
데이터 시각화 및 탐색적 데이터 분석을 통한 태양광 에너지 예측용 특징벡터 추출

정원석* · 함경선** · 박문규*** · 정영화* · 서정욱*

*남서울대학교 정보통신공학과

**전자부품연구원 에너지IT융합연구센터

***세종대학교 원자력공학과

Feature Vector Extraction for Solar Energy Prediction through Data Visualization and Exploratory Data Analysis

Wonseok Jung*, Kyung-Sun Ham**, Moon-Ghu Park***, Young-Hwa Jeong*, Jeongwook Seo*

*Department of Information and Communication Engineering, Namseoul University

**Energy IT Convergence Research Center, Korea Electronics Technology Institute

***Department of Nuclear Engineering, Sejong University

E-mail : jwseo@nsu.ac.kr

요 약

태양광 발전 시스템에서 전력 생산은 기상 상태에 따라 크게 영향을 받으므로 안정적인 부하 운용을 위해 태양광 에너지에 대한 예측이 필수적이다. 따라서 태양광 에너지 예측을 위한 기계학습 알고리즘의 입력으로 기상 상태에 대한 데이터가 필요하다. 본 논문에서는 알고리즘에 대한 입력 데이터로 표면의 3시간 동안 누적된 강수량, 상·하향 장파 복사선 평균, 상·하향 단파 복사선 평균, 지상 2m에서의 3시간 동안 온도, 표면에서의 온도 등 15가지 종류의 기상 데이터를 사용한다. 기상 데이터의 통계적 특성을 파악하고 상관관계를 분석하여 태양광 에너지와 70% 이상의 높은 상관성을 갖는 하향 단파 복사선 평균과 상향 단파 복사선 평균을 특징벡터의 주요 원소로 추출하였다.

ABSTRACT

In solar photovoltaic systems, power generation is greatly affected by the weather conditions, so it is essential to predict solar energy for stable load operation. Therefore, data on weather conditions are needed as inputs to machine learning algorithms for solar energy prediction. In this paper, we use 15 kinds of weather data such as the precipitation accumulated during the 3 hours of the surface, upward and downward longwave radiation average, upward and downward shortwave radiation average, the temperature during the past 3 hours at 2 m above from the ground and temperature from the ground surface as input data to the algorithm. We analyzed the statistical characteristics and correlations of weather data and extracted the downward and upward shortwave radiation averages as a major elements of a feature vector with high correlation of 70% or more with solar energy.

키워드

Exploratory Data Analysis, Data Visualization, Solar Energy, Feature Vector, Prediction

I. 서론

최근 화석연료, 원자력 등과 같은 기존 에너지원을 사용함으로써 발생하는 환경오염으로 인해 기후 변화, 자원 고갈에 따른 국제 유가 상승 등 날씨 및 에너지 문제가 기업이나 국가 심지어 개인의 일상생활, 경제활동에 크나큰 영향을 미치고 있다 [1-2]. 태양광발전시스템은 태양광으로부터 전력을 생산하는 발전기술로서 환경오염으로 발생하는 문제를 해결하기 위한 신재생 에너지 기술 중 가장 빠르게 성장하고 있다[3].

그러나 태양 에너지는 날씨 및 기상 상태에 따라 전력 생산이 불안정하기 때문에 안정적인 부하운용을 위해 전력 생산량에 대한 정확한 예측이 필요하다[1].

본 논문에서는 정확한 태양광 에너지 예측을 위해 사용하는 기계학습 알고리즘의 입력으로 표면의 3시간 동안 누적된 강수량, 상·하향 장파 복사선 평균, 상·하향 단파 복사선 평균, 지상 2m에서의 3시간 동안 온도, 표면에서의 온도 등 15가지 종류의 기상 데이터를 사용하여 특징벡터의 주요 원소를 추출하기 위해 기상 데이터의 통계적 특성을 파악하고 상관관계를 분석한다.

본 논문의 구성은 다음과 같다. 2장에서 기상 데이터에 대해서 살펴본다. 3장에서는 데이터의 통계적 특성을 파악하고 상관관계를 분석하며, 4장에서 분석한 상관관계를 바탕으로 특징벡터의 주요 원소를 추출하며, 5장에서 본 논문의 결론을 맺는다.

II. 기상 데이터

Kaggle은 예측모델 및 분석 대회 플랫폼으로, 기업 및 단체에서 데이터와 해결과제를 등록하면, 데이터 과학자들이 이를 해결하는 모델을 개발하고 경쟁한다. 본 논문에서는 미국 기상 학회 주최로 2013년부터 2014년까지 개최된 태양 에너지 예측 경연 대회의 데이터를 사용하였다[4]. 사용된 데이터는 1994년부터 2007년 총 14년 치의 미국 오클라호마 지역 중 Chickasha 지역의 기상정보 데이터로 그림 1과 같이 표현된다. 하루 단위의 15개 기상 요소로 이루어져 있으며, 데이터 수는 약 75,000개이다.

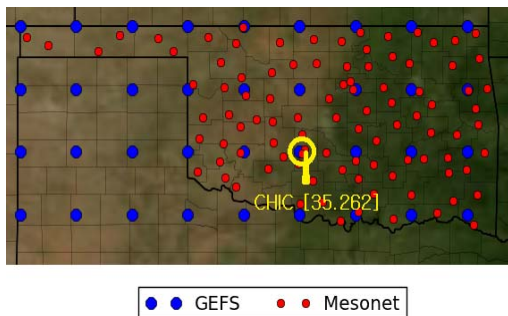


그림 1. 오클라호마 Chickasha 지역

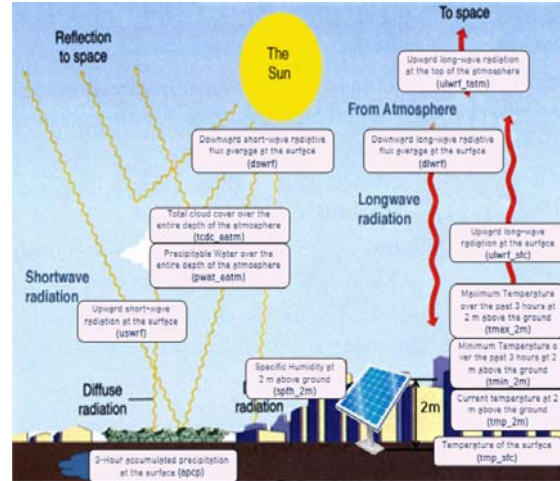


그림 2. 기상 요소

15가지 기상 요소는 그림 2에서와 같이 표면에서의 3시간 동안 누적된 강수량, 상·하향 장파 복사선 평균, 상·하향 단파 복사선 평균, 습도, 기압, 구름량, 지상 2m에서의 3시간 동안 최대·최소·현재 온도, 표면에서의 온도 등이 있으며 표 1에서 자세히 설명하였다.

표 1. 기상 요소 설명

Variable	Description
apcp_sfc	3-Hour accumulated precipitation at the surface
dlwrf_sfc	Downward long-wave radiative flux average at the surface
dswrf_sfc	Downward short-wave radiative flux average at the surface
pres_msl	Air pressure at mean sea level
pwat_eatm	Precipitable Water over the entire depth of the atmosphere
spfh_2m	Specific Humidity at 2 m above ground
tcdc_eatm	Total cloud cover over the entire depth of the atmosphere
tcopc_eatm	Total column-integrated condensate over the entire atmosphere
tmax_2m	Maximum Temperature over the past 3 hours at 2 m above the ground
tmin_2m	Minimum Temperature over the past 3 hours at 2 m above the ground
tmp_2m	Current temperature at 2 m above the ground
tmp_sfc	Temperature of the surface
ulwrf_sfc	Upward long-wave radiation at the surface
ulwrf_tatm	Upward long-wave radiation at the top of the atmosphere
uswrf_sfc	Upward short-wave radiation at the surface
total	the total daily incoming solar energy

III. 기상 데이터 분석

태양 에너지 예측을 위해 기계학습 알고리즘을 사용하기 전 데이터의 분석이 필요하다. 최소한의 기상 요소, 즉, 최소한의 특징벡터만을 사용하여 태양 에너지를 예측하기 위해서 특징벡터 추출이 필요하다. 최소한의 특징벡터만으로 예측한 값이 모든 특징벡터를 사용하여 예측한 값과 비교하였을 때 성능이 비슷하다면 추출된 특징벡터에 대한 최소한의 기상 관측 장비를 사용하여 예측하는 것이 효율적이기 때문이다.

본 논문에서는 위와 같은 이유로 탐색적 자료 분석(Exploratory Data Analysis; EDA)의 시각화를 통해 특징벡터 추출을 목적으로 데이터의 통계적 특성을 파악하고 특징벡터 간 상관관계를 분석하여 연관성을 파악한다. EDA의 시각화는 분석가가 고급의 데이터 분석 지식 없이도 손쉽게 데이터의 특성을 파악할 수 있는 장점이 있다. 데이터의 분포를 확인하기 위해 히스토그램을 그리는 것도 시각화 기법의 한가지이다[5].

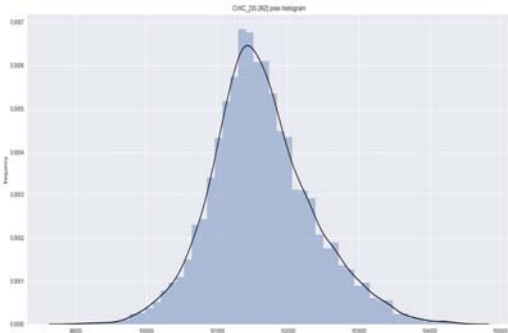


그림 3. 특징벡터 pres_msl에 대한 Histogram

그림 3은 15가지 특징벡터 중 하나인 pres_msl에 대한 Histogram으로 특징벡터 pres_msl의 데이터 분포도를 확인하기 위해 데이터를 시각화한 그림이며, 평균(mean), 중앙값(median), 최소값(minimum), 최대값(maximum), 표준편차(standard deviation)와 같은 산포의 측도 분석은 표 2와 같이 분석하였다.

표 2. 특징벡터 pres_msl 산포 측도 분석

	pres
count	5113.00
mean	101629.7489
std	720.1178
min	98999.0875
25%	101167.5563
50%	101561.1578
75%	102037.8219
max	104464.7797

특징벡터의 상관관계 그래프는 그림 4와 같이 표현되며 상관계수의 절대값이 클수록 좁은 타원형이 그려지며 작을수록 넓은 타원형이 그려진다. 또한, 빨간색이 진할수록 양의 상관관계를 가지며 진한 파란색일수록 음의 상관관계를 가진다.

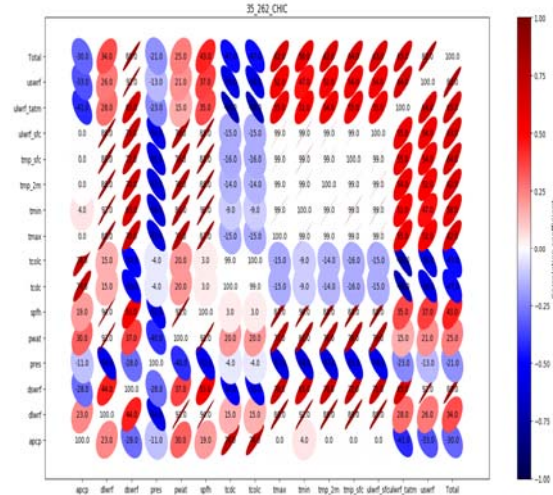


그림 4. 특징벡터 간 상관관계 그래프

IV. 특징벡터 주요 원소 추출

본 논문에서는 태양광 에너지 예측을 위해 특징벡터를 추출하는 것이므로 특징벡터와 일일 태양 에너지 간의 상관관계를 분석한다. 태양 에너지와 특징벡터 간의 상관관계 그래프를 분석한 결과 특징벡터 dswrf_sfc와 uswrf_sfc가 88%와 82%로 total과 가장 높은 상관관계를 가지고 있어 태양 에너지 예측을 위한 주요 원소로 추출하였다.

V. 결 론

태양 에너지 예측을 위한 특징벡터 주요 원소 추출 작업은 태양 에너지와 특징벡터 간의 상관관계를 바탕으로 태양 에너지 예측에 가장 영향을 끼치는 특징벡터들을 선별하는 작업으로 예측을 위한 기계학습 알고리즘의 성능 향상과 최소한의 기상 관측 장비를 사용하여 비용적 측면에서 효율성 향상을 기대할 수 있게 한다.

ACKNOWLEDGMENTS

This work was supported by the Korea Institute of Energy Technology Evaluation and Planning(KETEP) and the Ministry of Trade, Industry & Energy(MOTIE) of the Republic of Korea (No. 20171210201140).

참고문헌

- [1] 이기열, 문정태 and 정무영, "Economic analysis for the photovoltaic system," Korean Institute Of Industrial Engineers, pp.742~746, Nov. 2008.
- [2] Yangkoo Lee, Wontae Kim, Youngjin Jung, Kwangdeuk Kim and Keunho Ryu, "Cluster Analysis of Climate Data for Applying Weather Marketing," The Koean Association of Geographic Information Studies, Vol. 7, No. 3, pp.33-44, 2005
- [3] Donghun Yang, Nayoung Yeo, Sangcheol Kim, Jeongwoo Lee, Seontae Kim and Pyeongsoo Mah, "A Piecewise Prediction Model of Solar Radiation using eather Sensing Information for stand-alone Photovoltaic Systems," Korea Inormation Science Societ, pp.15-17, June. 2017
- [4] kaggle, "AMS 2013-2014 Solar Energy Prediction Contest", [Internet]. Available: <https://www.kaggle.com/c/ams-2014-solar-energy-prediction-contest>.
- [5] Sunghae Jun, "An Effective Usage of EDA for Data Mining," Korean Institute of Intelligent Systems, Vol. 19, No. 1, pp. 252-254, Apr. 2009.