

기상 및 전력 사용량 데이터를 이용한 인공지능 교육 모델 개발

김갑수*

서울교육대학교

A Development of AI Education Model Using Weather and Power Usage Data

Kim, Kapsu*

Seoul National University of Education

Abstract : The topic of artificial intelligence education is now emerging. It is being promoted not only in our country but also in other countries to systematically educate artificial intelligence from elementary school. In artificial intelligence education, the importance of data is paramount. In this study, we predict the relationship between maximum power usage due to climate change by applying it to AI learning models. As a result of this study, we found that the regression model was unsuitable because of its small determinant, and that the ensemble model, the Random Forest learning model, was suitable. When learning based on 365 data on weather and maximum power usage in 2019, we found that 175 trees, the factor in the Random Forest learning model, had the largest coefficient of determination.

Keywords : maximum power usage, weather data, regression, random forest, artificial intelligence

요약 : 현재 인공지능 교육에 대한 화두가 떠오르고 있다. 초등학교부터 인공지능 교육을 체계적으로 하는 것을 우리나라뿐만 아니라, 다른 나라에서도 추진 중에 있다. 인공지능 교육에서는 데이터의 중요성이 무엇보다도 중요하다. 본 연구에서는 기상 변화에 따른 최대 전력 사용량 간의 관계를 인공지능 학습 모델에 적용하여 예측해 본다. 본 연구의 결과, 회귀 모형은 결정계수가 적어서 적합하지 않다는 것을 알았고, 앙상블 모델인 램덤 포레스트 학습 모형이 적합한 것을 알았다. 2019년 기상 및 최대 전력 사용량에 대한 365 데이터를 기반으로 학습할 때에는 램덤 포레스트 학습 모델에서의 인자인 트리의 개수가 175개로 결정계수가 가장 크다는 것을 알았다.

주요어 : 최대 전력 사용량, 기상 데이터, 회귀 분석, 램덤 포레스트, 인공지능

1. 서론

4차 산업혁명에서 인공지능이 매우 좋은 이슈로 대두되고 있다. 인공지능 교육이 초등학교부터 체계적으로 이루어져야 한다고 이야기하고 있다. 기계가 인간의 노동을 대체하는 산업화에서 사람의 지적 활동을 인공지능이 대체하는 시대에 도달하였다. 우리나라를 비롯하여 전 세계는 인공지능을 위한 교육 시스템의 변화 등을 추진하고 있다.

국내 대부분의 산업에서 인공지능을 적용하고 있고, 인공지능 산업에는 향후 5년간 연평균 17.8%로 성장할 것(한국 IDC, 2020)을 예측할 정도로 성장하고 있다. 또한, 인공지능 국가 전략으로 세계 최고의

AI 인재 양성 및 전국 AI 교육(대한민국정부, 2019)을 추진 중이다. 한국판 뉴딜 정책으로 인공지능 및 소프트웨어 인재 10만 명 양성을 위한 국가 정책이 2020년부터 추진되고 있고(이병진, 2020), 전국 AI SW 교육 확산 방안을 제안하여 초등학교에서도 AI 교육의 필요성을 이야기하고 있다. 또한, 고등학교에서는 인공지능 기초와 인공지능 수학이라는 교과목을 신설하고 있다.

미국의 인공지능 교육은 유치원부터 고등학교까지 인식영역, 표현과 추론 영역, 학습 영역, 상호작용 영역 및 사회적 영향 영역으로 나누어 체계적으로 교육 과정을 만들고 있다(AI4K12, 2019).

중국의 경우에는 인공지능 교과서를 초등학교부터

*Corresponding author : 김갑수
E-mail : kskim@snu.ac.kr

체계적으로 가르치는 것을 중국 각 성 별로 만들어서 교육하고 있다.

우리나라도 2019년에 차세대 SW 교육표준 모델의 프레임(한국과학창의재단 2019)을 만들어 이 모델 안에 인공지능 영역을 추가하여 인공지능을 교육하게 하고 있고, 유치원 및 초등학교 1~2학년을 대상으로 인공지능 교육 프로그램에 관한 연구 등을 하고 있다(정영식, 2020).

인공지능 교육에는 데이터가 무엇보다도 중요하다. 따라서 국가 인공지능 교육을 위한 데이터를 기상 변화 데이터와 에너지 데이터를 이용하면 우리 학생들에게 인공지능 기술 교육과 더불어 에너지 및 기상 변화 교육을 바로 할 수 있다는 장점이 있다.

현대 사회에서는 인공지능 데이터와 더불어 에너지가 매우 중요하다는 것을 알 수 있다. 즉, 인공지능에서 데이터 축적이 중요하고, 또한 에너지가 사회를 움직이게 하는 동력 역할을 한다(Jennings, 2009). 에너지에 대한 소비는 양적으로 계속 증가하고 있다(Dias et al., 2004; World Bank, 1992). 에너지 사용이 증가하면서 많은 문제가 야기되기 때문에 초등학교 때부터 에너지를 잘 사용할 수 있는 절약 운동이 강조되어야 한다(Dias et al., 2004; Jennings, 2009).

에너지 교육은 초등학교부터 통합적인 접근 방법이 강조되어 왔다(이훈식, 1999; 인재천, 2012; Hungerford & Volk, 1990). 이것은 독립 교과로 만들 수는 없지만, 기본 교과에서 에너지의 중요성이 녹아 들어가야 한다는 것이다. 초등학교에서 소프트웨어 교육을 할 때에 5개의 소프트웨어 성취기준에 에너지 요건을 계산하는 프로그램을 만드는 것이 대표적인 사례가 될 수 있다(김갑수, 2019).

본 연구의 목표는 기상과 에너지 데이터를 이용하여 인공지능에 적용해 보는 교육 프로그램을 개발하는 것이다. 본 연구에서 개발한 교육 프로그램의 사례는 에너지 및 기상 변화 교육과 인공지능 교육에 동시에 만족할 수 있는 장점이 있다.

이 방법은 인공지능 학습 모델의 사례로 학생들이 기상변화 데이터와 에너지 데이터를 수집하고, 이를 인공지능 학습 모델에 적용해 봄으로써 에너지 사용 교육을 실천적으로 할 수 있다. 또한, 기상변화 데이터와 에너지 최대 사용량 데이터가 인공지능 교육의 기본 데이터 집합을 제공함으로써 인공지능 교육을 하는 표준 데이터로 기상 변화와 에너지의 중요성을

인지할 수 있다.

제2장에서는 본 연구와 관련 있는 연구로서 공공 데이터와 기계학습 모형에 대해서 설명한다. 제3장에서는 학생들이 에너지와 기상 데이터를 이용하여 인공지능을 학습하는 모형을 제안한다. 제4장에서는 학습 모형을 기반으로 기상 데이터를 통해 에너지 최대 소비량을 예측하는 실사례를 설명한다. 제4장은 연구의 결론이다.

II. 관련 연구

1. 개요

본 연구에서 인공지능 연구를 위한 기본적인 데이터를 획득하는 방법과 기계학습 모델에 많이 사용하고 있는 모델에 대해 간단히 설명한다.

2. 데이터 수집 방법

데이터를 수집하는 방법은 공개 데이터를 이용하는 방법과 직접 데이터를 수집하는 방법이 있다. 본 연구에서 공개 데이터를 수집하는 간단한 사이트에 대해 설명한다.

먼저 우리나라의 공개 데이터는 국가 통계포털(<https://kosis.kr>)이 대표적이다. 이 사이트에는 우리나라의 통계자료를 주제별 또는 기관별(정부 부처 및 산하 기관) 모든 통계자료를 수록하고 있다. 여기서 우리나라의 기본 통계자료를 확보할 수 있다. 기본적인 정보 중에 한 개의 사례는 우리나라 각 지역별 인구데이터를 연령별, 성별로 제공하고 있고, 이 인구데이터를 이용하여 각종 예측을 할 수 있다.

두 번째로, 우리나라의 공개 데이터는 공공데이터포털(<https://www.data.go.kr>)이다. 이 사이트에는 우리나라 953개 기관에서 데이터 파일과 오픈 API를 제공하고, 표준 데이터 집합도 제공하고 있다. 제공되는 사례는 서울시의 CCTV 위치 정보 등 많은 정보가 제공되고 있다.

세 번째는, 우리나라 인공지능 데이터를 구축하여 데이터를 제공하는 AI허브(aihub.or.kr) 사이트이다. 이 사이트는 우리나라 인공지능 데이터를 구축하는 2019년 데이터 집합을 제공하고 있고, 2020년에 구축할 데이터도 제공할 예정이다. 사례로는 우리나라 글자체에 대한 데이터를 제공하고 있다.

네 번째는, K-ICT 빅데이터 센터(<https://kbig.kr>)

이다. 이 사이트는 공공데이터들의 데이터 집합들과 연계되어 정보를 제공하고 있다. 이곳에는 지역 데이터를 효과적으로 제공하는 지역 정보제공 사이트(<http://www.localdata.go.kr/>)와 연계되어 있으며, 서울시 공공정보(<http://data.seoul.go.kr/>)도 연계되어 있고, 빅데이터 지도(<https://www.bigdata-map.kr/>)도 연계되어 있고, 마이크로 정보(<https://mdis.kostat.go.kr/>) 등도 제공하고 있다.

본 연구에서 사용할 데이터는 기상자료 개발 포털 사이트(<https://data.kma.go.kr/>)와 전력통계정보시스템(<http://epsis.kpx.or.kr/>)도 있다.

미국의 정부에서 공공 데이터를 매우 광범위하게 제공하고 있는 곳은 <https://www.data.gov/>이다. 이 사이트는 2020년 11월 29일 현재 218,168개를 제공하는 거대한 데이터 집합으로 PDF, CSV, HTML, JSON 등의 다양한 형태로 제공하고 있다. 기상 데이터, 에너지 데이터 등을 포함하여 정부 공공기관 데이터들을 광범위하게 제공하고 있다.

유럽 국가들의 공공데이터(<https://data.europa.eu/>)에서 제공하는 데이터 집합이 있고, 영국에서 제공하는 공공데이터(<https://data.gov.uk/>)가 있으며, 유럽 국가들이 공공데이터를 제공하고 있다.

오랫동안 기계학습 데이터를 제공한 곳은 캘리포니아 주립대학에서 제공하는 전통적이고 정제된 데이터(<https://archive.ics.uci.edu/ml/index.php>)가 있다. 구글에서 지원하는 캐글(<https://kaggle.com>)에서 인공지능 학습 데이터와 데이터를 이용한 경진 대회 등을 개최하고 있다.

3. 학습 모델

본 연구에서 사용할 수 있는 기계학습 모델은 많이 있다. 회귀 모델에서 사용될 수 있는 학습 모델에 대해서 간단히 정리하면 다음과 같다.

첫 번째 학습 모형은 회귀 모형이다. 회귀 모형에는 선형회귀 모형과 비선형 회귀 모형이 있다. 선형회귀 모형과 비선형 회귀 모형에는 독립변수를 여러 개 가질 수 있다.

먼저 독립변수 한 개를 갖는 경우를 살펴보면 식 1은 독립변수 한 개의 선형회귀이다.

$$y = w \times x + b \quad (\text{식 1})$$

식 1에서 독립변수 x 가 2차식, 3차식, 즉 n 차식으로 확장 가능하다. 그러면 비선형이 된다. 식 2는 독립변수 한 개에 3차식으로 확대한 비선형 회귀식이다.

$$y = w_1 \times x^3 + w_2 \times x^2 + w_3 \times x + b \quad (\text{식 2})$$

독립변수가 3개일 때의 선형식은 다음 식 3과 같다. 식 3과 같은 식에 각 상수를 예측하는 것이 선형 모델이 된다.

$$y = w_1 \times x_1 + w_2 \times x_2 + w_3 \times x_3 + b \quad (\text{식 3})$$

독립변수가 3개일 때의 비선형식은 다음 식 4와 같다.

$$y = w_1 \times x_1^2 + w_2 \times x_1 + w_2 \times x_2 + w_3 \times x_3 + b \quad (\text{식 4})$$

이와 같이 매우 다양한 식을 만들어서 계수를 정할 수 있다.

두 번째 학습 모형은 서포트 벡터 머신(support vector machine) 모형이다(Chris Burges, 1998, Nello Cristianini & John Shawe-Taylor, 2000). 이 모형은 다음 Fig. 1과 같이 벡터 개념을 이용하여 두 개로 분류하는 모형이다.

Fig. 1의 모형에서와 같이 벡터 개념을 이용하여 두 개로 분류하는 모형이고 성능이 매우 뛰어난 것으로 알려져 있다. 이 모형의 선택에는 선형 모형과 다항식 모형과 커널 모형이 있다. 각각에 대한 인자들을 수정하여 데이터에 대한 새로운 모

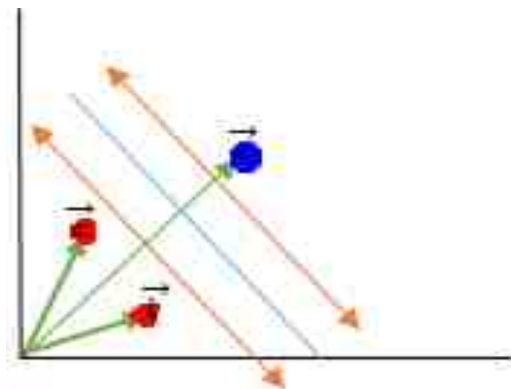


Fig. 1. SVM 모형

형들을 만들 수 있다.

세 번째 학습 모형은 앙상블 모형이 있다. 앙상블 모형 중에 대표적인 방법은 랜덤 포레스트 모형, 결정 트리 및 그레디언트 부스팅 방법 등이 있다. 먼저 랜덤 포레스트 방법은 의사 결정 트리를 랜덤하게 많이 만들어서 투표로 결정하는 방법이다. 이 방법에서 선택할 수 있는 인자는 트리의 수와 트리의 깊이 트리의 특성이다. 이 세 가지 인자들을 자유롭게 주어 결정계수가 가장 큰 것을 찾으면 좋은 방법이 될 것이다. 일반적으로 트리의 수는 많이 하면 좋다. 트리들의 평균을 계산하면 된다. 따라서 너무 많은 수행속도가 된다.

그레디언트 부스팅 방법은 트리의 가지치기를 많이 하여 성능을 높이는 방법이다. 트리의 오차를 보전하는 방법을 많이 사용한다. 이 방법은 학습률이라는 것이 추가되는 인자이다. 또 다른 방법은 에이다 부스팅 방법도 있다.

네 번째 모형은 K근접(KNN) 모형이고 두 지점까지의 거리를 계산하여 근접에 있는 것은 같은 그룹으로 묶는 방법이다. 거리 계산하는 방법은 Fig. 2와 같다. 이 모형에는 인자는 이웃 수를 결정하는 것과 가중치를 주는 방법 및 거리 계산 알고리즘을 정하는 방법에 따라 결정계수가 다르다.

다섯 번째 모형은 결정 트리 모형으로 분류 기준을 만들어 트리를 만드는 것이다. 이 모형의 인자는 최대 깊이를 결정하는 것과 분류 기준을 결정할 때 엔트로피와 기니 방법이 있다.

마지막으로 신경망 방법이다. 이 방법의 인자는 은닉층의 수, 활성화 함수, 최적화 선택 방법, 배치 크기, 학습률, 최대 반복횟수 등이 있다.

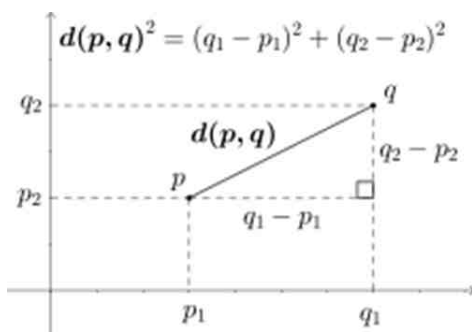


Fig. 2. 거리 계산하는 방법

III. 모 델

1. 개요

본 장에서는 학생들이 실험해 볼 수 있는 에너지 데이터에 영향을 미치는 기상 데이터를 이용하여 인공지능 교육 모델을 사용해 보는 교육 모델을 제안한다. 본 모델은 에너지 데이터에 영향을 미치는 데이터 만들기 단계, 학습 모델 만들기 단계, 적용해 보기 단계로 나눈다. 이 세 단계에는 일반적으로 학생들이 학습할 수 있는 데이터를 수집하여 인공지능 모델에 적용하고, 각 인자들을 수정하여 적합한 모델을 만들어 적용해 보는 것이다.

본 연구에서 제안하는 3단계 모형은 다음 Fig. 3과 같다. 단계별로 아래 각 절에서 상세히 설명한다.

2. 데이터 단계

인공지능의 핵심은 데이터이고, 데이터가 학습 모델을 만들기 때문에 표준 데이터를 만들어 보는 것이 매우 중요하다. 기본적인 표준 데이터는 통계학을 이론적으로 발전시킨 피셔가 만든 아이리스 붓꽃 데이터이다. 이 데이터를 기계학습 모델에 대표적으로 사용하는 것이다.

데이터 단계에서는 데이터를 수집하는 단계와 데이터를 정리해서 표준화하는 단계로 나눌 수 있다. 데이터 수집 단계에서는 공개 데이터를 이용하는 방법 또는 직접 수집하는 방법이 있다. 공개 데이터를 이용하는 방법은 데이터를 제공하는 기관에서 기관 특성에 따라 다양한 요소들의 데이터를 제공하고 있어서 원하는 데이터를 바로 찾을 수 없다. 일단 인공지능 모



Fig. 3. 모델

델에 적용할 데이터를 요소별로 찾는 것이 중요하다.

두 번째 단계에서는 수집 단계에서는 수집한 데이터를 정리하는 단계이다. 데이터를 수집하여 필요한 데이터를 제외 또는 특정 데이터를 통합하는 것이 필요하다.

이렇게 두 단계를 거쳐서 분석한 학습 데이터를 만드는 것이다.

3. 학습 모델

인공지능에서 학습 모듈은 매우 많은 학습 방법들이 있다. 이 학습 방법마다 다양한 특성이 있다. 따라서 학습 방법들을 가능한 한 많이 적용하여 적합한 모델을 만들어 볼 필요가 있다.

학습 모델의 세부 단계는 다음 Fig. 4와 같다.

Fig. 4의 각 단계에 대한 설명은 다음과 같다.

먼저 첫 번째는 훈련 데이터를 선정하는 것이다. 데이터 단계에서 정리한 데이터를 학습 모델에 사용할 데이터와 시험해 볼 데이터로 구별한다. 일반적으로 데이터 단계에서는 수집한 데이터의 70%는 학습 데이터로 사용하고, 나머지 30%는 시험 데이터로 사용한다. 그러나 꼭 비율을 정할 필요는 없다. 훈련 데이터를 데이터 단계에서 수집한 데이터로 임의로 선

택한다.

두 번째는 학습 모델을 선택한다. 학습 모델은 학생들이 개념적으로 이해할 수 있는 모델들을 선택하는 것으로 회귀 모형과 분류 모형으로 나누어져 있는데, 회귀 모형은 연속적인 데이터를 가질 때 선택하는 것이고, 분류 모형은 비연속적인 데이터를 선택한다. 이때 데이터의 특성에 따라 모델을 선택한다.

세 번째는 선택한 모델마다 인자가 있다. 보통 학습 모델이 디폴트값이 있지만, 대부분의 학습 모델들은 인자들을 직접 입력할 수 있다. 이때 인자들의 특성에 따라 잘 선택하여야 한다.

네 번째는 학습을 수행한다. 즉, 훈련 데이터를 학습 모델에서 선택한 학습 방법과 인자를 기반으로 학습을 수행한다. 학습을 수행하면 학습 모델이 만들어진다.

다섯 번째는 본 연구에서는 에너지 사용량을 기상 데이터를 이용하여 예측하기 때문에 연속적인 데이터이고, 이것은 회귀 모형이기 때문에 회귀 모형의 경우에는 결정계수가 중요하다. 따라서 결정계수를 계산한다.

분류 모델에서는 잘못 분류되는 것의 비율 등이 그 모델의 성능을 결정하는 것이다. 본 연구에서는 분류 모델보다 회귀 모델에 초점을 맞추기 때문에 분류에 대한 평가 기준은 고려하지 않는다.

여섯 번째는 결정계수가 원하는 것을 통과하는지를 정한다. 선택한 모델의 가장 큰 결정계수들을 같은 인자를 선택하는 것이다. 이때 일정한 학습을 하면 더 이상 결정계수가 증가하지 않거나 오히려 결정계수가 감소하는 경향이 있다. 오히려 결정계수가 감소할 때에는 과적합한 상태라고 한다.

마지막으로 학습 모델이 완성된다. 이 학습 모델은 각각의 상수 등이 결정되는 것이다. 만약 완성된 학습 모델이 만족스럽지 않을 때는 또 다른 학습 모델을 선택하여 Fig. 4의 3단계부터 차례대로 수행한다.

이런 과정을 거쳐 본 연구에서 기상 데이터와 에너지 데이터를 통합하여 에너지 최고 사용량을 예측하는 모형을 만들 수 있다.

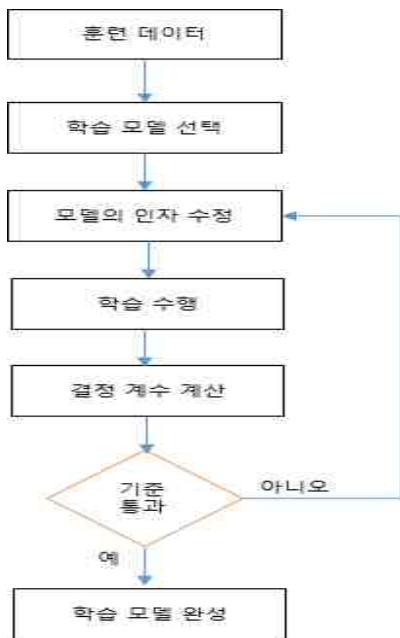


Fig. 4. 학습 모델

4. 적용 단계

3.2절에서 만든 학습 모델들을 실제 적용해 보는 것이다. 적용 단계는 Fig. 5와 같다.

Fig. 5에 대한 적용 단계의 3단계에 대한 자세한



Fig. 5. 적용 단계

설명은 다음과 같다.

첫 번째 단계인 시험 데이터 단계이다. 이 단계에서는 데이터 단계에서 수집한 데이터를 학습 모델을 만들 때 사용한 데이터를 제외한 것이 훈련 데이터이다.

두 번째 단계인 학습 모델 적용 단계에서는 적용할 수 있는 데이터 즉 시험 데이터를 기반으로 학습 모델에 직접 적용해 본다. 적용한 결과는 학습 모델에 따라 평가 요소들이 다르게 될 수 있지만, 기본적으로 실제 데이터와 학습 모델로 만든 데이터와의 차이 분석을 해 본다.

마지막 단계는 평가하기 단계이다. 시험 결과 등을 평가하여 학습 모형에 대한 유용성 또는 데이터 집합에 대한 유용성 등을 평가하여 이와 유사한 것에 사용하는 학습 모델로 이용할 수 있다.

IV. 적용 사례

1. 개요

데이터를 수집하기 위해서 공공 데이터에서 기상 데이터를 먼저 수집하고, 두 번째는 전기소비량 데이터를 수집하여 두 개의 파일을 한 개의 파일로 묶으면 간단히 처리할 수 있다.

2. 문제 정의하고 이해하기

문제를 정의하고 이해하기 단계가 문제 해결의 실마리가 된다.

본 문제는 기상 변화 데이터를 수집하여 에너지 소비량을 예측하는 인공지능 모델에서 기계학습 모형을 적용하여 필요한 에너지 소비량을 예측해 본다. 즉, 기상 데이터를 이용하여 에너지 소비량을 예측해 본다. 우리나라 기상 평년값 데이터를 구한다. 기상 값

은 평균온도, 평균습도, 바람, 일조시간의 정보를 구한다. 이런 기상 데이터가 전력 소비량과 어떤 관계가 있는지 기계학습을 통하여 알아본다.

3. 데이터 수집하기

본 연구에서는 데이터를 수집하기 위해서 공공 데이터에서 기상 데이터를 먼저 수집하고, 두 번째는 전기소비량 데이터를 수집한다. 세 번째는 두 개의 데이터를 한 개의 파일로 묶으면 간단히 처리할 수 있다.

먼저 기상 데이터를 수집하고 정리하는 과정을 살펴본다.

1단계: <https://data.kma.go.kr/cmmn/main.do> 사이트에 들어가면 다음과 같은 Fig. 6과 같은 화면이 나타난다.

2단계: 기상 통계 메뉴를 클릭하면 다음과 같은 Fig. 7과 같은 조건별 통계 화면을 찾는다.

3단계: 기상 통계 메뉴를 클릭하면 다음과 같은 Fig. 8과 같은 조건별 통계 화면을 찾는다.

검색조건에는 고려해야 할 사항으로 지역을 선택한다. 우리나라 전체 전력 최고 사용량을 예측하는 기계학습 모델을 만들어 보는 것이기 때문에 지역을 전국으로 선택하고, 요소는 기온을 선택하고, 기간은 일

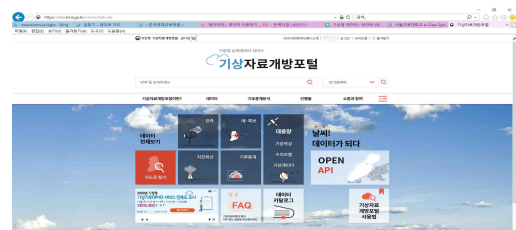


Fig. 6. 기상공개 자료



Fig. 7. 조건 통계 방법

■ 검색조건

• 분류 **지상** • 지역/지점 전국 **선택** • 요소 **기온**

• 기간 **일** 2019년 ~ 2019년

• 조건

요소 평균기온 < > 선택

월 01월 ~ 12월

일 01일 ~ 31일

계절 봄

> 검색

Fig. 8. 검색조건

단위로 2019년 1월 1일부터 2019년 12월 31일까지의 데이터를 기반으로 한다. 이에 대한 데이터를 차례대로 선택하여 데이터를 수집한다.

4단계: 위의 조건을 입력한 후에 조건 검색한 후에 다음 Fig. 9 및 Fig. 10과 같이 나타난다.

5단계: Excel 또는 csv 파일을 다운로드 받는다. 다운로드 받은 파일을 정리한다. 기상 데이터를 정리한 파일은 다음과 Fig. 11과 같다.



Fig. 9. 기온 데이터(그래픽)

| 지역 | 지역명 | 일시 | 평균기온 (°C) | 평균최고기온(°C) | 평균최저기온(°C) | 최고기온 관측지점 | 최저기온 관측지점 | 최대전력 (MW) | 최소전력 (MW) |
|----|-----|------------|-----------|------------|------------|-----------|-----------|-----------|-----------|
| 0 | 전국 | 2019-01-01 | -2.1 | 2.1 | 7 | 제주,서귀포 | -5.8 | -16.5 | 평화 |
| 0 | 전국 | 2019-01-02 | -2.5 | 3.2 | 8.4 | 서귀포 | -7 | -16.6 | 평화 |
| 0 | 전국 | 2019-01-03 | -2.1 | 5 | 10.6 | 서귀포 | -7.7 | -16.8 | 평화 |
| 0 | 전국 | 2019-01-04 | -0.7 | 5.3 | 11.8 | 서귀포 | -7.3 | -14.3 | 제천 |
| 0 | 전국 | 2019-01-05 | 0.2 | 5.3 | 13.2 | 서귀포 | -6.6 | -13.5 | 대관령 |
| 0 | 전국 | 2019-01-06 | -1.3 | 5.7 | 12.3 | 서귀포 | -6.1 | -15.8 | 대관령 |
| 0 | 전국 | 2019-01-07 | -0.5 | 5.5 | 12.8 | 서귀포 | -6.3 | -15.2 | 평화 |
| 0 | 전국 | 2019-01-08 | -0.7 | 4.2 | 12.1 | 서귀포 | -6.8 | -13.5 | 평화 |

Fig. 10. 기온 데이터(수치)

| 일시 | 일사합 | 평균기온 | 평균습도(%rh) | 평균풍속(m/s) |
|------------|-------|------|-----------|-----------|
| 2019-01-01 | 7.35 | -3.8 | 58 | 2.5 |
| 2019-01-02 | 10.54 | 5 | 53 | 2.1 |
| 2019-01-03 | 11.17 | 2.6 | 48 | 1.6 |
| 2019-01-04 | 6.66 | 0.5 | 53 | 1.4 |
| 2019-01-05 | 10.07 | 0.6 | 49 | 2.1 |
| 2019-01-06 | 10.28 | 3.7 | 46 | 1.6 |
| 2019-01-07 | 9.68 | 3.5 | 53 | 1.8 |
| 2019-01-08 | 10.26 | 2.1 | 46 | 2.8 |
| 2019-01-09 | 9.07 | 3.8 | 40 | 1.6 |
| 2019-01-10 | 7.16 | 3 | 51 | 2 |
| 2019-01-11 | 9.21 | 0.2 | 63 | 1.3 |
| 2019-01-12 | 6.33 | 1 | 77 | 1.3 |

Fig. 11. 기상 데이터

다음은 전력 데이터를 수집하고 정리하는 과정을 살펴본다.

1단계: 전력데이터 개방포탈 시스템(<https://bigdata.kepco.co.kr/>) 또는 전력통계정보시스템(<http://epsis.kpx.or.kr>)에서 전력 최대 사용량을 데이터를 찾아본다. 전력통계 정보 시스템에 가면 다음과 같은 화면이 Fig. 12와 같이 나타난다.

2단계: Fig. 12에서 최대 전력 수급 메뉴를 선택하면 다음 Fig. 13과 같이 나타난다.

Fig. 13에서 날짜를 지정하면 2019년 1월 1일부터 2019년 12월 31일까지의 최대 전력 소비량이 다음 Fig. 14와 같이 나타난다.

이를 엑셀로 다운로드한다. 그러면 기상 데이터의 엑셀 파일과 전력 데이터의 엑셀 파일에서 데이터를 통합하여 기상정보와 전력 사용량 정보가 있는 한 개의 파일을 만든다. 물론, 각각의 파일에서 해당 데이터 집합만 사용할 수도 있다.

Fig. 15 데이터가 기상 데이터와 에너지 최대 사용

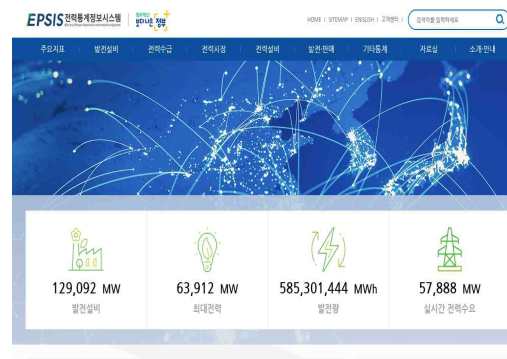


Fig. 12. 전력통계정보시스템



Fig. 13. 전력 데이터

| 년 | 월 | 일 | 발전량(MWh) | 평균능력(MW) | 최대전력(MW) | 평균예비용(MWh) | 평균예비용(MWh) | 최대전력(MWh) |
|------|----|----|----------|----------|----------|------------|------------|------------------|
| 2019 | 12 | 30 | 124,854 | 53,749 | 75,158 | 18,591 | 25 | 2019-12-30 10:30 |
| 2019 | 12 | 29 | 124,854 | 94,005 | 63,540 | 31,485 | 50 | 2019-12-29 19:00 |
| 2019 | 12 | 28 | 124,854 | 93,308 | 65,014 | 28,284 | 44 | 2019-12-28 01:00 |
| 2019 | 12 | 27 | 124,854 | 89,985 | 78,548 | 15,417 | 20 | 2019-12-27 19:00 |
| 2019 | 12 | 26 | 124,854 | 82,905 | 78,431 | 13,474 | 17 | 2019-12-26 11:00 |
| 2019 | 12 | 25 | 124,854 | 82,654 | 65,810 | 27,044 | 41 | 2019-12-25 19:00 |
| 2019 | 12 | 24 | 124,854 | 92,536 | 76,456 | 13,080 | 17 | 2019-12-24 10:00 |
| 2019 | 12 | 23 | 124,854 | 91,375 | 78,779 | 12,396 | 16 | 2019-12-23 19:00 |
| 2019 | 12 | 22 | 124,854 | 88,528 | 62,827 | 26,701 | 43 | 2019-12-22 19:00 |
| 2019 | 12 | 21 | 124,854 | 88,121 | 68,564 | 21,557 | 32 | 2019-12-21 11:00 |
| 2019 | 12 | 20 | 124,854 | 90,886 | 79,156 | 11,738 | 13 | 2019-12-20 10:00 |
| 2019 | 12 | 19 | 124,854 | 91,462 | 77,869 | 13,713 | 18 | 2019-12-19 10:00 |
| 2019 | 12 | 18 | 124,854 | 91,281 | 77,866 | 13,615 | 18 | 2019-12-18 10:00 |
| 2019 | 12 | 17 | 124,833 | 90,041 | 77,487 | 12,554 | 16 | 2019-12-17 19:00 |
| 2019 | 12 | 16 | 124,833 | 90,759 | 76,423 | 14,334 | 19 | 2019-12-16 10:00 |
| 2019 | 12 | 15 | 124,833 | 89,000 | 61,405 | 27,885 | 45 | 2019-12-15 19:00 |

Fig. 14. 전력 데이터

| 일시 | 일사합 | 평균기온 | 평균습도(%) | 평균풍속(m) | 최대전력(MW) |
|------------|-------|------|---------|---------|----------|
| 2019-01-01 | 7.35 | -3.8 | 58 | 2.5 | 65493 |
| 2019-01-02 | 10.54 | 5 | 53 | 2.1 | 82437 |
| 2019-01-03 | 11.17 | 2.6 | 48 | 1.6 | 84196 |
| 2019-01-04 | 6.66 | 0.5 | 53 | 1.4 | 84746 |
| 2019-01-05 | 10.07 | 0.6 | 49 | 2.1 | 69496 |
| 2019-01-06 | 10.28 | 3.7 | 46 | 1.6 | 66493 |
| 2019-01-07 | 9.68 | 3.5 | 53 | 1.8 | 84074 |
| 2019-01-08 | 10.26 | 2.1 | 46 | 2.8 | 83818 |
| 2019-01-09 | 9.07 | 3.8 | 40 | 1.6 | 85392 |
| 2019-01-10 | 7.16 | 3 | 51 | 2 | 83833 |
| 2019-01-11 | 9.21 | 0.2 | 63 | 1.3 | 81941 |
| 2019-01-12 | 6.33 | 1 | 77 | 1.3 | 68147 |

Fig. 15. 기상과 전력 최대사용량

량을 예측하는 학습 모델에 사용할 데이터이다.

4. 학습 모델 만들기

학습 모델을 만들기 위해서 기상 데이터와 에너지 최대 사용 데이터에서 훈련 데이터와 시험 데이터를

구별한다. 훈련 데이터는 학습 모델에 따라 70%로 훈련 데이터를 이용하는 경우가 있다. 첫 번째 단계로 훈련 데이터 정의하는 것을 보자. 본 연구에서 학생들이 쉽게 접근할 수 있는 파이썬이나 R 언어를 이용하여도 된다. 편의상 실습 프로그램으로 파이썬을 선택한다.

그러면 학습 모델을 만들어 본다. 독립변수가 기상 데이터인 ‘일사합’, ‘평균기온’, ‘평균습도’, ‘평균풍속’을 선택하고, 이 기상 데이터가 ‘최대전력’인 종속 변수에 선형 회귀 분석을 해 본다.

먼저 할 일은 다음과 같다.

독립변수와 종속 변수를 정한다.

```
x=newData[['일사합','평균기온','평균습도','평균풍속']]
y=newData['최대전력']
```

다음은 70%를 훈련데이터로 만든다. 그 결과는 다음과 같다.

```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test =
train_test_split(x,y,test_size=0.3,random_state=5)
```

다음은 학습 모델을 선택한다. 학습 모델은 회귀 모델을 선택한다.

```
from sklearn.linear_model import LinearRegression
model = LinearRegression()
```

다음은 인자를 선택한다. 선형 모형은 인자 없이 수행한다.

다음은 학습한다.

```
model.fit(x_train, y_train)
```

다음은 결정 계수 등을 정한다.

```
print('가중치 a: ', model.coef_)
print('y절편 b', model.intercept_)
print('\n')
relation_square = model.score(x_train, y_train)
print('결정계수 R:', relation_square)
print('\n')
```


이 모형은 결정계수가 0.059로 매우 적다. 따라서 4개항의 선형회귀는 원하는 학습 모형을 만들 수 없다. 다른 모형을 선택한다.

선택하는 모형은 2차식을 갖는 회귀 모형을 구하여 보아도 결정계수가 0.19566이다. 이 모형도 좋은 모형이 아니다. 단순회귀나 중회귀 모형은 기상데이터를 가지고 최대전력을 예측하는 모형을 만들 수 없다. 그 이유는 온도가 낮은 겨울에도 최대 전력량이 많고 온도가 높을 때도 최대 전력량이 많기 때문이다.

양상불 모형인 램덤포레스트 모형은 Fig. 16과 같고, 그 결과 Fig. 17은 램덤포레스트 방법으로 인자들이 트리의 개수가 50개부터 5개씩 추가되면 결정계수가 변경되는 것을 볼 수 있다. 결정계수가 가장 큰 인자가 170일 때이다.

양상불 모델인 램덤포레스트를 학습 모델로 선정하고, 인자는 트리의 개수가 175개인 것을 선택한다.

```

model = RandomForestRegressor
      (n_estimators=175,random_state=15)

from sklearn.ensemble import RandomForestRegressor

for i in range (50, 200, 5):

    model = RandomForestRegressor(n_estimators=i,random_state=15)
    model.fit(x, y)

    relation_square = model.score(x, y)
    print('결정계수 R:', relation_square)
    
```

Fig. 16. 램덤 포레스트

| | | |
|---------|-----|----------------------|
| 결정계수 R: | 50 | 0.08849919001450363 |
| 결정계수 R: | 55 | 0.08853905078472349 |
| 결정계수 R: | 60 | 0.08851731695574724 |
| 결정계수 R: | 65 | 0.08848459569806153 |
| 결정계수 R: | 70 | 0.08845587384423301 |
| 결정계수 R: | 75 | 0.08839128359782032 |
| 결정계수 R: | 80 | 0.08857245865869355 |
| 결정계수 R: | 85 | 0.0885634251559626 |
| 결정계수 R: | 90 | 0.08869225069906871 |
| 결정계수 R: | 95 | 0.0876902887313499 |
| 결정계수 R: | 100 | 0.08777619004498349 |
| 결정계수 R: | 105 | 0.08868491332697102 |
| 결정계수 R: | 110 | 0.08869860180342195 |
| 결정계수 R: | 115 | 0.08870690523400774 |
| 결정계수 R: | 120 | 0.08877477451469752 |
| 결정계수 R: | 125 | 0.08881579678835647 |
| 결정계수 R: | 130 | 0.08883410996209411 |
| 결정계수 R: | 135 | 0.08882455174762497 |
| 결정계수 R: | 140 | 0.08882418263442123 |
| 결정계수 R: | 145 | 0.08891106511398614 |
| 결정계수 R: | 150 | 0.08896572818611278 |
| 결정계수 R: | 155 | 0.0890154156489199 |
| 결정계수 R: | 160 | 0.08903607010603233 |
| 결정계수 R: | 165 | 0.08903732486895753 |
| 결정계수 R: | 170 | 0.08905462316431276 |
| 결정계수 R: | 175 | 0.08905090756500366 |
| 결정계수 R: | 180 | 0.08902778226393139 |
| 결정계수 R: | 185 | 0.08897212150971358 |
| 결정계수 R: | 190 | 0.08897176165544257 |
| 결정계수 R: | 195 | 0.089001746602609541 |

Fig. 17. 결정계수

이 모델을 이용하여 시험해 본다. 시험해 볼 때는 30%의 시험 데이터를 이용한다. 프로그램은 아래와 같다.

```

y_p=model.predict(x_test)
print(y_test)
print(y_p)

kskim.scatter(x_test['평균기온'],y_test,marker='+')
kskim.scatter(x_test['평균기온'],y_p,marker='o')
kskim.show()

ax1=sns.distplot(y_test, hist=False, label="y_실제")
ax2=sns.distplot(y_p,hist=False,label="y_예측",
ax=ax1)

kskim.show()
    
```

이 모델을 따라 예측한 값과 실제 값의 비교는 아래 Fig. 18과 같다.

V. 결 론

본 연구에서는 기상 변화 데이터를 기반으로 에너지 최대 사용량을 예측하는 인공지능 학습 모델을 이용하는 방법을 제안하였다.

본 연구에서 제안한 모델은 데이터 단계와 학습 단계 그리고 적용하기 단계로 나누었다. 데이터 단계에서 기상 데이터와 에너지 최대 사용량 데이터를 수집하여 정리하였다. 학습 모델을 만드는 단계에서 인공지능에 여러 가지 학습 모델들을 선택하여 수집한 데

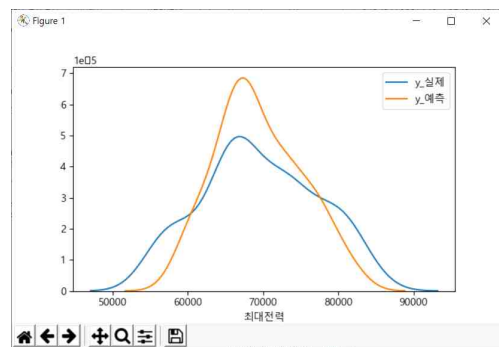


Fig. 18. 예측값 비교

이터를 적용하여 최적인 학습 모델과 인자를 결정하였다. 본 연구에서는 회귀 모델은 결정계수가 너무 적어 타당하지 않고 앙상블 모델인 랜덤포레스트가 적합하다는 것을 알았다. 랜덤포레스트에서 각 인자를 변화시켜 데이터의 특성에 따라 인자들을 조정하였는데 365일의 기상 데이터와 에너지 최대 사용량 데이터를 기반으로 결정계수가 가장 큰 것은 175개의 트리를 만드는 것을 알았다.

기상 데이터와 에너지 데이터의 표준 데이터 집합을 만들어서 인공지능 교육에 데이터를 제공함으로써 기상변화와 에너지 교육의 중요성을 인식할 수 있는 계기가 될 것이다.

본 연구에서 개발한 기상 데이터와 에너지 데이터를 이용한 인공지능 교육 프로그램을 초등학교 현장에 적용하여 학생들이 인공지능에 대한 이해 능력, 인공지능에 대한 적용 능력 등을 평가하면 인공지능 교육과 에너지와 기후 변화 교육에 많은 도움이 될 것이다.

참고문헌

- 교육부, 2015, 실과(기술가정)/정보과 교육과정, 교육부고시 제 2015-74호[별책 10].
- 김소희, 정영식, 2020, 유치원 및 초등학교 1-2학년을 위한 AI 기반 교수학습활동 탐색, 정보교육학회 논문지, 24(5), 413-421.
- 대한민국정부, 2019, AI 국가 전략.
- 이병진, 2020, SW AI 교육 활성화 정책 방안, 한국정보교육학회 학술발표논문집, 2020, 3-24.
- 이춘식, 1999, 기술과에서의 환경교육 체계화 방안, 환경교육, 12(1), 134-149.
- 인재천, 2012, STS접근법을 활용한 통합적 유아 에너지 교육이 환경 친화적 태도와 과학적 태도에 미치는 영향, 한국보육학회지, 12(3), 83-101.
- 한국 IDC, 2020, 국내 인공지능 시장전망.
- 한국과학창의재단, 2019, 차세대 소프트웨어 교육 표준 모델 개발.
- AI4K12, 2019, 5 Big Ideas of AI, Retrieved from <https://github.com/touretzkzyds/ai4k12/wiki>.
- Chris Burges, 1998, A tutorial on support vector machines for patternrecognition, Data Mining and Knowledge Discovery, 2(2), 121-167.
- Dias, R. A., Mattos, C. R., & Balestieri, J. A., 2004, Energy education: Breaking up the rational energy use barriers, Energy Policy, 32(11), 1339-1347.
- Hungerford, H. R., & Volk, T. L., 1990, Changing learner behavior through environmental education, The Journal of Environmental Education, 21(3), 8-21.
- Jennings, P., 2009, New directions in renewable energy education, Renewable Energy, 34(2), 435-439.
- Cristianini, N., & Shawe-Taylor, J. 2000, An Introduction to Support Vector Machines: And Other Kernel-Based Learning Methods. New York, NY: Cambridge University Press.
- World Bank, 1992, World Development Report 1992: Development and the Environment, Washington D. C.: World Bank.

2020년 12월 2일 접수

2020년 12월 23일 수정원고 접수

2020년 12월 24일 채택

* 김갑수, 서울교육대학교 교수(Kim, Kapsu; Professor, Seoul National University of Education).