR의 표준화된 수식은 다음 식 (3)과 같다(Cristianini *et al.*, 2000).

$$\begin{aligned} \max_{w,b,\xi,\xi^*} \quad & \frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i + C \sum_{i=1}^n \xi_i^* \\ \text{subject to} \quad & w^T \phi(x_i) + b - y_i \leq \epsilon + \xi_i, \\ & y_i - w^T \phi(x_i) - b - y_i \leq \epsilon + \xi_i^* \\ & \xi_i, \xi_i^* \geq 0, \quad i = 1, \dots, n \end{aligned} \quad (3)$$

쌍대문제는 다음 식 (4)와 같이 표현된다.

$$\begin{aligned} \min_{\alpha,\alpha^*} \quad & \frac{1}{2} (\alpha - \alpha^*)^T Q (\alpha - \alpha^*) \\ & + \epsilon \sum_{i=1}^n (\alpha - \alpha^*) + \epsilon \sum_{i=1}^n y_i (\alpha - \alpha^*) \\ \text{subject to} \quad & \sum_{i=1}^l (\alpha - \alpha^*) = 0, \quad 0 \leq \alpha_i, \\ & \alpha_i^* \leq C, \quad i = 1, \dots, n \end{aligned} \quad (4)$$

SVR도 SVC에서와 같이 커널함수를 사용하여 학습 데이터를 특징공간의 점으로 변화시킨 다음 특징공간에서 학습을 수행하게 되는데 이때 특징공간에서 나타나는 $Q_{ij} = \phi(x_i)^T \phi(x_j)$ 는 커널함수 $K(x_i, x_j)$ 를 사용하여 계산하게 된다.

3.2 특징추출 및 변수선택

다양한 변수를 예측 모델에 적용할 경우 변수간의 상관성이나 중복된 정보로 인하여 예측 모델에 잡음(Noise)이 발생하여 정교한 예측이 불가능하거나 복잡하고 무거운 모델을 구성한다는 단점이 있다. 이러한 잡음을 제거하기 위한 방법으로 특징추출과 변수선택 방법을 적용한다. 특징추출은 다양한 변수가 존재할 때 입력변수로부터 출력변수를 표현하는 공통적인 성질을 추출하여 선형 또는 비선형 조합으로 새로운 변수를 만드는 방법이다. 특징추출의 대표적인 방법론으로는 Principal Component Analysis(PCA), Kernel Principal Component Analysis(KPCA), Nonlinear Principal Component Analysis(NLPCA)가 있다. PCA는 데이터를 한 개의 축으로 사상시켰을 때 분산의 크기를

기준으로 새로운 좌표계를 형성하여 데이터를 선형변환 시키는 방법이다. 이들 주성분은 고유벡터(Eigen Vector)로 나타내어지며 많은 수의 입력변수들을 일련의 소수의 요인들에 의해 설명이 가능한 알고리즘이다. KPCA는 PCA와 동일하게 주어진 데이터의 분포를 반영한 새로운 기저(basis)를 찾아내어 데이터를 표현하는 동일한 원리를 이용하지만, 입력공간(Input Space)상이 아닌 특징공간(Feature Space)에서 주성분변수를 얻는다는 차이점이 있다. NLPKA의 방법론인 자기연상신경망(Auto-Association Neural Network, AANN)은 신경망(Neural Network)의 일종으로 입력층과 출력층, 그리고 은닉층으로 구성된다. AANN은 입력층의 데이터가 출력층에 재구성되도록 하는 동일 사상을 수행토록 신경망을 학습시켜, 실제 데이터 분포에 닮은 특징공간을 생성함으로써 차원을 축소시키는 방법이다 (Park et al., 2011). 변수선택은 다수의 특징 중에서 출력변수를 표현하는데 필요 충분하다고 생각되는 소수의 변수를 선택하는 방법이다. 대표적인 방법으로는 단계적 회귀분석(Stepwise Regression Analysis)과 상관분석을 이용한 변수선택방법이 있다. 단계적 회귀분석은 출력변수와 상관관계가 가장 높으면서 다른 입력변수와 상관이 낮은 변수를 우선적으로 선택하고, 다음으로 선택되는 변수는 이미 선정된 변수가 예측하지 못하는 출력 변수의 변량을 예측해 주는 변수를 선택하는 방법이다. 이러한 과정을 계속하여 새로운 선정 변수가 포함되었을 때, 이미 얻어진 R^2 를 의미 있게 증가시키지 못할 경우 변수의 선택은 완료된다. 상관분석을 활용한 변수선택 방법은 출력변수와 입력변수와의 상관계수(Correlation Coefficient)를 이용하여 변수를 선택하는 방법으로 상관분석을 통해 얻어진 상관계수를 이용한다. 상관계수는 두 변수 사이의 관련성의 강한 정도를 나타내며 -1과 +1 사이의 값으로 나타낸다(Hall, 1999). R^2 는 상관계수를 제곱한 값으로 표현되기 때문에 모든 입력변수가 독립일 경우 두 방법의 효과는 동일하게 나타나지만 다중공선성이 존재하는 경우 변수의 선택에 차이가 있다.

3.3 시계열 데이터 처리방법

전력 수요 데이터는 시계열 데이터이므로 데이터마이닝 모델에 적용하기에 앞서 데이터에 대한 처리방법이 필요하다. 일반적인 다변량 시계열 모델의 형태는 $y_t = (x_1, x_n, y_{t-1}, y_{t-k})$ 로 표현할 수 있으며 선형모델을 가정할 경우 식 (5)와 같은 형태로 확장할 수 있다.

$$y_t = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n + \beta_{n+1} y_{t-1} + \beta_{n+2} y_{t-2} + \dots + \beta_{n+k} y_{t-k} \quad (5)$$

즉, 출력값은 각각의 입력변수와 가중치의 곱의 총합과 자기회귀값의 합으로 나타낼 수 있으며, 각각의 변수에 대한 가중치와 자기 회귀의 차수를 결정해 주어야 한다.

식 (5)에서 x 로 표현되는 입력변수는 모델에 적합한 형태로

입력되기 위해서 전처리과정을 거쳐야 한다. 입력변수로 활용되는 각 데이터들은 변수별로 모두 단위가 다르기 때문에 정확한 전력 수요 예측에 불안정한 요소로 작용할 수 있다. 따라서 모든 입력변수와 관련된 데이터는 각각의 단위에 따른 영향을 제거하여 데이터의 범위를 일치시키고 분포를 유사하게 만들어 척도를 맞추어 주기 위하여 최대-최소 정규화(Max-Min Normalization)를 수행한다. 정규화를 위한 수식은 식 (6)과 같이 나타낼 수 있다.

$$x'_t = \frac{(x - \min_{i=1, \dots, n} x_i)}{(\max_{i=1, \dots, n} x_i - \min_{i=1, \dots, n} x_i)} \quad (6)$$

식 (6)에서 max와 min은 각 변수의 최소, 최대값이며 정규화를 통해 각 변수는 0~1사이의 값을 갖도록 변환된다. 식 (5)에서 y 로 표현되는 출력변수는 정규화된 입력변수에 비하여 매우 큰 값을 가지게 되므로 입력변수의 변화가 출력변수에 미미한 영향만을 미치게 된다. 따라서 출력변수의 척도를 맞추어 주기 위한 방법으로 전력 수요량을 증가율의 형태로 나타내어 입력변수와 출력변수의 척도를 유사하게 변환한다. 이를 위한 수식은 다음 식 (7)과 같이 나타낼 수 있다.

$$y'_t = \frac{y_t}{y_{t-1}} \quad (7)$$

여기서 y'_t 는 변환된 출력변수를 나타내며 y_t 는 변환되기 전의 출력변수를 나타낸다. 이러한 변환을 통해 목적변수인 출력변수가 증가율의 형태로 변환되면 예측 값 역시 증가율의 형태로 출력된다. 따라서 변환된 형태로 출력된 예측 값을 실제의 값과 같은 척도로 재변환하여야 한다. 이를 위한 수식은 다음과 같이 표현할 수 있다. 이때, f'_t 는 변환된 예측 값을 나타내며 f_t 는 변환되기 전의 예측값을 나타낸다.

$$f'_t = f_t \times y_{t-1} \quad (8)$$

또한 식 (5)에서 $\beta_{n+1} y_{t-1} + \beta_{n+2} y_{t-2} + \dots + \beta_{n+k} y_{t-k}$ 로 표현되는 자기회귀시점을 결정하여야 한다. 시계열 데이터의 특성상 이전 시점의 데이터에 영향을 받으므로 현재 시점의 전력 수요를 예측 할 때, 이전 시점의 데이터를 변수에 포함하게 되면 최근 추세를 반영할 수 있을 뿐만 아니라 좀 더 정확한 예측을 수행 할 수 있다. 따라서 본 연구에서는 자기 회귀변수를 고려하여 분석을 수행하였으며 입력변수에 이전 시점의 데이터들을 포함하였다. 고려할 시점의 길이는 평가용 데이터셋을 통하여 결정하며 이는 제 4.3절에서 언급한다.

3.4 시계열 모델의 정확도 평가방법

예측모델의 성능검증을 위한 성능지표로 일정한 트렌드를

반영하는 시계열값에 적합한 정확도의 척도로서 Mean Absolute Percentage Error(MAPE)를 사용한다. 예측 값의 정확도는 예측 오차에 의해서 결정되며 오차가 작으면 작을수록 정확도가 높아진다. 일반적으로 시계열 모델에서 정확도를 측정하는 방법으로 다양한 방법이 존재하나, 이 중 Mean Absolute Percentage Error(MAPE)는 모델 간의 예측오차의 비교가 용이하고, 그 신뢰도가 높기 때문에 자주 사용되고 있다. MAPE는 실제값에 대한 예측 값의 평균 절대 에러율을 표현하며 수식은 다음 식 (9)과 같이 정의된다.

$$MAPE = \frac{1}{n} \sum_i \left| \frac{y_i - f_i}{y_i} \right| \times 100 \quad (9)$$

수식에서 y_i 는 실제 값, f_i 는 예측 값을 나타내며 n 은 성능 지표를 통해 검증할 기간에서 관측되는 데이터의 개수를 의미하며 MAPE는 실제 값과 예측 값의 차이의 절대 값을 실제 값으로 나누어 준 후, 에러율의 형태로 변환한 값을 의미한다. 본 연구에서는 비교 실험 및 특징추출 기법과 변수선택 기법을 적용한 각 모델에서의 성능 지표로 MAPE를 사용하여 성능을 평가한다. 일반적으로 MAPE는 낮을수록 우수한 성능을 나타내나 낮은 MAPE 만으로 시계열 모델의 성능을 평가하기에는 한계가 있다. 시계열 예측 값이 이전 시점의 값을 계속해서 따라가는 형태를 보인다고 해도 MAPE로는 이를 식별할 수 없다는 단점이 존재하기 때문이다. 즉, 실제에서는 사용할 수 없는 예측값을 산출하는 모델도 낮은 MAPE 값을 갖게 된다. 이러한 모델은 그래프 상에서 실제값을 나타내는 실선을 우측으로 한 시점 이동시킨 형태를 띠게 된다. 따라서 이러한 단점을 보완하기 위하여 본 연구에서는 이를 식별할 수 있는 평가 척도로서 시프트인덱스를 제안한다. 시프트인덱스의 수식은 아래의 식 (10)과 같이 표현된다.

$$\text{shift index} = \frac{|y_t - f_t|}{|y_{t-1} - f_t|} \quad (10)$$

예측값 f_t 와 이전 시점 y_{t-1} 과의 차이가 현시점 y_t 와의 차이보다 클 경우 1 이하의 값을 갖게 되고 반대의 경우 1 이상의 값을 갖게 된다. 따라서 1 이상의 값을 갖게 될 경우 전 시점의 값을 따라간다고 할 수 있으므로 모델이 시프트 되는 경향이 있다는 것을 알 수 있다.

4. 실험 및 결과

4.1 데이터

본 실험에서는 에너지경제연구원에서 제공하는 2000년부터 2008년까지의 월별 데이터를 사용하였다. 전력관련 변수는 한국지역난방공사 중앙사업소의 데이터를 사용하였으며 사회지표관련변수는 통계청의 성남 및 경기지역의 데이터를 사용하였다. 기온관련변수는 기상청 서울관측소의 월별 통계량을 사용하였다. 학습용 데이터셋으로는 5년 치 데이터인 60개 데이터를, 평가용 데이터셋으로는 1년 치 데이터인 12개를, 검증용 데이터셋으로는 3년 치인 36개 데이터를 사용하였다. 총 데이터의 개수는 108개이며 총 입력, 출력변수의 수는 23개이다. 정리된 데이터는 아래의 <표 1>과 같으며 연도별, 월별로 정리하였다.

전력 수요는 다양한 사회·경제적 요소, 시간 계절적 요소, 기후적 요소, 전력시장 요소, 기타 비선형적 요소에 의한 영향을 받는다. 따라서 이러한 다양한 비선형적인 특성을 고려한 예측 모델을 설계하고자 관련 변수들을 총 3개의 카테고리로 분류하였다. 각각의 카테고리는 전력관련변수, 사회지표 관련변수, 기온관련변수로 구성되며 각 카테고리에 해당되는 변수들의 데이터를 수집하여 예측 모델에 적용하였다. 입력변수와 출력변수를 포함한 총 14개의 변수를 사용하였으며 분석에 사용된 변수의 목록은 아래의 <표 2>를 통해 명시하였다. 출력변수는 전력 수요량($t+1$ 시점)이며 표에 있는 나머지 변수들은 예측변수(t 시점)로 사용되었다. 각각의 변수는 전처리과정을 거쳐 예측 모델에 적합한 형태로 변환하여 적용한다.

표 1. 입력데이터

년	월	전력관련변수		기온관련변수				사회지표관련변수		
		전력 수요량	판매 단가	기온	증발량	...	기압	인구수	...	경제활동 인구
		중앙 사업소	중앙 사업소	월 평균기온	월 평균증발량	...	월 평균기압	분당구	...	경기
2000	1	3,844,623	90.1	-2.0223	0.9194	...	1015.15	393,429	...	4,106,000
	2	4,088,291	91.48	-1.7062	1.5034	...	1012.05	393,429	...	4,106,000
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2002	9	3,755,383	88.27	21.5433	3.6167	...	1002.29	407,540	...	4,678,000
	10	3,537,398	89.84	12.7995	2.7968	...	1001.02	407,540	...	4,690,000
	11	3,358,728	93.33	4.0261	1.8367	...	997.15	407,540	...	4,690,000
2009	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	11	5,019,723	97.16	6.8985	1.7100	...	998.71	460,688	...	5,800,000
	12	5,188,454	97.16	-0.9976	1.1839	...	1002.74	460,688	...	5,800,000

표 2. 사용변수

구 분	변 수 명
전력관련변수	전력 수요량, 전기판매단가
사회지표 관련변수	인구수, GRDP*, 경제활동인구, 1인당 연간소비지출액
기온관련변수	기온, 증발량, 습도, 풍속, 일조량, 강수량, 전운량, 기압

* GRDP : Gross Regional Domestic Production.

4.2 실험 설정

본 연구에서 제안하는 변수선택을 적용한 SVR 모델의 타당성 검증에 위하여 다음과 같은 비교실험 절차를 구상하였다. 전체적인 실험의 흐름으로 예측 모델에 적합하게 전 처리한 데이터를 기반으로 각각의 최적의 파라미터를 선정 한 후 예측을 수행하는 순서로 진행된다. 비교를 위한 AR 모델은 시계열 데이터 분석을 위하여 사용되는 대표적인 통계기법으로 널리

알려져 있다(Akaike *et al.*, 1994; Kitagawa *et al.*, 1996). 또한 ANN 모델 역시 데이터마이닝 분야의 대표적인 알고리즘으로서 분류문제 및 시계열 데이터를 예측하는데 많이 사용되고 있다(Haykin, 1999; Bishop, 2006; Nrgaard *et al.*, 2006). 따라서 시계열 데이터인 전력 수요의 예측을 위해 세 모델들을 비교한 후 가장 우수한 성능을 보이는 모델을 기반으로 특징추출 및 선택 기법을 적용한다. 특징추출기법으로는 PCA와 NLPCA, KPCA의 세 가지 기법을 적용하며 특징선택기법의 기준으로는 R2와 Correlation의 두 가지 기준을 적용한다. 전체적인 프로세스를 간단한 그림으로 나타내면 다음 <그림 2>와 같다.

예측 모델을 위해 활용될 자기회귀변수의 시점은 총 12시점까지의 자기회귀 데이터를 예측 모델에 적용한 결과를 바탕으로 하여 결정하였으며 12시점 중에서 가장 낮은 MAPE를 기록한 시점을 실제모델에 적용하였다. 실험의 결과는 <그림 3>을 통해 나타내었다. 실험결과 12시점 중 9번째 시점에서 4.5418로 가장 낮은 MAPE를 기록하였으며 따라서 예측 모델에 들어갈 자기회귀변수의 시점으로 yt-9시점으로 결정하였다.

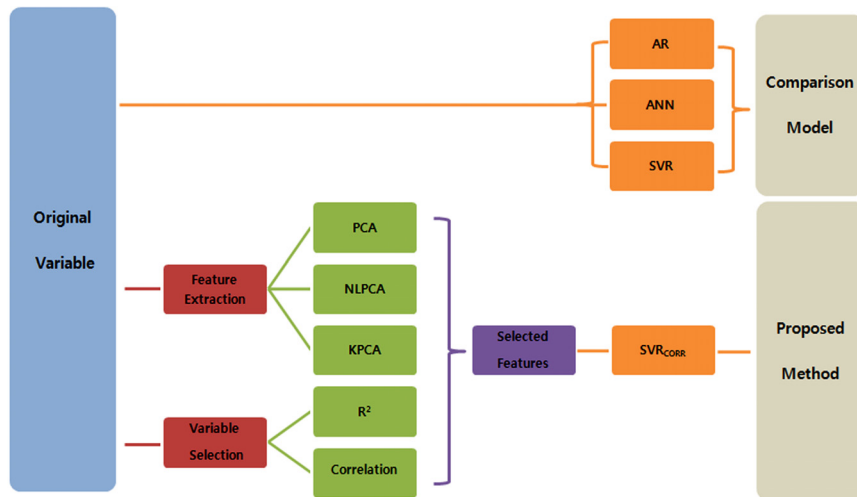


그림 2. 실험 절차

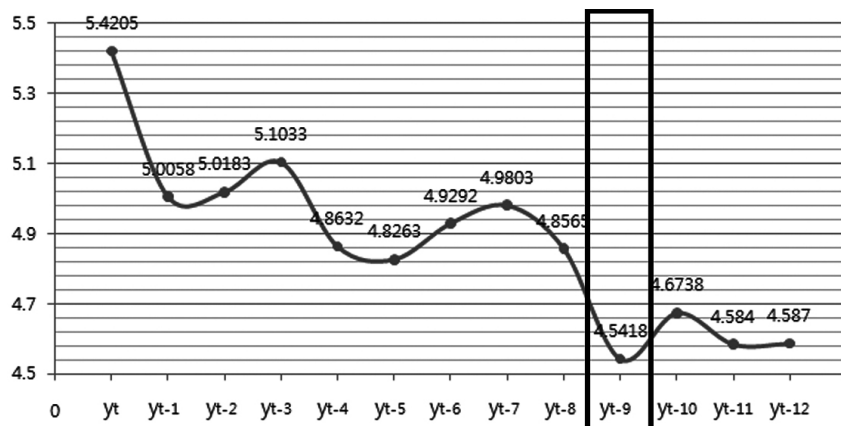


그림 3. 평가용 데이터셋을 이용한 자기회귀 차수 결정

예측 모델을 위한 총 14개의 변수를 선택하기 위한 기준으로는 R^2 , Correlation을 사용하였다. R^2 기준에서는 Squared Correlation이 0.005의 이상이거나 Stepwise R^2 improvement가 0.005 이상인 변수를 입력변수로 활용하였다. Correlation 기준에서는 Correlation이 0.15 이상이거나 P-Value가 0.2 이하인 변수를 입력변수로 활용하였다. 아래의 <표 3>은 변수선택에 대한 결과를 보여준다.

표 3. 변수선택 결과

	SVR _{R²}	SVR _{CORR}
변수의 수	6	4

예측 모델을 위한 총 14개의 특징을 추출하기 위한 기준으로는 R^2 , Correlation을 사용하였다. 예측 모델에 활용하기 위한 축약된 주성분 수는 고유값이 1 이상인 주성분만을 선택하는 것으로 기준을 정하였다. 아래의 <표 4>는 특징추출에 대한 결과를 보여준다.

표 4. 특징추출 결과

	PCA_SVR	KPCA_SVR	NLPCA_SVR
변수의 수	4	10	3

이 중 SVR_{CORR}이 MAPE 3.85로 가장 우수한 성능을 나타내었

으며 따라서 최종 결과에서는 SVR_{CORR}만을 포함하였다. 각각의 모델에 대한 최적의 파라미터는 각각의 파라미터의 값을 변경시켜 가면서 평가용 데이터에서 가장 오류가 적은 파라미터를 선택하는 파라미터 선택 과정을 통해 이루어진다. 각각의 최적 파라미터는 모델에 있어 최적의 성능을 나타내는 값으로 선택된다. 즉, 최적 파라미터는 각각의 예측 모델을 수행하였을 때 각 모델에 가장 적합한 값, 가장 낮은 MAPE 값을 나타내는 파라미터로 선택되며 최적 파라미터로 선정된 값들은 다음 <표 5>와 같다.

표 5. 최적 파라미터

모델	파라미터	값
AR	자기회귀시점	9
ANN	은닉 노드수	1
SVR	오차허용률(ϵ)	0.2
	RBF 커널 파라미터(σ)	0.6
SVR _{CORR}	오차허용률(ϵ)	0.3
	RBF 커널 파라미터(σ)	1

4.3 결과

AR, ANN, SVR 모델과 변수선택을 적용한 SVR_{CORR} 모델을 통해 얻어진 검증용 데이터셋의 예측 결과는 다음 <그림 4>와

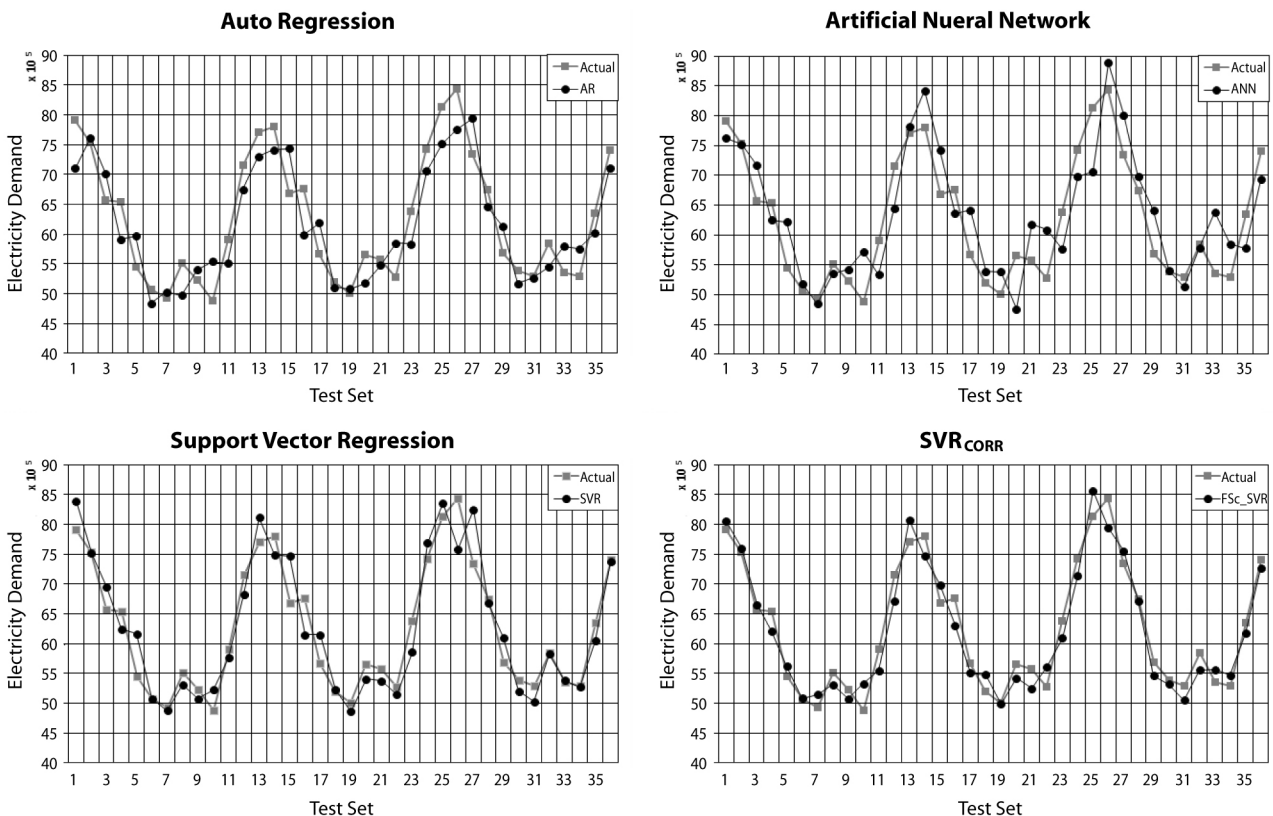


그림 4. 모델 별 시계열 적합

같이 보여진다. 회색으로 표현된 값은 검증용 데이터셋의 실제 전력 수요량을 나타내며, 검은색으로 표현된 값은 각각의 예측모델의 예측값을 나타낸다.

MAPE를 통한 성능 비교

AR, ANN 모델보다 SVR을 기반으로 한 모델이 실제의 전력 수요량 패턴을 잘 따라가는 것을 확인할 수 있으며, AR과 ANN 모델은 전반적으로 예측값이 오른쪽으로 시프트되는 경향을 확인할 수 있다. 검증용 데이터셋의 각각의 시점에 대한 절대 백분율오차(Absolute Percentage Error)와 MAPE, T-TEST에 관한 결과는 다음의 <표 6>과 같다.

전반적으로 AR과 ANN에 비하여 SVR을 기반으로 한 모델의 오차가 적은 것을 확인할 수 있으며, 특히 SVR_{CORR} 모델에서 오차가 가장 적은 것을 확인할 수 있다. 또한 MAPE도 SVR_{CORR}에서 가장 낮게 나타났으며 각각에 대한 표준편차 역시 SVR_{CORR}이 2.04로 가장 낮은 값으로 안정적인 성능을 나타내고 있다. 제한한 SVR_{CORR}이 가장 좋은 성능을 보이므로 다른 세 모델과의 성능의 차이를 T-TEST를 통해 통계적으로 검증하였다. T-TEST 결과로 SVR_{CORR}과 AR, ANN 모델과는 p-value가 0.00으로 유의한 값을 나타내어 모델 간의 성능 차이가 있음을 확인할 수 있다. 그러나 SVR 모델과는 P-value가 0.27로 유의하지 않은 결과를 보여 SVR_{CORR}이 MAPE에 있어서는 좀 더 우수한 성능을 나타내지만, 두 모델간의 큰 성능차이는 없음을 확인할 수 있다. 또한 <그림 5>를 통해 보이는 바와 같이 SVR_{CORR}에서 가장 낮은 MAPE를 보이는 것을 확인할 수 있다.

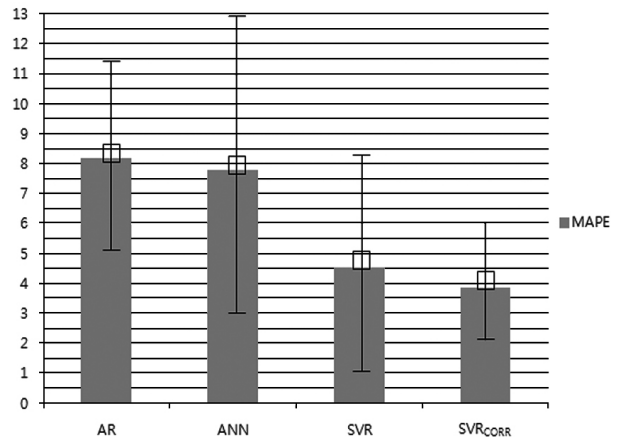


그림 5. MAPE를 통한 성능 비교

시프트지표를 통한 성능 비교

아래의 <그림 6>은 $y_{t-1} - f_t$ 를 x 축으로 $y_t - f_t$ 를 y 축으로 한 그래프를 나타낸다. x 축으로 산점이 많을 경우 예측 값이 실제의 값과 비슷한 경향이 크며 y 축으로 산점이 많은 경우에는 예측 값이 전 시점의 값과 비슷한 경향이 크다. 즉, x 축으로 산점이 많을수록 예측 값과 실제 값의 차이가 크지 않다는 것을 의미한다. y축에 가까운 산점이 많을 경우에는 MAPE 상의 정확도가 높다고 하더라도 실제로 예측 값이 시프트 될 가능성이 높다. 또한 그래프에서 각각의 산점은 l_a 아래쪽에 위치할수록 정확한 예측을 수행한 것이며, l_b 의 위쪽에 위치할 경우 시프트되는 경향이 크다.

표 6. 모델 비교 결과

	AR	ANN	SVR	SVR _{CORR}		AR	ANN	SVR	SVR _{CORR}
1	10.19	3.62	6.13	1.80	19	1.61	7.61	2.91	0.41
2	0.97	0.27	0.22	0.75	20	8.53	15.83	4.44	4.25
3	6.50	8.99	5.65	1.19	21	1.70	10.83	3.54	5.86
4	9.56	4.38	4.57	5.02	22	10.93	15.41	2.40	6.40
5	9.31	13.95	12.93	3.02	23	8.65	9.85	8.31	4.61
6	4.78	2.14	0.13	0.06	24	4.96	5.94	3.61	3.76
7	1.96	1.58	0.98	4.36	25	7.61	13.18	2.72	5.24
8	9.81	2.89	3.63	3.52	26	7.94	5.42	10.04	5.77
9	3.08	3.50	3.08	3.14	27	8.18	9.14	12.33	2.90
10	13.45	17.14	7.03	9.19	28	4.12	3.59	0.77	0.33
11	6.74	9.66	2.48	6.04	29	7.78	12.64	7.32	3.99
12	5.74	9.84	4.62	6.08	30	4.12	0.34	3.50	1.23
13	5.30	1.42	5.23	4.75	31	0.58	3.06	5.01	4.59
14	5.03	7.89	4.03	4.36	32	6.91	1.09	0.48	4.97
15	11.13	10.93	11.60	4.49	33	8.09	19.20	0.40	3.84
16	11.26	5.84	8.98	6.63	34	8.75	10.55	0.39	3.46
17	9.19	13.31	8.46	2.71	35	5.22	8.88	4.66	2.69
18	1.68	3.66	0.54	5.33	36	3.97	6.37	0.44	1.79
		AR	ANN	SVR	SVR _{CORR}				
MAPE		8.19±(3.31)	7.78±(5.12)	4.54±(3.58)	3.85±(2.04)				
T-TEST		0.00	0.00	0.269	.				

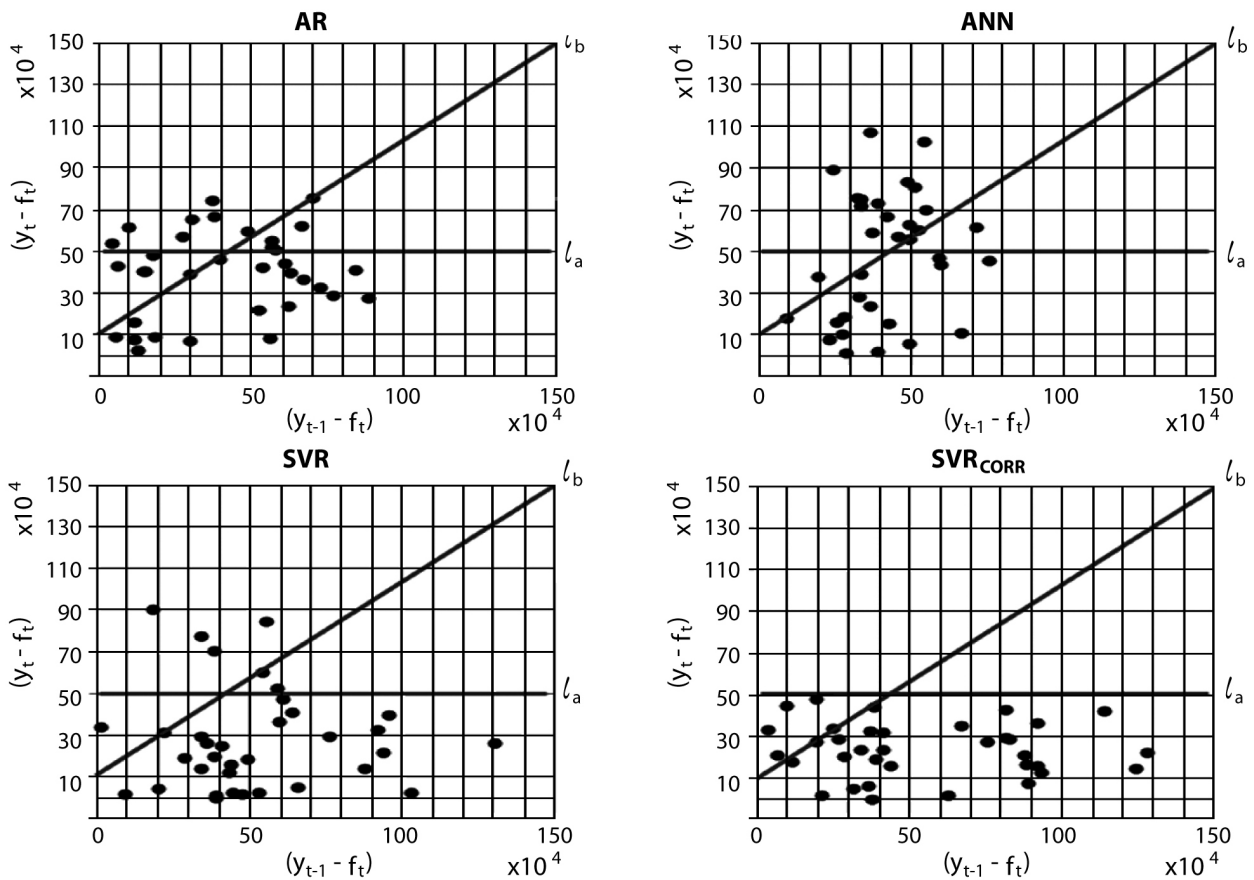


그림 6. 시프트 지표를 통한 성능 비교

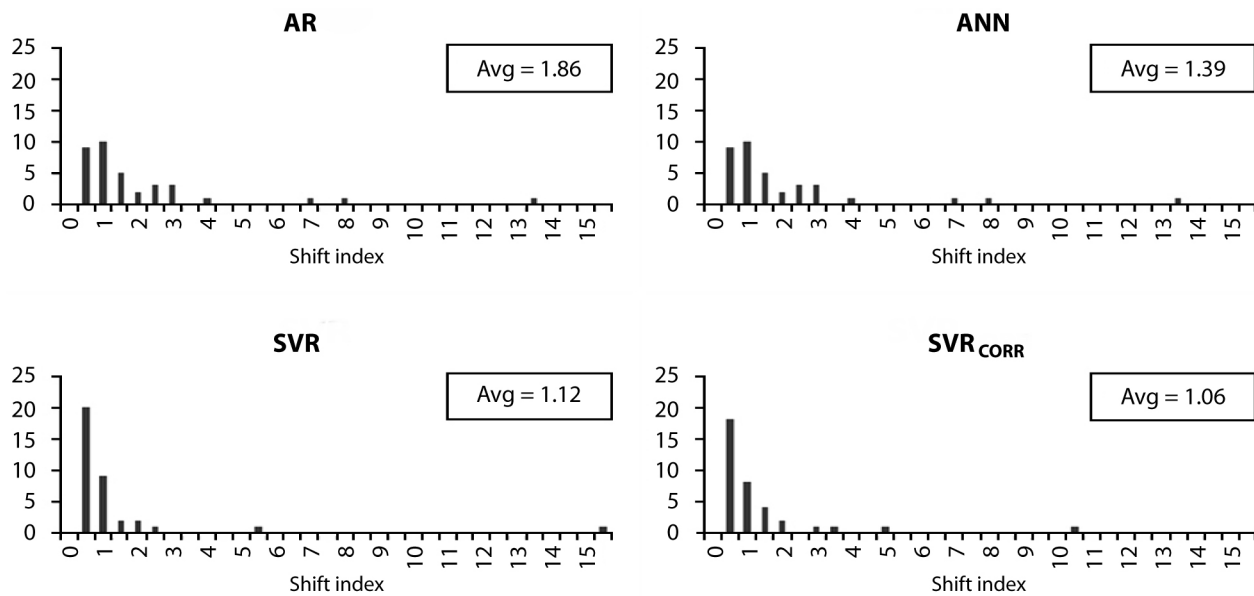


그림 7. 히스토그램

그래프를 통해 AR과 ANN 모델은 SVR을 기반으로 하는 모델보다 시프트되는 경향이 큰 것을 확인할 수 있으며 SVR_{CORR}

모델이 가장 정확한 예측을 수행하는 것을 확인할 수 있다. 다음의 <그림 7>은 시프트 지표를 통해 얻어진 값을 히스토그

램으로 표현한 그래프이다.

그래프를 통해 AR과 ANN에서는 큰 값이 많이 분포하므로 예측값이 시프트되는 경향이 크다는 것을 볼 수 있다. SVR과 SVR_{CORR}은 1보다 거의 대부분의 값들이 1보다 낮은 값을 포함하므로 앞에서 언급한 두 모델에 비해 시프트되는 경향이 낮다는 것을 확인할 수 있다.

SVR과 SVR_{CORR}은 정확도를 나타내는 지표인 MAPE 상으로 통계적으로 유의한 차이가 없고 시프트되는 정도 역시 비슷하다고 할 수 있다. 하지만 같은 성능을 보일 경우 취득이 용이한 입력 변수를 사용하며 적은 수의 변수를 사용하는 모델이 편의성과 즉각적인 수요예측을 수행하는데 유리하다. 따라서 24개의 변수를 사용하는 SVR에 비해 취득이 용이한 기온, 증발량, 풍속, 기압, 총 4개의 변수를 사용하는 SVR_{CORR} 모델이 더 우수하다고 할 수 있다.

5. 결론

전력 수요 예측은 전력거래소의 급전계획을 위한 필수적인 요소로서, 소비자에게 안정적인 전력 공급을 하기 위한 예비전력의 마련을 위해서는 정확한 전력 수요 예측이 필요하다. 이를 위해서는 시간대별 전력 수요 예측을 수행하여야 한다. 한편으로 지역난방모형의 사업 허가신청 및 경제성 분석 시에는 중장기 전력 수요 예측 및 열 수요 예측을 필요로 한다. 이러한 경우의 전력 수요 예측은 급전계획에서와 같은 시간대별 예측보다는 월별 예측이 적절하다. 따라서 본 연구에서는 이러한 상황 하에서 필요한 월별 전력 수요 예측 모델을 제안하였다. 기술적인 측면으로는 최근 이상기후로 인한 기후의 변화 하에서 적은 수의 입력변수를 활용하여 즉각적이고 정확한 예측을 수행하고자 하였다. 본 연구에서는 특징추출 기법과 변수선택 기법을 통하여 예측에 중요한 요인이 되는 입력 변수들을 선별하고 이를 SVR에 적용하여 분석을 수행하였다.

분석을 위해 제안한 모델과 다른 다양한 모델을 비교한 결과, SVR 모델에 Correlation 기준으로 변수선택을 수행한 모델이 가장 우수한 성능을 나타내었다. 특징추출을 통한 방법에서는 추출된 변수의 수가 원래의 변수의 수보다 적다고 하더라도 이를 계산하기 위해서는 원래 변수의 값이 모두 필요하다. 즉, 엄밀히 말해 관리해야 할 변수의 수가 줄었다고 할 수는 없으며, 단지 모델에 들어가는 데이터의 차원만 줄어든 것이다. 하지만 변수 선택의 경우 변수의 변환 없이 입력변수로 활용되는 변수의 개수만 줄어든 것이기 때문에 원래의 변수의 값이 모두 필요하지 않다. 본 연구에서 가장 우수한 성능을 나타낸 SVR_{CORR} 모델은 변수 선택을 통해 얻어진 변수를 입력변수로 사용하기 때문에 최종 선정된 4개의 변수만을 사용하여 예측이 가능하다.

또한 SVR_{CORR} 모델의 Support Vector의 수는 42개로 학습용 데이터셋의 총 개수인 60개 중 차지하는 비율이 70%이다. 전체

학습용 데이터 대비 Support Vector의 수가 너무 많아질 경우, 예측 모형이 무거워지고 단순한 시계열 예측 기법에 비해서 장점이 반감 될 수 있다. 하지만 본 연구에서는 학습용 데이터의 총 개수가 많지 않고, 과거 데이터가 누적되더라도 현재의 학습용 데이터셋으로 활용한 과거 5년 치의 데이터를 Time Window의 크기로 고정시킨다면 시프트되는 경향이 있는 시계열 모델에 비해서 정확도가 높은 결과를 보이기 때문에 효용성이 있다. SVR_{CORR} 모델은 MAPE 3.85, 즉 3.85%의 평균오차를 보였으며, MAPE의 표준편차가 2.04로 다른 모델에 비해 우수하면서 안정적인 성능을 나타내었다. 시프트지표를 이용한 분석에서도 예측 값이 시프트되지 않고 패턴을 잘 따라가는 경향을 확인할 수 있다.

결론적으로 본 연구에서는 적은 변수만을 가지고 최적의 결과를 산출하는 결과를 얻을 수 있었다. 전력 수요 예측을 위해 최종 선정된 입력변수의 수는 4개로 기온, 증발량, 풍속, 기압이 선정되었다. 선정된 변수는 그 개수가 적을 뿐 아니라 데이터의 취득이 용이하므로 즉각적인 전력 수요 예측을 가능하게 한다. 본 연구에서 제안한 모델은 적은 수의 입력변수만으로도 우수한 성능을 보여주었다.

나아가 본 연구에서 제안된 모델을 통하여 얻어진 변수와 예측 모델은 지역난방공사의 열병합발전소 입지 선정을 위한 사업 허가신청 및 경제성 분석을 위한 기초 자료로 활용될 수 있을 것으로 예상된다. 또한 본 연구에서 제안된 알고리즘은 시간대별 전력 수요 데이터를 적용하는 방향으로 확장 가능하다. 이러한 모델은 발전량 계획을 효과적으로 수행하고, 전력시스템의 안정화와 지역 전력시장운영을 위한 전력거래소의 급전계획 및 지역난방사업의 운영에도 활용될 수 있을 것이다.

추후에는 전력 공급자와 소비자가 양방향으로 실시간 정보를 교환함으로써 에너지 효율을 최적화하는 차세대 지능형 전력망인 스마트 그리드가 도래함에 따라 실시간으로 전력가격이 변화하고 이에 따른 수요가 계속해서 변화될 전망이다. 이러한 상황 하에서 즉각적인 전력 수요 예측의 중요성이 점차 증대될 것으로 예상되며 앞으로도 전력 수요 예측에 대한 연구가 보다 활성화 될 것으로 예상된다. 따라서 향후에는 본 연구에서 제안한 알고리즘을 통해 시간대별 데이터에 적용하여 급전 계획 및 지역난방 모형의 운영상에 필요한 시간대별 전력 수요를 예측하고 선정된 변수와 예측 모델을 기반으로 전력 수요 예측을 수행 할 수 있는 전력 수요 예측 프로그램을 개발하고자 한다. 전력 수요 예측 프로그램은 사용자가 간편하게 변수를 입력하고 예측을 수행함으로써 즉각적인 전력 수요 예측을 가능하게 할 것으로 기대된다.

참고문헌

- Bishop, C. M. (2006), Pattern recognition and machine learning, Springer.
Choi, N. H., Son, K. M., and Lee, T. G. (2001), Daily peak load forecasting

- considering the load trend and temperature, *Journal of the Korean Institute of Illuminating and Electrical Installation Engineers*, 6(6), 35-42.
- Cristianini, N. and Taylor, J. S. (2000), An Introduction to support vector machines and other kernel based learning methods, Cambridge University Press.
- Hall, M. A. (1999), Correlation-based Feature Selection for Machine Learning, The University of Waikato Press.
- Han, C. H., Lee, J. W., and Lee, K. K. (2009), Analyzing Information Value of Temperature Forecast for the Electricity Demand Forecasts, *The Korean Operations Research and Management Science Society*, 26(1), 79-91.
- Han, S. M. (2009), A study on the development of the generation expansion planning system using multi-criteria decision making rule, Ph.D diss, Hongik University.
- Harvey, A. C. and Koopmand, S. J. (1993), Forecasting Hourly Electricity Demand using Time-varying splines, *Journal of the American Statistical Association*, 88(424), 1228-1236.
- Haykin, S. (1999), Neural networks: a comprehensive foundation, Prentice-Hall, Inc.
- Hong, W. C. (2009), Electric Load Forecasting by Support Vector Model, *Applied Mathematical Modelling*, 33(5), 2444-2454.
- Khotanzad, A., Davis, M. H., Abaye, A., and Maratukulam, D. J. (1996), An Artificial Neural Network Hourly Temperature Forecaster with Applications in Load Forecasting, *Journal of Power Systems*, 11(2), 870-876.
- Kim, J. H., Lee, J. G., and Cha, J. M. (2009), Load forecasting and demand management considering with renewable energy, *The Korean Institute of Electrical Engineers*, 2260-2261.
- Kitagawa, G. and Gersch, W. (1996), Smoothness Priors Analysis of Time Series, Lecture Notes in Statistics, Springer-Verlag.
- Kong, D. S., Kwak, Y. H., and Huh, J. H. (2010), Artificial Neural Network Based Energy Demand Prediction for the Urban District Energy Planning, *Journal of the Architectural Institute of Korea*, 26(2), 221-230.
- Kwon, S. K. (2004), The Present Situation of District Heating in Korea and Foreign Countries, *The Society of Air-conditioning and Refrigerating Engineers of Korea*, 1217-1222.
- Mirasgedis, S., Sarafidis, Y., Georgo Poulou, E., Lalas, D. P., Moschovits, M. Karagiannis, F., and Papakonstantinou, D. (2006), Models for mid-term electricity demand forecasting incorporating weather influences, *Energy*, 31(2-3), 208-227.
- Nam, B. W., Song, K. B., Kim, K. H., and Cha, J. M. (2008), The Spatial Electric Load Forecasting Algorithm using the Multiple Regression Analysis Method, *Journal of the Korean Institute of Illuminating and Electrical Installation Engineers*, 22(2), 63-70.
- Nrgaard, M. and Norgaard, P. M. (2006), Neural Networks for Modelling and Control of Dynamic Systems : A Practitioner's Handbook(Advanced Textbooks in Control and Signal Processing), Springer.
- Pai, P. F. and Hong, W. C. (2005), Support vector machines with simulated annealing algorithms in electricity load forecasting, *Energy Conversion and Management*, 46(17), 2669-2688.
- Pappas, S. Sp., Ekonomou, L., Karampelas, P., Karamousantas, D. C., Katsikas, S. K., Chatzarakis, G. E., and Skafidas, P. D. (2010), Electricity demand load forecasting of the Hellenic power system using an ARMA model, *Electric Power Systems Research*, 80(3), 256-264.
- Park, C. K. (2006), Estimating Software Development Cost USING Support Vector Regression, *The Korean Operations Research and Management Science Society*, 23(1), 75-91.
- Park, K., Hou, T., and Shin, H. (2011), Oil Price Forecasting Based on Machine Learning Techniques, *Journal of the Korean Institute of Industrial Engineers*, 37(1), 64-73.
- Romera, E. G., Moran, M. A. J., and Fernandez, D. C. (2008), Monthly electric energy demand forecasting with neural networks and Fourier series, *Energy Conversion and Management*, 49(11), 3135-3142.
- Shin, J. H. and Hong, Y. C. (2006), GMDH Algorithm with Data Weighting Performance and Its Application to Power Demand Forecasting, *Journal of Control Automation and Systems Engineering*, 12(7).
- Sholkopf, B. and Smola, A. (2002), Learning with Kernels, MIT Press, Cambridge MA.
- Srinivasan, D. (2008), Energy Demand Prediction using GMDH networks, *Neurocomputing*, 72(1-3), 625-629.
- Thissen, U., van Brakel, R., de Weijer A. P., Melsse, W. J., and Buydens, L. M. C. (2003), Using Support Vector Machines for time series prediction, *Chemometrics and Intelligent Laboratory Systems*, 69(1-2), 35-49.
- Wenwu, H., Zhizhong, W., and Hui, J. (2008), Model Optimizing and Feature Selection for Support Vector Regression in time series forecasting, *Neurocomputing*, 72(1-3), 600-611.
- Wi, Y. M., Moon, G. H., Lee, J. H., Joo, S. K., and Song, K. B. (2007), Load Forecasting for the Holidays using a Polynomial Regression Incorporating Temperature Effect, *The Korean Institute of Electrical Engineers*, 29-30.
- Yang, J., Rivard, H., and Zmeureanu, R. (2005), On-line building energy prediction using adaptive artificial neural networks, *Energy and Buildings*, 37(12), 1250-1259.



이형로

아주대학교 산업공학과 학사
현재: 아주대학교 산업공학과 석사과정
관심분야: 데이터마이닝, 경영과학



신현정

서울대학교 박사
독일 막스플랑크 연구소(Max-Planck-Institute)
연구원
서울대학교 의과대학 연구교수
현재: 아주대학교 산업정보시스템공학부 교수
관심분야: 데이터마이닝, 경영과학, 바이오
메디컬 인포매틱스