

# 2017 (사)ICT플랫폼학회 하계학술대회 및 (사)ICT대연합 지능정보행정 토론회 논문집



ICTPS



**일시** 2017년 6월 16일(금) 오후 1시~17일(토) 11시

**장소** 백석대학교



- 주 최 : (사)ICT플랫폼학회, (사)ICT대연합, 동국대 링크사업단
- 주 관 : (사)ICT플랫폼학회
- 후 원 : 백석대학교, (주)세림티에스지, (주)시스원, (주)콤텍정보통신, (주)세림소프트, (주)굿모닝아이텍, (주)대유플러스, (사)아시아행정학회, (주)티맥스소프트

목차 | CONTENTS

기업맞춤형 교육 및 연구윤리

사회맞춤형 산학협력 선도대학 육성사업(산학협력 고도화형) ..... 1  
 주해중(동국대학교)

연구윤리규정 개정을 위한 연구윤리지침 해설 내용정리 ..... 11  
 박동희(한국교통대학교)

연구출판윤리규정 ..... 17  
 (사)ICT플랫폼학회

초청 논문

한국 공공부문의 제4차 산업혁명 대응현황과 과제 ..... 23  
 홍길표(백석대학교)

지능정보기술을 활용한 5대 스마트 행정과제 ..... 43  
 고대식(목원대학교)

SESSION 1

E134 기반 EDA 시스템을 사용한 스마트팩토리 구축 ..... 49  
 감윤희(백석대학교), 남성현(두원공과대학교), 강경우(백석대학교)

4차 산업혁명을 위한 유망기술 도출 연구 ..... 52  
 나영식, 정효남, 조재혁(한국과학기술기획평가원)

이미지 처리를 이용한 차량의 사각 지대 내 보행자 인식 ..... 59  
 문지혜, 위지원, 이현병, 구본근(한국교통대학교)

효율적인 신재생 에너지 시스템의 오류 모니터링을 위한 기존 데이터 분석 방법과 K-means  
 알고리즘 비교를 통한 데이터 분석 ..... 62  
 이진협, 장우성, 박보경, 김영철(홍익대학교)

4차 산업혁명 선도를 위한 센서 측정 기술개발의 국제공동연구 전략 ..... 67  
 조재혁, 나영식(한국과학기술기획평가원)

# 효율적인 신재생 에너지 시스템의 오류 모니터링을 위한 기존 데이터 분석 방법과 K-means 알고리즘 비교를 통한 데이터 분석

## Analyzing Data through Comparing The Previous Data Analysis Technique and K-means Algorithm for Error Monitoring of The Effective New & Renewable Energy System

이진협\*, 장우성\*, 박보경\*, 김영철\*\*

Jin-hyub Lee\*, Woo-sung Jang\*, Bo-kyung Park\* and R. Young Chul Kim\*\*

### 요 약

최근 전 세계적으로 대기오염 및 화석연료 고갈 문제로 신재생에너지에 대한 관심이 높아졌다. 신재생에너지 산업 중 태양광에너지 발전 산업은 상대적으로 설치비용이 낮고, 설치장소, 유지보수면에서 용이하여 많은 비중을 차지하고 있다. 관련 연구에서 태양광 발전량 현황을 쉽게, 수시로 파악할 수 있도록 신재생 에너지 통합 관리 모니터링 시스템을 구축하였다. 이 논문에서는 이 시스템 데이터베이스에 축적된 기존 데이터를 활용한 시계열 그래프와 상관관계 그래프를 생성한다. 이를 통해 데이터간 특징들을 찾아내고, 오류 데이터를 분석한다. 또한 텐서플로우를 이용해 데이터 전처리 과정을 거친 후 K-means 알고리즘을 적용하여 오류 데이터 분석을 시도한다.

Key words : Photovoltaic Energy Monitoring System, Power Generation, Solar Insolation, Error Data, K-means Algorithm

### 1. 서론

최근 전 세계적으로 화석연료 고갈, 지구 온난화 문제 등으로 대체할 수 있는 신재생 에너지에 대한 관심이 높아지고 있다. 설치비용, 장소, 유지보수 등에 용이하여 태양광 에너지 산업의 비중이 높다. 2014년 기준, 국내 신재생에너지 전체 대비 태양광 산업의 비율을 보면 기업체 수가 28%, 고용인원 52%, 매출액 63%, 수출액 79%에 달한

다[1].

그러나 기존 태양광 에너지 산업에서 전기 생산은 매달 지료를 통해서만 확인했다. 만약 이달 초에 태양광 패널에 문제가 생겨 발전 불가 시 손해를 본다. 이런 문제를 해결하고자, 수시로 확인가능한 신재생 에너지 통합 관리 모니터링 시스템을 개발하였다[2].

이 논문에서는 신재생 에너지 통합 관리 모니터링 시스템 서버에 1년간 축적된 데이터를 이용

\* 홍익대학교 소프트웨어공학 연구실(ljh@selab.hongik.ac.kr)

\*\* 교신저자: 홍익대학교 소프트웨어공학 연구실(bob@hongik.ac.kr)

· 제1저자 (First Author) : 이진협

하여 발전소 별 특징, 오류 데이터 분석 및 머신러닝의 K-means 알고리즘 적용으로 분석하였다.

이 논문의 구성은 다음과 같다. 2장에서는 관련 연구로 신재생 에너지 통합 관리 모니터링 시스템과 K-means 알고리즘에 대해 언급한다. 3장에서는 태양광 발전 데이터를 분석하고, 4장에서는 K-means 알고리즘을 이용한 데이터 분석을 수행한다. 마지막으로 5장에서 결론에 대해 언급한다.

## II. 관련 연구

### 1. 신재생(태양광) 에너지 모니터링 시스템

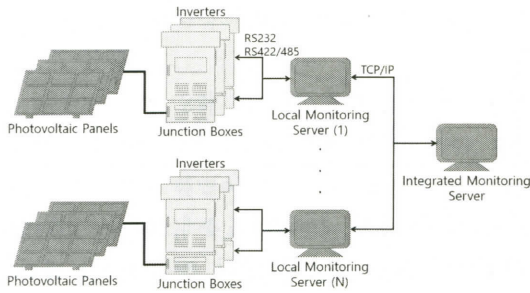


그림 1. 신재생 에너지 통합 관리 모니터링 시스템 구성도

그림 1은 신재생 에너지 통합 관리 모니터링 시스템 구성도이다. 여러 태양광 패널에 모인 빛 에너지를 전기 에너지로 변환하여 Junction Box에 연결한다. Junction Box와 인버터가 연결되고, 인버터에서는 발전된 에너지, 수평 일사량, 경사 일사량, 모듈 온도 등이 수치화된다. 수치화된 데이터를 로컬 모니터링 서버에 저장하고, 여러 로컬 서버들이 통합 모니터링 서버에 데이터를 전송한다. 통합 모니터링 서버에서는 모든 데이터를 저장하고, 이를 기반으로 웹 페이지와 어플리케이션에 현재 발전량, 누적 발전량, 일사량 등을 모니터링 한다[2]. 이 논문에서는 통합 모니터링 서버에 저장된 데이터들을 기반으로 분석하였다.

### 2. K-means 알고리즘

K-means 알고리즘은 클러스터링 방법 중 하나로, 라벨링되지 않은 입력 데이터를 k개의 클러

스터로 묶고, 라벨링하는 알고리즘이다. 각 클러스터와 거리 차이의 분산을 최소화하는 방식으로 동작한다[3]. 알고리즘의 순서로는, 가장 먼저 모든 점들을 랜덤 또는 휴리스틱 기법을 이용하여 k개의 집합으로 나눈다. 각 집합의 무게 중심을 구한다. 이 무게 중심을 중심으로 다시 k개의 집합으로 나눈다. 이 작업을 모든 점들이 각각 소속된 집합이 바뀌지 않거나, 무게 중심이 변하지 않을 때까지 반복한다.

이 알고리즘을 수학적 표현으로 나타내면 다음과 같다[4].

$$V = \sum_{i=1}^k \sum_{x_j \in S_i} |x_j - \mu_i|^2 \quad \dots(1)$$

$$S_i^{(t)} = \{x_p : |x_p - \mu_i^{(t)}|^2 \leq |x_p - \mu_j^{(t)}|^2 \forall j, 1 \leq j \leq k\} \dots(2)$$

$$\mu_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j \quad \dots(3)$$

- 전체 분산

공식(1)로  $\mu_i$ 는  $i$ 번째 클러스터의 중심,  $S_i$ 는 클러스터에 속하는 점의 집합을 의미하고, 전체 분산  $V$ 를 최소화하는  $S_i$ 를 찾는 것이 K-means 알고리즘의 목표이다.

- 클러스터

공식(2)로 각 데이터로부터 각 클러스터들의  $\mu_i$ 까지의 유클리드 거리를 계산하여, 해당 데이터에서 가장 가까운 클러스터를 찾아 데이터를 포함시킨다.

- 클러스터 중심

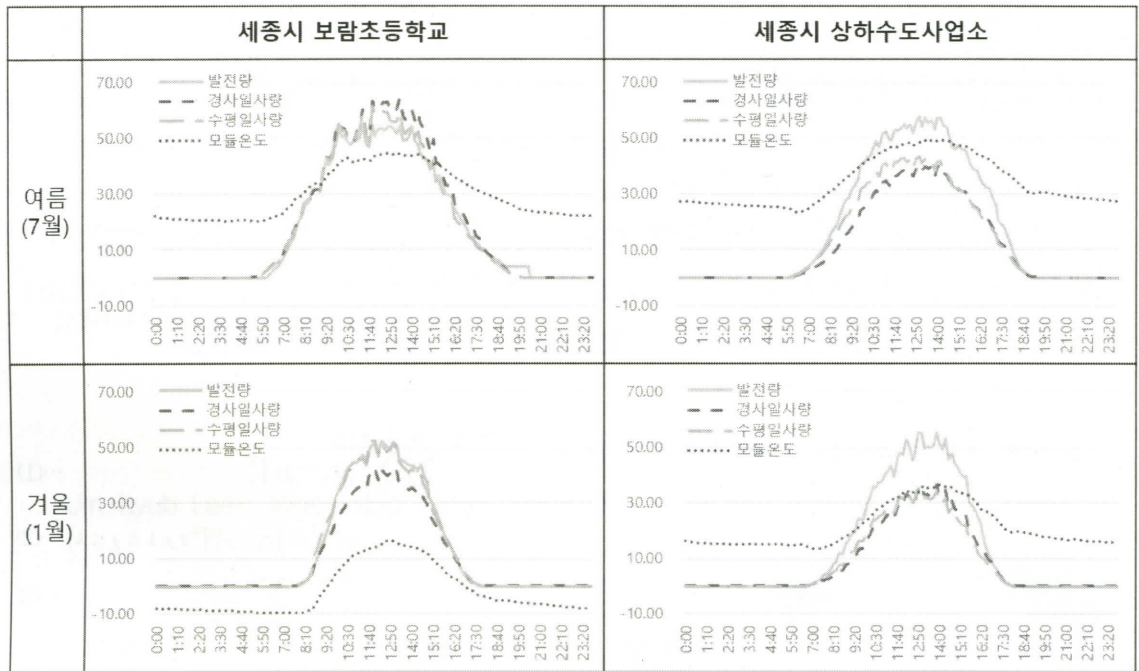
공식(3)으로  $\mu_i$ 를 각 클러스터에 있는 데이터들의 무게중심 값으로 변경한다.

## III. 태양광 발전 데이터 및 오류 분석

### 1. 시계열 그래프

위의 표 1은 세종시 보람초등학교 발전소와 세종시 상하수도사업소 발전소의 발전량, 경사 일사량, 수평 일사량, 모듈 온도를 7월 평균과 1월 평균을 시계열 그래프로 나타낸 표이다.

표 1. 세종시 보람초등학교 발전소와 세종시 상하수도사업소 발전소의 데이터별 7월, 1월 평균 시계열 그래프



주의할 점으로는 일사량 수치는 보통 0~1600 사이의 수치를 갖고, 발전량은 설치 용량에 따라 차이나므로 하나의 그래프 내에서 수치들을 비교하기 어렵다. 따라서 일사량은 1/10을 곱해주고, 발전량은 용량을 모든 발전소가 동일하게 100으로 맞춰 계산해 준다. 이때, 발전량은 보통 0~70 사이의 수치가 나온다. 따라서 이렇게 변환시켜 준 데이터들을 하나의 그래프에 나타내면, 수치적 데이터보다는 선의 흐름, 기울기에 집중하였다.

첫째로, 세종시 보람초등학교 발전소를 보면, 모듈온도를 제외한 발전량, 경사 일사량, 수평 일사량이 시간에 대비해 각 선의 기울기 변화가 거의 같음을 알 수 있다. 이는 모듈온도보다는 일사량들이 발전량에 가장 큰 영향을 미치는 것으로 볼 수 있다.

둘째로, 세종시 보람초등학교 발전소의 그래프는 정규분포와 유사한 형태로 좌우가 비교적 대칭되는 반면, 세종시 상하수도사업소 발전소의 그래프는 다소 오른쪽으로 치우쳐 있다. 이는 보람초등학교 발전소는 태양광 패널이 남향으로 설치되어 해가 뜰 때부터 질 때까지 골고루 발전하였다. 반면, 상하수도사업소의 태양광 패널은 서쪽

으로 설치되었다. 그래서 표 1에 세종시 상하수도사업소의 두 그래프와 같이 해질녘까지 발전량 기울기가 상대적으로 천천히 감소하다가 해가 짐으로써 발전량 기울기가 가파르게 감소한다.

## 2. 상관관계 그래프

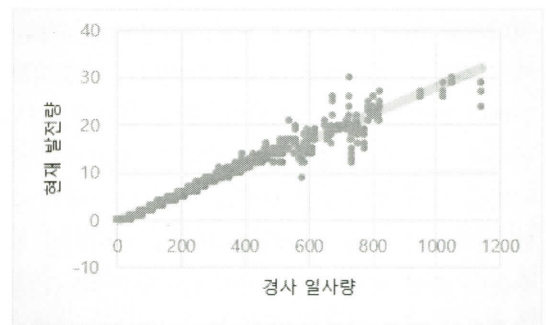


그림 2. 세종시 보람초등학교 발전소의 발전량과 경사 일사량 상관관계 그래프

그림 2는 세종시 보람초등학교 발전소의 발전량과 경사 일사량 사이의 상관관계 그래프이다. 이 그래프에서는 일사량 증가 대비 발전량 증가를 볼 수 있는데, 추세선과 같이 선형관계가 있다. 이는 일사량과 발전량의 직접적인 관계를 직접 확인할 수 있다.

이처럼 대체로 선형관계인 반면에 상관관계 그래프를 이용하면 오류 데이터도 확인할 수 있다.

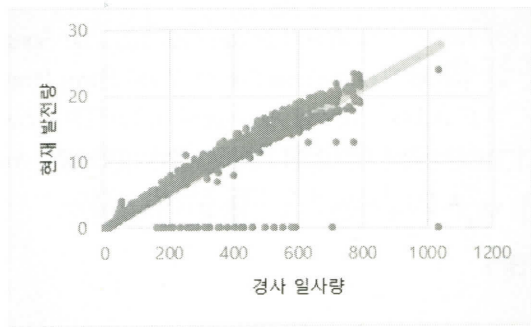


그림 3. 세종시 연서중학교 발전소의 발전량과 경사 일사량 상관관계 그래프

그림 3은 세종시 연서중학교 발전소의 발전량과 경사 일사량 사이의 상관관계 그래프이다. 동일하게 경사 일사량과 발전량이 선형관계로 확인할 수 있다.

하지만 그림 2와 다르게 경사 일사량 수치에 비해 발전량이 0인 점들을 확인할 수 있다. 이는 태양광 패널에 햇빛이 쬐는데도 발전이 일어나지 않음을 의미한다. 즉, 고장 또는 오류 데이터인 것이다. 이 데이터는 발전기 고장, 데이터 전송 시 에러 등 예상 가능하지만 다양한 원인이 있을 수 있다.

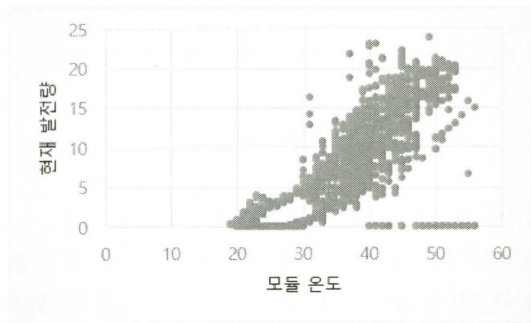


그림 4. 세종시 연서중학교 발전소의 발전량과 모듈 온도 상관관계 그래프

그림 4는 발전량과 모듈 온도간의 상관관계를 나타낸 그래프이다. 어떤 특별한 특징을 찾기 어렵다. 즉, 모듈 온도는 발전량에 큰 영향을 끼치지 않는다.

#### IV. K-means 알고리즘 적용

그림 3에서 확인된 오류 데이터를 통해 발전소의 고장진단 또는 예측을 할 수 있다. 이를 위해

신재생 통합 관리 모니터링 시스템 서버 내 데이

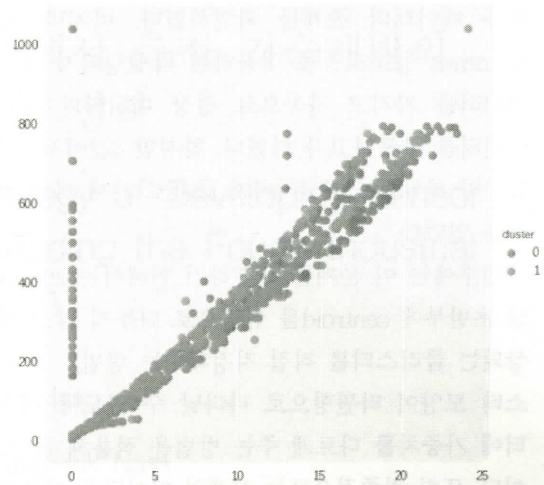


그림 5. 세종시 연서중학교 발전소의 경사 일사량과 발전량 데이터를 이용한 K-means clustering 결과를 전처리 후 텐서플로우를 이용하여 K-means 알고리즘을 적용하여 정상 데이터와 오류 데이터를 클러스터링 하였다.

그림 5는 클러스터링 결과이다. 그림에서 y축은 경사 일사량, x축은 발전량이다. 오류 데이터 클러스터와 정상 데이터 클러스터로 구분하여 k에 2를 입력하여 실행하였다. 그래서 x축이 0이면서 y축은 높은 수치인 부분을 하나의 클러스터, 정상 부분을 다른 하나의 클러스터로 예상하였다. 하지만 그림 4와 같이 정상 데이터와 오류 데이터를 구분 할 수 없었다. 그 이유로는 첫째, 정상 데이터의 개수에 비해 오류 데이터 개수가 거의 없다. 둘째, 표준 K-means 알고리즘은 데이터의 가중치를 동일하게 하여 클러스터의 모양이 둥글게 생성된다. 때문에 전체적으로 큰 원 하나와 작은 원 하나가 생성되어 겹쳐서 나타난 것이다.

#### V. 결론

이 논문에서는 관련연구에서 구현한 신재생 에너지 통합 관리 모니터링 시스템 서버에 1년간 축적된 데이터를 이용하여 다양한 분석을 시도하였다.

첫째, 데이터를 시계열 그래프에 나타내 각 데이터별 기울기를 비교하였고, 태양광 패널 설치 방향에 따라 그래프 모양이 다름을 확인하였다. 둘째, 발전량과 경사 일사량간의 상관관계 그래프

를 이용하여 서로 선형관계에 있음을 확인하였고, 오류 데이터의 존재를 확인하였다. 마지막으로 K-means 알고리즘을 적용하여 라벨링되지 않은 데이터를 가지고 자동으로 정상 데이터와 오류 데이터를 구분하고자 하였다. 하지만 4장에서 언급한바와같이 크게 두 가지 문제로 인해 다른 결과를 얻었다.

향후에는 이 문제를 해결하기 위해서는 프로그램 초반부에 centroid를 랜덤으로 나누지 않고 예상되는 클러스터를 직접 지정해주는 방법, 클러스터 모양이 타원형으로 나타날 수 있도록 데이터에 가중치를 다르게 주는 방법을 적용해 볼 것이다. 또한 최종적으로는 시계열 데이터를 이용한 발전량 예측도 시도할 예정이다[5].

## ACKNOWLEDGMENT

이 논문은 2015년 교육부와 한국연구재단의 지역혁신창의인력양성사업의 지원을 받아 수행된 연구임(NRF-2015H1C1A1035548).

## 참고문헌

- [1] 2016 신·재생에너지 백서, 신·재생에너지센터 (<https://www.renewableenergy.or.kr>), Oct. 2016.
- [2] Hyun Seung Son, R. Young Chul Kim, "Modeling a Photovoltaic Monitoring System based on Maintenance Perspective for New & Renewable Energy", The 2nd International Joint Conference on Convergence, AACL 07, pp. 144-147, Jan. 2016.
- [3] Tapas Kanungo, David M. Mount, Nathan S. Netanyahu, Christine D. Piatko, Ruth Silverman, Angela Y. Wu, "An Efficient k-Means Clustering Algorithm: Analysis and Implementation", IEEE Transactions on pattern analysis and machine intelligence, vol. 24, no. 7, pp. 881-892, July 2002.
- [4] Wikipedia, website, [https://ko.wikipedia.org/wiki/K-%ED%8F%89%EA%B7%A0\\_%EC%95%8C%EA%B3%A0%EB%A6%AC%E](https://ko.wikipedia.org/wiki/K-%ED%8F%89%EA%B7%A0_%EC%95%8C%EA%B3%A0%EB%A6%AC%E)

C%A6%98#.ED.91.9C.EC.A4.80\_.EC.95.8C.EA.B3.A0.EB.A6.AC.EC.A6.98.

- [5] Andrew Kusiak, Haiyang Zheng, Zhe Song, "Short-Term Prediction of Wind Farm Power: A Data Mining Approach", IEEE TRANSACTIONS ON ENERGY CONVERSION, vol. 24, no. 1, pp. 125-136, March 2009.