



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0066927
(43) 공개일자 2023년05월16일

(51) 국제특허분류(Int. Cl.)
H02J 3/00 (2006.01) G01W 1/02 (2022.01)
G06F 17/18 (2006.01) G06N 3/08 (2023.01)
(52) CPC특허분류
H02J 3/004 (2020.01)
G01W 1/02 (2022.01)
(21) 출원번호 10-2021-0152453
(22) 출원일자 2021년11월08일
심사청구일자 2021년11월08일

(71) 출원인
세종대학교산학협력단
서울특별시 광진구 능동로 209 (군자동, 세종대학교)
(72) 발명자
장윤
서울특별시 동대문구 정릉천동로 16, 103동 401호
(용두동, 용두두산위브)
박혜연
서울특별시 광진구 능동로32길 28-27, 301호 (능동)
정성민
경기도 하남시 미사강변한강로 290-2, 304동 2901호 (망월동, 미사강변도시씨3단지)
(74) 대리인
특허법인엠에이피에스

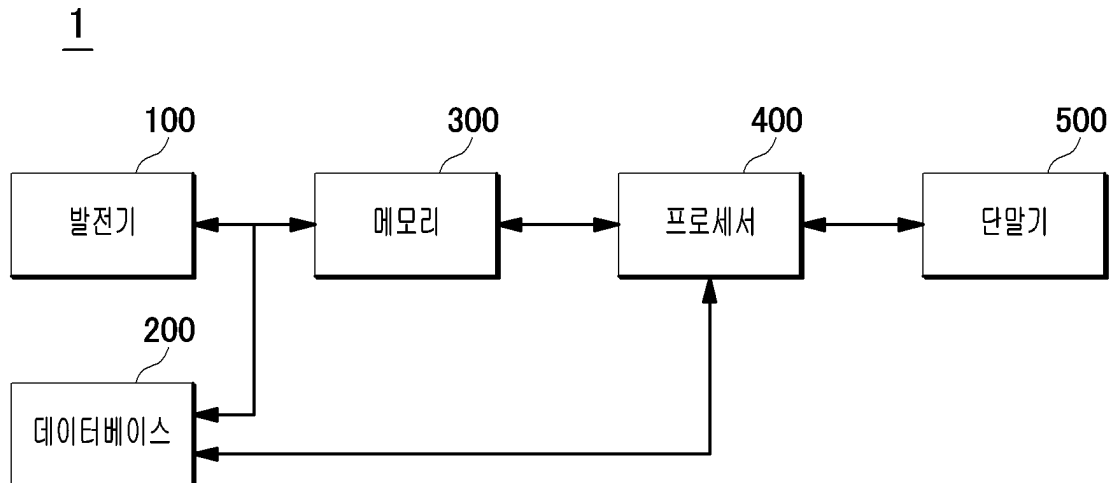
전체 청구항 수 : 총 18 항

(54) 발명의 명칭 발전량 예측 방법 및 장치

(57) 요약

실시예는 발전량 예측 방법 및 장치에 관한 것으로서, 입력 변수의 그래프의 형태적 특징을 이용하여 입력 변수를 복수의 특징 데이터 셋으로 분류하여 발전량을 예측하며, 입력 변수와 예측 발전량의 관계를 그래프를 이용하여 시각적으로 표현한, 발전량 예측 방법 및 장치를 제공한다.

대표도 - 도1



(52) CPC특허분류

G06F 17/18 (2013.01)

G06N 3/08 (2023.01)

H02J 2300/24 (2020.01)

이 발명을 지원한 국가연구개발사업

과제고유번호 1711125889

과제번호 2019-0-00374-003

부처명 과학기술정보통신부

과제관리(전문)기관명 정보통신기획평가원

연구사업명 빅데이터핵심기술

연구과제명 빅데이터 및 AI기반 에너지신산업형 분산자원(신재생에너지발전단지) 중개거래 플랫폼 개발

기 여 율 1/2

과제수행기관명 (주)엘시스

연구기간 2019.04.01 ~ 2021.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호 1711134451

과제번호 2021-0-00469-001

부처명 과학기술정보통신부

과제관리(전문)기관명 정보통신기획평가원

연구사업명 빅데이터핵심기술

연구과제명 이중 융합 데이터 탐지 및 추적 기술 개발

기 여 율 1/2

과제수행기관명 (주)피씨엔

연구기간 2021.04.01 ~ 2023.12.31

명세서

청구범위

청구항 1

복수의 입력 변수 데이터 중 어느 하나 이상의 입력 변수 데이터에 대한 제1 그래프를 생성하는 단계,

상기 제1 그래프의 형태를 기준으로 상기 입력 변수 데이터를 복수의 특징 데이터 셋으로 분류하는 단계, 그리고

상기 복수의 특징 데이터 셋을 이용하여 학습된 발전량 예측 모델에 상기 입력 변수 데이터를 입력하여 발전량을 예측하는 단계

를 포함하는, 발전량 예측 방법.

청구항 2

제1항에 있어서,

상기 입력 변수 데이터에 대한 그래프를 생성하는 단계는,

상기 복수의 입력 변수 데이터 중 어느 하나 이상을 독립 변수 및/또는 종속 변수로 하는 제2 그래프를 생성하는 단계,

상기 제2 그래프를 이용하여 상기 독립 변수와 상기 종속 변수의 상관도를 도출하는 단계,

상기 상관도가 미리 설정된 기준값 이상인 제3 그래프를 선택하는 단계, 그리고

상기 제3 그래프에 대응하는 입력 변수 중 어느 하나 이상을 이용하여 상기 제1 그래프를 생성하는 단계

를 포함하는, 발전량 예측 방법.

청구항 3

제2항에 있어서,

상기 입력 변수 데이터를 복수의 특징 데이터 셋으로 분류하는 단계는,

상기 제1 그래프의, 최대값, 최소값, 변곡점의 위치, 변곡점의 개수, 기울기, 분포도, 평균값 중 하나 이상을 이용하여 상기 입력 변수 데이터를 상기 특징 데이터 셋으로 분류하는 단계

를 포함하는, 발전량 예측 방법.

청구항 4

제3항에 있어서,

상기 발전량을 예측하는 단계는,

상기 복수의 입력 변수 데이터 중 어느 하나 이상을 독립 변수로, 상기 발전량 예측 결과를 종속 변수로 설정한 제 4그래프를 생성하는 단계

를 포함하는, 발전량 예측 방법.

청구항 5

제 4항에 있어서,

상기 제 4그래프를 생성하는 단계는,

상기 입력 변수 데이터와 상기 발전량 예측 결과 간의 기여도를 도출하고, 상기 기여도를 이용한 제 5그래프를 생성하는 단계

를 포함하는, 발전량 예측 방법.

청구항 6

제5항에 있어서,

상기 제 4그래프를 생성하는 단계는,

상기 입력 변수 데이터와 상기 발전량 예측 결과 간의 의존도를 도출하고, 상기 의존도를 이용한 제6 그래프를 생성하는 단계

를 더 포함하는, 발전량 예측 방법.

청구항 7

제3항 또는 제6항에 있어서,

상기 발전량을 예측하는 단계는,

상기 복수의 특징 데이터 셋을 이용하여, 앙상블 모델을 학습하고, 상기 앙상블 모델을 이용하여 발전량 예측 결과를 도출하는 단계

를 더 포함하는, 발전량 예측 방법.

청구항 8

제7항에 있어서,

상기 발전량을 예측하는 단계는,

RNN(Recurrent Neural Networks), LSTM(Long Short Term Memory), DENSE 구조 중 어느 하나 이상을 이용하여, 상기 특징 데이터 셋의 학습 및 발전량 예측을 수행하는 단계

를 더 포함하는, 발전량 예측 방법.

청구항 9

제8항에 있어서,

상기 발전량을 예측하는 단계는,

상기 특징 데이터 셋에 대하여, 결측값 처리, 여-존슨 변화(Yeo-Johnson Transformation), 최소값-최대값 스케일링(min-max scaling) 중 어느 하나 이상을 수행하여 상기 특징 데이터 셋의 전처리를 수행하는 단계

를 더 포함하는, 발전량 예측 방법.

청구항 10

발전량 예측 프로그램을 저장하는 메모리, 그리고

상기 메모리에 저장된 발전량 예측 프로그램을 실행하는 프로세서를 포함하며,

상기 프로세서는 상기 발전량 예측 프로그램을 실행하여, 복수의 입력 변수 데이터 중 어느 하나 이상의 입력 변수 데이터에 대한 제1 그래프를 생성하고, 상기 제1 그래프의 형태를 기준으로 상기 입력 변수 데이터를 복수의 특징 데이터 셋으로 분류하며, 상기 복수의 특징 데이터 셋을 이용하여 학습된 발전량 예측 모델에 상기 입력 변수 데이터를 입력하여 발전량을 예측하는, 발전량 예측 장치.

청구항 11

제10항에 있어서,

상기 프로세서는 상기 발전량 예측 프로그램을 실행하여, 상기 복수의 입력 변수 중 어느 하나 이상을 독립 변수 및/또는 종속 변수로 하는 제2 그래프를 생성하고, 상기 제2 그래프를 이용하여 상기 독립 변수와 상기 종속 변수의 상관도를 도출하며, 상기 상관도가 미리 설정된 기준값 이상인 제3 그래프를 선택하고, 상기 제3 그래프

에 대응하는 입력 변수 중 어느 하나 이상을 이용하여 상기 제1 그래프를 생성하는, 발전량 예측 장치.

청구항 12

제11항에 있어서,

상기 프로세서는 상기 발전량 예측 프로그램을 실행하여, 상기 제1 그래프의 최대값, 최소값, 변곡점의 위치, 변곡점의 개수, 기울기, 분포도, 평균값 중 하나 이상을 이용하여 상기 입력 변수 데이터를 상기 특징 데이터셋으로 분류하는, 발전량 예측 장치.

청구항 13

제12항에 있어서,

상기 프로세서는 상기 발전량 예측 프로그램을 실행하여, 상기 복수의 입력 변수 데이터 중 어느 하나 이상을 독립 변수로, 상기 발전량 예측 결과를 종속 변수로 설정한 제 4그래프를 생성하는, 발전량 예측 장치.

청구항 14

제13항에 있어서,

상기 프로세서는 상기 발전량 예측 프로그램을 실행하여, 상기 입력 변수와 상기 발전량 예측 결과 간의 기여도를 도출하고, 상기 기여도를 이용한 제 5그래프를 생성하는, 발전량 예측 장치.

청구항 15

제14항에 있어서,

상기 프로세서는 상기 발전량 예측 프로그램을 실행하여, 상기 입력 변수와 상기 발전량 예측 결과 간의 의존도를 도출하고, 상기 의존도를 이용한 제6 그래프를 생성하는, 발전량 예측 장치.

청구항 16

제12항 또는 제 15항에 있어서,

상기 프로세서는 상기 발전량 예측 프로그램을 실행하여, 상기 복수의 특징 데이터 셋을 이용하여, 앙상블 모델을 학습하고, 상기 앙상블 모델을 이용하여 발전량 예측 결과를 도출하는, 태양광 발전량 예측 장치.

청구항 17

제16항에 있어서,

상기 발전량 예측 프로그램은,

RNN(Recurrent Neural Networks), LSTM(Long Short Term Memory), DENSE 구조 중 어느 하나 이상을 포함하는, 발전량 예측 장치.

청구항 18

제17항에 있어서,

상기 프로세서는 상기 발전량 예측 프로그램을 실행하여, 상기 특징 데이터 셋에 대하여, 결측값 처리, 여-존슨 변화(Yeo-Johnson Transformation), 최초값-최대값 스케일링(min-max scaling) 중 어느 하나 이상을 수행하여 상기 특징 데이터 셋의 전처리를 수행하는, 발전량 예측 장치.

발명의 설명

기술 분야

본 발명의 실시예는 발전량 예측 방법 및 장치에 관한 것으로서, 보다 상세하게는 입력 데이터의 그래프를 도출하고, 그래프의 형태를 이용하여 발전량을 예측하는 발전량 예측 방법 및 장치에 관한 것이다.

[0001]

배경 기술

- [0002] 태양광 발전은 화석연료의 유력한 대체재이자 스마트 그리드(smart grid)의 핵심 역할로서 주목 받고 있다. 태양광 발전량은 기상조건과 발전 단지의 물리적 조건에 의해 발전량의 편차가 심하여 예측이 어렵다. 이는 에너지 생산 및 분배의 불균형 문제로 연결되며, 전력계통의 안정성을 감소시킨다. 따라서, 안정적인 스마트 그리드의 운용을 위해 정확도 높은 태양광 발전량의 예측이 필요하다.
- [0003] 태양광 에너지를 생산하는 태양광 패널에서는 다변량 시계열 데이터가 수집된다. 다변량 시계열 데이터는 비선형성과 비정상성을 가진다. 시간에 따라 평균, 분산, 공분산과 같은 통계적 독립변수들이 변하며 불규칙한 변동들이 나타나기 때문에 다변량 시계열 데이터의 정확한 예측이 어렵다.
- [0004] 특히, 입력 변수들 간에 강한 상관관계가 존재하는 경우, 학습 과정에서 다중공선성의 문제가 발생할 수 있다. 다중공선성이란 예측 모델의 학습을 위한 입력 변수들 간에 독립적인 관계가 아닌 서로 강한 상관관계가 존재하는 경우를 의미한다.
- [0005] 따라서, 데이터에 다중공선성(Multicollinearity)이 존재하는 경우, 예측 모델의 계수 추정이 불가하거나, 데이터의 미세한 변화에 따라 예측 모델의 예측 결과가 크게 달라질 수 있다는 문제점이 있다.
- [0006] 또한, 다변량 시계열 데이터의 예측을 위해, 다변량 시계열 데이터에서 과거와 현재 간의 종속관계들을 분석하여 향후 데이터 변화를 예측하기 위한 다양한 연구들이 수행되었다. 하지만, 종래의 예측 방법은 예측 모델의 훈련 시, 학습 데이터의 순서와 연속성을 보장하지 않으므로 시계열 예측에 한계가 존재한다.
- [0007] 또한, 과거의 상태를 보존하는 순환 구조를 가지므로, 예측 모델의 훈련이 반복됨에 따라 복잡도가 증가하여, 모델의 내부 구조 또는 예측 데이터 도출 과정을 설명하기 어렵다는 문제점이 있다.

발명의 내용

해결하려는 과제

- [0008] 본 발명의 실시예에 따른 발전량 예측 방법 및 장치는 입력 데이터 그래프의 형태를 이용하여 발전량을 예측하기 위한 발전량 예측 방법 및 장치를 제공하기 위한 것이다.
- [0009] 또한, 본 발명의 실시예에 따른 발전량 예측 방법 및 장치는 입력 변수 간의 다중공선성이 발생하지 않도록 하기 위한 발전량 예측 방법 및 장치를 제공하기 위한 것이다.
- [0010] 또한, 본 발명의 실시예에 따른 발전량 예측 방법 및 장치는 발전량 예측 결과의 도출 과정을 시각화하여 제공하기 위한 발전량 예측 방법 및 장치를 제공하기 위한 것이다.
- [0011] 다만, 본 실시예가 이루고자 하는 기술적 과제는 상기된 바와 같은 기술적 과제로 한정되지 않으며, 또 다른 기술적 과제들이 존재할 수 있다.

과제의 해결 수단

- [0012] 상술한 기술적 과제를 달성하기 위한 기술적 수단으로서, 발전량 예측 방법은, 복수의 입력 변수 데이터 중 어느 하나 이상의 입력 변수 데이터에 대한 제1 그래프를 생성하는 단계, 상기 제1 그래프의 형태를 기준으로 상기 입력 변수 데이터를 복수의 특징 데이터 셋으로 분류하는 단계, 상기 복수의 특징 데이터 셋을 이용하여 학습된 발전량 예측 모델에 상기 입력 변수 데이터를 입력하여 발전량을 예측하는 단계를 포함한다.
- [0013] 또한, 실시예에 따른 제1 그래프를 생성하는 단계는, 상기 복수의 입력 변수 데이터 중 어느 하나 이상을 독립 변수 및/또는 종속 변수로 하는 제2 그래프를 생성하는 단계, 상기 제2 그래프를 이용하여 상기 독립 변수와 상기 종속 변수의 상관도를 도출하는 단계, 상기 상관도가 미리 설정된 기준값 이상인 제3 그래프를 선택하는 단계, 그리고 상기 제3 그래프에 대응하는 입력 변수 중 어느 하나 이상을 이용하여 상기 제1 그래프를 생성하는 단계를 포함한다.
- [0014] 또한, 실시예에 따른 발전량을 예측하는 단계는, 상기 제1 그래프의, 최대값, 최소값, 변곡점의 위치, 변곡점의 개수, 기울기, 분포도, 평균값 중 하나 이상을 이용하여 상기 입력 변수 데이터를 상기 특징 데이터 셋으로 분류하는 단계를 포함한다.
- [0015] 또한, 실시예에 따른 발전량 예측 장치는, 발전량 예측 프로그램을 저장하는 메모리, 그리고 상기 메모리에 저

장된 발전량 예측 프로그램을 실행하는 프로세서를 포함하며, 상기 프로세서는 상기 발전량 예측 프로그램을 실행하여, 복수의 입력 변수 데이터 중 어느 하나 이상의 입력 변수 데이터에 대한 제1 그래프를 생성하고, 상기 제1 그래프의 형태를 기준으로 상기 입력 변수 데이터를 복수의 특징 데이터 셋으로 분류하며, 상기 복수의 특징 데이터 셋을 이용하여 학습된 발전량 예측 모델에 상기 입력 변수 데이터를 입력하여 발전량을 예측한다.

발명의 효과

- [0016] 본 발명의 실시예에 따른 발전량 예측 방법 및 장치는 입력 데이터 그래프의 형태를 이용하여 발전량을 예측하기 위한 발전량 예측 방법 및 장치를 제공할 수 있다.
- [0017] 또한, 본 발명의 실시예에 따른 발전량 예측 방법 및 장치는 입력 변수 간의 다중공선성이 발생하지 않도록 하기 위한 발전량 예측 방법 및 장치를 제공할 수 있다.
- [0018] 또한, 본 발명의 실시예에 따른 발전량 예측 방법 및 장치는 발전량 예측 결과의 도출 과정을 시각화하여 제공하기 위한 발전량 예측 방법 및 장치를 제공할 수 있다.

도면의 간단한 설명

- [0019] 도 1은 실시예에 따른 발전량 예측 장치의 구성도이다.
- 도 2는 실시예에 따른 발전량 예측 방법의 흐름도이다.
- 도 3은 실시예에 따른 제1 그래프 생성 방법의 흐름도이다.
- 도 4는 실시예에 따른 제2 그래프의 예시도이다.
- 도 5는, 실시예에 따른 발전량 그래프의 예시도이다.
- 도 6은 실시예에 따른 제1 그래프의 예시도이다.
- 도 7은 실시예에 따른 특징 데이터 셋 분류의 예시도이다.
- 도 8은 실시예에 따른 특징 데이터 셋 분류의 예시도이다.
- 도 9는 실시예에 따른 앙상블 모델의 예시도이다.
- 도 10은 실시예에 따른 앙상블 모델의 데이터 처리 개념도이다.
- 도 11은 실시예에 따른 발전량 예측 방법의 흐름도이다.
- 도 12는 실시예에 따른 제4 그래프 생성 방법의 흐름도이다.
- 도 13은 실시예에 따른 제5 그래프의 예시도이다.
- 도 14는 실시예에 따른 제5 그래프의 예시도이다.
- 도 15는 실시예에 따른 제5 그래프의 예시도이다.
- 도 16은 실시예에 따른 제6 그래프의 예시도이다.
- 도 17 은 실시예에 따른 발전량 예측 결과의 예시도이다.
- 도 18은 실시예에 따른 발전량 예측 결과의 예시도이다.
- 도 19는 실시예에 따른 발전량 예측 결과의 예시도이다.

발명을 실시하기 위한 구체적인 내용

- [0020] 아래에서는 첨부한 도면을 참조하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 본 발명의 실시예를 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.
- [0021] 명세서 전체에서, 어떤 부분이 다른 부분과 "연결"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우뿐 아니라, 그 중간에 다른 소자를 사이에 두고 "전기적으로 연결"되어 있는 경우도 포함한다. 또한 어떤 부분이

어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다.

- [0022] 본 명세서에 있어서 '장치'란, 하드웨어에 의해 실현되는 유닛(unit), 소프트웨어에 의해 실현되는 유닛, 양방을 이용하여 실현되는 유닛을 포함한다. 또한, 1 개의 유닛이 2 개 이상의 하드웨어를 이용하여 실현되어도 되고, 2 개 이상의 유닛이 1 개의 하드웨어에 의해 실현되어도 된다. 한편, '유닛'은 소프트웨어 또는 하드웨어에 한정되는 의미는 아니며, '유닛'은 어드레싱할 수 있는 저장 매체에 있도록 구성될 수도 있고 하나 또는 그 이상의 프로세서들을 재생시키도록 구성될 수도 있다. 따라서, 일 예로서 '유닛'은 소프트웨어 구성요소들, 객체지향 소프트웨어 구성요소들, 클래스 구성요소들 및 태스크 구성요소들과 같은 구성요소들과, 프로세스들, 함수들, 속성들, 프로시저들, 서브루틴들, 프로그램 코드의 세그먼트들, 드라이버들, 펌웨어, 마이크로코드, 회로, 데이터, 데이터베이스, 데이터 구조들, 테이블들, 어레이들 및 변수들을 포함한다. 구성요소들과 '유닛'들 안에서 제공되는 기능은 더 작은 수의 구성요소들 및 '유닛'들로 결합되거나 추가적인 구성요소들과 '유닛'들로 더 분리될 수 있다. 뿐만 아니라, 구성요소들 및 '유닛'들은 디바이스 내의 하나 또는 그 이상의 CPU들을 재생시키도록 구현될 수도 있다.
- [0023] 또한, 첨부된 도면은 본 명세서에 개시된 실시예를 쉽게 이해할 수 있도록 하기 위한 것일 뿐, 첨부된 도면에 의해 본 명세서에 개시된 기술적 사상이 제한되지 않으며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.
- [0024] 제1, 제2 등과 같이 서수를 포함하는 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되지는 않는다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다.
- [0025] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는, 중간에 다른 구성요소가 존재하지 않는 것으로 이해되어야 할 것이다.
- [0026] 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다.
- [0027] 본 출원에서, "포함한다" 또는 "가지다" 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [0028] 이하, 도 1을 참조하여 실시예에 따른 발전량 예측 장치(1)의 구성을 설명한다.
- [0029] 도 1은 실시예에 따른 발전량 예측 장치의 구성도이다.
- [0030] 도 1을 참조하면, 발전량 예측 장치(1)는 입력 변수 데이터 및 발전량 데이터를 이용하여 발전량을 예측한다. 따라서, 발전기(100)로부터 발전량 데이터를 수신하거나, 데이터베이스(200)에 누적하여 저장된 발전량 데이터, 입력 변수 데이터를 수신할 수 있다.
- [0031] 즉, 발전기(100)는 발전량 데이터를 데이터베이스(200) 또는 메모리(300)로 전송할 수 있다. 발전기(100)는 하나 또는 복수개의 발전기를 포함할 수 있다. 또한, 데이터베이스(200)는 발전기(100)에서 수신한 발전량 데이터 또는 발전량과 입력 변수에 관한 빅데이터를 수집 및 저장하며, 이를 메모리(300)로 전송할 수 있다.
- [0032] 입력 변수 데이터는 기상에 관련된 다변량 시계열 데이터를 포함할 수 있다. 예를 들어, 발전기(100)는 복수의 태양광 발전기를 포함하는 경우, 입력 변수 데이터는 온도, 습도, 일조량, 강수량, 풍속, 일출몰 시각 등에 관한 시계열 데이터를 포함할 수 있다. 또한, 입력 변수 데이터는 측정된 데이터뿐만 아니라 사용자가 임의로 설정한 설정값을 포함할 수 있다.
- [0033] 상술한 태양열 발전기 및 기상 데이터는 단순히 설명의 편의를 위해 설명한 예시에 불과할 뿐, 실시예가 이에 한정되는 것은 아니다. 발전기(100)는 풍력 발전기, 화력 발전기, 수력 발전기, 원자력 발전기 등을 포함할 수 있으며, 입력 변수 데이터는 기상 데이터뿐만 아니라 발전량을 예측하기 위한 다양한 시계열 데이터 또는 그래프로 표현 가능한 데이터를 입력 변수 데이터로 이용할 수 있다.
- [0034] 메모리(300)는 발전량 예측 프로그램을 저장한다. 발전량 예측 프로그램은 설명의 편의를 위해 설정된 것으로,

명칭 그 자체로 프로그램의 기능을 제한하는 것은 아니다. 메모리(300)는 발전기(100)로부터 생성되거나 측정된 데이터, 데이터베이스(200)에 저장된 데이터, 프로세서(400)에 의해 수행되는 기능에 필요한 정보 및 데이터, 프로세서(400)의 실행에 따라 생성된 데이터 중 적어도 어느 하나 이상을 저장할 수 있다.

- [0035] 메모리(300)는 전원이 공급되지 않아도 저장된 정보를 계속 유지하는 비휘발성 저장장치 및 저장된 정보를 유지하기 위하여 전력을 필요로 하는 휘발성 저장장치를 통칭하는 것으로 해석되어야 한다. 또한, 메모리(300)는 프로세서(400)가 처리하는 데이터를 일시적 또는 영구적으로 저장하는 기능을 수행할 수 있다.
- [0036] 메모리(300)는 저장된 정보를 유지하기 위하여 전력이 필요한 휘발성 저장장치 외에 자기 저장 매체(magnetic storage media) 또는 플래시 저장 매체(flash storage media)를 포함할 수 있으나, 본 발명의 범위가 이에 한정되는 것은 아니다.
- [0037] 더불어, 데이터베이스(200)는 발전량 예측 프로그램의 결과 누적 데이터, 발전량 예측을 위해 부수적으로 생성된 데이터 등을 메모리(300)로부터 수신하여 저장할 수 있다. 데이터베이스(200)는 메모리(300)의 일부를 구성할 수 있으나, 반드시 태양광 발전량 예측 장치(1)의 내부에 위치하는 것이 아니라 외부에 위치할 수도 있다.
- [0038] 프로세서(400)는 메모리(300)에 저장된 발전량 예측 프로그램을 실행하도록 구성된다. 프로세서(400)는 데이터를 제어 및 처리하는 다양한 종류의 장치들을 포함할 수 있다. 프로세서(400)는 프로그램 내에 포함된 코드 또는 명령으로 표현된 기능을 수행하기 위해 물리적으로 구조화된 회로를 갖는, 하드웨어에 내장된 데이터 처리 장치를 의미할 수 있다. 일 예에서, 프로세서(400)는 마이크로프로세서(microprocessor), 중앙처리장치(central processing unit: CPU), 프로세서 코어(processor core), 멀티프로세서(multiprocessor), ASIC(application-specific integrated circuit), FPGA(field programmable gate array) 등의 형태로 구현될 수 있으나, 본 발명의 범위가 이에 한정되는 것은 아니다.
- [0039] 프로세서(400)는, 발전량 예측 프로그램을 실행하여 다음과 같은 기능 및 절차들을 수행하도록 구성된다.
- [0040] 프로세서(400)는 발전량 예측 프로그램을 실행하여, 발전량에 영향을 끼치는 입력 변수 데이터를 시각화하고, 시각화된 데이터의 형태적 특징을 이용하여 입력 변수 데이터를 복수의 특징 데이터 셋으로 분류한다.
- [0041] 입력 변수 데이터를 이용한 그래프를 생성하는 것은 시각화의 한 방법으로 이용될 수 있다. 따라서, 특징 데이터 셋은 입력 변수 데이터의 그래프를 이용하여, 패턴 분리, 형태적 특징 등 시각적 특징을 이용하여 분류된 데이터를 의미한다. 이 때, 형태적 특징이란 그래프의 최대값, 최소값, 변곡점의 위치, 변곡점의 개수, 기울기, 분포도, 평균값 등 시각적으로 인식되는 특징 중 어느 하나 이상을 의미한다.
- [0042] 프로세서(400)는 발전량 예측 프로그램을 실행하여, 특징 데이터 셋을 이용한 머신 러닝의 학습 및 발전량 예측을 수행한다.
- [0043] 구체적으로, 프로세서(400)는 발전량 예측 프로그램을 실행하여 발전량 예측 모델을 생성할 수 있다. 발전량 예측 모델은 특징 데이터 셋을 학습 데이터로 하여 발전량 예측 결과를 도출하도록 학습된 인공지능 모델일 수 있다.
- [0044] 이때, 발전량 예측 모델은 앙상블 모델을 이용한다. 따라서, 복수의 입력 변수 데이터 각각에 대응하는 특징 데이터 셋을 이용하여, 각각의 입력 변수 데이터에 대한 발전량 예측을 수행한다.
- [0045] 또한, 프로세서(400)는 발전량 예측 프로그램을 실행하여, 특징 데이터 셋에 대한 결측값 처리, 여-존슨 변화(Yeo-Johnson Transformation), 최소값-최대값 스케일링(min-max scaling) 중 어느 하나 이상을 수행하여 상기 특징 데이터 셋의 전처리를 수행할 수 있다.
- [0046] 또한, 프로세서(400)는 발전량 예측 프로그램을 실행하여, RNN(Recurrent Neural Networks), LSTM(Long Short Term Memory), DENSE 구조 중 어느 하나 이상을 이용하여 입력 변수 데이터에 대한 학습을 수행할 수 있다.
- [0047] 또한, 프로세서(400)는 발전량 예측 프로그램을 실행하여, 발전량 예측 결과의 도출 과정을 설명하기 위해, 입력 변수 데이터 또는 특징 데이터 셋과 발전량 예측 결과 사이의 인과관계에 관한 그래프를 생성할 수 있다.
- [0048] 즉, 프로세서(400)는 발전량 예측 프로그램을 실행하여, 발전량 예측 결과 도출 과정에서 도출되는 중간 결과 데이터를 이용하여, 발전량 예측 결과를 도출하는 과정을 시각적 수단으로 표현할 수 있다.
- [0049] 발전량 예측 프로그램을 이용한 발전량 예측 방법은, 후술하는 도 2 내지 도 19를 참조하여 상세히 설명한다.
- [0050] 또한, 프로세서(400)는 통신부를 이용하여 단말기(500)와 정보 송수신을 수행할 수 있다. 따라서, 단말기(500)

로부터 입력 변수 데이터를 수신하고나, 발전량 예측 프로그램의 결과를 단말기(500)로 전송하여 사용자에게 디스플레이할 수 있다.

- [0051] 단말기(500)는 예를 들어, 웹 브라우저(WEB Browser)가 탑재된 노트북, 데스크톱(desktop), 랩톱(laptop), 휴대성과 이동성이 보장되는 무선 통신 장치 또는 스마트폰, 태블릿 PC 등과 같은 모든 종류의 핸드헬드(Handheld) 기반의 무선 통신 장치를 의미할 수 있다. 또한, 통신망은 근거리 통신망(Local Area Network; LAN), 광역 통신망(Wide Area Network; WAN) 또는 부가가치 통신망(Value Added Network; VAN) 등과 같은 유선 네트워크나 이동 통신망(mobile radio communication network) 또는 위성 통신망 등과 같은 모든 종류의 무선 네트워크로 구현될 수 있다.
- [0052] 이하, 도 2를참조하여 실시예에 따른 발전량 예측 방법을 설명한다.
- [0053] 도 2는 실시예에 따른 발전량 예측 방법의 흐름도이다.
- [0054] 도 2를 참조하면, 제1 그래프를 생성 단계(S100)에서는, 복수의 입력 변수 데이터 각각을 시각화한 하기 위해 제1 그래프를 생성한다. 그리고, 특징 데이터 셋 분류 단계(S200)에서는, 제1 그래프의 형태적 특징을 이용하여 입력 변수 데이터를 복수의 특징 데이터 셋으로 분류한다. 학습 및 예측 결과 도출 단계(S300)에서는, 특징 데이터 셋을 이용하여 발전량 예측 프로그램의 학습을 수행하여 발전량 예측 결과를 도출한다.
- [0055] 이하, 도 3 및 도 8을 참조하여 실시예에 따른 제1 그래프 생성 방법을 설명한다.
- [0056] 도 3은 실시예에 따른 제1 그래프 생성 방법의 흐름도이고, 도 4는 실시예에 따른 제2 그래프의 예시도이고, 도 5는 실시예에 따른 발전량 그래프의 예시도이다.
- [0057] 도 3을 참조하면, 제2 그래프 생성 단계(S110)에서는, 복수의 입력 변수 데이터를 이용하여 입력 변수 중 어느 하나 이상을 독립 변수 및/또는 종속 변수로 하는 제2 그래프를 생성한다.
- [0058] 도 4는, 태양광 발전량을 예측하기 위하여, 입력 변수가 강수량(Rain), 온도(Temprt), 풍속(Wnd_spd), 습도(Humdt), 구름량(Cloud), 태양광 발전량(Pow)인 경우 제2 그래프의 예시이다.
- [0059] x축은 독립 변수, y축은 종속 변수에 해당하며 독립 변수 및 종속 변수는 상술한 입력 변수 중 어느 하나를 선택하여 페어 플롯(PairPlot)을 생성한다. 따라서, 강수량(Rain), 온도(Temprt), 풍속(Wnd_spd), 습도(Humdt), 구름량(Cloud), 태양광 발전량(Pow)을 각각이 독립 변수 및 종속 변수로 설정된 36개의 그래프가 생성된다.
- [0060] 도 4에 도시된 페어 플롯은, 독립 변수와 종속 변수간의 상관관계를 시각적으로 표현할 수 있다. 즉, 제2 그래프는 입력 변수들 간의 상관관계를 나타내는 그래프이다. 구체적으로, 페어 플롯 그래프의 형태가 $y=x$ 그래프의 형태와 유사할수록, 독립 변수와 종속 변수간의 강한 양의 상관관계 존재한다. 하지만, 페어 플롯 그래프의 형태가 $y=-x$ 그래프의 형태와 유사할수록, 독립 변수와 종속 변수가 강한 음의 상관관계를 가진다.
- [0061] 또한, 도 4에 도시된 바와 같이 실시예에 따른 제2 그래프는, 특정 구간 데이터의 상관관계를 확인하기 위해, 브러싱 기능을 제공한다. 브러싱된 데이터 구간은 파랑색으로 표시되며, 특정 구간 데이터의 상관관계를 색상을 이용하여 즉각적으로 파악할 수 있다.
- [0062] 따라서, 제2 그래프를 이용하여, 입력 변수 중 어느 하나를 타겟 변수로 선택하고, 타겟 변수와 상관관계가 높은 입력 변수를 파악할 수 있다. 더불어, 강한 상관관계를 가지는 입력 변수를 선택하여 발전량 예측 장치(1)의 학습을 수행할 수 있다.
- [0063] 타겟 변수가 아닌 두 입력 변수들 간에 강한 상관관계가 존재하는 경우, 학습 과정에서 다중공선성의 문제가 발생할 수 있다. 다중공선성이란 예측 모델의 학습을 위한 입력 변수들 간에 독립적인 관계가 아닌 서로 강한 상관관계가 존재하는 경우를 의미한다.
- [0064] 따라서, 데이터에 다중공선성이 존재하는 경우, 예측 모델의 계수 추정이 불가하거나, 데이터의 미세한 변화에 따라 예측 모델의 예측 결과가 크게 달라질 수 있다. 그러므로, 입력 변수 중 어느 하나를 타겟 변수로 선택하고, 타겟 변수와 높은 상관관계가 존재하는 다른 하나의 입력 변수에 대한 학습을 수행함으로써, 다중공선성의 발생을 감소시킬 수 있다.
- [0065] 제1 그래프를 생성하기 위해, 제2 그래프 중 그래프의 형태가 $y=x$ 그래프 또는 $y=-x$ 그래프의 형태와 유사한 그래프를 제3 그래프로 선택할 수 있다. 그리고, 제3 그래프에 대응하는 입력 변수를 이용하여 제1 그래프를 생성할 수 있다. 또한, 제2 그래프의 형태를 이용하여 독립 변수와 종속 변수간의 상관도를 도출하고, 상관도가 미

리 설정된 기준값이 상인 그래프를 제3 그래프로 설정할 수 있다. 그리고, 제3 그래프에 대응하는 입력 변수를 이용하여 제1 그래프를 생성할 수 있다.

- [0066] 도 5는 타겟 변수 그래프의 예시로서, 시계열적 변화에 따른 태양광 발전량 그래프이다. 타겟 변수가 태양광 발전량으로 선택된 경우, 도 5와 같이 태양광 발전량의 라인 차트를 생성 할 수 있다.
- [0067] 구체적으로, 도 5의 (a)는 시간(HOUR)단위, 도 5의 (b)는 년(year)단위, 도 5의 (c)는 계절(season) 단위, 도 5의 (d)는 월(month) 단위로 태양광 발전량을 도시한다.
- [0068] 하지만, 실시예가 이에 한정되는 것은 아니며, 입력 변수에 따라 효과적으로 데이터를 시각화 할 수 있는 다양한 그래프 및 방법을 이용하여 표현될 수 있다.
- [0069] 도 6은 제1 그래프의 예시도이다.
- [0070] 도 6는 타겟 변수가 태양광 발전량인 경우, 풍속, 습도, 태양 복사(solar radiation), 지면 온도(Ground temperature), 구름량에 관한 제1 그래프가 각각 생성된 경우이다. 즉, 제 2 그래프 중 태양광 발전량이 독립 변수이고, 풍속, 습도, 태양 복사(solar radiation), 지면 온도(Ground temperature), 구름량을 종속 변수로 하는 그래프가 제3 그래프로 선택되고, 풍속, 습도, 태양 복사(solar radiation), 지면 온도(Ground temperature), 구름량에 대한 제1 그래프가 생성된 예시이다.
- [0071] 제1 그래프는, 입력 변수 데이터의 특징을 도출하기 위한 것으로서, 제1 그래프의 형태적 특징을 이용하여 입력 데이터를 복수의 특징 데이터 셋으로 분류할 수 있다.
- [0072] 이하, 도 7을 참조하여 실시예에 따른 특징 데이터 셋 분류 방법을 설명한다.
- [0073] 도 7은 실시예에 따른 특징 데이터 셋 분류의 예시도이다.
- [0074] 도 7의 (a) 및 도 7의 (b)는 상술한 복수의 제1 그래프 중 지면 온도에 관한 입력 변수 데이터를 복수의 특징 데이터 셋으로 분류한 데이터 분할 그룹을 나타낸다. 즉, 지면 온도에 관한 입력 변수 데이터를 4개의 그룹(Group0, Group1, Group2, Group3)으로 분할하는 것을 나타낸다. 각각의 그룹의 구분을 시각적으로 명확히 표현하기 위해 각각의 그룹은 색을 달리하여 표시될 수 있다.
- [0075] 구체적으로, 도 6에 도시된 지면 온도에 관한 제1 그래프를 선택하는 경우, 제1 그래프의, 최대값, 최소값, 변곡점의 위치, 변곡점의 개수, 기울기, 분포도, 평균값 등과 같은 형태적 특징을 하나 이상 이용하여 지면 온도 데이터를 네 개의 특징 데이터 셋으로 분류한다.
- [0076] 도 7의 (c)는 선택한 입력 변수인 지면 온도의 라인차트를 나타내며, 도 7의 (d)는 타겟 변수인 태양광 발전량의 라인 차트를 나타낸다. 따라서, 입력 변수가 그래프의 형태적 특징에 따라 복수의 특징 데이터 셋으로 분류되는 것을 시각적으로 표현할 수 있다.
- [0077] 이 때, 도 7에는 입력 변수를 지면 온도를 선택하여 특징 데이터 셋을 분류한 예시만 도시되었지만, 본 발명의 실시예가 이에 한정되는 것은 아니며 입력 변수는 모두 선택될 수 있다. 또한, 입력 변수 데이터를 4개의 그룹으로 분류한 예시가 도시되어 있으나, 본 발명의 실시예가 이에 한정되는 것은 아니며 제1 그래프의 형태적 특성에 따라 분류되는 그룹 및 특징 데이터 셋의 개수는 변화할 수 있다.
- [0078] 도 8은 실시예에 따른 특징 데이터 셋 분류의 예시도이다.
- [0079] 도 8은 입력 변수가 각각 기상 패턴, 계절기후, 발전량인 경우 특징 데이터 셋의 예시를 나타낸다.
- [0080] 입력 변수가 기상 패턴인 경우, 특징 데이터 셋은 맑음(SUNNY DATA), 강수(RAINY DATA), 흐림(CLOUDY DATA), 강풍(WINDY DATA)로 분류될 수 있다. 또한, 입력 변수가 계절 기후인 경우, 특징 데이터 셋은 봄(Spring DATA), 여름(Summer DATA), 가을(Autumn DATA), 겨울(Winter DATA)로 분류될 수 있다. 입력 변수가 발전량인 경우, 특징 데이터 셋은 저전력(LOW-POWER DATA), 고전력(HIGH-POWER DATA)로 구분될 수 있다.
- [0081] 이하, 도 9 및 도 10을 참조하여 실시예에 따른 앙상블 모델의 구조를 설명한다.
- [0082] 도 9는 실시예에 따른 앙상블 모델의 예시도이다.
- [0083] 도 9를 참조하면, 앙상블 모델은 복수의 입력 변수 데이터 각각에 대하여 병렬적으로 학습을 수행하고, 각각의 입력 변수 데이터를 이용하여 발전량 예측 결과를 도출하는 학습 모델을 의미한다.
- [0084] 예를 들어, 서브 데이터 셋 1(Sub-dataset 1)은 도 8에 도시된 바와 같이, 기상패턴에 따라 분리된 특징 데이터

셋인, 맑음(SUNNY DATA), 강수(RAINY DATA), 흐림(CLOUDY DATA), 강풍(WINDY DATA)에 대응될 수 있다.

- [0085] 서버 데이터 셋 1은 앙상블 모델의 하위 모델인 서버 모델 1(Sub-model 1)에 입력되며, 서버 모델 1은 기상패턴에 따라 분리된 특징 데이터 셋을 이용하여 학습을 수행한다. 그리고, 기상패턴에 따라 분리된 특징 데이터 셋을 이용한 발전량 예측 결과(Prediction 1)를 도출한다.
- [0086] 또한, 서버 데이터 셋 2(Sub-dataset 2)는 입력 변수가 계절 기후인 경우, 특징 데이터 셋인 봄(Spring DATA), 여름(Summer DATA), 가을(Autumn DATA), 겨울(Winter DATA)에 대응될 수 있다.
- [0087] 서버 데이터 셋 2는 앙상블 모델의 하위 모델인 서버 모델 2(Sub-model 2)에 입력되며, 서버 모델 2는 입력 변수가 계절 기후인 경우 분류된 특징 데이터 셋을 이용하여 학습을 수행한다. 그리고, 계절 기후에 따라 분리된 특징 데이터 셋을 이용한 발전량 예측 결과(Prediction 2)를 도출한다.
- [0088] 서버 데이터 셋 3(Sub-dataset 3)은 입력 변수가 발전량인 경우, 특징 데이터 셋인 저전력(Low-Power DATA), 고전력(High-Power DATA)에 대응될 수 있다.
- [0089] 서버 데이터 셋 3은 앙상블 모델의 하위 모델인 서버 모델 3(Sub-model 3)에 입력되며, 서버 모델 3은 입력 변수가 발전량인 경우 분류된 특징 데이터 셋을 이용하여 학습을 수행한다. 그리고, 발전량에 따라 분리된 특징 데이터 셋을 이용한 발전량 예측 결과(Prediction 3)를 도출한다.
- [0090] 더불어, 앙상블 모델은 N개의 입력 변수에 대한 학습을 수행하기 위해, N개의 하위 모델이 병렬적으로 구비되는 구조를 가질 수 있다. 따라서, 각각의 하위 모델은 분류된 특징 데이터 셋을 이용하여 입력 데이터에 대한 학습을 수행하고, 각 특징 데이터 셋에 대한 발전량 예측 결과를 도출한다.
- [0091] 따라서, 실시예에 따른 발전량 예측 방법 및 장치는 입력 변수를 분리하여 학습을 수행하므로, 입력 변수 간의 다중공선성이 발생하지 않도록 학습 및 발전량 예측을 수행한다.
- [0092] 각각의 하위 모델(Sub-model 1, 2, 3, ..., N)에서 도출된 발전량 예측 결과(Prediction 1, 2, 3, ..., N)는 메모리(300) 및/또는 데이터베이스(200)에 저장될 수 있다. 그리고, 프로세서(400)는 발전량 예측 프로그램을 실행하여 각각의 예측 결과(Prediction 1, 2, 3, ..., N)에 대한 손실율(loss) 또는 예측 정확도를 도출할 수 있다.
- [0093] 손실율 또는 에러(Error Rate)는 편향 분산 트레이드오프(bias-variance tradeoff)를 이용하여 도출할 수 있으며, 손실율 또는 에러가 가장 낮은 예측 결과 또는 예측 정확도가 가장 높은 예측 결과를 최종 발전량 예측 결과로 도출할 수 있다.
- [0094] 도 10은 실시예에 따른 앙상블 모델의 데이터 처리 개념도이다.
- [0095] 도 10을 참조하면, 도 10의 (a)는 전처리 단계를 나타낸다. 따라서, 앙상블 모델의 하위 구조로 입력된 특징 데이터 셋은 전처리 단계를 거친다. 전처리 단계는 특징 데이터 셋에 대하여, 결측값 처리, 여-존슨 변화(Yeo-Johnson Transformation), 최소값-최대값 스케일링(min-max scaling) 중 어느 하나 이상을 수행할 수 있다.
- [0096] 도 10의 (b)는 학습 및 결과 예측을 위한 레이어(layer) 구조를 나타낸다. 전처리 단계를 거친 특징 데이터 셋은 레이어 구조를 이용하여 학습 및 발전량 예측을 수행한다. 레이어는 RNN(Recurrent Neural Networks), LSTM(Long Short Term Memory), DENSE 구조 중 어느 하나 이상을 포함할 수 있다.
- [0097] 또한, 도 10의 (c)는 하이퍼파라미터 설정을 도시화하여 나타낸다. 하이퍼파라미터 설정 단계에서는, 형상(shape), 유닛(units), 활성화 함수(activation functions), 편향(bias) 등을 설정할 수 있다. 이 때, 일반적으로 활성화 함수(activation functions)에는 tanh 또는 ReLU이 사용될 수 있다.
- [0098] 도 10의 (d)는 레이어 구조를 시각화하여 나타낸다. 따라서, 전처리 단계(Preprocessing)에서는 앙상블 모델의 하위 구조로 입력된 특징 데이터 셋에 대한 전처리를 수행한다. 전처리된 특징 데이터 셋은, 레이어 계층(LSTM, DENSE)를 거쳐 학습 및 발전량 예측에 사용된다.
- [0099] 그리고, 각각의 하위 모델에서 도출된 발전량 예측 값은 메모리(300) 및/또는 데이터베이스(200)에 저장될 수 있다. 또한, 프로세서(400)는 편향 분산 트레이드오프(bias-variance tradeoff)를 이용하여 각각의 예측 결과(Prediction 1, 2, 3, ..., N)에 대한 손실율(loss) 또는 에러(Error Rate)를 도출할 수 있다. 손실율 또는 에러가 가장 낮은 예측 결과 또는 예측 정확도가 가장 높은 예측 결과를 최종 발전량 예측 결과로 도출할 수 있다.

- [0100] 이하, 도 11 내지 도 19를 참조하여 실시예에 따른 제4 그래프 생성 방법을 설명한다.
- [0101] 도 11은 실시예에 따른 발전량 예측 방법의 흐름도이다.
- [0102] 도 11을 참조하면, 실시예에 따른 발전량 예측 방법은 제4 그래프 생성 단계(S400)를 더 포함할 수 있다. 제 4 그래프는 입력 변수로부터 발전량 예측 결과를 도출하는 과정을 도시적으로 표현하기 위한 그래프를 의미한다. 즉, 제4 그래프는 입력 변수 데이터와 발전량 예측 결과간의 인과관계를 나타내는 그래프를 의미한다.
- [0103] 도 12는, 실시예에 따른 제4 그래프 생성 방법의 흐름도이다.
- [0104] 도 12를 참조하면, 입력 변수와 예측 결과간 기여도 도출 단계(S410)에서 기여도를 도출한다. 그리고, 제 5그래프 생성 단계(S420)에서, 입력 변수와 예측 결과간 기여도를 이용하여 제5 그래프를 생성한다.
- [0105] 구체적으로, 입력 변수와 예측 결과간 기여도는 SHAP(shapley additive explanations) 밸류(Value)를 이용하여 도출할 수 있다. SHAP 밸류를 이용한 제 5 그래프는 도 13, 도 14 및 도 15와 같이 도출될 수 있다.
- [0106] 도 13은 SHAP 서머리 플롯(Summary plot)의 예시도를 나타낸다. 도 13의 x축은 입력 변수의 기여도를 나타내는 SHAP 값을 의미한다. y축은 입력 변수를 나타내며, 입력 변수의 순서는 기여도에 따라 내림차순으로 정렬된다.
- [0107] 또한, 도 13의 각 점은 입력 변수 내 인스턴스이며, 점의 색상은 인스턴스 값의 크기를 나타낸다. 따라서, 붉은 색 점일수록 높은 값에 해당하며, 푸른색 점일수록 낮은 값의 인스턴스에 해당한다. 더불어, 점들이 뭉쳐있는 것은, SHAP 값의 밀집된 분포를 보여준다.
- [0108] 따라서, 도 13은 복수의 입력 변수 중 온도(temprt), 습도(humdt), 구름량(cloud), 강수량(raisin)에 대한 SHAP 서머리 플롯이다. 그리고, 최상단에 위치한 온도가 발전량에 가장 기여도가 크며, 강수량이 가장 낮은 기여도를 가진다.
- [0109] 도 14는 결정 플롯(Decision plot)의 예시도를 나타낸다. 결정 플롯은 예측까지 도달하는 경로를 도시한다. 경로는 그래프의 아래에서 위를 향하는 방향으로 진행된다. 이 때, 기여도는 선의 이동 방향과 거리로 표현된다. 그래프의 중앙에서 오른쪽 방향은 양의 기여도, 왼쪽 방향은 음의 기여도를 의미한다.
- [0110] 이동 거리는 기여도의 크기를 나타낸다. 또한, 선의 색상은 모든 인스턴스의 예측 결과들 내에서 해당 인스턴스의 분위수에 따라 지정된다. 따라서, 선의 이동 거리가 가장 긴 온도(temprt)의 기여도가 가장 크며, 선의 이동 거리가 가장 짧은 강수량(rain)이 가장 낮은 기여도를 가진다.
- [0111] 도 15는, 클러스터링 SHAP 플롯(Clustering SHAP plot)의 예시도를 나타낸다. 클러스터링 SHAP 플롯은, 전체 입력 데이터셋에 대한 예측 결과의 변화량을 나타낸다.
- [0112] x축은 데이터의 인스턴스를 나타내며, 인스턴스는 예측 결과가 높은 순으로 정렬된다. y축은 예측 결과와 기여도의 크기를 나타낸다. 각 인스턴스의 예측 결과는 붉은 색과 푸른색이 접해있는 지점에 해당한다. 붉은 색은 양의 기여도를 의미하고, 푸른 색은 음의 기여도를 나타낸다. 따라서, 인스턴스의 모든 입력 변수들의 기여도 합이 양인 경우, 기준 값보다 예측 결과가 높게 나타난다.
- [0113] 또한, 도 12에 도시된 바와 같이, 실시예에 따른 발전량 예측 방법은, 기여도에 따른 제5그래프를 생성하는 것뿐만 아니라, 입력 변수와 발전량 예측 결과간의 의존도를 이용한 제6 그래프를 생성할 수 있다.
- [0114] 구체적으로, 입력 변수와 예측 결과간 의존도 도출 단계(S430)에서는, 다른 모든 입력 변수들(marginal feature set)의 조건이 동일하다고 가정한 후, 한가지 입력 변수(selected feature set)의 변화가 예측 결과에 미치는 한계 효과(Marginal effect)로부터 의존도를 도출한다. 그리고, 제6 그래프 도출 단계(S440)에서는 입력 변수와 발전량 예측 결과 사이의 의존도를 이용한 제6 그래프를 생성한다. 제6 그래프는, 도 16 과 같이 도출될 수 있다.
- [0115] 도 16은, ICE(Individual Conditional Expectation) 플롯의 예시도를 나타낸다. ICE 플롯은 입력 변수의 값이 변할 때 발전량 예측 결과 값의 변화를 모든 인스턴스에 대하여 도시한다. 따라서, 도 16에는 입력 변수가 발전량(power), 풍속(wind), 습도(Humidity), 조도(Irradiance), 온도(Temperature)인 경우의 ICE 플롯의 예시이다.
- [0116] ICE 플롯에서 의존도는 입력 변수 값과 예측 결과 사이의 변화 관계를 의미한다. 따라서, 인스턴스 각각의 의존도에 대응하는 선의 이용하여 의존도를 도시한다.

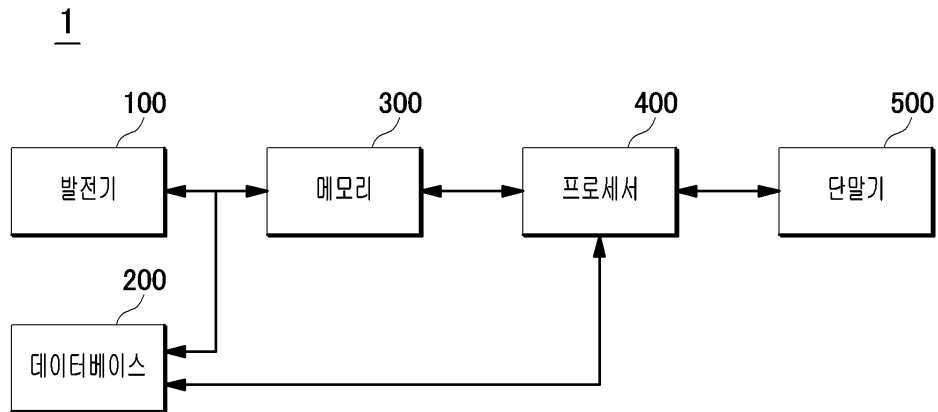
- [0117] 굵은 선은 모든 인스턴스의 평균 의존도를 나타낸다. 평균 의존도는 PDP(Partial Dependence Plot)와 동일한 결과를 나타낸다. 또한, ICE 플롯에서 x축은 입력 변수들의 값을 나타내고, y축은 발전량 예측 결과 값을 나타낸다. 더불어, 다수의 모델들의 의존도를 색상으로 구분하여 하나의 ICE 플롯에 표시할 수 있다.
- [0118] 도 17은 내지 도 19는 실시예에 따른 발전량 예측 결과의 예시도이다.
- [0119] 도 17은 실제 예측해야하는 발전량 값과 실시예에 따른 발전량 예측 장치 및 방법을 이용하여 예측한 발전량 예측 값을 도시한 그래프이다. x축은 시간축(epoch)을 나타내며, y축은 태양광 발전량을 나타낸다.
- [0120] L1은 모든 입력 변수를 이용하여 예측한 태양광 발전량 예측 값을 나타낸다. L2는 실제 예측해야 하는 태양광 발전량 값을 나타낸다. L3 및 L4는 하이퍼파라미터 설정을 변경하여 예측한 태양광 발전량 예측 값을 나타낸다.
- [0121] 도 18 및 도 19는 편향 분산 트레이드오프(bias-variance tradeoff)를 확인하기 위한 그래프이다. x축은 에러(error, bias 혹은 variance)를 나타내며, y축은 손실(loss)을 나타낸다. L5는 편향(Bias), L6는 분산(variance)을 나타낸다. 시간(epoch)에 따른 오버피팅(overfitting), 언더피팅(underfitting)여부를 분석하여 편향 분산 트레이드오프를 확인할 수 있다.
- [0122] 또한, 전체 에러는 아래의 수학적 식 1과 같이 정의될 수 있다.
- [0124] [수학적 식 1]
- [0125]
$$\text{Total error} = \text{bias}^2 + \text{Variance} + \text{Irreducible Error}$$
- [0127] 본 발명의 일 실시예는 컴퓨터에 의해 실행되는 프로그램 모듈과 같은 컴퓨터에 의해 실행가능한 명령어를 포함하는 기록 매체의 형태로도 구현될 수 있다. 컴퓨터 판독 가능 매체는 컴퓨터에 의해 액세스될 수 있는 임의의 가용 매체일 수 있고, 휘발성 및 비휘발성 매체, 분리형 및 비분리형 매체를 모두 포함한다. 또한, 컴퓨터 판독 가능 매체는 컴퓨터 저장 매체를 포함할 수 있다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 모두 포함한다.
- [0128] 본 발명의 방법 및 시스템은 특정 실시예와 관련하여 설명되었지만, 그것들의 구성 요소 또는 동작의 일부 또는 전부는 범용 하드웨어 아키텍처를 갖는 컴퓨터 시스템을 사용하여 구현될 수 있다.
- [0129] 전술한 본 발명의 설명은 예시를 위한 것이며, 본 발명이 속하는 기술분야의 통상의 지식을 가진 자는 본 발명의 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다. 예를 들어, 단일형으로 설명되어 있는 각 구성 요소는 분산되어 실시될 수도 있으며, 마찬가지로 분산된 것으로 설명되어 있는 구성 요소들도 결합된 형태로 실시될 수 있다.
- [0130] 본 발명의 범위는 상기 상세한 설명보다는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

부호의 설명

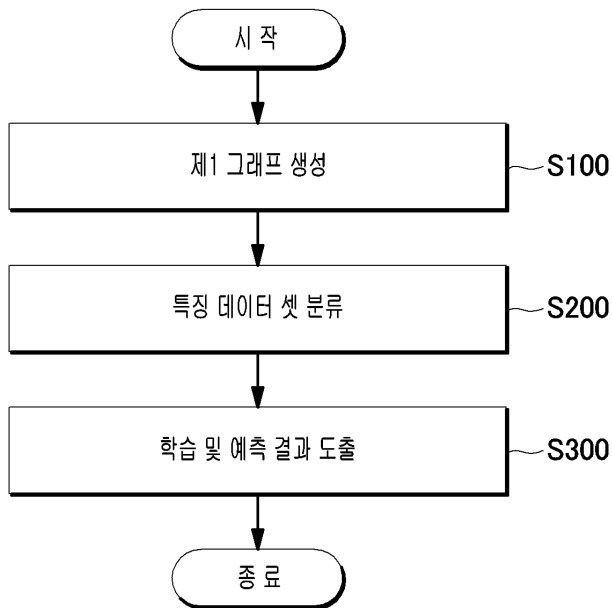
- | | |
|---------------------|----------|
| [0131] 1: 발전량 예측 장치 | 100: 발전기 |
| 200: 데이터베이스 | 300: 메모리 |
| 400: 프로세서 | 500: 단말기 |

도면

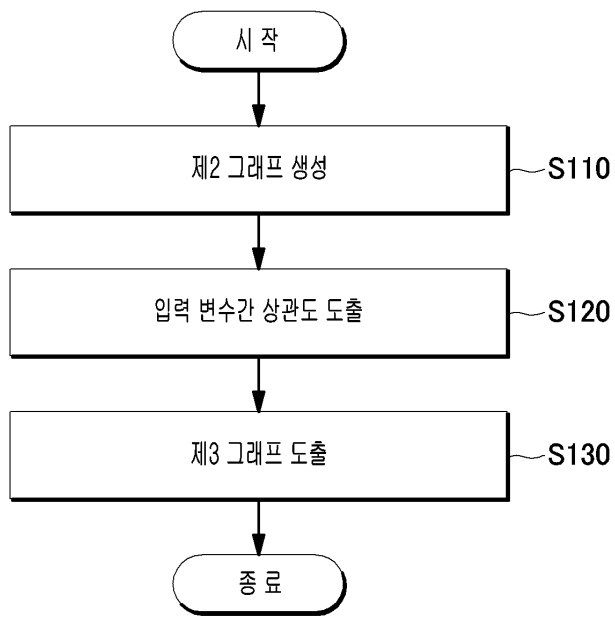
도면1



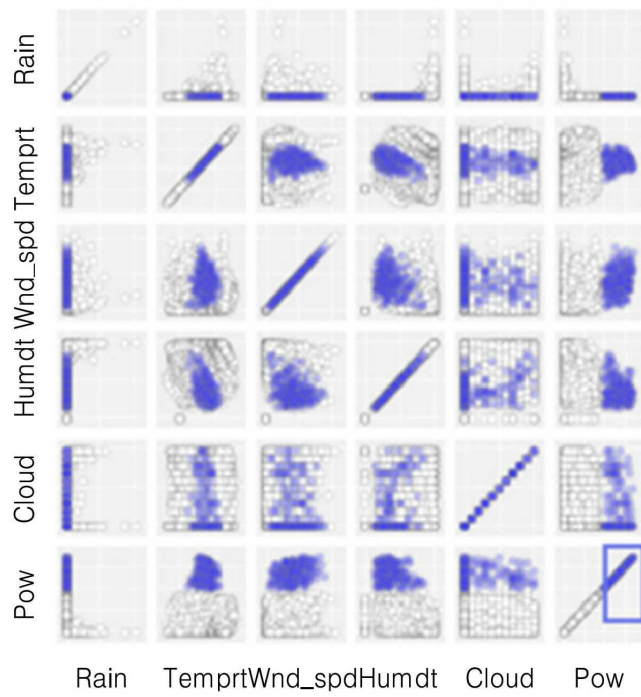
도면2



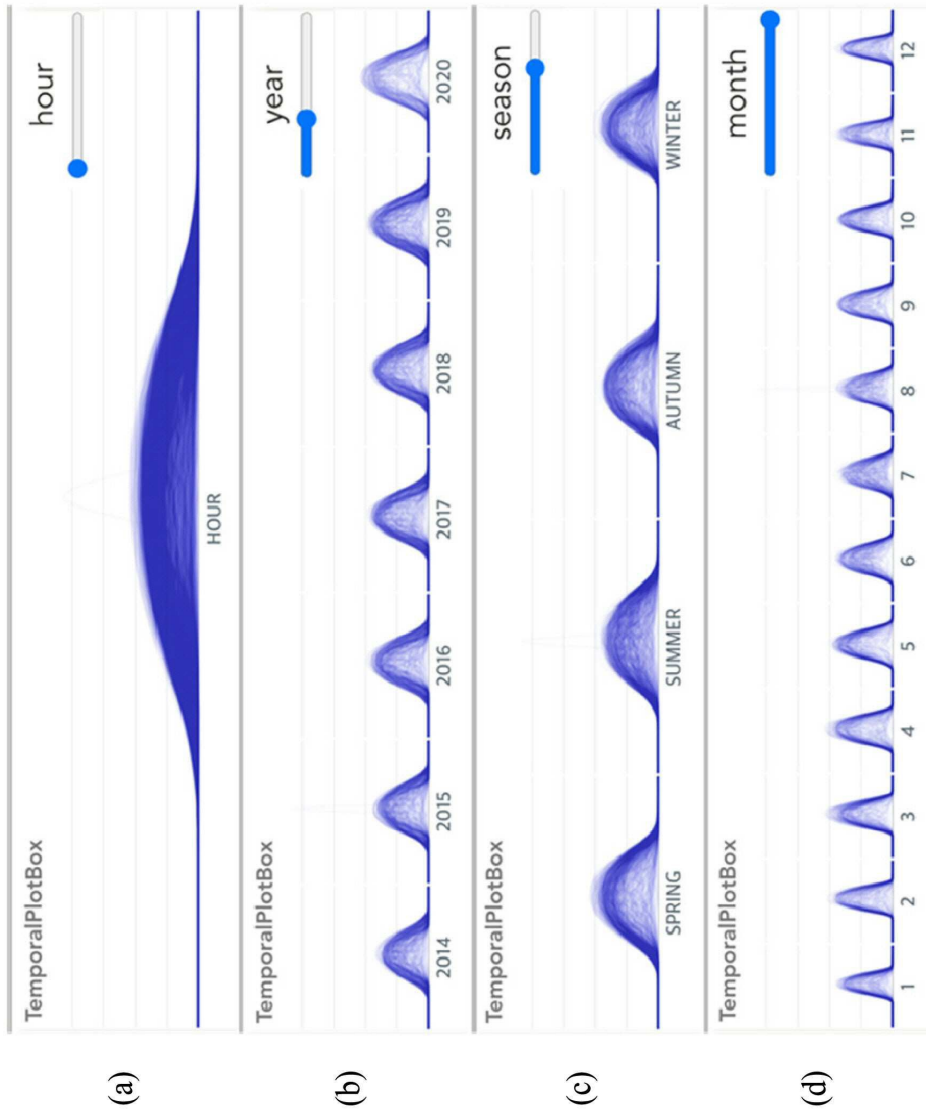
도면3



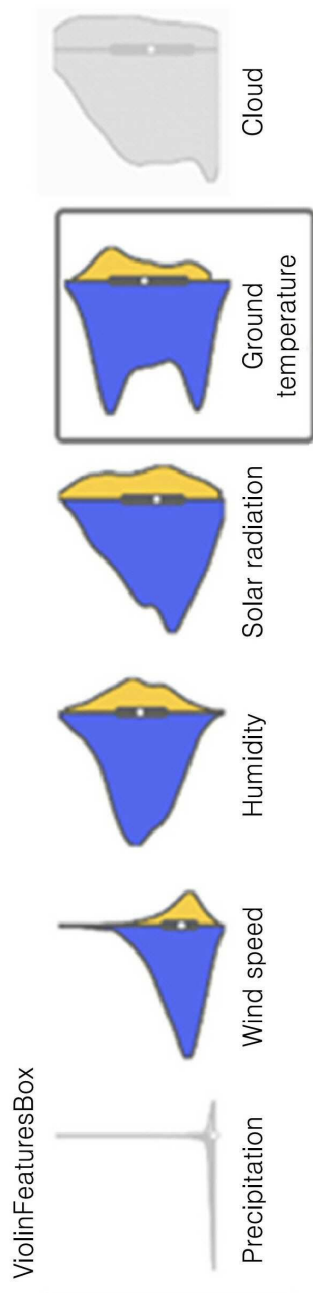
도면4



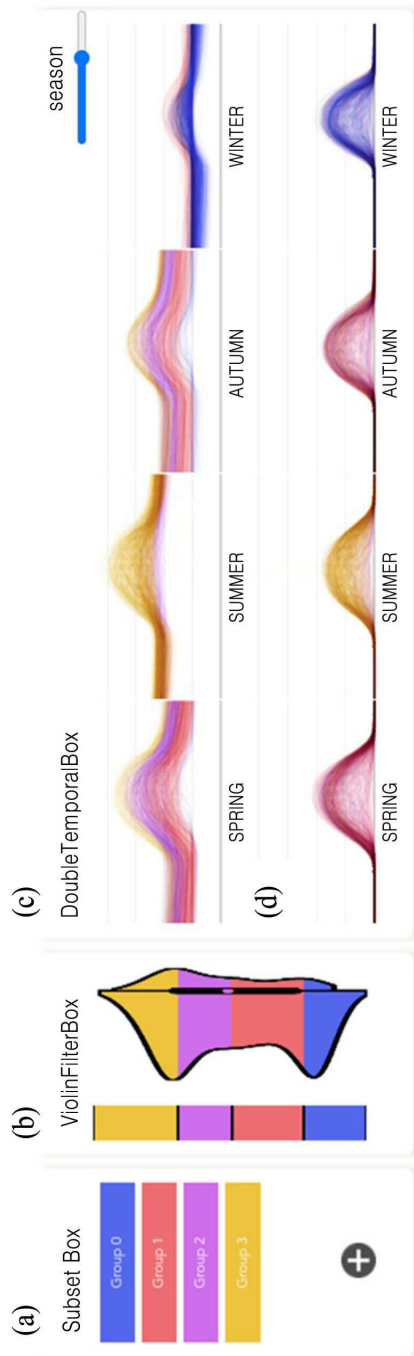
도면5



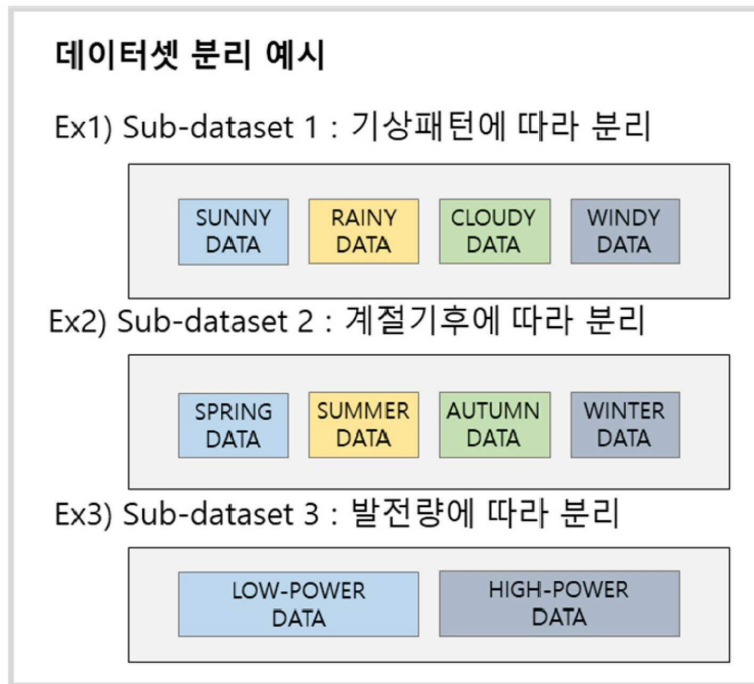
도면6



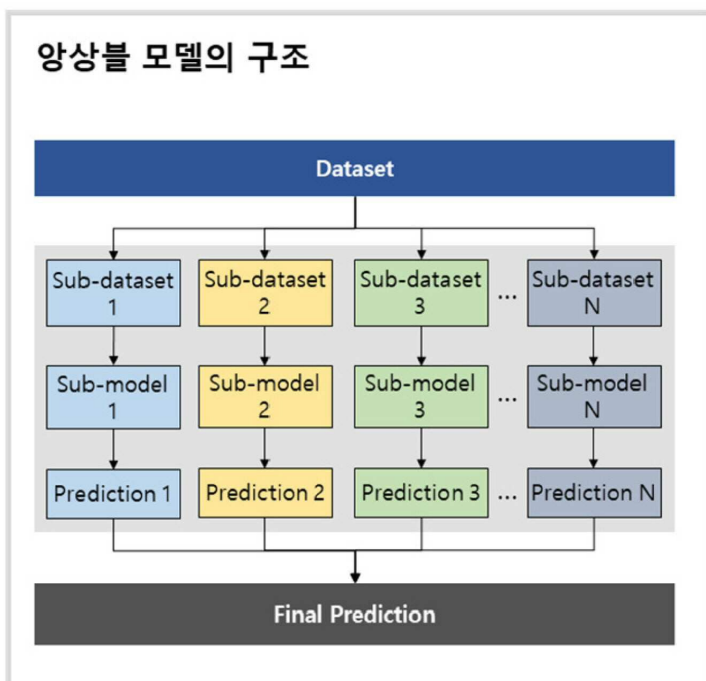
도면7



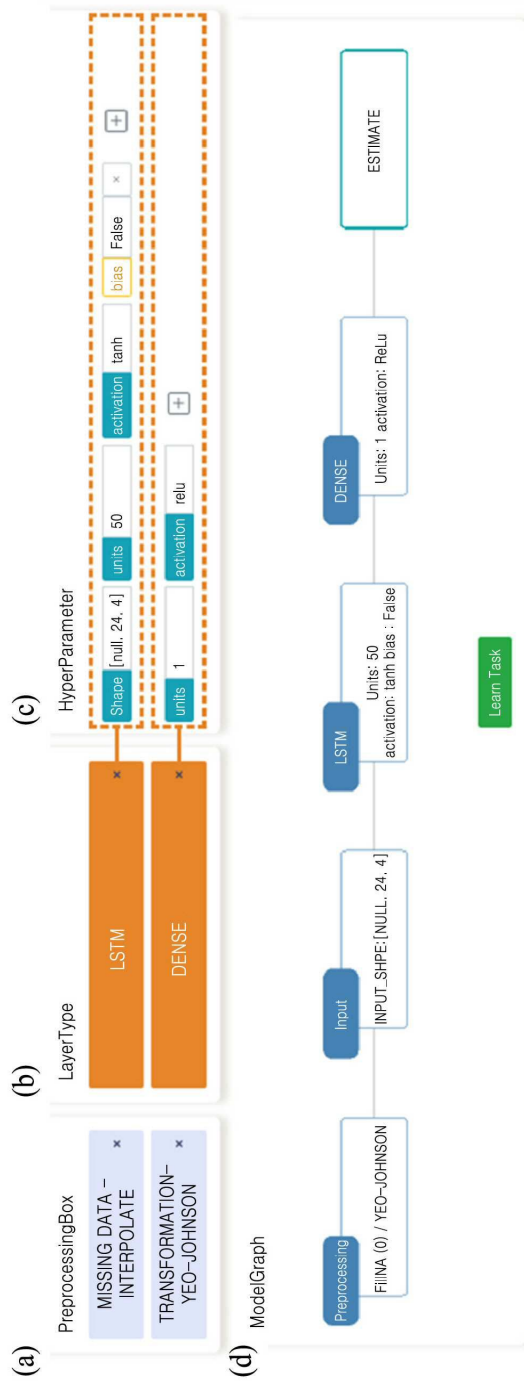
도면8



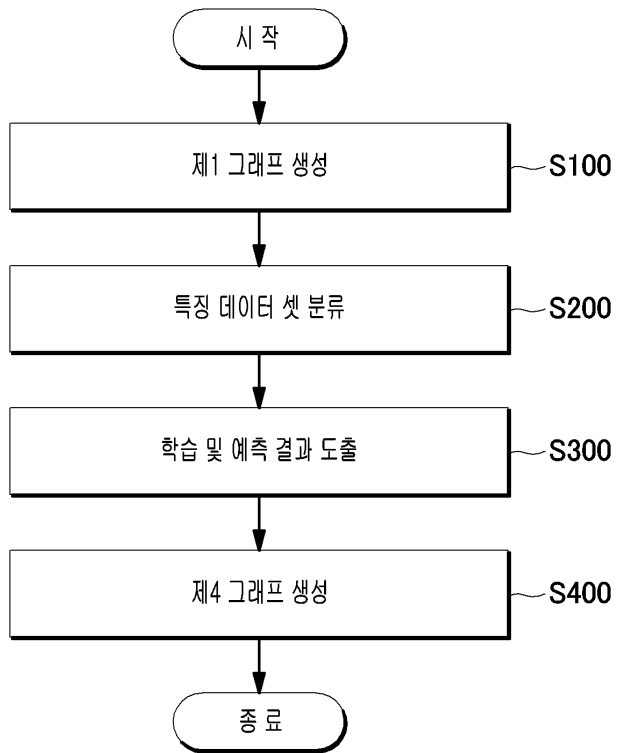
도면9



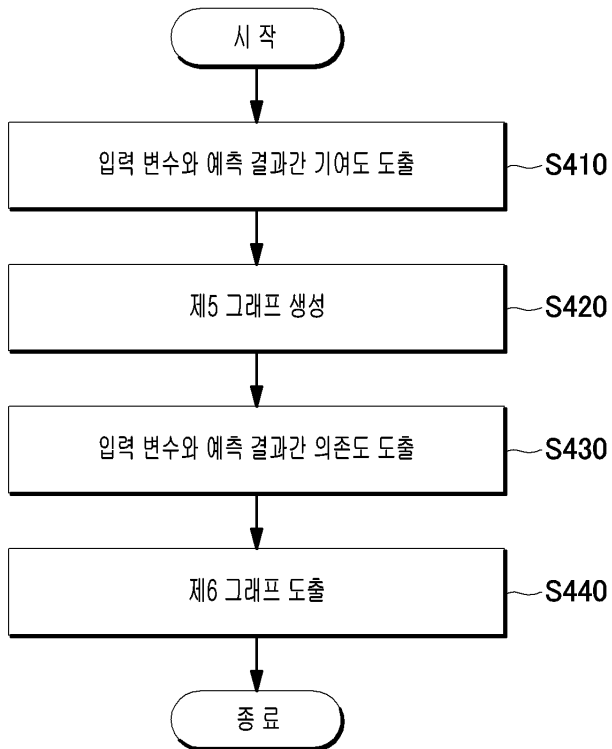
도면10



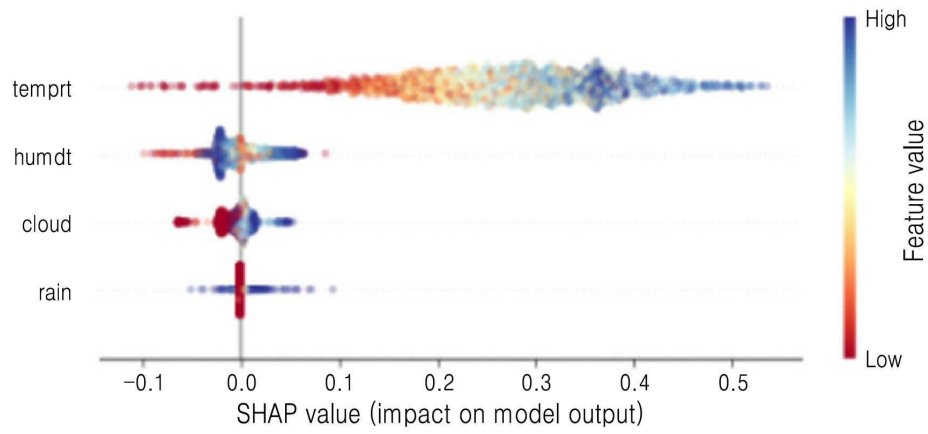
도면11



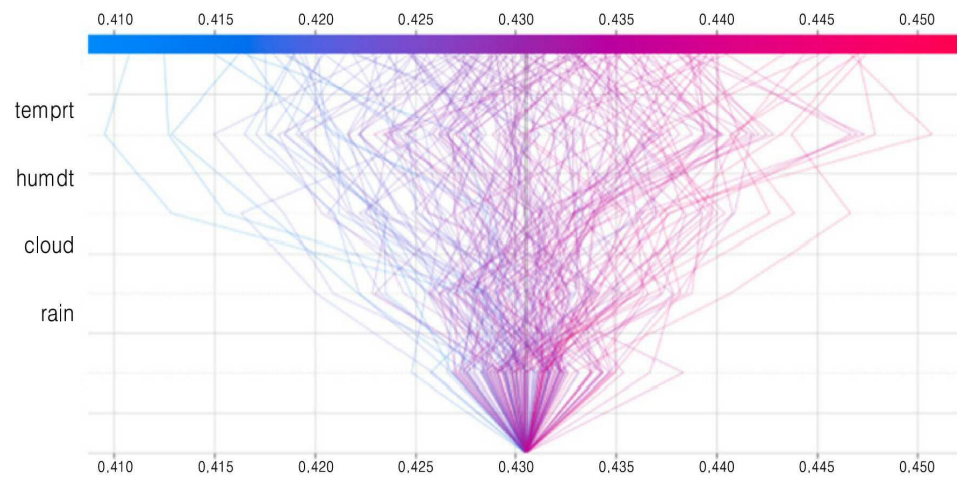
도면12



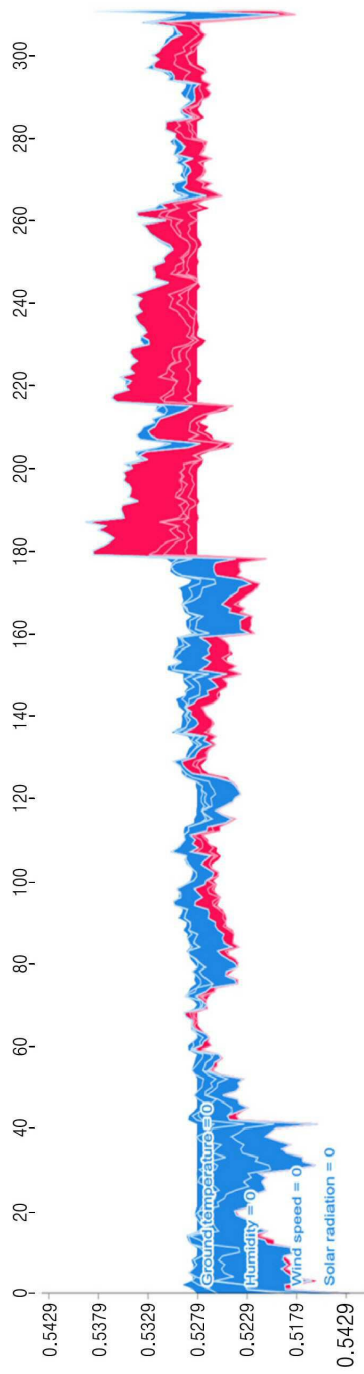
도면13



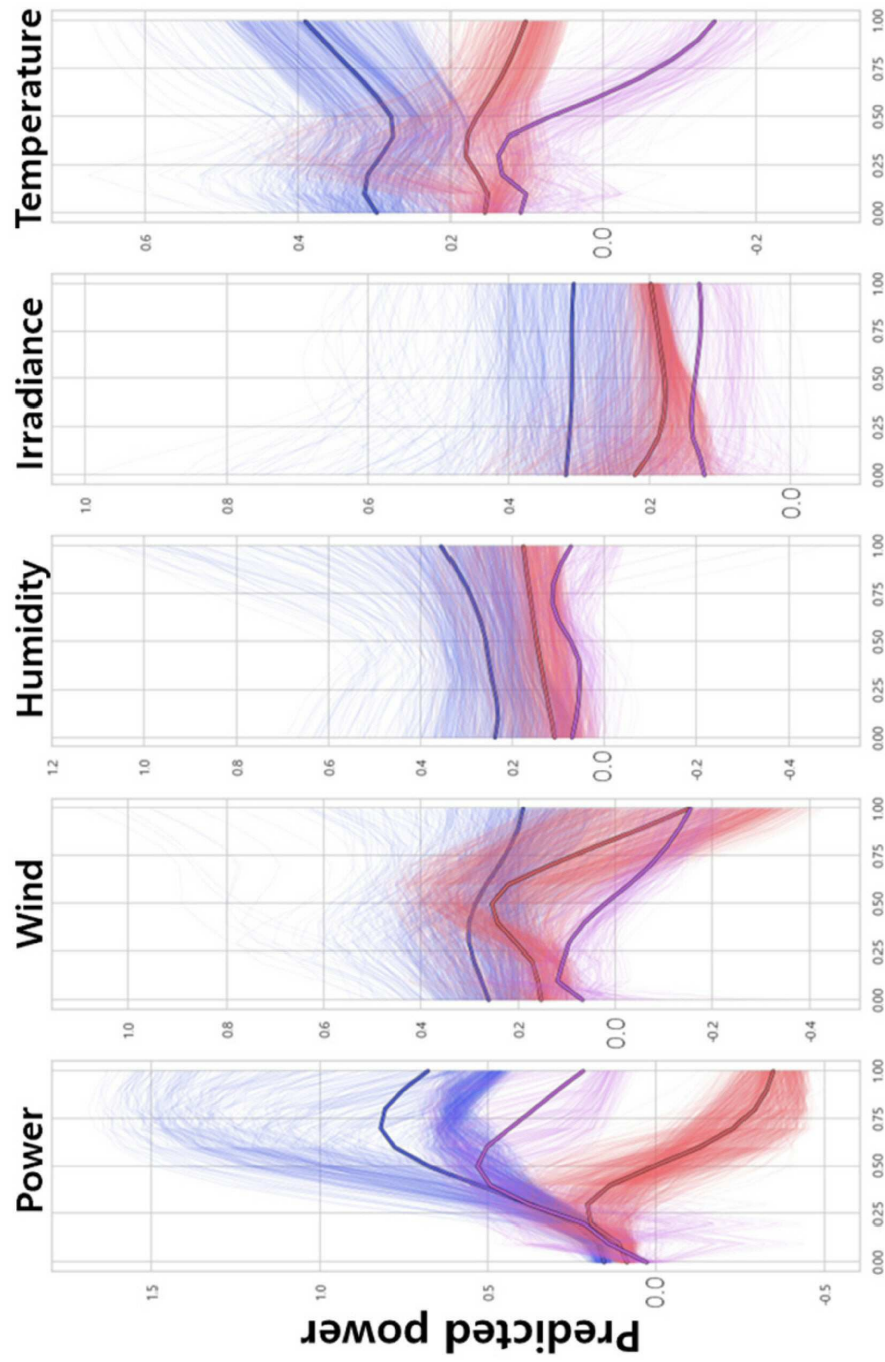
도면14



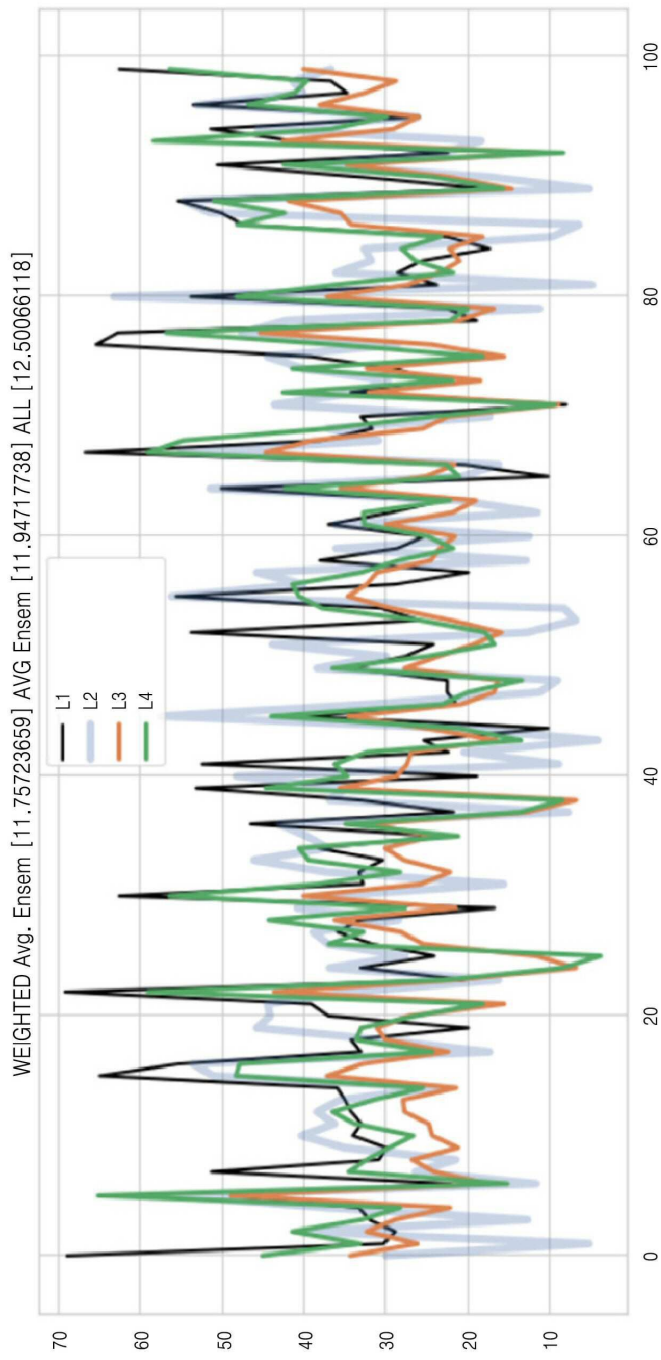
도면15



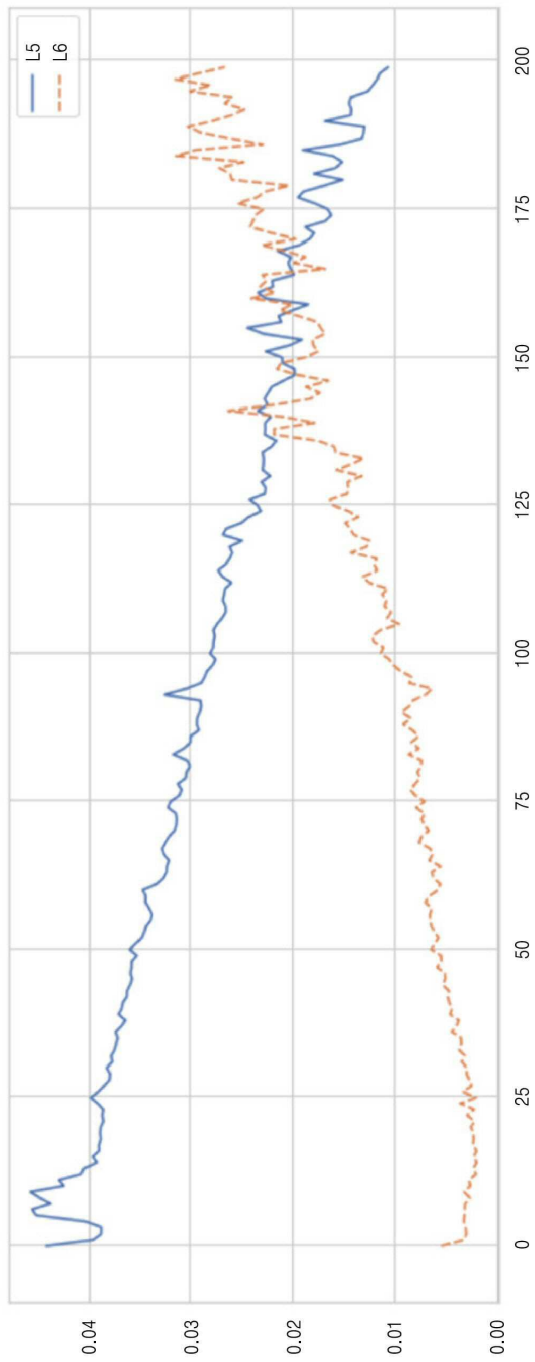
도면16



도면17



도면18



도면19

