

# 다양한 feature를 이용하여 태양광 발전량 예측

유재상<sup>o</sup> 장유진<sup>o</sup> 이영구

경희대학교 컴퓨터공학과

jsyoo1996@khu.ac.kr 2015104214@khu.ac.kr yklee@khu.ac.kr

## Photovoltaic prediction by using various features

Jaesang Yoo<sup>o</sup> Yujin Jang<sup>o</sup>, Youngkoo Lee

Department of Computer Science and Engineering, Kyung Hee University

### 요약

최근 신재생 에너지의 중요성이 커짐에 따라 정부도 이에 대한 지원이 커지고 있다. 그 중에서 태양광 발전은 다른 신재생 에너지와 다르게 설치하기 용이하여 다양한 공기업, 사기업, 공장 등의 건물 옥상에 설치되고 있다. 현재 태양광 발전을 하는 고객과 앞으로 참여할 잠재 고객의 태양광 발전량을 예측하는 딥러닝 모델을 구현한다. 예측된 태양광 발전량은 주식회사 시너지의 태양광발전 관리 시스템에 적용할 수 있고, 이는 소규모 분산 자원 거래 시장에서 입찰 정확성을 높일 수 있고 이를 이용하여 추가 인센티브를 확보할 수 있다.

### 1. 서론

화석연료의 고갈로 인한 자원확보 경쟁 및 고유가의 지속으로 인해 신재생 에너지로의 전환은 이제 필수에 준하는 상황에 왔다. 한국을 포함해 세계적인 추세는 기존의 석탄, 석유와 더불어 원자력 발전까지 신재생 에너지로의 전환이 필요한 자원으로 보고 있으며, 각종 에너지 전환 정책들과 함께 친환경 발전을 장려하고 있다. 태양광, 태양열, 풍력 등 많은 종류의 신재생 에너지가 있지만, 현재로서는 대부분이 설치 과정에 까다로운 조건이 붙거나, 설치 비용 대비 발전효율이 떨어지고 소규모의 작업장에는 적용하기 힘든 것이 대부분이다. 태양광 발전은 다른 발전 방법에 비해 설치에 제한 사항이 적으며, 기업을 포함해 규모가 작은 일반 가정에서도 쉽게 접할 수 있다.

태양광 발전이 다른 신재생에너지에 비해 설치 조건이 까다롭지 않은 것은 맞지만, 발전의 효율을 높이기 위해서는 여러가지 요소들을 만족해야 한다. 일조량이 곧 전력생산량과 비례하기 때문에 가장 우선적으로

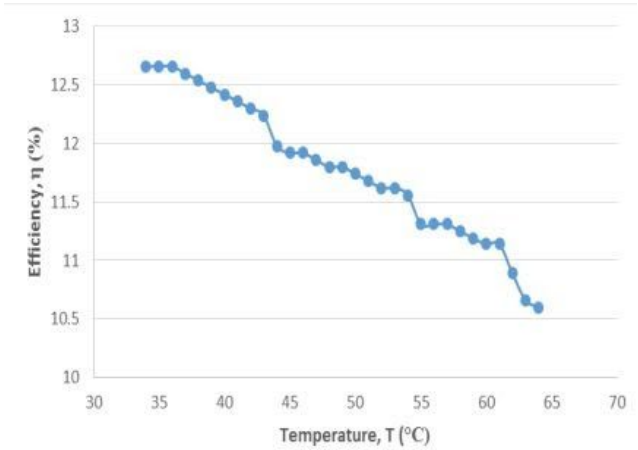
고려해야 하는 것이며, 패널의 온도가 너무 높거나 낮아도 발전의 효율이 떨어지며, 초기 투자비가 많이 소요된다. 이러한 요소들로 인해, 이미 운용 중인 고객 층과 잠재 고객 층 모두에게 특정 부지 (설치 후보지) 및 조건에서의 발전량 예측은 투자 대비 효율을 높일 수 있는 효과적인 방법이 될 수 있다.

현재까지 발전량 데이터와 해당 기간의 기상 정보를 딥러닝 모델에 학습시킨 후, 차후의 기상 정보가 있다면 발전량을 예측할 수 있다. 또한, 태양광 발전에 영향을 미치는 요소들 각각의 가중치를 책정할 수 있고, 이를 관리 시스템에 적용해 발전 효율을 높일 수 있다. 아직 설치하지 않은 잠재고객의 경우, 설치 후보지에 일정 수준 이상의 데이터가 존재한다면 유의미한 수준의 예측이 가능할 것으로 예상되며, 초기 설치 비용의 감소와 투자 대비 효율을 높이는 파생효과를 기대할 수 있을 것이다.

### 2. 관련연구

### 2.1. 태양광 발전과 온도의 관계

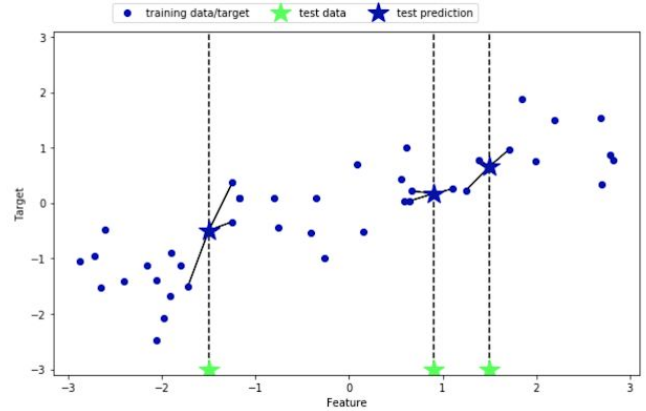
일조시간이 길어지면 일조 표면의 온도는 점진적으로 상승하는데, 태양광 패널 역시 이 영향을 받는다. 패널의 온도에 따라 발전 효율이 변화하는데, 일정 수준을 초과하게 될 경우 [그림1]과 같이 효율이 떨어지는 모습을 보인다. 1°C 상승마다 0.27-0.77% 정도 효율이 점점 떨어지게 되며, 이는 태양광 패널의 소재나 모듈의 설계에 따라 편차가 있다[1]. 때문에 여름이 일사량이 가장 많지만, 평균 발전량은 봄, 가을, 여름, 겨울 순으로 높다. 이를 위한 해결 방안으로 온도 조절을 위해 모듈의 배치를 조정하거나 냉각수를 분사하는 등의 방법을 통해 패널의 온도를 낮추는 시도를 하고 있으며, IBM 사의 sunflower 같은 다양한 방법들이 연구되고 있다.



[그림 1] 온도와 태양광 발전 효율의 관계 그래프

### 2.2 K-NN Regression

K-최근접 이웃(K-Nearest Neighbor, KNN)은 지도 학습 알고리즘 중 하나이다. 어떤 데이터가 주어지면 그 주변의 데이터를 살펴본 뒤 더 많은 데이터가 포함되어 있는 범주로 분류하는 방식이다. K-NN의 특징은 훈련이 따로 필요 없다는 것이다. 훈련 데이터를 기반으로 모델을 만들고 테스트 데이터로 테스트를 하는 방식이다. K-NN은 분류가 아니라 회귀로도 사용할 수 있다.



[그림 2] K-NN Regression을 이용한 예측

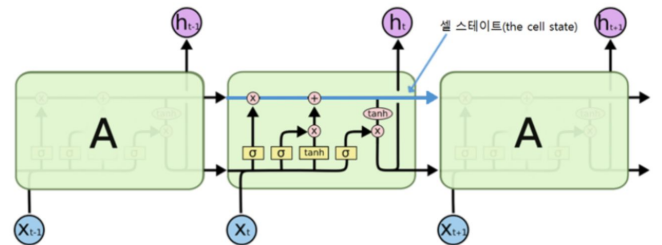
그림[2]를 보면 분류할 때와 마찬가지로 주변의 데이터를 참고하여 그 주변 데이터의 값과 거리 등을 고려하여 예측한다.

### 2.2 SVM

서포트 벡터 머신(support vector machine SVM)은 기계학습의 분야 중 하나로 패턴 인식, 자료분석을 위한 지도 학습 모델이며, 주로 분류와 회귀 분석을 위해 사용한다. 회귀를 위한 SVM은 SVR(support vector regression)이라고 불린다. SVR에 의해 생성된 모델은 손실 함수가 모델 예측에 가까운 모든 학습 데이터를 무시하기 때문에 학습 데이터의 하위 집합에만 의존한다.

### 2.2.3 LSTM

RNN의 경우, 데이터간의 거리가 멀어질수록 데이터의 상관관계가 불명확해지는데 이를 개선한 모델이 바로 LSTM(Long Short Term Memory Network)이다. 가장 기본적인 RNN 모델의 경우, 학습이 거듭될수록 초기의 weight값이 유지되지 않아 학습률이 저하되는 Long-Term Dependency 현상이 생긴다.



[그림 3] LSTM 구조와 셀 스테이트

[그림 3]과 같이 LSTM에서는 Cell State라는 layer를 추가해서 weight를 계속 기억할 것인지 결정하여 이 문제를 개선했다.[2] 이를 통해 직전의 데이터만 참고하는 Vanilla RNN과 달리 LSTM은 cell state에서 다른 weight를 참고할 수 있게 되었다.

### 3. 다양한 feature를 이용하여 태양광 발전량 예측

#### 3.1. 시계열 데이터를 통한 근미래의 태양광 발전량 예측 모델

태양광 발전량, 날씨, 온도, 일조량 등의 feature로 사용할 데이터를 수집한다. 데이터의 완성도를 높이기 위해 누락된 구간과 오류값을 확인하고 대체값으로 수정한다. 이후 해당 feature 데이터들을 같은 시간에 축정한 데이터끼리 맞춰서 전처리한다.

다양한 feature를 이용하여 하나의 feature의 이후 값을 예측하는 lstm 모델을 설계한다. 그리고 이 모델에 전처리한 데이터들을 학습하여 근미래의 태양광 발전량을 예측하는 모델을 만든다.

#### 3.2. 실제 발전소의 데이터를 바탕으로 가상 발전소의 발전량 예측 모델

실제 발전소의 발전량 데이터와 발전소 주변의 기상 데이터 등을 기반으로 그 features에 기반한 초평면(hyperplane)을 생성한다.

SVM 기반의 모델의 경우 결정 경계(decision boundary)를 생성하게 되는데, 최적의 결정 경계를 위해서 마진(margin)의 조정이 필요하다. 마진이 최대일 경우가 최적이라고 해서 너무 큰 마진을 주는 soft margin의 경우 학습의 정확도가 떨어져 underfitting이 발생하고, 너무 적게 주는 hard margin의 경우 모든 데이터들을 다 붙잡고 가는 overfitting의 문제가 발생한다.

KNN 기반의 모델의 경우 실제 발전소의 데이터 집단에 가상 발전소의 데이터를 삽입하여 발전량을 예측하게 된다. 이때, 몇 개의 훈련 표본을 확인할 것인지 정하는 상수 K값을 지정하고, 가상 발전소의

데이터로부터 가까운 K개의 데이터에서 발전량을 도출해낸다. K의 값이 너무 크거나 작을 경우, 적절한 예측이 불가능하기 때문에, K값을 잘 지정해 주는 것이 관건이다. 일반적으로 총데이터셋의 제곱근값을 사용한다.

### 4. 결론 및 향후연구

본 논문에서는 태양광 발전량 예측을 통해 시너지의 태양광 관리 시스템에 접목시켜 태양광 발전 효율의 증대를 도모한다. 이를 통해 전력거래 시장에서의 입찰 정확성을 올릴 수 있으며, 추가적인 이익으로 귀결된다. 태양광 설치시에 정부가 많은 지원을 해주고 있지만, 그럼에도 여전히 부담되는 가격인 것은 사실이며, 적합한 지역에 설치하지 않으면 기대만큼의 효과를 볼 수 없다. 그렇기에 부지 선정에 만전을 기해야 하는데 잠재고객에게 이 정보를 토대로 구체적이고 가시적인 발전량 예측을 제공하면 부지 선정의 정확성을 높일 수 있으며 신규 고객 유치가 용이하다.

다만 기상청 같이 기상 데이터를 전문적으로 다루어야 근미래에 비 또는 눈이 온다는 것이나, 태풍이 온다는 것은 알 수 있는 것이기 때문에 이와 같은 상황은 이 모델이 예측할 수 없다. 따라서 이 모델은 현재까지의 날씨 데이터로 근미래의 태양광 발전량을 예측하는 것이기 때문에 비나 눈 같이 구름이 많이 끼는 날씨 상황에 대응하는 발전량 예측은 할 수 없다.

#### 참고문헌

- [1] Li Wah Thong, Sharweeni Murugan, Poh Kiat Ng and Cha Chee Sun, Analysis of Photovoltaic Panel Temperature Effects on its Efficiency, 2016.
- [2] Christopher Olah, Understanding LSTM Networks, 2015.  
<https://colah.github.io/posts/2015-08-Understanding-LSTM/>
- [3] Andreas C.Muller and Sarah Guido, Introduction to machine learning with python, 2017.