

An Electric Load Forecasting Scheme for University Campus Buildings Using Artificial Neural Network and Support Vector Regression

Jihoon Moon[†] · Sanghoon Jun^{**} · Jinwoong Park[†] · Young-Hwan Choi^{***} · Eenjun Hwang^{****}

ABSTRACT

Since the electricity is produced and consumed simultaneously, predicting the electric load and securing affordable electric power are necessary for reliable electric power supply. In particular, a university campus is one of the highest power consuming institutions and tends to have a wide variation of electric load depending on time and environment. For these reasons, an accurate electric load forecasting method that can predict power consumption in real-time is required for efficient power supply and management. Even though various influencing factors of power consumption have been discovered for the educational institutions by analyzing power consumption patterns and usage cases, further studies are required for the quantitative prediction of electric load. In this paper, we build an electric load forecasting model by implementing and evaluating various machine learning algorithms. To do that, we consider three building clusters in a campus and collect their power consumption every 15 minutes for more than one year. In the preprocessing, features are represented by considering periodic characteristic of the data and principal component analysis is performed for the features. In order to train the electric load forecasting model, we employ both artificial neural network and support vector machine. We evaluate the prediction performance of each forecasting model by 5-fold cross-validation and compare the prediction result to real electric load.

Keywords : Electric Load Forecasting, Educational Institution, Support Vector Regression, Artificial Neural Network

인공 신경망과 지지 벡터 회귀분석을 이용한 대학 캠퍼스 건물의 전력 사용량 예측 기법

문지훈[†] · 전상훈^{**} · 박진웅[†] · 최영환^{***} · 황인준^{****}

요약

전기는 생산과 소비가 동시에 이루어지므로 필요한 전력 사용량을 예측하고, 이를 충족시킬 수 있는 충분한 공급능력을 확보해야만 안정적인 전력 공급이 가능하다. 특히, 대학 캠퍼스는 전력 사용이 많은 곳으로 시간과 환경에 따라 전력 변화폭이 다양하다. 이러한 이유로, 효율적인 전력 공급 및 관리를 위해서는 전력 사용량을 실시간으로 예측할 수 있는 모델이 요구된다. 국내의 대학 건물에 대해서는 전력 사용 패턴과 사례 분석을 통해 전력 사용에 영향을 주는 요인들을 파악하기 위한 다양한 연구가 진행되었으나, 전력 사용량의 정량적 예측을 위해서는 더 많은 연구가 필요한 상황이다. 본 논문에서는, 기계 학습 기법을 이용하여 대학 캠퍼스의 전력 사용량 예측 모델을 구성하고 평가한다. 이를 위해, 대학 캠퍼스의 주요 건물 클러스터에 대해 전력 사용량을 15분마다 1년 이상 수집한 데이터 셋을 사용한다. 수집된 전력 사용량 데이터는 수열 형태의 시계열 데이터로 기계 학습 모델에 적용 시 주기성 정보를 반영할 수 없으므로, 2차원 공간의 연속적인 데이터로 증강함으로써 주기성을 반영하였다. 이 데이터와 교육기관의 특성을 반영하기 위한 요일과 공휴일로 구성된 8차원 특성 벡터에 대해 주성분 분석(Principal Component Analysis) 알고리즘을 적용한다. 이어, 인공 신경망(Artificial Neural Network)과 지지 벡터 회귀분석(Support Vector Regression)을 이용하여 전력 사용량 예측 모델을 학습시키고, 5겹 교차검증(5-fold Cross Validation)을 통하여 적용된 기법의 성능을 평가하여, 실제 전력 사용량과 예측 결과를 비교한다.

키워드 : 전력 사용량 예측, 교육기관, 지지 벡터 회귀분석, 인공 신경망

※ 본 연구는 산업통상자원부(MOTIE)와 한국에너지기술연구원(KETEP)의 지원을 받아 수행한 연구 과제입니다(No. 20152010103060).

※ 이 연구를 위해 전력 사용량 데이터를 제공한 고려대학교 에너지-안전팀에 감사의 말씀을 전합니다.

† 준 회원 : 고려대학교 전기전자공학과 석·박사통합과정

** 비 회원 : 서울아산병원 의료영상로봇연구실 박사후연구원

*** 준 회원 : 귀뚜라미 연구기획실 수석연구원

**** 중신회원 : 고려대학교 전기전자공학과 교수

Manuscript Received : August 9, 2016

Accepted : August 29, 2016

* Corresponding Author : Eenjun Hwang(ehwang04@korea.ac.kr)

1. 서 론

안정적인 전력 공급을 위해서는 생산과 소비가 동시에 이루어지는 전력을 정확히 예측하는 것이 중요하다. 특히, 단기 수요 예측을 통해 전력계통설비의 안정성을 보장하고, 전력예비율을 능동적으로 조절함으로써 정전 및 용량 초과로 인한 손실에 대비하고, 중장기 수요 예측을 통해 설비 구축 및 장기적 수급계획, 소비 효율 구간 탐색 등 기관 또는 사회적 비용을 최소화할 수 있는 근거가 될 수 있다. 이 뿐만 아니라, 최근 전력의 수요와 공급이 서로 정보를 주고 받을 수 있는 양방향성을 통해 에너지 효율을 최적화하는 스마트 그리드 시스템(Smart Grid System)의 운영에서도 정확한 전력 사용량 예측 기술이 필요하다.

전력 사용의 패턴은 수요 주체인 건물 또는 건물 클러스터의 특성에 따라 상이한 차이를 보인다. 예를 들어, 대학 캠퍼스는 전력 사용량이 많은 건물에 속하며[1], 학기 또는 방학이나 요일에 따라 전력 수요의 소비형태가 다르다. 현재 국내 대학 건물의 전력 사용 패턴과 사례 분석을 통하여 전력 사용에 영향을 미치는 요소들에 대한 연구는 활발히 진행되고 있지만[2-6], 예측을 위한 연구 사례는 아직 미흡한 실정이다. 대학의 전력 사용은 건물 용도와 외부의 복잡한 요인으로 다양한 패턴들을 가지고 있다. 이러한 패턴들을 고려한 기계 학습 예측 모델링을 구성한다면 더욱 정확한 전력 사용량 예측이 가능할 것이다.

본 연구는 국내 대형 대학 캠퍼스에서 발생하는 방대한 양의 전력 사용량 데이터를 대상으로 한다. 특히, 15분 단위의 단기 전력 사용량의 예측을 위해 연간, 월간, 일간, 시간대별 정보와 공휴일, 기상 정보 등 다양한 정보를 토대로 예측 모델을 설계한다. 제안하는 대학 캠퍼스 단기 전력 사용량 예측 기법은 크게 데이터의 전처리, 모델링, 평가로 구성된다. 먼저, 입력 데이터의 전처리 과정에서는 주기성을 반영할 수 있는 데이터 유형으로 변환하고, 주성분 분석(Principal Component Analysis; PCA)을 적용한다. 정확한 전력 사용량 예측을 위해 본 논문에서는 다양한 분야에서 예측에 활용되고 있는 인공 신경망(Artificial Neural Network: ANN)과 지지 벡터 회귀분석(Support Vector Regression: SVR)을 기반으로 두 가지 예측 모델을 구성하고, 대학 캠퍼스 건물에 대한 15분 단위의 전력 사용량을 예측하여 그 정확도를 비교 평가한다.

본 논문의 구성은 다음과 같다. 2장은 기계 학습 기반의 예측 모델과 국내 교육기관 전력 사용에 관한 연구들을 살펴본다. 3장은 본 연구에서 적용한 기계 학습 알고리즘의 이론적 배경과 예측 모델에 대해 알아본다. 4장은 앞서 살펴본 알고리즘으로 예측 모델 시스템 구성 및 성능 지표를 제시한다. 5장은 제안한 기법들의 전력 사용량 예측 정확성을 통한 실험 및 평가를 수행한다. 마지막으로 6장에서는 실험 결과에 대한 평가와 함께 결론을 맺는다.

2. 관련 연구

이 장에서는 먼저 기계 학습 알고리즘을 기반으로 한 전력 사용량 예측 모델을 소개하고, 특히 국내 교육기관 전력 사용량 데이터를 대상으로 수행된 연구에 관해 기술한다.

2.1 기계 학습 기반 전력 사용량 예측 연구

예전부터 다양한 기계 학습을 사용하여 전력 사용량 예측 모델을 구성하는 연구[7-9]가 활발히 진행되고 있으며, 최근에는 ANN을 활용한 다양한 전력 사용량 예측이 연구되고 있다. 입력 변수로 과거 전력 사용량 이외에 평일/휴일 여부, 날씨 정보 등의 외부 요인을 추가 변수로 도입하여 예측 정확성을 높이는 연구가 진행되고 있다. Kangji[10]는 빠른 속도로 최적해를 탐색하는 데 효과적인 기법으로 검증된 PSO(Particle Swarm Optimization) 알고리즘을 ANN에 결합한 IPSO-ANN(Improved Particle Swarm Optimization-Artificial Neural Network)을 제안하고, 기존 ANN이나 GA-ANN(Genetic Algorithm-Artificial Neural Network)[11]과 전력 사용량 예측에 대한 성능 비교를 수행하였다. Bagnasco[12]은 ANN을 기반으로 토리노의 첼리니 병원 건물 전력 사용량을 예측하였으며, Grolinger[13]는 기존 외부 요인에 이벤트 정보를 추가한 ANN과 SVR 모델을 구성하여 캐나다의 대형 엔터테인먼트 건물의 전력 사용량을 예측 및 비교하였다. Chitsaz[14]는 매개 변수의 최적화 프로세스가 간편한 SRWNN(Self-Recurrent Wavelet Neural Network)을 구성하고 MLP(Multi-Layer Perceptron), WNN(Wavelet Neural Network) 알고리즘을 이용하여, 용도가 다른 두 건물의 시간대별 전력 사용량과 과거 24시간 기준 피크치 예측의 성능 비교를 수행하였다.

ANN을 비롯하여 다양한 기계 학습 알고리즘을 사용한 전력 사용량 예측에 관한 연구가 보고되고 있다. Amber[15]는 다중 회귀 모형과 유전자 프로그래밍 모형을 구성하여 런던 사우스뱅크 대학 캠퍼스에 있는 건물의 일간 전력 사용량을 예측 및 비교하였다. Ghelardoni[16]는 시계열들을 분해하는 EMP(Empirical Mode Decomposition) 알고리즘을 적용한 SVM(Support Vector Machine) 모델을 이용하여 슬로바키아 동부 전력 사용량과 이탈리아 사무용 건물 전력 사용량 예측을 수행하였다. Jurado[17]는 다양한 기계 학습 방법으로 RF(Random Forest), NN(Neural Network), FIR(Fuzzy Inductive Reasoning), ARIMA(AutoRegressive Integrated Moving Average)를 통해 바르셀로나에 있는 3개의 건물을 시간대별 전력 사용량 예측 및 비교하였다.

2.2 교육기관 전력 사용에 관한 연구

국내에 속한 교육기관 중 전력 사용이 높은 대학을 중심으로 수행한 연구가 보고되었다. [2]는 대전지역 “D” 대학의 전력 사용량 실측 데이터와 조사를 바탕으로 대학 건물의 사용 특성을 분석하였으며, [3]은 대구지역 “K” 대학 건물 중 태양광 발전 시스템이 설치된 건물의 연간 전력 사용량

데이터와 관련 부서 방문조사를 바탕으로 전력 사용 패턴을 분석하였다. [4]는 에너지 효율화 방안 도출을 위한 목적으로, 경기도 안양시에 있는 “Y” 대학 건물의 신재생에너지 이용과 조명밀도 및 조명설계 방식을 적용하여 전력 사용 현황을 분석하였다. [5]는 경상북도 경산시에 있는 “Y” 대학 건물의 사례분석 및 전력 사용량 조사를 통해 에너지 성능 개선 방안만을 제시하였다. [6]은 전력 사용을 줄이기 위한 목적으로, 서울시 동대문구에 있는 “U” 대학 건물 중 강의실과 연구실로 구성된 건물을 대상으로 전력 사용량에 대한 비교분석을 통해 설계 및 운영방식, 입주가 이루어진 학과 특성 등에 따라 전력 사용 패턴이 달라짐을 확인하였다.

기존의 기계 학습을 사용한 연구들을 살펴보면, 전력 사용량 예측의 주요 범위가 일별, 시간대별로 이루어져 있다. 또한, 국내 교육기관 관련된 연구들을 살펴보면, 전력 사용의 현황과 패턴 분석에만 머물러 있다.

본 논문은 정확한 전력 사용량 예측을 위한 목적으로, 대표적인 기계 학습 알고리즘인 ANN과 SVR을 적용하여 예측 모델을 구성하였으며, 15분 단위의 세분화된 전력 사용량 예측을 수행한다.

3. 이론적 배경

이 장에서는 본 연구에 사용된 기계 학습 기법들을 소개한다. 먼저 ANN에 대해 설명하고, SVR을 기술한다.

3.1 Artificial Neural Network

ANN은 인간의 두뇌를 모델로 하여, 다양한 데이터를 두뇌와 비슷한 방식으로 처리하기 위해 구성된 알고리즘이다. 두뇌의 기본 구조 조직인 뉴런(Neuron)이 상호 연결되어 네트워크를 형성할 때, 이를 신경망(Neural Network: NN)이라 한다. 이러한 뉴런을 모방한 퍼셉트론(Perceptron)이 구조 단위로 구성된 인위적인 네트워크를 ANN이라 한다. 단층 퍼셉트론(Single Layer Perceptron)은 입력 벡터를 두 부류로 구분하는 선형 분류기이다. MLP는 다층 퍼셉트론으로 단층 퍼셉트론이 해결할 수 없는 비선형적인 분류가 가능하도록, 입력층과 출력층 사이에 은닉층(Hidden Layer) 추가와 역전파(Backpropagation) 알고리즘을 이용하여 학습한다. 하나의 은닉층으로 모든 함수 표현이 가능하지만, 일반적으로 여러 개의 은닉층을 사용하는 것이 더욱 높은 정확성을 가진다. 은닉층에 속한 각 뉴런의 출력 값을 결정하는 공식은 Equation (1)과 같다.

$$y_j = \varphi \left(\sum_{i=1}^N w_{ij} x_i + w_{i0} \right) \quad (1)$$

x_i 가 뉴런의 입력일 때, w_{ij} 는 입력층의 i 번째 뉴런과 은닉층의 j 번째 뉴런 사이의 Synaptic Weight로 지식을 축적하기 위해 사용된다. φ 는 주로 기본 함수나 시그모이드(Sigmoid) 함수를 모델링하는 활성화 함수이다. 이와 같은

방법으로, 출력층 뉴런의 출력이 모델링되며, Weight는 은닉층과 출력층 사이의 연결을 말한다.

ANN은 각 뉴런이 독립적으로 동작하는 처리기의 역할을 하므로 병렬성(Parallelism)이 뛰어나다. 또한, 많은 연결선에 정보가 분산되어 있으므로 몇몇 뉴런에 문제가 발생하더라도 전체 시스템에 큰 영향을 주지 않는 결함 허용(Fault Tolerance) 능력이 있으며, 주어진 환경에 대한 학습 능력이 있다. 이와 같은 특성 때문에 인공 지능 분야를 비롯하여 문자 인식, 화상 처리, 자연 언어 처리, 음성 인식 등 다양한 분야[18]에서 이용되고 있다.

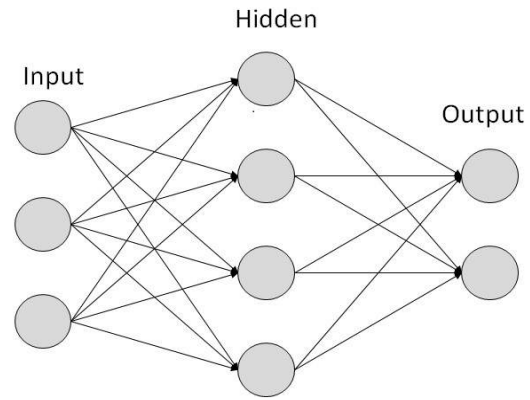


Fig. 1. Typical Architecture of Artificial Neural Network

3.2 Support Vector Regression

SVM은 분류(Classification)와 회귀분석(Regression)에 응용할 수 있는 기계 학습 알고리즘이다. 크게 분류를 위한 SVC(Support Vector Classification)와 예측을 위한 SVR로 나누어진다.

SVR은 임의의 실숫값을 예측할 수 있도록 SVM의 회귀 모형에 ϵ -무감도 손실함수(ϵ -insensitive loss function)를 도입하여 비선형 회귀 영역까지 확장하였다. 먼저 입력 값에 고차원(High dimension)의 특성 공간(Feature space)에 비추어(Mapped)지고, 그 후 출력 값과 연관된 선형함수를 찾는다. SVR의 선형 추정함수는 Equation (2)와 같다.

$$f(x) = w^T x + b \quad (2)$$

즉, 저차원 입력 공간(x)에 있는 비선형 회귀 문제가 고차원 특성 공간(F)에 있는 선형 회귀로 전환한다. ϵ -무감도 손실함수 L_ϵ 은 일반적으로 SVR에 사용되는 비용함수로 Equation (3)과 같이 나타낸다.

$$L_\epsilon(F(x), q) = \max(0, |f(x) - q| - \epsilon) = \max(0, |(w^T x + b) - q| - \epsilon) \quad (3)$$

여기서 ϵ 은 회귀함수 $f(x)$ 의 주변에 있는 튜브의 반지름을 나타내는 정밀도수(Precision parameter)이다. SVR에서는 실제 값과 예측 값을 가능한 한 ϵ 이내로 유지하면서 마

진을 최대화한다. 이를 $x_i \in R^m$ 이 입력변수이고, $y_i \in R^1$ 이 출력변수인 주어진 데이터 집합 $(x_1, y_1), \dots, (x_i, y_i)$ 에서 SVR의 최소화 수식은 Equation (4)와 같다.

$$\begin{aligned} \text{Minimize : } & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\ \text{subject to : } & y_i - wx_i - b \leq \epsilon + \xi_i \\ & wx_i + b - y_i \leq \epsilon + \xi_i^* \\ & \xi_i, \xi_i^* \geq 0 \end{aligned} \quad (4)$$

Equation (4)의 첫 번째 항은 회귀모형의 복잡성과 정확성의 균형을 조정하는 정규화 항이며, ξ_i 와 ξ_i^* 는 ϵ 의 경계선을 벗어난 잔차와의 거리를 말한다. C 는 경험적 위험과 정규화 항의 균형을 맞추는 데 사용되는 정규화 매개 변수이다. 또한, C 와 ϵ 은 사용자가 정해야 하는 모수가 된다.

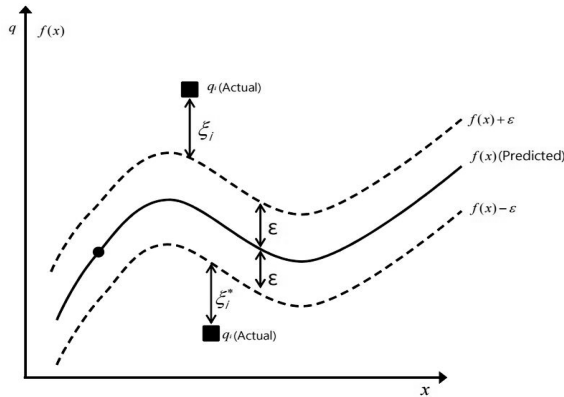


Fig. 2. The Notation of ϵ -insensitive Loss Function in SVR

SVR도 SVM과 마찬가지로 커널함수를 사용하여 학습 데이터를 특정 공간의 점으로 변화시킨 다음 특정 공간에서 학습을 수행한다. 대표적 커널함수인 RBF(Radial Basis Function) 커널의 표현식은 Equation (5)와 같다.

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (5)$$

이 수식에서 γ 을 통해 각 데이터 포인트의 가중치를 적용하게 된다.

4. 전력 사용량 예측 모델

이 장에서는, 교육기관의 전력 사용량 예측 모델을 구성하기 위한 과정을 자세하게 서술한다. 본 논문에서 제안하는 예측 모델링의 구조는 Fig 3과 같다. 일시별 전력 사용량 데이터와 날씨 정보를 입력하여 학습 및 테스트 데이터로 나눈다. 학습 데이터는 전처리 과정을 통해 특징 벡터로 추출되고 SVR 및 ANN으로 학습된다. 테스트 데이터를 이용해 다양한 평가 지표를 통해 성능을 평가한다.

4.1에서는 사용된 데이터 셋에 대한, 4.2에서는 사용되는 특성에 대해 설명을 한다. 4.3은 데이터 전처리 과정에 사용된 기법을 설명한다. 4.4에서는 사용된 예측 모델에 대한 내용을 다룬다. 4.5에서는 예측 기법의 정확성 비교를 위한 성능 지표에 관해 설명한다.

4.1 데이터 구성

본 연구에서 대상으로 하는 데이터는 서울특별시 성북구에 위치한 “K”대학교의 전력 사용량이다. 전력 사용량 예측 모델을 구성하기 위해, 과거 전력 사용량 데이터와 기상 관련 데이터를 수집한다. 전력 사용은 “K”대학교의 에너지 안 전팀이 한국전력공사 iSmart를 통하여 실시간으로 관리하고 있다. Fig 4에서 보이듯이, “K”대학교의 전력 사용량 데이터는 세 군데 지역으로 구성되어 있다. Cluster A는 중앙광장, 본관, 중앙도서관을 비롯하여 인문·사회계 캠퍼스가 위치한 지역이다. Cluster B와 C는 자연계 캠퍼스를 말하며, 연구실이 밀집해 있어 건축면적에 비해 전력 사용이 많다. Cluster C는 B보다 주로 신축 건물들이 자리 잡고 있다. 기상 관련 데이터는 기상청에서 제공하는 기상자료개방포털[19]을 통해, 서울 지역의 중관기상관측 데이터를 사용한다.

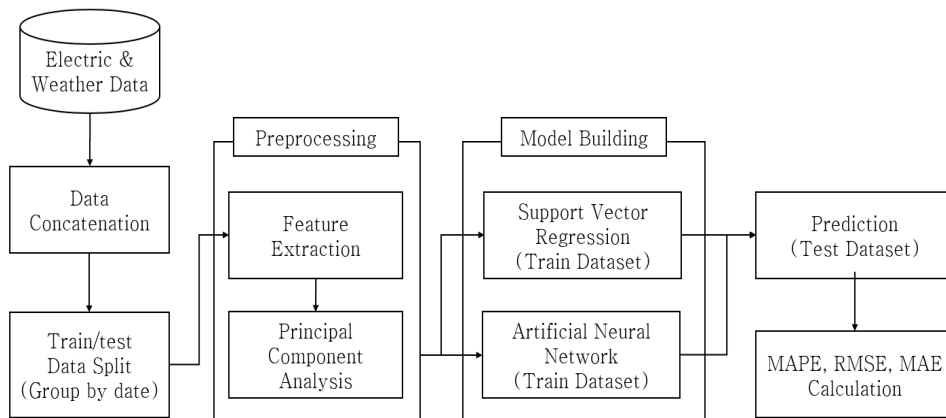


Fig. 3. Overall Architecture of Electric Load Forecasting Model

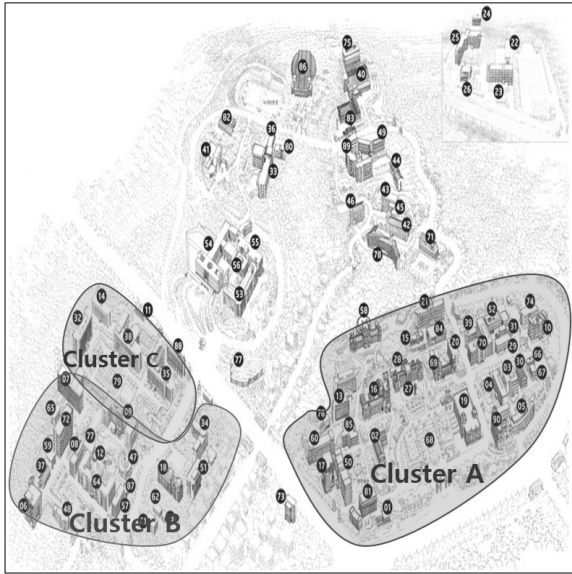


Fig. 4. Building Clusters

수집된 데이터는 한국전력공사 iSmart에서 제공하는 15분 단위의 전력 사용량이다. Fig. 5와 같이, Cluster A, B의 수집 기간은 2015년 1월 1일부터 2016년 7월 17일까지이다. 다른 클러스터보다 iSmart가 늦게 설치된 Cluster C는 2015년 9월 1일부터 2016년 7월 17일까지 데이터가 수집되었다. 기상 정보는 중관기상관측 데이터 중 전력 사용량과 밀접한 관련이 있는 온도와 습도를 예측 변수에 적용하였다. 데이터의 수집 기간은 전력 사용량 데이터의 기간과 같은 2015년 1월 1일부터 2016년 7월 17일까지이다.

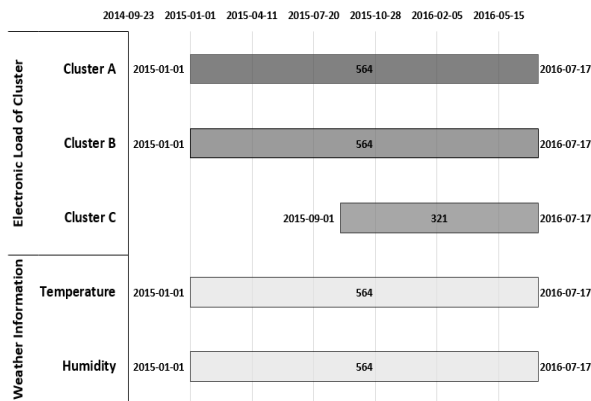


Fig. 5. Collected Dataset of 3 Clusters

Fig. 6은 Cluster A의 11월 전력 사용을 요일별로 나타낸 예시이다. 11월은 가을 학기가 운영되고 있으며, 총 5주로 구성된 공휴일이 없는 달이다. 마지막 주는 일요일과 월요일이 포함되어 있어, 월요일의 전력 사용량이 다른 주중의 요일보다 상대적으로 많은 전력 사용량을 보였다. 주중과 주말에 따른 전력 사용량 및 패턴이 다를 수 있다.

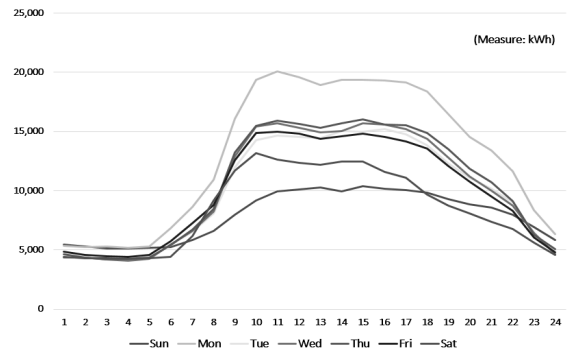


Fig. 6. Daily Electric Load of November

4.2 특징 추출

전력 사용량 데이터는 시계열 데이터이므로 기계 학습 모델을 적용하기에 앞서, 데이터에 관하여 적합한 처리 방법으로 주기성을 반영한다.

월, 일, 시간, 분 등의 시간 정보 데이터는 수열 형태를 보인다. 수열 형태의 데이터는 기계 학습 모델에 적용 시 주기성 정보를 반영할 수 없다. 예를 들어 23시와 0시는 실제 인접한 정보이지만, 수열로 표현 시 23의 차이를 갖게 된다. 따라서 본 논문에서는 다음과 같은 식을 통해 수열 형태의 1차원 데이터를 2차원 공간의 연속적인 데이터로 증강함으로써 주기성을 반영한다. Equation (11), (12)에서 EOM_{month} 은 End of Month로 속한 달의 마지막 날을 나타낸다. 예를 들어, 1월은 31, 2월은 28, 3월은 31 등을 말한다.

$$minute_x = \sin((360/60) \times minute) \quad (6)$$

$$minute_y = \cos((360/60) \times minute) \quad (7)$$

$$hourmin = hour + (minute/60) \quad (8)$$

$$hour_x = \sin((360/60) \times hourmin) \quad (9)$$

$$hour_y = \cos((360/60) \times hourmin) \quad (10)$$

$$day_x = \sin((360/EOM_{month}) \times day) \quad (11)$$

$$day_y = \cos((360/EOM_{month}) \times day) \quad (12)$$

$$month_x = \sin((360/12) \times month) \quad (13)$$

$$month_y = \cos((360/12) \times month) \quad (14)$$

교육기관은 대부분 평일에 수업이나 학사 운영 등이 이루어지며, 주말이나 공휴일에는 특별한 이벤트를 제외하고는 운영되지 않으므로, 일반적으로 평일과 주말은 서로 다른 전력 사용 패턴을 보인다. 이러한 요소를 반영하기 위해 명목척도인 요일 정보를 예측 모델에 반영한다. 이를 위해, 7가지의 요일과 공휴일을 포함한 8차원 특성 벡터를 구성하였으며, 해당 요일은 1을 갖고, 그렇지 않은 경우 0을 갖도록 변환하여 표현하였다.

일반적으로 여름철에는 기온이 높아질수록 에어컨 사용으로 인한 전력 사용량이 많아지고, 겨울철에는 기온이 낮아

질수록 히터 사용으로 인한 전력 사용량이 많아진다. 이는 전력 사용이 환경 요인과 밀접한 관련이 있음을 보여준다.

본 논문에서는 이러한 특성 정보들을 종합하여 ANN 및 SVR로 각각 구성된 예측 모델의 시간 정보를 표현한다. 각 특성은 0의 평균과 1의 표준편차를 갖도록 표준화(Standardization)를 수행하였다.

$$X = \left\{ \begin{array}{l} minute_x, minute_y, hour_x, hour_y, \\ day_x, day_y, month_x, month_y, \\ monday, tuesday, wednesday, \\ thursday, friday, \\ saturday, sunday, holiday, \\ temperature, humidity \end{array} \right\} \quad (15)$$

4.3 주성분 분석

Equation (15)에서의 X는 차원이 많아 감쇄의 필요성을 보인다. 본 논문에서는 수집된 데이터들에 대해 PCA 알고리즘[20]을 적용하여 입력 변수의 차원을 큰 표준 편차를 갖는 방향으로 사상(Projection)한다. PCA는 많은 수의 입력 변수들을 일련의 소수의 요인에 의해 설명이 가능한 알고리즘이다. 데이터를 한 개의 축으로 사상시켰을 때, 분산의 크기를 기준으로 새로운 좌표계를 형성하여 이를 선형 변환한다. 이들 주성분은 고유벡터(Eigenvector)로 나타내어지며, 이는 ANN과 SVR의 입력 변수로 사용된다.

4.4 모델 구성

본 논문에서는 ANN과 SVR 기법을 전력 사용량 예측 모델로 사용한다. SVR의 학습 시, Grid search algorithm을 통해 가장 최적의 모수를 선택하여 사용한다. Grid search 결과 일반적으로 최적의 성능을 보이는 모수는 RBF 커널상의 C=0.1, ε=0.1 이었다. ANN의 학습 시 overfitting을 피하고자 100번의 epoch상에서 학습하였으며, 은닉층의 구성은 검증 결과 대체로 좋은 성능을 보이는 1개 내의 7개 유닛을 사용하였다. Cluster A의 전력 사용량 예측에 대해 평균 절대 백분율 오차를 이용하여 유닛의 수에 따른 성능의 정확성을 Table 1에 나타내었다.

Table 1. Accuracy by Number of Hidden Layer Units

	1	2	3	4	5	6	7
MAPE(%)	0.138	0.120	0.109	0.108	0.107	0.099	0.096

4.5 성능 평가 지표

본 논문의 성능 평가 지표는 기존 연구에서 예측의 정확성을 평가하기 위해 사용되는 MAPE(Mean Absolute Percentage Error)[12-13, 16-17], RMSE(Root Mean Square Error)[14-15], MAE(Mean Absolute Error)[14-16]을 각 예측 모델에 대해 계산하여 비교한다.

1) Mean Absolute Percentage Error

MAPE는 평균 절대 백분율 오차라 불리며, 정확도를 오

차의 백분율로 표시한다. 이 숫자는 백분율이기 때문에 다른 통계량보다 쉽게 이해할 수 있는 장점을 가진다. MAPE의 계산법은 Equation (16)과 같다.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad (16)$$

여기서, A_t 는 실제 값이고, F_t 는 예측 값, n 은 관측치의 수이다. 예를 들어 MAPE가 5이면 예측 값은 평균 5% 벗어나는 것으로 볼 수 있다.

2) Root Mean Square Error

RMSE는 평균 제곱근 오차라 불리며, 예측 값과 실제 환경에서 관찰되는 값의 차이를 다룰 때 흔히 사용하는 측도이다. 정밀도(Precision)를 표현하는 데 적합하다. 각각의 차이 값은 잔차(Residual)라고도 하며, 평균 제곱근 오차는 잔차들을 하나의 측도로 종합할 때 사용된다. RMSE의 수식은 Equation (17)과 같다.

$$RMSE(\hat{\theta}) = \sqrt{MSE(\hat{\theta})} = \sqrt{E((\hat{\theta} - \theta)^2)} \quad (17)$$

추정치 θ 에 대한 추정량 $\hat{\theta}$ 의 평균 제곱근 편차를 평균 제곱 오차의 제곱근으로 정의된다. 편차 추정량에서 평균 제곱근 오차는 분산의 제곱근, 즉 표준 오차가 된다.

3) Mean Absolute Error

MAE는 평균 절댓값 오차라 불리며, 주로 시계열에서 예측 값의 오류를 측정하는 데 사용된다. 이 값이 낮을수록 예측 값의 정확도가 높다고 판단할 수 있다. MAE의 수식은 Equation (18)과 같다.

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - y_i| = \frac{1}{n} \sum_{i=1}^n |e_i| \quad (18)$$

여기서, f_i 는 예측 값이고, y_i 는 실제 값이다. 예측 값은 가중치와 구간에 좌우됨으로 평균 절댓값 오차를 통해 최적의 가중치를 구하는 작업이 필요하다.

5. 실험 결과

본 논문은 Python 환경에서 데이터 셋의 전처리를 수행하였고, 기계 학습 공개 라이브러리인 scikit-learn[21]과 pybrain[22]을 사용하여 예측 모델링을 수행하였다. 5겹 교차검증을 수행하여 Training set과 Test set의 비율을 80 : 20으로 설정하였으며 5회 반복 후 평균치를 계산하였다. 15분마다 짧은 데이터 특성상 정확한 교차검증을 위해, 일자를 기반으로 상호배타적인 데이터 셋을 나누어 수행하였다.

5.1 대학 건물 클러스터의 전력 사용 분포

Fig. 7은 클러스터별로 수집한 15분 단위의 전력 사용량 데이터를 사용량에 따른 샘플 개수의 분포로 나타낸 것이다. Cluster A는 인문·사회계 캠퍼스가 위치한 지역으로, 주말이나 공휴일, 방학 등의 야간 시간대에 전력 사용량이 적음을 알 수 있다. 하지만 클러스터의 규모가 B와 C보다 크므로, 시험 기간이나 행사 등 학사 이벤트에 따라 전력 사용이 커짐을 확인할 수 있다. Cluster B와 C는 자연계 캠퍼스가 위치한 지역으로, 방학에도 운영하는 연구실이 밀집해 있어 Cluster A보다 높은 통상 전력 사용량을 보인다. 또한, Cluster B와 C의 전력 사용이 비슷한 분포를 보임을 확인할 수 있다. Cluster C는 A, B보다 클러스터 크기가 작아, 짧은 시간에 많은 전력 사용을 보이지 않았다.

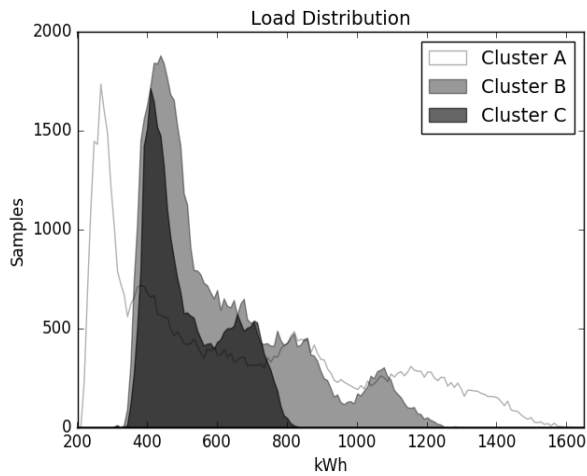


Fig. 7. Distribution of Electric Load

5.2 수집된 특성과 전력 사용량 연관성 분석

본 논문에서 수집된 특성은 요일, 시간, 온도, 습도 등이다. 이러한 특성이 실제 전력 사용량과 연관되어 있는지 피어슨 상관계수와 p-value를 통해 나타내었다.

일반적으로, 피어슨 상관계수는 1에 가까울수록 양적 선형관계를 가지며, -1에 가까울수록 음적 선형관계를 가진다. Table 2에서 보이는 바와 같이, 전력 사용량과 대부분 특성이 유의한 상관성(p-value < 0.01)을 가짐을 알 수 있다.

5.3 주성분 분석 적용 및 예측 기법의 비교

데이터의 5겹 교차검증을 통해, 모든 특성이 반영된 Model 3을 이용하여, SVR과 ANN을 대상으로 PCA 사용 여부에 따른 네 가지 경우에 대한 성능 평가를 수행하였다. PCA는 입력 벡터의 차원 수까지 적용하여 가장 좋은 성능을 보이는 감쇄된 차원을 사용하였다. SVR과 ANN은 전처리 과정으로 주성분 분석을 이용하지 않고, 기존의 알고리즘으로 전력 사용량을 예측한 것이다. SVR/ANN+PCA는 주성분 분석을 거친 후, SVR/ANN을 적용하여 전력 사용량을 예측한 것이다.

Table 2. Correlation Between Collected Feature and Electric Consumption

	Pearson Correlation to Electric Load		
	Cluster A	Cluster B	Cluster C
Monday	0.085**	0.094**	0.078**
Tuesday	0.087**	0.107**	0.103**
Wednesday	0.086**	0.102**	0.108**
Thursday	0.089**	0.101**	0.125**
Friday	0.045**	0.073**	0.063**
Saturday	-0.125**	-0.196**	-0.179**
Sunday	-0.268**	-0.283**	-0.295**
Holiday	-0.345**	-0.417**	-0.418**
Minute(x)	0.001	0.000	0.007
Minute(y)	0.002	0.001	0.004
Hour(x)	-0.417**	-0.431**	-0.515**
Hour(y)	-0.669**	-0.546**	-0.579**
Day(x)	-0.002	-0.082**	0.006
Day(y)	-0.028**	-0.011	0.038**
Month(x)	-0.101**	-0.222**	-0.005
Month(y)	0.010*	0.084**	-0.067**
Temperature	0.165**	0.129**	0.202**
Humidity	-0.318**	-0.331**	-0.373**

(* p-value<0.05 **p-value<0.01)

Table 3. 5-fold Cross-Validation

MAPE	SVR	SVR+PCA	ANN	ANN+PCA
Cluster A	0.100	0.100	0.096	0.101
Cluster B	0.059	0.060	0.064	0.058
Cluster C	0.042	0.042	0.038	0.043

RMSE	SVR	SVR+PCA	ANN	ANN+PCA
Cluster A	82.774	84.173	79.930	82.766
Cluster B	49.971	50.280	51.722	47.484
Cluster C	28.691	29.060	25.658	29.237

MAE	SVR	SVR+PCA	ANN	ANN+PCA
Cluster A	59.562	60.222	58.015	60.434
Cluster B	35.839	35.991	38.685	35.138
Cluster C	20.673	20.887	19.032	21.498

Table 3에서 보이는 바와 같이, 예측의 정확성을 나타낸 MAPE와 오차의 정밀도를 나타낸 RMSE, MAE 측면에서 SVR보다 ANN이 전체적으로 좋은 성능을 보였다.

PCA의 적용 측면에서는 대부분 사용하는 정보가 주기성을 띄기 때문에, 큰 편차를 갖지 않고 분포가 고른 특성을 가지므로, PCA를 사용하지 않는 경우가 모든 특성 정보의 손실 없이 대체로 좋은 결과를 나타내었다.

5.4 학습 데이터별 예측 성능 비교

본 실험에서는 학습에 사용되는 데이터 모델을 다르게 함으로써, 실제 예측 성능상의 차이를 비교한다. Table 4에 제시한 바와 같이, 각기 특성 반영의 정도가 다른 세 가지 모델을 정의한다.

Table 4. Data Models

Data model	Features
Model 1	Month, Day, Hour, Minute
Model 2	Model 1 + Weekend/Holiday
Model 3	Model 2 + Weather information

Model 1은 월, 일, 시, 분의 데이터만 적용한 데이터 모델이다. Model 2는 Model 1에 평일과 주말 및 공휴일의 정보를 추가한 데이터 모델이다. Model 3은 Model 2에 날씨 정보를 추가로 반영한 모델이다.

앞서 실험에서 정확한 예측을 보인 SVR과 ANN을 이용하여, 3가지의 데이터 모델별로 예측 모델을 학습하였다. Table 5에서 볼 수 있듯이, 모든 특성을 반영한 Model 3이 가장 좋은 정확성을 가짐을 알 수 있다. 이를 통해, 전력 사용에 영향을 미치는 다양한 변수들을 고려할수록, 예측의 정확도가 향상됨을 확인할 수 있다.

5.5 실제 전력 사용량과 예측 결과 분석

본 연구에서 구성한 SVR과 ANN 기법을 적용하여, Model 3을 Cluster A의 전력 사용량 예측에 적용한 결과 일부분을 Fig. 8에 나타내었다.

15년 1월 10일은 1월 둘째 주 토요일이며, 방학 기간과 휴일이 포함된 전력 사용 패턴을 나타낸 것이다. 외부 기온이 낮아 히터 등의 전력 사용으로 인해 15년 4월 5일보다

높은 전력 사용량을 알 수 있다. 15년 4월 5일은 4월 둘째 주 일요일이며, 봄 학기가 운영되고 있는 휴일을 나타낸 것이다. 여름과 겨울보다 휴일의 영향을 더 크게 받으며 전력 사용이 전체적으로 낮음을 알 수 있다. 15년 7월 3일은 7월 첫째 주 금요일을 나타낸 것이며, 주중인 방학 기간을 나타낸 것이다. 이는 계절학기 등의 학사 일정 기간에 에어컨 사용으로 인한 전력 사용이 높음을 알 수 있다. 15년 10월 1일은 10월 첫째 주 목요일이며, 학기 운영 중 주중을 나타낸 것이다. 대학의 일반적인 전력 사용 패턴을 보이며, ANN이 SVR보다 실제 전력 사용과 비슷한 패턴으로 예측함을 알 수 있다.

6. 결 론

본 논문은 전력 사용이 많은 곳 중 하나인 대학 캠퍼스의 효율적인 전력 공급 및 관리를 목적으로, 15분 단위의 1년 이상 축적된 대학 캠퍼스 전력 사용량 데이터를 이용하여 예측 모델 학습 및 예측을 수행하였다. 예측 모델은 대표적인 예측 알고리즘으로 알려진 ANN과 SVR을 이용하여 학습하고 성능을 평가하였다. 실험은 서로 다른 특성을 가지는 건물 클러스터 A, B, C에 대해 각각 독립적으로 수행하였으며, MAPE, RMSE 그리고 MAE를 이용하여 예측 모델의 성능을 비교하였다. 예측 성능을 비교한 결과, ANN을 적용한 모델이 모든 클러스터에서 MAPE, RMSE, MAE 측면에 SVR보다 정확한 예측 성능을 보였다. 또한, 예측 성능 부분에서 PCA의 적용 여부와 성능 차이가 크지 않았으며, 실제 예측 결과를 비교해 봤을 때, PCA를 적용하지 않은 ANN이 전반적으로 우수한 성능을 확인할 수 있었다.

향후 전력 사용량 예측의 정확도를 향상하기 위해 교육기관에서 발생하는 이벤트 및 학교 일정 정보와 전력 사용에

Table 5. 5-fold Cross-Validation Using SVR/ANN

	SVR			ANN		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
MAPE						
Cluster A	0.209	0.113	0.100	0.201	0.108	0.096
Cluster B	0.143	0.070	0.059	0.155	0.076	0.064
Cluster C	0.097	0.048	0.042	0.088	0.043	0.038
RMSE						
Cluster A	194.655	98.163	82.774	187.970	94.790	79.930
Cluster B	127.154	59.781	49.971	131.610	61.876	51.722
Cluster C	74.260	32.803	28.691	66.410	29.335	25.658
MAE						
Cluster A	121.836	68.937	59.562	118.672	67.147	58.015
Cluster B	81.299	42.734	35.839	87.755	46.128	38.685
Cluster C	47.923	23.987	20.673	44.119	22.083	19.032

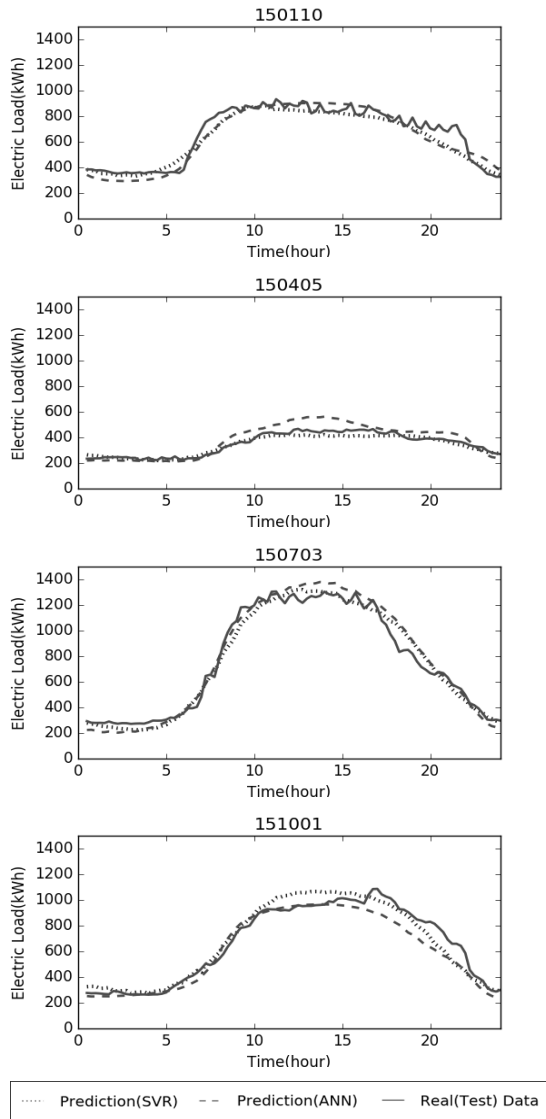


Fig. 8. Prediction Result Example

영향을 주는 추가적인 환경 요인들 분석하고, 다양한 데이터를 반영하여 학습 예측 기법에 적용하는 연구를 수행할 계획이다. 또한, 에너지 운영 전략 측면에서의 피크 부하 방지 및 예측 등을 추가로 연구함으로써 더욱 정확하고 신뢰성 높은 예측 기법을 실현할 수 있을 것으로 기대한다.

References

[1] M. H. Chung and E. K. Rhee, "Potential opportunities for energy conservation in existing buildings on university campus: A field survey in Korea," *Energy and Buildings*, Vol.78, pp.176-182, 2014.

[2] W. J. Lee, D. W. Lee, J. B. Lee, J. H. Yoon, and U. C. Shin, "A Case Study of Electric Power Consumption Characteristics in University Building," *Journal of the Korean Solar Energy Society*, Vol.32, No.4, pp.90-95, 2012.

[3] B. K. Koo, W. H. Hong, and K. M. Kim, "A Study on the Energy Reduction Effect Using Renewable Energy Through the Analysis of Energy Consumption Structure in the University Buildings," *Journal of the Architectural Institute of Korea Planning & Design*, Vol.29, No.9, pp.203-210, 2013.

[4] N. S. Youn and J. T. Kim, "Survey and Analysis of Power Energy Usage of University Buildings," *Journal of the Korea Institute of Ecological Architecture and Environment*, Vol.13, No.2, pp.27-32, 2013.

[5] J. W. Jung, D. W. Kim, J. M. Lee, J. H. Yang, and H. T. Seok, "The Survey and Analysis of Electric Power Consumption in University Building by Analyzing Case Study," *Journal of the Korean Society of Living Environmental System*, Vol.17, No.1, pp.1-9, 2010.

[6] K. C. Noh, S. M. Lee, T. G. Lee, M. D. Oh, and Y. J. Lee, "Comparison of Electricity Consumption in University Buildings for Low Energy Consumption Benchmarking," in *Proceedings of the SAREK Summer Annual Conference*, PyeongChang, pp.823-825, 2013.

[7] A. S. Ahmad, M. Y. Hassan, M. P. Abdullah, H. A. Rahman, F. Hussin, H. Abdullah, and R. Saidur, "A review on applications of ANN and SVM for building electrical energy consumption forecasting," *Renewable and Sustainable Energy Reviews*, Vol.33, pp.102-109, 2014.

[8] M. Q. Raza and A. Khosravi, "A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings," *Renewable and Sustainable Energy Reviews*, Vol.50, pp.1352-1372, 2015.

[9] L. Hernandez, C. Baladron, J. M. Aguiar, B. Carro, A. J. Sanchez-Esguevillas, J. Lloret, and J. Massana, "A survey on electric power demand forecasting: future trends in smart grids, microgrids and smart buildings," *IEEE Communications Surveys & Tutorials*, Vol.16, No.3, pp.1460-1495, 2014.

[10] K. Li, C. Hu, G. Liu, and W. Xue, "Building's electricity consumption prediction using optimized artificial neural networks and principal component analysis," *Energy and Buildings*, Vol.108, pp.106-113, 2015.

[11] K. Li, H. Su, and J. Chu, "Forecasting building energy consumption using neural networks and hybrid neuro-fuzzy system: A comparative study," *Energy and Buildings*, Vol.43, No.10, pp.2893-2899, 2011.

[12] A. Bagnasco, F. Fresi, M. Saviozzi, F. Silvestro, and A. Vinci, "Electrical consumption forecasting in hospital facilities: An application case," *Energy and Buildings*, Vol.103, pp.261-270, 2015.

[13] K. Grolinger, A. L'Heureux, M. A. Capretz, and L. Seewald, "Energy forecasting for event venues: Big Data and prediction accuracy," *Energy and Buildings*, Vol.112, pp.222-233, 2016.

[14] H. Chitsaz, H. Shaker, H. Zareipour, D. Wood, and N. Amjady, "Short-term electricity load forecasting of buildings in microgrids," *Energy and Buildings*, Vol.99, pp.50-60, 2015.

[15] K. P. Amber, M. W. Aslam, and S. K. Hussain, "Electricity consumption forecasting models for administration buildings of the UK higher education sector," *Energy and Buildings*, Vol.90, pp.127-136, 2015.

[16] L. Ghelardoni, A. Ghio, and D. Anguita, "Energy Load Forecasting Using Empirical Mode Decomposition and Support Vector Regression," *IEEE Transactions on Smart Grid*, Vol.4, No.1, pp.549-556, 2013.

[17] S. Jurado, À. Nebot, F. Mugica, and N. Avellana, "Hybrid methodologies for electricity load forecasting: Entropy-based feature selection with machine learning and soft computing techniques," *Energy*, Vol.86, pp.276-291, 2015.

[18] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, Vol.61, pp.85-117, 2015.

[19] Korea Meteorological Administration [Internet], <https://data.kma.go.kr/cmmn/main.do>.

[20] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, Vol.2, No.4, pp.433-459, 2010.

[21] Scikit-learn [Internet], <http://scikit-learn.org/stable/>.

[22] PyBrain [Internet], <http://pybrain.org/>.



문 지 훈

e-mail : johnny89@korea.ac.kr
 2015년 한성대학교 정보통신공학과(학사)
 2015년~현 재 고려대학교 전기전자공학과
 석·박사통합과정
 관심분야: 스마트 그리드, 데이터 마이닝,
 수요 예측, 기계 학습



전 상 훈

e-mail : sjun.korea@gmail.com
 2008년 고려대학교 전기공학과(학사)
 2010년 고려대학교 전자전기공학과(석사)
 2015년 고려대학교 전기전자전파공학과
 (박사)
 2015년~현 재 서울아산병원 의료영상
 로봇연구실 박사후연구원
 관심분야: 멀티미디어 시스템, 음악 구조 분석, 기계 학습,
 웹/모바일 응용 프로그램, 데이터 시각화, 데이터
 마이닝



박 진 응

e-mail : timeless@korea.ac.kr
 2013년 조선대학교 제어계측로봇공학과
 (학사)
 2013년~현 재 고려대학교 전기전자공학과
 석·박사통합과정
 관심분야: 수요 예측, 에너지 운영 계획,
 스마트 그리드



최 영 환

e-mail : work48@krib.co.kr
 2007년 고려대학교 전자공학과(학사)
 2014년 고려대학교 전자전기공학과(박사)
 2014년~2016년 귀뚜라미 연구기획실
 선임연구원
 2016년~현 재 귀뚜라미 연구기획실
 수석연구원
 관심분야: 기계공학, 보일러, 열역학, U-Healthcare, 영상 처리,
 바이오 인포매틱스



황 인 준

e-mail : ehwang04@korea.ac.kr
 1988년 서울대학교 컴퓨터공학과(학사)
 1990년 서울대학교 컴퓨터공학과(석사)
 1998년 Univ. of Maryland at College
 Park 전산학과(박사)
 1998년~1999년 Bowie State Univ. 조교수
 1999년 Hughes Research Lab. 연구교수
 1999년~2004년 아주대학교 정보통신전문대학원 조/부교수
 2004년~현 재 고려대학교 전기전자공학과 교수
 관심분야: 데이터베이스, 빅데이터 처리, 영상 처리, 음악 검색,
 멀티미디어 검색, 소셜 네트워크 데이터 분석