

빅데이터를 활용한 머신러닝 기반 태양에너지 발전량 예측 모델*

김 종 민*, 이 준 형**

요 약

본 연구는 기후변화에 따른 전남 영암지역의 기상변화와 태양광 에너지 생산량의 빅데이터 분석을 통해 상관관계를 도출하여 태양광 에너지 생산 예측 모델을 제시하였다. 사용된 데이터는 공공데이터에서 제공하는 2016년 1월부터 2019년 12월까지의 영암지역의 날씨와 태양에너지 생산량 데이터를 사용하였다. 머신러닝 기법을 활용하여 기상변화와 태양광 에너지 생산량의 회귀분석을 통하여 지역의 날씨와 태양광 에너지 생산량과의 상관 관계식을 도출 하였다. 도출된 예측식을 적용하여 지역의 태양에너지 생산을 계산하였으며, 이를 생산지수로 표현하여 3단계로 구분하였다. 이 같이 구분된 3단계의 생산지수를 통해 향후 기후변화에 따른 태양에너지 생산량을 예측하고, 농업 활동에 있어 중요한 바로미터로 활용될 것이라 판단된다.

Implementation of machine learning-based prediction model for solar power generation

Jong-Min Kim*, Joon-hyung Lee**

ABSTRACT

This study provided a prediction model for solar energy production in Yeongam province, Jeollanam-do. The model was derived from the correlation between climate changes and solar power production in Yeongam province, Jeollanam-do, and presented a prediction of solar power generation through the regression analysis of 6 parameters related to weather and solar power generation. The data used in this study were the weather and photovoltaic production data from January in 2016 to December in 2019 provided by public data. Based on the data, the machine learning technique was used to analyzed the correlation between weather change and solar energy production and derived to the prediction model. The model showed that the photovoltaic production can be categorized by the three-stage production index and will be used as an important barometer in the agriculture activity and the use of photovoltaic electricity.

Key words : Machine learning, Climate change, Weather, photovoltaic system, solar energy generation, Prediction, correlation analysis

접수일(2022년 05월 01일), 수정일(2022년 06월 10일),
게재확정일(2022년 06월 30일)

* 동신대학교 정보보안학과

** 동신대학교 신재생에너지학과(교신저자)

★이 논문은 2021년도 동신대학교 학술연구비에 의하여 연구
되었음

1. 서론

최근 세계 각국은 지구온난화가 심각한 문제로 떠오름에 따라 이를 대응하기 위해 2016년 파리기후변화협정을 통해 단계별 탄소 중립을 선언하고 이를 이행하기로 약속하였다. 여기에 맞추어 우리나라는 2050년 탄소 중립을 목표로 기존 화석연료 발전 중심의 전력공급 체계를 재생에너지 중심으로 전환하고 있다. 전력거대소에서 낸 '2020 발전설비 현황에 의하면 전체발전설비용량 중 신재생에너지의 비중은 9.3%(2017년), 11.2%(2018년), 12.6%(2019년), 16%(2020년)으로 점진적으로 증가하고 있다. 2020년 기준 신재생에너지 발전설비 비중을 보면 16%이고 이중 태양광 발전량은 71%의 비중을 차지하고 있어 신재생에너지 중 주요한 역할을 하고 있다[1]. 전력공급에 있어 신재생에너지 발전량의 날씨에 따른 불확실성은 기존 전력계통에 위험요소로 작용할 수 있다. 이 문제를 해결하기 위해 신재생에너지 중 태양광 발전은 다음날 발전량을 예측하여 발전량을 미리 파악하여 기존의 전력계통에 위험요소를 낮추려 많은 노력과 연구가 진행되고 있다. 태양광 발전에는 다음날 발전량을 예측하기 위해서 태양광 발전량에 주요한 영향을 미치는 기상 변수를 파악하고 선정하는 것이 필요하다. 따라서 본 논문에서는 기상요소들의 변화에 따라 태양열 생산량에 대해 상관관계가 분석 및 회귀분석을 통해 태양열 생산 예측식을 도출하고자 한다.

2. 관련연구

2.1 상관관계분석

2개의 변수 x 와 y 가 있을 때에, x 의 변화에 따라서 y 도 변화하는 관계를 상관관계라고 한다.

변수 간 상관관계가 있는지를 수치적으로 판단하는 데는 상관계수라고 불리는 지표를 이용한다.

상관계수는 통상 r 이라는 기호로 표기되고, -1에서 1까지의 값을 취한다.

$$-1 \leq r \leq 1$$

상관계수의 부호가 양(+)일 때에는 양의 상관관계가 있다는 것을, 음(-)일 때에는 음의 상관관계가 있다는 것을 나타내고 있다. 상관관계의 강도는 상관계수의 절대 값 $|r|$ 또는 자승 값 r^2 으로 평가한다. 어느 쪽도 1에 가까울수록 상관이 강하다는 것을 의미한다. 상관관계가 존재하지 않을 때에는 상관계수의 값은 0에 가까운 값이 된다.

변수 x 와 y 의 상관계수는 다음과 같은 순서로 산출할 수 있다[2][3].

① x 의 편차제곱의 합 $S(xx)$ 를 계산한다.

$$S(xx) = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 / n$$

② y 의 편차제곱의 합 $S(yy)$ 를 계산한다.

$$S(yy) = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2 / n$$

③ x 와 y 의 편차곱의 합 $S(xy)$ 를 계산한다.

$$S(xy) = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right) / n$$

④ 상관계수 r 을 계산한다.

$$r = \frac{S(xy)}{\sqrt{S(xx)S(yy)}}$$

2.3 회귀분석

회귀분석은 데이터 분석에서 가장 널리 쓰이는 통계적 분석방법의 하나로서, 종속변수와 독립변수들간의 관계를 밝히는 분석방법으로 선형모형을 통해서 모형을 적합하고, 적합된 모형을 통해서 관련된 현상을 연구할 수 있다. 회귀분석에서 독립변수가 여러개인 경우를 다중회귀분석이라고 한다. 이런 다중회귀분석은 시간의 흐름에 따른 변화를 적용하여 예측기법

에 많이 사용되고 있다.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

여기서 β_0 는 절편을 나타내는 회귀계수이고, β_0, \dots, β_k 는 종속변수와 독립변수간의 기울기를 나타내는 회귀계수로 다른 독립변수들의 고정되었을 경우 X_1 이 Y 에 미치는 영향을 나타낸다고 할 수 있다[4][5][6].

3. 제안하는 방법

3.1 구현 환경

<표 1>은 예측 알고리즘 구현을 위한 분석 환경이며, .csv 데이터를 핸들링 하기 위해 pandas를 이용하고 데이터 탐색과 시각화를 위해 seaborn, matplotlib를 이용하여 데이터분석을 진행한다. 또한 데이터 분석 전에 정교한 분석을 위해 정규화를 진행한다.

<표 1> 구현 환경

| | |
|----------|---|
| 운영체제 | macOS Monterey |
| 프로그래밍 언어 | Python3.8 |
| 에디터 | jupyter notebook |
| 라이브러리 | - tensorflow - pandas - keras - matplotlib - seaborn - numpy 등 |

<표 2>는 분석에 사용된 데이터의 변수이다.

<표 2> 데이터 구성

| | |
|-------|---------------------|
| 기상 변수 | DHI, DNI, WS, RH, T |
| 발전량 | TARGET |

<표 3> 데이터의 변수 소개

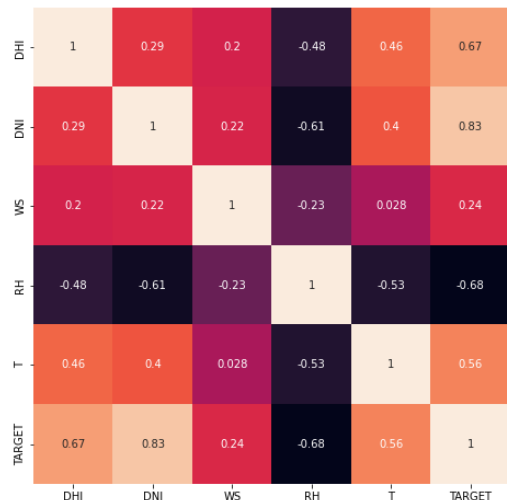
| 변수 | 변수 소개 |
|--------|------------|
| Hour | 시간 |
| Minute | 분 |
| DHI | 수평면 산란 일사량 |
| DNI | 직달 일사량 |
| WS | 풍속 |
| RH | 상대습도 |
| T | 기온 |
| TARGET | 태양광에너지 발전량 |

4. 연구결과

본 연구에서는 머신러닝을 활용하여 기상요소와 태양광에너지 발전량과의 상관분석과 회귀분석을 실시하였다.

4.1 상관관계 분석

태양광 발전량과 기상요소의 변수 간 상관계수를 분석하였다. 데이터의 구성은 DHI, DNI, WS, RH, T, TARGET(태양광에너지 생산량) 순서이며, 상관계수 최대값은 1 최소값을 -1로 지정하였고 시각화 된 발전량과 기상요소 상관도는 (그림 1)과 같다.



(그림 1) 기상요소와 태양광 에너지 발전량의 상관관계

<표 4> 기상요소와 태양광 에너지 발전량의 상관관계

| 변인 | 1 | 2 | 3 | 4 | 5 | 6 |
|-------|-------|-------|-------|-------|-------|-------|
| 1.DHI | 1.000 | | | | | |
| 2.DNI | 0.288 | 1.000 | | | | |
| 3.WS | 0.203 | 0.220 | 1.000 | | | |
| 4.RH | 0.479 | 0.611 | 0.230 | 1.000 | | |
| 5.T | 0.458 | 0.402 | 0.028 | 0.533 | 1.000 | |
| 6.생산량 | 0.667 | 0.834 | 0.239 | 0.677 | 0.562 | 1.000 |

4.2 예측모형

기상요소가 태양광 에너지 생산량에 미치는 영향에 대해 회귀분석 기법을 적용시켜 예측식을 제시하였다.

4.2.1 태양광 에너지 생산량 예측

태양광 발전량 예측 연구 모형을 구성하기 위해 회귀분석을 통한 요소별 생산량에 미치는 영향정도(R Square)를 산출하였고, 영향력의 크기를 표준화하여 표준화된 결정계수(Adjusted R Square)를 산출하였다. 또한 표준화된 결정계수를 상대적 영향력을 백분율로 환산하였다. 그 결과 직달일사량(DNI)이 가장 높은 영향력을 나타냈다.

<표 5> 태양광에너지 생산량에 대한 요소별 영향력

| 구분 | R ² | Adjusted R ² | 상대적 영향력(%) | 영향 순위 |
|-----|----------------|-------------------------|------------|-------|
| DHI | .445 | .445 | 22.566 | 3 |
| DNI | .695 | .695 | 35.243 | 1 |
| WS | .057 | .057 | 2.890 | 5 |
| RH | .459 | .459 | 23.276 | 2 |
| T | .316 | .316 | 16.024 | 4 |

4.2.2 추정식

요소들에 의한 생산량 발생을 예측하기 위해 요소에 따른 회귀식에 영향도를 곱하여 합한 값을 태양열 예측식으로 하고 이를 식으로 나타내면 (식 1)과 같다.

$$\begin{aligned}
 \text{태양광에너지 생산량}(Q_1) = & \\
 & -1.0560 + [(DHI \times 0.1053 \times 0.22566) \\
 & + (DNI \times 0.0488 \times 0.35243) \quad (\text{식 1}) \\
 & + (WS \times -0.0117 \times 0.0289) \\
 & + (RH \times -0.0255 \times 0.23276) \\
 & + (T \times 0.2265 \times 0.16024)]
 \end{aligned}$$

태양광 생산량 예측의 빈도수가 (-)의 값을 가질 경우는 0으로 산정한다.

본 연구에서 (식 1)에서 계산된 예측 값이 22이상 일 때는 일 생산량이 매우 높은 것으로 하고, 14미만 일 때는 아주 낮은 것으로 하여 이를 3가지의 범위로 나누어 지수화 하여 <표 6>로 나타내었다.

<표 6> 태양광에너지 생산량에 대한 생산지수

| 예측 값(Q _i) | 생산지수 |
|-----------------------|------|
| 22이상 | 많음 |
| 14이상-22미만 | 보통 |
| 1이상-14미만 | 낮음 |

(그림 2)는 생산량 지수에 대해 예측 할 수 있는 계산기의 코드이며, 코드를 실행하면 (그림 3)과 같이 나타난다.

```

from tkinter import *

def calculate(): # pre를 계산하는 함수
    DHI = float(e1.get())
    DNI = float(e2.get())
    WS = float(e3.get())
    RH = float(e4.get())
    T = float(e5.get())
    pre = (DHI*0.1053)*0.22566+(DNI*0.0488)*0.35243+(WS*-0.0117)*0.0289+(RH*-0.0255)
    pre = round(pre, 1) # pre를 소수점 둘째자리에서 반올림

중략

b1 = Button(window, text = "계산하기", command=calculate)
# 클릭 시 함수 retry를 실행한다.
b2 = Button(window, text = "다시하기", command=retry)
b1.place(x=140, y=250)
b2.place(x=225, y=250)

window.mainloop()
    
```

(그림 2) 태양광에너지 생산량 예측 코드

발전량 예측 계산기

| | | |
|------|-------|----------|
| DHI: | 800 | W/m2 |
| DNI: | 934.2 | W/m2 |
| WS: | 4.4 | m/s |
| RH: | 56.8 | % |
| T: | 16.3 | Degree C |

계산하기 다시하기

금일 발전량 예측: kW

(그림 3) 태양광에너지 생산량 예측 계산기

5. 결론

본 연구에서 태양광에너지 생산량과 기상요소(DHI, DNI, WS, RH, T)에 대해 상관관계가 있는지 알 수 있었으며, 회귀분석을 통해 태양광에너지의 예측식을 도출하여 생산량을 예측할 수 있는 예측식을 도출하였다. 도출된 예측식을 통해 구하여진 예측값에 따라 22 이상이면 많음, 14이상 22미만이면 보통, 14미만이면 낮음으로 생산지수를 3단계로 구분하여 생산량에 대한 예방활동에 활용할 수 있도록 하였다.

본 연구에 사용되었던 모델을 지속적인 연구가 수행된다면 더욱더 신뢰도 있는 예측값을 바탕으로 생산량에 대해 적절한 정책을 수립함으로써 태양열 발전을 적절하게 사용함으로써, 기여할 수 있을 것이라 기대된다. 또한, 도출한 예측값들을 토대로 기상청 및 발전소 등과 연계하여 실시간으로 확인 가능한 시스템을 설계하게 되면 전력소비활동의 정책을 도입할 때 유용하게 사용할 수 있을 것으로 본다.

참고문헌

- [1] 이석호, 김지홍 .태양광 발전량에 주요한 영향을 미치는 기상변수 선정을 위한 탐색 및 시각화. 대한전기학회학술대회 논문집, pp 464-465. 2021.
- [2] Seong S. Chae, Chansoo Kim, Jong-Min Kim, William D. Warde, "Cluster analysis using different correlation coefficients", Statistical Papers Vol. 49, No.4, 2006, pp.715-727
- [3] Ali Abbas, "Statistical Signal Processing Technique for Identification of Different Infected Sites of the Diseased Lungs", Journal of medical systems Vol. 36, No.3, 2010, p p.1537-1541.
- [4] Özbayoğlu G, Özbayoğlu ME (2006) A new approach for the prediction of ash fusion temperatures: a case study using Turkish lignites. Fuel 85:545-552.
- [5] Mehmet Bilgili, "Prediction of soil temperature using regression and artificial neural network models", Meteorology and Atmospheric Physics Vol. 110, Numbers.1-2, 2010, pp.59-70
- [6] K. T. Chau, J. E. Chan, "Regional bias of landslide data in generating susceptibility maps using logistic regression: Case of Hong Kong Island", Landslides Vol. 2, No. 4, 2005, pp.280-290.
- [7] <http://www.kma.go.kr/>

————— [著 者 紹 介] —————



김 종 민 (Jong-Min Kim)
2015년 산업보안학박사
현 재 동신대학교 정보보안학과
교수

email : dyuo1004@dsu.ac.kr



이 준 형 (Joon-Hyung Lee)
2002년 UC Berkely, 화학공학박사
현 재 동신대학교 신재생에너지학과
교수

email : jhlee@dsu.ac.kr