



등록특허 10-2685543



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2024년07월16일

(11) 등록번호 10-2685543

(24) 등록일자 2024년07월11일

(51) 국제특허분류(Int. Cl.)
G06Q 50/06 (2024.01) *G06F 17/16* (2006.01)
G06F 17/17 (2006.01) *G06N 7/00* (2023.01)
H02J 3/00 (2006.01)
(52) CPC특허분류
G06Q 50/06 (2024.01)
G06F 17/16 (2013.01)
(21) 출원번호 10-2021-0158737
(22) 출원일자 2021년11월17일
심사청구일자 2021년11월17일
(65) 공개번호 10-2023-0072219
(43) 공개일자 2023년05월24일
(56) 선행기술조사문헌
JP2019211922 A*
(뒷면에 계속)

(73) 특허권자
주식회사 에이치에너지
경상북도 포항시 남구 청암로 87, 609호(지곡동, 체인지업그라운드)
포항공과대학교 산학협력단
경상북도 포항시 남구 청암로 77 (지곡동)
(72) 발명자
함일한
서울특별시 마포구 월드컵북로 501, 915동 603호
(상암동, 상암월드컵파크9단지)
임성빈
서울특별시 서대문구 신촌로1길 29, 601호(창천동, 레디앙아파트)
(뒷면에 계속)
(74) 대리인
김명진

전체 청구항 수 : 총 11 항

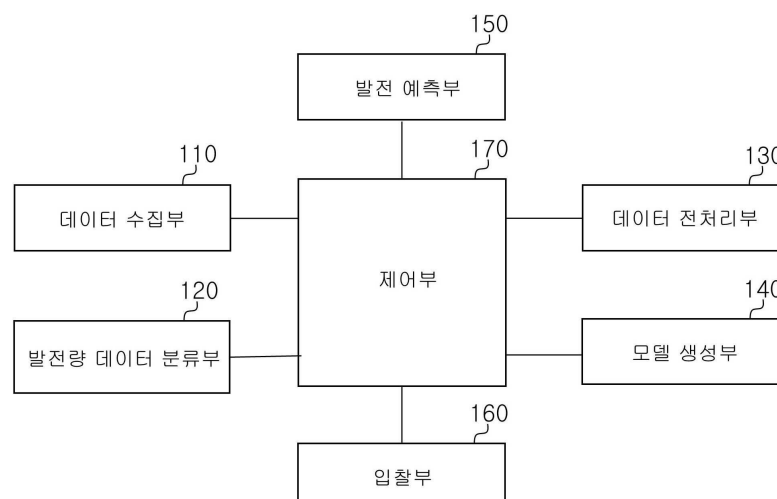
심사관 : 유선중

(54) 발명의 명칭 날씨 예측 데이터를 이용한 확률적 시나리오 기법의 최적 입찰량 결정 시스템

(57) 요약

본 발명에 따른 날씨 예측 데이터를 이용한 확률적 시나리오 기법의 최적 입찰량 결정 시스템은, 기상예측 데이터 및 태양광 발전량 데이터를 수집하는 데이터 수집부; 수집된 기상예측 데이터를 리스케일링하고 차원 축소 알고리즘을 통해 기상예측 데이터의 차원을 축소한 후, 차원 축소된 기상예측 데이터와 합해진 발전량 데이터에 대해 2차 리스케일링을 수행한 뒤에 차원 축소된 기상예측 데이터와 합하는 데이터 전처리부; 리스케일링된 기상예측 데이터 및 발전량 데이터를 EM알고리즘으로 가우시안 믹스처 모델 피팅을 수행하고 혼합기상모델을 생성하는 모델 생성부; 혼합기상모델에 의해 학습된 가우시안 믹스처 모델의 파라미터를 이용하여 발전량을 예측하는 발전 예측부; 및 혼합기상모델에 기초하여 최적 입찰량을 산출하는 입찰부;를 포함할 수 있다.

대표도 - 도1



(52) CPC특허분류

G06F 17/17 (2013.01)
G06N 7/01 (2023.01)
H02J 3/004 (2023.08)
H02J 3/008 (2023.08)
Y04S 50/10 (2013.01)

(72) 발명자

김병철

서울특별시 종로구 난계로 29길 33-4, 1208호 (승인동, 종로UK201오피스텔)

지승배

서울특별시 강남구 논현로115길 18-8 (논현동)

고영명

경상북도 포항시 남구 지곡로 155, 5동 1404호 (지곡동)

김민석

경상북도 포항시 남구 효자동길5번길 23, 308호 (효자동, 포스빌)

(56) 선행기술조사문헌

KR1020210115955 A*
 KR1020210132446 A*
 Jinxia Zhang 외 2명, “Solar Power Generation Forecast Based on LSTM”, 2018.11.25.*
 KR1020210123594 A
 KR1020210077474 A
 KR1020210071583 A
 KR1020210056667 A
 KR101856320 B1
 KR102283487 B1
 JP2011200040 A
 *는 심사관에 의하여 인용된 문헌

이 발명을 지원한 국가연구개발사업

과제고유번호	1415169187
과제번호	20209210200180
부처명	산업통상자원부
과제관리(전문)기관명	한국에너지기술평가원
연구사업명	에너지기술수용성제고및사업화촉진(R&D)
연구과제명	공유옥상 및 전력020 혁신을 통한 시민가상발전소 성장촉진
기 여 율	1/1
과제수행기관명	(주)에이치에너지
연구기간	2020.06.01 ~ 2021.02.28

명세서

청구범위

청구항 1

기상예측 데이터 및 태양광 발전량 데이터를 수집하는 데이터 수집부;

수집된 기상예측 데이터를 리스케일링하고 차원 축소 알고리즘을 통해 상기 기상예측 데이터의 차원을 축소한 후, 상기 발전량 데이터에 대해 2차 리스케일링을 수행하고 2차 리스케일링된 발전량 데이터를 차원 축소된 기상예측 데이터와 병합하는 데이터 전처리부;

리스케일링된 기상예측 데이터 및 발전량 데이터를 EM(Expectation-Maximization) 알고리즘으로 가우시안 믹스처 모델 피팅을 수행하고 혼합기상모델을 생성하는 모델 생성부;

상기 혼합기상모델에 의해 학습된 가우시안 믹스처 모델의 파라미터를 이용하여 발전량을 예측하는 발전 예측부; 및

상기 혼합기상모델에 기초하여 최적 입찰량을 산출하는 입찰부;를 포함하고,

상기 데이터 전처리부는 민맥스 스케일러(MinMax Scaler)를 이용하여 기수집된 복수의 발전량의 발전단위가 0과 1 사이의 값이 되도록 리스케일링하고,

로버스트 스케일러(Robust Scaler)를 이용하여 지역과 시간을 조건으로 상기 기수집된 복수의 기상예측 데이터를 리스케일링하는, 날씨 예측 데이터를 이용한 확률적 시나리오 기법의 최적 입찰량 결정 시스템.

청구항 2

제1항에 있어서,

수집된 상기 기상예측 데이터 및 태양광 발전량을 기초로 통합 예측할 지역을 결정하는 발전량 데이터 분류부를 포함하는, 날씨 예측 데이터를 이용한 확률적 시나리오 기법의 최적 입찰량 결정 시스템.

청구항 3

제2항에 있어서,

상기 발전량 데이터 분류부는 클러스터링 기법을 기반으로 위도값 및 경도값을 이용하여 동일 지역의 관측소를 선정하며, 설정된 지역의 개수는 상기 혼합기상모델의 하이퍼파라미터로 사용하는, 날씨 예측 데이터를 이용한 확률적 시나리오 기법의 최적 입찰량 결정 시스템.

청구항 4

삭제

청구항 5

제1항에 있어서,

상기 모델 생성부는,

상기 발전량 데이터로 상기 가우시안 믹스처 모델의 파라미터를 초기화한 후, 상기 EM알고리즘의 E스텝 및 M스텝을 순차적으로 수행하여 상기 가우시안 믹스처 모델을 고도화하는, 날씨 예측 데이터를 이용한 확률적 시나리오 기법의 최적 입찰량 결정 시스템.

청구항 6

제5항에 있어서,

상기 E스텝은 잠재변수에 대한 기댓값을 산출하고, 상기 M스텝은 상기 잠재변수에 대한 기댓값을 최대화하는 π ,

μ , Σ 를 산출하고,

상기 π 는 가우시안 분포가 선택될 확률, 상기 μ 는 가우시안 분포의 평균벡터, 상기 Σ 는 가우시안 분포의 공분산 행렬을 나타내는, 날씨 예측 데이터를 이용한 확률적 시나리오 기법의 최적 입찰량 결정 시스템.

청구항 7

제5항에 있어서,

상기 E스텝 및 M스텝은 하기 수학적식을 만족할 때까지 반복되는, 날씨 예측 데이터를 이용한 확률적 시나리오 기법의 최적 입찰량 결정 시스템.

(수학적식)

$$\Delta L < \epsilon_{EM}$$

(여기서, L 은 최대로그우도, ϵ_{EM} 는 EM알고리즘에서의 오차범위(tolerance)를 나타냄)

청구항 8

제3항에 있어서,

상기 모델 생성부는 상기 혼합기상모델의 평균 및 표준편차를 기초로 상기 하이퍼파라미터 별로 태양광 발전량 시나리오를 생성하고,

상기 하이퍼파라미터는 PCA차원(p), 지역(r), 라쏘 정규화 하이퍼파라미터(Lasso regularization hyperparameter)(λ , λ) 및 GMM클러스터 개수(c)를 포함하는, 날씨 예측 데이터를 이용한 확률적 시나리오 기법의 최적 입찰량 결정 시스템.

청구항 9

제8항에 있어서,

상기 태양광 발전량 시나리오는 트리 형태인, 날씨 예측 데이터를 이용한 확률적 시나리오 기법의 최적 입찰량 결정 시스템.

청구항 10

제1항에 있어서,

상기 발전 예측부는 L1 error KPI 및 L2 error KPI를 에너지표로 사용하여 발전량 분포 예측값의 정확도를 판단하는, 날씨 예측 데이터를 이용한 확률적 시나리오 기법의 최적 입찰량 결정 시스템.

청구항 11

제10항에 있어서,

상기 L1 error KPI 및 L2 error KPI의 값이 적을수록 상기 발전량 분포 예측값 또는 발전량 예측 구간의 정확도가 높은, 날씨 예측 데이터를 이용한 확률적 시나리오 기법의 최적 입찰량 결정 시스템.

청구항 12

데이터 수집부를 통해 기상예측 데이터 및 태양광 발전량 데이터를 수집하는 단계;

데이터 전처리부를 통해 수집된 기상예측 데이터를 리스케일링하고 차원 축소 알고리즘을 통해 상기 기상예측 데이터의 차원을 축소한 후, 상기 발전량 데이터를 2차 리스케일링하고 2차 리스케일링된 발전량 데이터를 차원 축소된 기상예측 데이터와 병합하는 단계;

모델 생성부를 통해 리스케일링된 기상예측 데이터 및 발전량 데이터가 EM(Expectation-Maximization) 알고리즘을 통해 가우시안 믹스처 모델 피팅을 수행하고 혼합기상모델을 생성하는 단계;

발전 예측부를 통해 상기 혼합기상모델에 의해 학습된 가우시안 믹스처 모델의 파라미터를 이용하여 발전량을

예측하는 단계; 및

입찰부를 통해 상기 혼합기상모델에 기초하여 최적 입찰량을 산출하는 단계를 포함하고,

상기 데이터 전처리부는 민맥스 스케일러(MinMax Scaler)를 이용하여 기수집된 복수의 발전량의 발전단위가 0과 1 사이의 값이 되도록 리스케일링하고,

로버스트 스케일러(Robust Scaler)를 이용하여 지역과 시간을 조건으로 상기 기수집된 복수의 기상예측 데이터를 리스케일링하는, 날씨 예측 데이터를 이용한 확률적 시나리오 기법의 최적 입찰량 결정 방법.

발명의 설명

기술 분야

- [0001] 본 발명은 날씨 예측 데이터를 이용한 확률적 시나리오 기법의 최적 입찰량 결정 시스템에 관한 것으로서, 보다 상세하게는 날씨 데이터를 활용하여 발전 불확실성을 최소화함으로써 발전 예측 시스템의 신뢰도를 향상시킬 수 있는 태양광 발전 시나리오에 따른 최적 입찰량 결정 시스템에 관한 것이다.

배경 기술

- [0002] 전 세계적으로 전기에너지 소비량은 지속적으로 증가하고 있다. 그러나, 화석연료를 기반으로 한 에너지 자원의 양은 한정되어 있어, 태양광발전, 풍력발전과 같은 재생에너지 발전원의 전력생산 비중을 증가시키려는 추세를 보이고 있다.
- [0003] 이러한 재생에너지 발전을 활용하여 전기에너지를 생산할 경우 온실가스와 미세먼지 등의 환경오염 물질을 감소시킬 수 있을 뿐만 아니라 지속가능한 미래에너지의 관점에서 에너지 공급 안정성을 증대시킬 수 있어 그 중요성이 날로 커지고 있다.
- [0004] 그러나, 재생에너지를 이용한 발전 시스템의 출력전력(이하, 발전량)은 눈, 비 등과 같은 외부적 요인의 영향으로 하루 종일 변화하고 계절적인 요인에 의해서도 많은 영향을 받는다.
- [0005] 이처럼 변동성이 큰 재생에너지를 이용한 발전 시스템으로 인한 불확실성을 최소화하고, 전력계통의 신뢰성을 향상시키며, 전력품질을 향상시키기 위하여 재생에너지 발전량 예측제도가 도입되었다. 이는 20mW 이상의 태양광, 풍력을 비롯한 재생에너지 예상 발전량을 1일 전 미리 제출하고 당일 날 일정 기준 오차율 내 발전량이 충족하면 정산금을 지급하는 방식으로 진행된다.
- [0006] 이에 따라, 거래 당일 실제 발전량(실제 값)과 1일 전 제출한 예상 발전량(예측 값)이 비슷하면 발전사업자가 수익을 얻게 된다.
- [0007] 따라서, 발전사업자의 최적의 수익(이하, 입찰량)을 위해서는 발전량 예측을 위해 활용되는 기상예측 데이터의 복잡성 등을 해결하여 보다 정확한 재생에너지의 발전량 예측이 가능한 입찰량 결정 시스템의 개발이 필요하다.

발명의 내용

해결하려는 과제

- [0008] 본 발명이 해결하고자 하는 과제는 다양한 관측소로부터 수집되는 방대한 양의 기상예측 데이터의 단위 및 범위가 서로 상이하기 때문에, 이들을 일정 범위 내로 통일시킬 수 있도록 리스케일링하고 차원 축소 알고리즘을 이용하여 최적 입찰 시나리오 생성을 위한 학습 모델에 사용되는 피처의 개수를 감소시켜 모델 학습의 과적합 문제를 방지할 수 있는 최적 입찰량 결정 시스템을 제공하는 것이다.
- [0009] 또한, 본 발명이 해결하고자 하는 과제는 리스케일링하고 차원 축소된 기상예측 데이터에 태양광 발전량을 합하여 추가적인 리스케일링 과정을 진행함으로써 모델 파라미터를 최적화할 수 있다.
- [0010] 또한, 본 발명이 해결하고자 하는 과제는 리스케일링된 태양광 발전량으로 모델의 하이퍼파라미터와 관련있는 인덱스 값을 찾은 뒤, 각 지역별로 리스케일링 되고 차원 축소된 기상예측 데이터를 합한 데이터의 인덱스를 하이퍼파라미터 탐색에 사용함으로써 발전예측 값의 불확실성을 최소화할 수 있다.
- [0011] 또한, 본 발명이 해결하고자 하는 과제는 예측 구간의 정확도까지 판단할 수 있는 에러지표를 적용함으로써 학습 모델을 이용한 발전량 분포 예측 시 예측의 정확도뿐만 아니라 발전량 예측 구간의 정확도까지 높일 수

있다. 따라서, 정확도 높은 태양광 발전량 예측값 및 예측 구간을 기반으로 예측을 수행하여 최적의 입찰량을 결정할 수 있는 최적 입찰량 결정 시스템을 제공하는 것이다.

과제의 해결 수단

- [0012] 전술한 바와 같은 과제를 해결하기 위하여 본 발명에 따른 날씨 예측 데이터를 이용한 확률적 시나리오 기법의 최적 입찰량 결정 시스템은, 기상예측 데이터 및 태양광 발전량 데이터를 수집하는 데이터 수집부; 수집된 기상예측 데이터를 리스케일링하고 차원 축소 알고리즘을 통해 상기 기상예측 데이터의 차원을 축소한 후, 차원 축소된 기상예측 데이터와 합해진 상기 발전량 데이터에 대해 2차 리스케일링을 수행하고 2차 리스케일링된 발전량 데이터를 차원 축소된 기상예측 데이터와 병합하는 데이터 전처리부; 리스케일링된 기상예측 데이터 및 발전량 데이터를 EM알고리즘으로 가우시안 믹스처 모델 피팅을 수행하고 혼합기상모델을 생성하는 모델 생성부; 상기 혼합기상모델에 의해 학습된 가우시안 믹스처 모델의 파라미터를 이용하여 발전량을 예측하는 발전 예측부; 및 상기 혼합기상모델에 기초하여 최적 입찰량을 산출하는 입찰부;를 포함할 수 있다.
- [0013] 또한, 수집된 상기 기상예측 데이터 및 태양광 발전량을 기초로 통합 예측할 지역을 결정하는 발전량 데이터 분류부를 포함할 수 있다.
- [0014] 또한, 상기 발전량 데이터 분류부는, 상기 발전량 데이터 분류부는 위도값 및 경도값의 유사도가 높은 관측소를 동일 지역으로 선정하며, 설정된 지역의 개수는 상기 혼합기상모델의 하이퍼파라미터로 사용하고, 상기 유사도는 클러스터링 기법을 기반으로 측정될 수 있다.
- [0015] 또한, 상기 데이터 전처리부는 민맥스 스케일러(MinMax Scaler)를 이용하여 기수집된 복수의 발전량의 발전 단위가 0과 1 사이의 값으로 균일해지도록 리스케일링하고, 로버스트 스케일러(Robust Scaler)를 이용하여 지역과 시간을 조건으로 상기 기수집된 복수의 기상예측 데이터를 리스케일링할 수 있다.
- [0016] 또한, 상기 모델 생성부는, 상기 발전량 데이터로 상기 가우시안 믹스처 모델의 파라미터를 초기화한 후, 상기 EM알고리즘의 E스텝 및 M스텝을 순차적으로 수행하여 상기 가우시안 믹스처 모델을 고도화할 수 있다.
- [0017] 또한, 상기 E스텝은 잠재변수에 대한 기댓값을 산출하고, 상기 M스텝은 상기 잠재변수에 대한 기댓값을 최대화하는 π , μ , Σ 를 산출하고, 상기 π 는 가우시안 분포가 선택될 확률, 상기 μ 는 가우시안 분포의 평균벡터, 상기 Σ 는 가우시안 분포의 공분산 행렬을 나타낼 수 있다.
- [0018] 또한, 상기 E스텝 및 M스텝은 하기 수학적식을 만족할 때까지 반복될 수 있다. (수학식) $\Delta L < \epsilon_{EM}$ (여기서, L 은 최대로그우도, ϵ_{EM} 는 EM알고리즘에서의 오차범위(tolerance)를 나타냄)
- [0019] 또한, 상기 모델 생성부는 상기 혼합기상모델의 평균 및 표준편차를 기초로 하이퍼파라미터 별로 태양광 발전량 시나리오를 생성하고, 상기 하이퍼파라미터는 PCA차원(p), 지역(r), 라쏘 정규화 하이퍼파라미터(Lasso regularization hyperparameter)(lambda, λ) 및 GMM클러스터 개수(c)를 포함할 수 있다.
- [0020] 또한, 상기 태양광 발전량 시나리오는 트리 형태일 수 있다.
- [0021] 또한, 상기 발전 예측부는 L1 error KPI 및 L2 error KPI를 에너지표로 사용하여 발전량 분포 예측값의 정확도를 판단할 수 있다.
- [0022] 또한, 상기 L1 error KPI 및 L2 error KPI의 값이 적을수록 상기 발전량의 예측값의 정확도가 높을 수 있다.
- [0023] 전술한 바와 같은 과제를 해결하기 위하여 본 발명에 따른 날씨 예측 데이터를 이용한 확률적 시나리오 기법의 최적 입찰량 결정 방법은, 데이터 수집부를 통해 기상예측 데이터 및 태양광 발전량 데이터를 수집하는 단계; 데이터 전처리부를 통해 수집된 기상예측 데이터를 리스케일링하고 차원 축소 알고리즘을 통해 상기 기상예측 데이터의 차원을 축소한 후, 상기 발전량 데이터를 2차 리스케일링하여 차원 축소된 기상예측 데이터와 합 하는 단계; 모델 생성부를 통해 리스케일링된 기상예측 데이터 및 발전량 데이터가 가우시안 믹스처 모델이라는 가정 하에서 EM알고리즘으로 가우시안 믹스처 모델 피팅을 수행하고 혼합기상모델을 생성하는 단계; 발전 예측부를 통해 상기 혼합기상모델에 의해 학습된 가우시안 믹스처 모델의 파라미터를 이용하여 발전량을 예측하는 단계; 및 입찰부를 통해 상기 혼합기상모델에 기초하여 최적 입찰량을 산출하는 단계를 포함할 수 있다.
- [0024] 기타 실시예의 구체적인 사항들은 상세한 설명 및 도면들에 포함되어 있다.

발명의 효과

- [0025] 본 발명의 실시예에 따른 최적 입찰량 결정 시스템은, 다양한 관측소로부터 수집되는 방대한 양의 기상예측 데이터를 리스케일링하여 단위 및 범위가 일정 범위 이내로 통일되도록 하고 차원 축소 알고리즘을 통해 최적 입찰 시나리오 생성을 위한 학습 모델에 사용되는 피처의 개수를 감소시켜 모델 학습의 복잡성 또는 과적합 등의 문제를 해결할 수 있다.
- [0026] 또한, 본 발명의 실시예에 따른 최적 입찰량 결정 시스템은, 리스케일링하고 차원 축소된 기상예측 데이터에 태양광 발전량을 합하여 추가적인 리스케일링 과정을 진행함으로써 발전 예측 값의 불확실성을 최소화할 수 있다.
- [0027] 또한, 본 발명의 실시예에 따른 최적 입찰량 결정 시스템은, 예측 구간의 정확도까지 판단할 수 있는 에너지표를 적용함으로써 학습 모델을 이용한 발전량 분포 예측 시 예측의 정확도뿐만 아니라 발전량 예측 구간의 정확도까지 높일 수 있다.
- [0028] 즉, 리스케일링된 태양광 발전량으로 모델의 하이퍼파라미터와 관련있는 인덱스 값을 찾은 뒤, 각 지역별로 리스케일링 되고 차원 축소된 기상예측 데이터를 합한 데이터의 인덱스를 하이퍼파라미터 탐색에 사용함으로써 발전예측 값의 불확실성을 최소화할 수 있다.
- [0029] 따라서, 정확도 높은 태양광 발전량 예측값 및 예측 구간을 기반으로 예측을 수행하여 최적의 입찰량을 산출함에 따라 사용자의 이익을 극대화할 수 있다.
- [0030] 본 발명에 따른 효과는 이상에서 예시된 내용에 의해 제한되지 않으며, 더욱 다양한 효과들이 본 명세서 내에 포함되어 있다.

도면의 간단한 설명

- [0031] 도 1은 본 발명의 실시예에 따른 날씨 예측 데이터를 이용한 확률적 시나리오 기법의 최적 입찰량 결정 시스템의 블록도이다.
- 도 2는 본 발명의 실시예에 따른 최적 입찰량 결정 방법을 설명하기 위한 순서도이다.
- 도 3은 본 발명의 실시예에 따른 가우시안 믹스처 모델의 피팅 단계를 설명하기 위한 순서도이다.
- 도 4는 본 발명의 실시예에 따라 발전 예측부가 클러스터 개수 및 기상예측 데이터의 PCA 컴포넌트 개수와 기상예측 데이터가 묶이는 지역의 개수와 라쏘 정규화 하이퍼파라미터(Lasso regularization hyperparameter)(λ , λ)에 따른 예측 값과 예측 구간을 산출한 그래프 예시도이다.
- 도 5는 본 발명의 실시예에 따른 발전량 분포 예측을 기반으로 한 태양광 발전량 시나리오의 예시도이다.
- 도 6은 본 발명의 실시예에 따라 수집된 기상예측 데이터의 예시도이다.
- 도 7은 본 발명의 실시예에 따라 리스케일링된 기상예측 데이터의 예시도이다.
- 도 8은 본 발명의 실시예에 따라 차원이 축소된 기상예측 데이터의 예시도이다.
- 도 9 내지 도 10은 본 발명의 비교예에 따른 에너지표를 이용하여 하이퍼파라미터 별로 생성된 시나리오의 편차를 나타낸 예시도이다.
- 도 11 내지 도 12는 본 발명의 실시예에 따른 에너지표를 이용하여 하이퍼파라미터 별로 생성된 시나리오의 편차를 나타낸 예시도이다.

발명을 실시하기 위한 구체적인 내용

- [0032] 이하, 첨부된 도면들을 참조하여 본 발명의 실시예를 상세히 설명하면 다음과 같다. 그러나 본 발명은 이하에서 개시되는 실시예에 한정되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있는 것으로, 이하의 실시예는 본 발명의 개시가 완전하도록 하며, 통상의 지식을 가진 자에게 발명의 범주를 완전하게 알려주기 위해 제공되는 것이다. 또한, 설명의 편의를 위하여 도면에서는 구성 요소들이 그 크기가 과장 또는 축소될 수 있다.
- [0033] 그러나, 이하의 실시예는 이 기술분야에서 통상적인 지식을 가진 자에게 본 발명이 충분히 이해되도록 제공되는 것으로서 여러 가지 다른 형태로 변형될 수 있으며, 본 발명의 범위가 다음에 기술되는 실시예에 한정되는 것은 아니다.

- [0034] 한편, 명세서 전체에서, 어떤 부분이 어떤 구성 요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성 요소를 제외하는 것이 아니라 다른 구성 요소를 더 포함할 수 있는 것을 의미한다.
- [0035] 상술한 목적, 특징 및 장점은 첨부된 도면과 관련한 다음의 상세한 설명을 통하여 보다 분명해질 것이며, 그에 따라 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 발명의 기술적 사상을 용이하게 실시할 수 있을 것이다.
- [0036] 본 명세서에서, '파라미터'는 최적 입찰량을 결정할 수 있는 확률적 시나리오를 위한 학습 모델(이하, '혼합기상모델'이라고도 함) 설정을 위해 사용되는 매개 변수를 의미한다. 다시 말해, 수집되는 기상예측 데이터(이하, '날씨예측 데이터' 또는 '관측소 데이터'라고도 함), 발전량 데이터로부터 추정된 값을 의미한다.
- [0037] 본 명세서에서, '하이퍼파라미터(Hyper parameter)'는 모델링을 위해 설정해주는 값을 의미한다. 본 발명에서는 지역(Region), PCA 차원(PCA component, p), GMM 클러스터 개수(Cluster, c), Lambda(Lasso regularization hyperparameter, λ)가 하이퍼파라미터로서 사용된다.
- [0038] 본 명세서에서, '가우시안 믹처 모델(Gaussian Mixture Model)'은 가우시안 모델의 혼합 형태를 의미할 수 있고, GMM 모델이라고 지칭될 수 있다.
- [0039] 본 명세서에서, '리스케일링(Rescaling)'은 데이터 값의 범위를 조정하는 것을 의미한다. 이때, 스케일링을 수행하기 위해서는 다양한 스케일러(Scaler)가 사용될 수 있다. 본 발명은 다양한 스케일러 중, 데이터가 0과 1 사이에 위치하도록 스케일링 하는 민맥스 스케일러(MinMax Scaler) 또는 데이터의 중앙값이 0, IQE(interquartile range)가 1이 되도록 스케일링 하여 모든 피처(Feature)들이 동일 스케일을 갖게 되는 로버스트 스케일러(Robust Scaler)를 사용한다. 다만, 실시예에 따라 다양한 스케일러가 사용될 수 있다.
- [0040] 도 1은 본 발명의 실시예에 따른 날씨 예측 데이터로 발전 불확실성을 통해 확률적 시나리오 기법의 최적 입찰량 결정 기법 시스템의 블록도이다.
- [0041] 도 1을 참조하면, 날씨 예측 데이터로 발전 불확실성을 통해 확률적 시나리오 기법의 최적 입찰량 결정 기법 시스템은 데이터 수집부(110), 발전량 데이터 분류부(120), 데이터 전처리부(130), 모델 생성부(140), 발전 예측부(150), 입찰부(160) 및 제어부(170)를 포함할 수 있다.
- [0042] 데이터 수집부(110)는 발전량 및 기상예측 데이터를 수집하는 구성으로서, 태양광 발전량 데이터를 발전 계량기를 통해 일정 시간 간격으로 수집하고 이를 무선 통신으로 서버에 전송하여 데이터베이스에 저장할 수 있다.
- [0043] 구체적으로, 다양한 날씨 예보 서비스로부터 42시간 이상의 기상예측 데이터를 온라인으로 수집하여 데이터베이스에 저장할 수 있다. 여기서, 기상 데이터는 발전소 주변 최소 2개 이상의 지점에서 수집하며 최소 2개 이상의 날씨 예보 서비스를 이용하여 수집할 수 있다.
- [0044] 발전량 데이터 분류부(120)는 수집된 상기 기상예측 데이터 및 태양광 발전량을 기초로 통합 예측할 지역을 선정하는 구성이다. 구체적으로, 위도값 및 경도값의 유사도가 높은 관측소들을 동일 지역 내에 위치한 관측소라고 설정하고, 이들이 위치한 지역을 통합 예측할 지역으로 선정하는 것이다. 여기서, 유사도는 k-means 클러스터링 기법 등을 기반으로 측정된다. 또한, 발전량 데이터 분류부(120)에 의해 통합 예측할 지역으로 설정된 지역의 개수는 본 발명의 혼합기상모델의 하이퍼파라미터로 사용된다.
- [0045] 데이터 전처리부(130)는 수집된 발전량에 대해 누락데이터를 처리하고 시간대별 발전량을 합산하여 리스케일링(Rescaling)하는 구성이다. 구체적으로, 데이터 전처리부는 발전량 수집 간격이 등간격이 아닌 경우에 등간격의 발전 수집 간격으로 리샘플링(Resampling)을 수행할 수 있다.
- [0046] 또한, 데이터 전처리부(130)는 발전량 데이터 분류부(120)를 통해 분류된 지역들의 1시간 단위의 발전량을 합치고 오전 7시에서부터 오후 18시 사이의 핵심 시간대의 날씨 예측값들과 함께 데이터베이스에 저장할 수 있다.
- [0047] 또한, 데이터 전처리부(130)는 기상예측 데이터를 속성에 맞게 리스케일링(Rescaling)한 후 데이터베이스에 저장할 수 있다. 이때, 데이터 전처리부(130)는 MinMax Scaler 또는 Robust Scaler와 같은 스케일러(Scaler)를 이용하여 기상예측 데이터를 리스케일링한다.
- [0048] 이후, 데이터 전처리부(130)는 리스케일링된 기상예측 데이터의 차원 축소 단계를 수행한다. 보다 상세하게, 기상예측 데이터의 경우 성분이 많아지게 되면 차원이 높아져서 모델 학습 시 과적합을 유발하기 때문에, 데이터 전처리부(130)는 기상예측 데이터를 주성분 분석(Principal Component Analysis, PCA)을 통해 각 지역의 기상

예측 데이터의 피처(Feature)를 2-5개로 줄여 데이터 셋(Set)의 차원을 축소할 수 있다.

- [0049] 모델 생성부(140)는 데이터 전처리부(130)에서 전처리된 과거 데이터(기상예측 데이터 및 발전량 데이터)를 이용하여 기상예측 데이터와 발전량 데이터가 각 시간대별로 여러 개(K개)의 가우시안 모델의 혼합형태(Gaussian Mixture Model, 이하, '가우시안 믹스처 모델(GMM)'이라고도 함)로 있다고 가정하고, EM(Expectation-Maximization) 알고리즘을 통해 가우시안 믹스처 모델(GMM)의 파라미터를 산출한다. 보다 상세하게, 모델 생성부(140)는 EM알고리즘을 통해 리스케일링된 태양광 발전량으로 단변량 가우시안 믹스처 모델(Multivariate GMM)의 하이퍼파라미터와 관련있는 인덱스 값을 찾은 뒤, 각 지역별로 리스케일링하고 차원 축소된 기상예측 데이터와 리스케일링된 태양광 발전량 데이터가 합쳐진 데이터에 상기 인덱스를 부여한다. 다변량 가우시안 믹스처 모델피팅(GMM fitting)을 Covariance Lasso가 포함된 EM알고리즘으로 합쳐진 데이터로 모델피팅을 수행하고 혼합 기상모델을 생성한다.
- [0050] 또한, 모델 생성부(140)는 파라미터 중 공분산(Covariance)의 경우, 과적합 방지를 위해 Sparse precision graphical lasso 알고리즘을 통해 Sparse covariance estimation 방법론을 수행할 수 있다.
- [0051] 또한, 모델 생성부(140)는 후술할 발전 예측부에 의해 확률 모델의 평균과 표준편차를 이용하여 하이퍼파라미터(Hyper parameter)별로 태양광 발전량 시나리오를 생성한다. 또한, 생성된 태양광 발전량 시나리오를 기반으로 태양광 발전과 기상예측 데이터간의 관계를 학습하여 클러스터에 따른 혼합기상모델을 생성한다. 이에 대한 상세한 설명은 도 3 내지 도 5 등을 참조하여 후술하기로 한다.
- [0052] 발전 예측부(150)는 모델 생성부(140)로부터 생성된 혼합기상모델에 의해 학습된 GMM 파라미터를 이용하여 발전량을 예측한다. 구체적으로, 모델 생성부(140)에서 학습된 GMM 파라미터를 이용하여 기상예측 데이터가 주어졌을 때, 발전량의 분포를 산출하여 예측 발전량의 기댓값과 표준편차 예측을 수행할 수 있다. 보다 상세하게, 발전 예측부(150)는 기댓값과 표준편차를 통해 클러스터의 개수와 기상예측 데이터의 PCA 컴포넌트 개수와 기상예측 데이터가 묶이는 지역의 개수와 Lambda(Covariance Lasso Coefficient)에 따른 예측 값과 예측 구간(Prediction Interval)을 산출할 수 있다.
- [0053] 입찰부(160)는 태양광 발전량 예측 모형인 혼합기상모델을 기반으로 최적 입찰량을 산출하는 구성이다.
- [0054] 제어부(170)는 상술한 최적 입찰량 결정 시스템들의 구성들을 제어하는 구성이다.
- [0055] 이하에서는 도 2 내지 도 8을 참조하여 날씨 예측 데이터를 이용한 확률적 시나리오 기법의 최적 입찰량 결정 시스템의 최적 입찰량 결정 방법을 설명하기로 한다.
- [0056] 도 2는 본 발명의 실시예에 따른 최적 입찰량 결정 방법을 설명하기 위한 순서도이다. 도 3은 본 발명의 실시예에 따른 가우시안 믹스처 모델의 피팅 단계를 설명하기 위한 순서도이다. 도 4는 본 발명의 실시예에 따라 발전 예측부가 클러스터 개수 및 기상예측 데이터의 PCA 컴포넌트 개수와 기상예측 데이터가 묶이는 지역의 개수와 라쏘 정규화 하이퍼파라미터(Lasso regularization hyperparameter)(λ)에 따른 예측 값과 예측 구간을 산출한 그래프 예시도이다. 도 5는 본 발명의 실시예에 따른 발전량 분포 예측을 기반으로 한 태양광 발전량 시나리오의 예시도이다. 도 6은 본 발명의 실시예에 따라 수집된 기상예측 데이터의 예시도이다. 도 7은 본 발명의 실시예에 따라 리스케일링된 기상예측 데이터의 예시도이다. 도 8은 본 발명의 실시예에 따라 차원이 축소된 기상예측 데이터의 예시도이다.
- [0057] **최적 입찰량 결정 방법**
- [0058] **1. 데이터 수집 및 지역 선정 과정**
- [0059] 먼저, 데이터 수집부(110)를 통해 기상예측 데이터 및 태양광 발전량 데이터를 확보한다(S100 단계).
- [0060] 구체적으로, 데이터 수집부(110)는 날씨 예보 서비스로부터 최소 1시간 단위로 최소 9시간 후부터 42시간 이상의 기상예측 데이터를 수집하고, 발전 계량기로부터 15분 단위로 태양광 발전량 데이터를 수집할 수 있다. 이렇게 수집된 데이터들은 S200단계에서 발전량 데이터 분류부(120)가 통합 예측할 지역을 선정하는데 활용될 수 있다.
- [0061] 여기서, 기상예측 데이터는 국내 기상청 이외에도 오픈 웨더 맵(Open Weather Map) 또는 아큐 웨더(Accu Weather)와 같은 다양한 날씨 예보 서비스로부터 수집될 수 있다.
- [0062] 예를 들어, 이러한 날씨 예보 서비스로부터 수집되는 기상예측 데이터는 기상데이터의 예측시간(fcst_time), 관측시간(time), 기온(temperature(℃)), 최저기온(temperature_min), 최고기온(temperature_max), 체감온도

(real_fell_temperature), 습구온도(wet_bulb_temperature), 이슬점 온도(dew_point)), 습도(humidity(%)), 상대습도(relative_humidity), 최저운고(ceiling(m)), 자외선강도(uv_idx(UV)), 기압(groud_press(hPa)), 풍속(wind_speed(m/s)), 돌풍속도(wind-gust_speed), 풍향(wind_direction(degress)), 강우량(rain(mm)), 적설량(snow(mm)), 적빙량(ice(mm)) 등 중 적어도 하나의 파라미터를 포함할 수 있다.

[0063] 그러나, 날씨 예보 서비스별로 상술한 파라미터들을 측정하는 관측소의 위치가 상이할 수도 있고, 관측된 파라미터들(이하, '파라미터'라고도 함)의 계수가 상이할 수 있고, 날씨 예보 서비스 마다 관측하는 파라미터들의 종류/개수가 상이할 수 있고, 또는 각 파라미터들을 예측하는데 사용하는 알고리즘이 다를 수도 있다. 예컨대, 파라미터의 계수가 상이하다는 것은 서로 상이한 위치에 위치한 관측소 중 어느 하나는 강우량과 강수량을 동시에 표기하고, 다른 하나는 이를 각각 표시한다면 두 관측소로부터의 파라미터의 계수가 상이해질 수 있다는 것을 의미한다.

[0064] 이처럼, 날씨 예보 서비스로부터 수집되는 파라미터들의 단위 및 범위가 서로 상이하기 때문에 이를 일정 범위 이내로 통일시켜주는 과정이 수반되어야 한다. 이러한 과정은 데이터 전처리부(130)에 의해 수행되며 S300 단계 내지 S400 단계에서 후술하기로 한다.

[0065] 또한, 데이터 수집부(110)는 발전 계량기로부터 발전량 데이터를 일정 시간 간격으로 수집한다. 이때, 동일 지역 내에 포함되는 발전 계량기로부터 수집된 발전량 데이터는 합산하여, 해당 지역에 대한 태양광 발전량은 합산된 값으로 사용할 수 있다.

[0066] 또한, 데이터 수집부(110)는 지역별 날씨 예측값들을 합산한 값이 아닌 독립적인 특성 변수(feature)로 사용할 수 있다.

[0067] 이어서, 발전량 데이터 분류부(120)에 의해 수집된 기상예측 데이터와 태양광 발전량 데이터를 기초로 통합 예측할 지역을 선정한다(S200 단계).

[0068] 통합 예측할 지역은 발전량에 따른 기상예측 데이터를 많이 활용하는 지역일 수 있으며, 1개의 지역을 선정하는 것을 기본으로 한다. 그러나, 실시예와 같이 복수의 지역을 선정할 수도 있다. 실시예에 따라, 통합 예측할 지역을 복수 개 선정하는 경우에는 발전량 데이터를 기준으로 지역을 묶어서 구역을 선정할 수도 있다.

[0070] 2. 전처리 과정

[0071] 2-1. 1차 전처리 과정(PCA)

[0072] S200 단계 이후에, 데이터 전처리부(130)를 통해 기상예측 데이터를 리스케일링(Rescaling)하고 PCA(Principal Component Analysis) 알고리즘을 통해 차원을 축소한다(S300단계).

[0073] 기상예측 데이터(이하, '관측소 데이터'라고도 함)는 다양한 관측소에서 수집되므로 데이터의 성분이 많아질 수밖에 없다. 보다 상세하게, 5개의 관측소에서 8개의 파라미터가 수집된다고 할 때, 총 수집되는 관측소 데이터는 40개이다. 이처럼 많은 관측소 데이터를 활용하여 학습 모델을 만들 경우 과적합(Overfitting) 문제가 발생할 수 있다. 뿐만 아니라, 기온, 기압, 강수량 등과 같이 관측소 데이터 마다 단위가 다르기 때문에 데이터의 속성 단위가 다를 경우 있기 때문에 모델 학습 시 과적합을 유발할 수 있다.

[0074] 이에 따라, 본 발명의 데이터 전처리부(130)는 관측소 데이터들의 단위가 균일해지도록 리스케일링 과정을 수행한 후, 차원축소 과정을 거쳐 방대한 양의 관측소 데이터 성분의 개수를 줄임으로써 학습 모델(이하, '혼합기상 모델'이라고도 함)이 과적합 되는 것을 방지할 수 있다. 이러한 전 과정을 '1차 전처리 과정'이라고 한다. 예를 들어, 도 6에 도시된 바와 같이, 관측소(20)에서 26개의 기상예측 데이터(10)가 수집되었을 때, 대략 2,500만개의 관측소 데이터(10)가 수집될 수 있다.

[0075] 이후, 데이터 전처리부(130)는 기수집된 기상예측 데이터(10)의 속성 단위가 0에서 1사이의 값으로 균일해지도록 로버스트 스케일러(Robust Scaler)를 이용하여 지역(region)과 시간(hour)을 조건으로 기상예측 데이터(10)를 리스케일링할 수 있다.

[0076] 보다 상세하게, 도 7에 도시된 바와 같이, 데이터 전처리부(130)에 의해 기상예측 데이터(10)가 7시 'r00지역'을 기준으로 리스케일링되고(리스케일링 1), 7시 'r01지역'을 기준으로 리스케일링되고(리스케일링 2), 18시 'r02지역'을 기준으로 리스케일링된 것을 알 수 있다. 즉, 1,250개의 관측소가 3개의 지역으로 분류된 것을 알 수 있다.

[0077] 리스케일링 후, 데이터 전처리부(130)는 차원 축소(dimension reduction) 알고리즘인 PCA를 이용하여 PCA차원

수만큼 피처를 선택(Feature selection) 함으로써 주요 날씨 정보들에 대한 피처를 추출할 수 있다. 보다 상세하게는, PCA 알고리즘을 이용한 차원 축소 과정을 통해 관측소 데이터셋(즉, pca_00|r00, pca_00|r01, pca_00|r02 와 같은 3개 지역의 기상예측 데이터를 통칭함)의 차원을 6개로 줄일 수 있다. 다시 말해, 도 8에 도시된 바와 같이, 3개의 지역(pca_00|r00, pca_00|r01, pca_00|r02) 별로 기상예측 데이터의 피처를 PCA차원 수만큼(예: 6개) 줄임으로써 데이터셋(Data set)의 차원을 6개로 축소시킬 수 있다. 이처럼 PCA 알고리즘을 이용하여 주요 날씨 정보를 추출하는 과정은 '차원 축소 과정'이라고 할 수 있다.

[0079] **2-2. 2차 전처리 과정 : 리스케일링(rescaling)**

[0080] S300 단계 이후에, 데이터 전처리부(130)를 통해 발전량 데이터를 리스케일링(Rescaling)한다(S400 단계).

[0081] 구체적으로, 데이터 전처리부(130)는 민맥스 스케일러(MinMax Scaler)를 이용하여 기수집된 복수의 발전량 데이터의 발전단위가 0과 1 사이의 값으로 균일해지도록 리스케일링한다. 이때, 발전량 데이터는 전체 태양광 사이트들이 묶여 리스케일링된다.

[0083] **3. EM알고리즘을 이용한 모델링 과정: 가우시안 믹스처 모델(GMM) 피팅**

[0084] S400 단계 이후에, 모델 생성부(140)를 통해 가우시안 믹스처 모델 피팅(Gaussian Mixture Model(GMM) fitting)을 수행한다(S500 단계). 모델링 과정을 설명하기 위해, 설명의 편의를 위해 각 인덱스 및 파라미터들을 먼저 정의하기로 한다.

[0085] 먼저, 다양한 수식에 사용되는 인덱스(index)들을 정의하기로 한다.

[0086] D 는 데이터가 속한 날짜의 수를 의미한다. T 는 하루 시간 길이를 의미하는 것으로서, 오전 07:00 에서 오후 18:00까지의 14시간을 의미한다. 다만, T 는 실시예에 따라 12시간으로 정의될 수도 있다. j, k 는 클러스터의 인덱스를 의미한다. N 은 총 요일의 수를 의미한다. K 는 클러스터(Cluster)의 개수를 의미한다. W_j 는 모든 지역의 기상 데이터를 의미한다. Y 는 발전량을 의미한다. $X = (W, Y)$ 는 기상데이터와 발전량을 묶은 벡터를 의미한다. Cap 은 총 발전최대용량을 의미한다.

[0087] 이어서, 다양한 수식에 사용되는 파라미터(parameter)들을 정의하기로 한다.

[0088] $\pi_{k,t}$ 는 t시간대의 k번째 가우시안 분포가 선택될 확률을 나타낸다. $\mu_{k,t}$ 는 t시간대의 k번째 가우시안 분포의 평균벡터를 나타낸다. $\Sigma_{k,t}$ 는 t시간대의 k번째 가우시안 분포의 공분산 행렬을 나타낸다. L 은 최대로그우도(Complete loglikelihood)를 나타낸다. $z_{k,t}^d$ 는 d일의 t 시간대 샘플이 k번째 가우시안 분포에 속하는 경우에 대한 잠재변수를 나타낸다. $Z_{i,k,t}$ 는 i 날짜의 k 클러스터의 t시간의 가우시안 분포일 확률을 나타낸다. $Y_{i,k,t}$ 는 i 날짜의 k 클러스터의 t시간의 책임 벡터(responsibility vector, 이하, $p(Z_{ik} = 1 | X_i)$ 라고도 표현함)를 나타낸다. ϵ_{EM} 은 EM알고리즘에서의 오차범위(tolerance)를 나타낸다. ϵ_{cov} 은 Sparse Covariance Estimation 알고리즘에서 오차범위(tolerance)를 나타낸다.

[0090] S500 단계에서, 모델 생성부(140)는 전처리된 기상예측 데이터와 발전량 데이터가 각 시간대별로 여러 개의 가우시안 모델이 혼합된 형태로 있다고 가정하고, EM알고리즘을 이용하여 가우시안 믹스처 모델(GMM)의 파라미터를 산출한다. 이때, 모델 생성부(140)가 가우시안 믹스처 모델 피팅(GMM fitting)을 수행하는 단계(S500 단계)는 도 3에 도시된 바와 같이 3단계에 거쳐 수행될 수 있다.

[0091] 첫번째 단계에서는 발전량 데이터로 기본 모델 파라미터(즉, GMM 모델 파라미터)를 초기화한다(S501 단계). 초기화를 수행하지 않을 경우, 최적 파라미터 도출에 장시간이 소요되거나 데이터가 수렴하지 않는 경우가 발생할 수 있기 때문에 S501 단계와 같이 초기화를 수행하여 EM알고리즘으로 최적의 파라미터를 찾는 과정은 필수적이다.

[0092] 보다 상세하게, S501단계는 다음과 같은 과정을 통해 수행될 수 있다.

[0093] 발전량 데이터로 단변량 GMM 모델을 훈련하고, 단변량 GMM 모델로 발전량 데이터(row) 마다 속하는 클러스터 인덱스를 도출한 후, 날씨예측과 발전량이 전처리된 데이터(row) 마다 도출된 클러스터 인덱스를 부여하고, 각 클

러스터의 파라미터를 계산한 후, 계산된 파라미터를 다변량 GMM 모델의 초기 파라미터 값으로 설정한다.

[0094] 두번째 단계에서는 EM알고리즘의 E스텝을 수행한다(S502 단계). E스텝에서는 최대로그우도(Complete loglikelihood, L)와 책임 벡터(responsibility vector, Y_{ik})를 계산한다.

[0095] 세번째 단계에서는 EM알고리즘의 M스텝을 수행한다(S503 단계). M스텝에서는 책임 벡터(Y_{ik})가 주어졌을 때의 최대우도(Complete likelihood)를 최대화(maximize)하는 π , μ , Σ 를 계산한다. 이때, EM스텝(S502 단계 내지 S503 단계)은 $\Delta L < \epsilon_{EM}$ 일 때까지 반복한다. 다시 말해, 최대로그우도가 EM알고리즘에서의 오차범위 내로 수렴할 때까지 반복한다.

[0096] 이하에서는, 수학적 등을 이용해 GMM 피팅 과정에 대해 보다 상세하게 설명하기로 한다.

[0097] 각 시간의 확률분포가 믹스처 형태의 다변량 가우시안 분포(Multivariate Gaussian Distribution)을 따른다고 가정할 경우, t시간대에서 샘플 x가 발생할 확률(p(x))은 하기 (수학적 1)처럼 표현될 수 있다.

$$p(x) = \sum_{k=1}^K \pi_k \mathcal{N}(x | \mu_k, \Sigma_k) \\ = \sum_{k=1}^K \pi_k \frac{\exp(-\frac{1}{2}(x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k))}{\sqrt{(2\pi)^K |\Sigma_k|}} \quad \dots \text{ (수학적 1)}$$

[0099] 여기서, π_k 는 k번째 가우시안 분포(Gaussian distribution)가 선택될 확률을 의미하기 때문에, 아래 수학적 2를 따른다.

$$\sum_{k=1}^K \pi_k = 1 \quad (0 \leq \pi_k \leq 1) \quad \dots \text{ (수학적 2)}$$

[0101] 모든 샘플에 대한 최대우도(Complete likelihood)를 찾기 위해 아래 수학적 3을 계산해야 한다.

$$\ln p(X | \pi, \mu, \Sigma) = \sum_{t=1}^T \sum_{d=1}^D \ln \left\{ \sum_{k=1}^K \pi_{k,t}^d \mathcal{N}(x_t^d | \mu_{k,t}, \Sigma_{k,t}) \right\} \quad \dots \text{ (수학적 3)}$$

[0103] 그러나, 상기 수학적 3을 계산하기에는 복잡성, 과적합 등의 이슈가 있어서, 샘플 X가 k번째 가우시안 분포에 속하는 경우를 의미하는 변수 z_k 를 도입한다. 변수 z_k 는 관측할 수 없기 때문에 잠재변수(latent variable)라 불리며, k차원으로 구성된 이진변수이다. 이에, 샘플 X가 k번째 가우시안 분포에 속하면 $z_k=1$, 속하지 않으면 $z_k=0$ 의 값을 갖는다. 따라서, 아래 (수학적 4a), (수학적 4b), (수학적 4c), (수학적 4d)의 특징을 보일 수 있다.

$$p(z_k = 1 | x) = \pi_k \quad (z \in \{0, 1\}) \quad \dots \text{ (수학적 4a)}$$

$$p(z) = \prod_{k=1}^K \pi_k^{z_k} \quad \dots \text{ (수학적 4b)}$$

$$p(x | z_k = 1) = \mathcal{N}(x | \mu_k, \Sigma_k) \quad \dots \text{ (수학적 4c)}$$

$$p(x | z) = \prod_{k=1}^K \mathcal{N}(x | \mu_k, \Sigma_k)^{z_k} \quad \dots \text{ (수학적 4d)}$$

[0108] 잠재변수 z 를 이용하여 샘플 x가 발생할 확률(p(x))을 나타내면 하기 (수학적 5)와 같으며, 이는 상기 (수학적 1)과 같은 수식임을 알 수 있다.

$$\begin{aligned}
 p(x) &= \sum_z p(z)p(x|z) \\
 &= \sum_{z=1}^K \prod_{k=1}^K (\pi_k \mathcal{N}(x|\mu_k, \Sigma_k))^{z_k} \\
 &= \sum_{k=1}^K \pi_k \mathcal{N}(x|\mu_k, \Sigma_k)
 \end{aligned}$$

[0109] ... (수학식 5)

[0110] 즉, t 시간대에서 샘플 x가 발생할 확률(수학식 1)은 잠재변수 z를 이용하여 샘플 x가 발생할 확률과 같음을 알 수 있다.

[0111] 이에 따라, 잠재변수 z를 좀 더 활용하여, 특정 데이터가 주어졌을 때 z가 1일 확률을 정의한다. 이는 베이즈 이론(Bayes' theorem)을 적용하여 아래 (수학식 6)으로 도출할 수 있다.

$$\begin{aligned}
 \gamma(z_{k,t}^d) &\equiv p(z_{k,t}^d = 1|x_t^d) = \frac{p(z_{k,t}^d = 1)p(x_t^d|z_{k,t}^d = 1)}{\sum_{j=1}^K p(z_{j,t}^d = 1)p(x_t^d|z_{j,t}^d = 1)} \\
 &= \frac{\pi_{k,t}^d \mathcal{N}(x_t^d|\mu_{k,t}, \Sigma_{k,t})}{\sum_{j=1}^K \pi_{j,t}^d \mathcal{N}(x_t^d|\mu_{j,t}, \Sigma_{j,t})} \quad \dots \text{(수학식 6)}
 \end{aligned}$$

[0112] ... (수학식 6)

[0113] 상기 (수학식 3)의 최대우도에서는 각 파라미터 π, μ, Σ 에 대해 미분했을 때 0의 값을 가져야 하는데, 상기 (수학식 6)을 이용하여 아래 (수학식 7a~7c)를 도출할 수 있다.

$$\mu_{k,t} = \frac{1}{N_{k,t}} \sum_{d=1}^D \gamma(z_{k,t}^d) x_t^d \quad (N_{k,t} \equiv \sum_{d=1}^D \gamma(z_{k,t}^d)) \quad \dots \text{(수학식 7a)}$$

[0114] ... (수학식 7a)

$$\Sigma_{k,t} = \frac{1}{N_{k,t}} \sum_{d=1}^D \gamma(z_{k,t}^d) (x_t^d - \mu_{k,t})(x_t^d - \mu_{k,t})^T \quad \dots \text{(수학식 7b)}$$

[0115] ... (수학식 7b)

$$\pi_{k,t} = \frac{N_{k,t}}{D} \quad \dots \text{(수학식 7c)}$$

[0116] ... (수학식 7c)

[0117] 파라미터가 주어졌을 때, 데이터가 관측될 확률은 잠재변수 Z에 대한 주변 합과 같다. 그리고 모든 샘플은 잠재변수 Z를 갖고 있으며, 잠재변수 Z를 알 때의 우도(likelihood)는 아래 (수학식 8a)와 같고 로그우도함수(loglikelihood function)는 (수학식 8b)와 같다. 즉, (수학식 8b)는 (수학식 8a)에 자연로그를 취하고 잠재변수를 곱한 형태를 갖는다.

$$p(X, Z|\mu, \Sigma, \pi) = \prod_{d=1}^D \prod_{t=1}^T \prod_{k=1}^K (\pi_{k,t}^d \mathcal{N}(x_t^d|\mu_{k,t}, \Sigma_{k,t}))^{z_{k,t}^d} \quad \dots \text{(수학식 8a)}$$

[0118] ... (수학식 8a)

$$\ln p(X, Z|\mu, \Sigma, \pi) = \sum_{d=1}^D \sum_{t=1}^T \sum_{k=1}^K z_{k,t}^d \ln(\pi_{k,t}^d \mathcal{N}(x_t^d|\mu_{k,t}, \Sigma_{k,t})) \quad \dots \text{(수학식 8b)}$$

[0119] ... (수학식 8b)

[0120] 그러나, 실제로는 데이터 집합만 주어지기 때문에 잠재변수 Z에 대한 지식은 사후 분포 $p(Z|X, \theta)$ 를 통해서만 주어진다. 따라서, 잠재변수까지 알고 있는 우도 함수를 사용하는 것은 불가능하기 때문에 잠재변수(Z)의 사후 분포($p(Z|X, \theta)$)를 바탕으로 한 기댓값을 고려해야만 한다. 즉, 데이터 X가 주어졌을 때 잠재변수 z에 대한 기댓값($E_{Z|X}$)은 아래 (수학식 9b)와 같다. 여기서, (수학식 9b)는 (수학식 9a)를 고려하여 단순화될 수 있다.

$$E_{Z|X}[z_{k,t}^d] = p(z_{k,t}^d = 1|x_t^d) = \gamma(z_{k,t}^d) \quad \dots \text{(수학식 9a)}$$

[0121] ... (수학식 9a)

$$\begin{aligned}
 E_{Z|X}[\ln(p(X, Z|\mu, \Sigma, \pi))] &= E_{Z|X} \left[\sum_{d=1}^D \sum_{t=1}^T \sum_{k=1}^K z_{k,t}^d \ln(\pi_{k,t}^d \mathcal{N}(x_t^d|\mu_{k,t}, \Sigma_{k,t})) \right] \\
 &= \sum_{d=1}^D \sum_{t=1}^T \sum_{k=1}^K \gamma(z_{k,t}^d) (\ln \pi_{k,t}^d + \ln(\mathcal{N}(x_t^d|\mu_{k,t}, \Sigma_{k,t}))) \quad \dots \text{(수학식 9b)}
 \end{aligned}$$

[0122] ... (수학식 9b)

[0123] 상술한 바와 같이, (수학식 1) 내지 (수학식 9b)를 이용하여 잠재변수 z 에 대한 기댓값($E_{z|x}$)을 구하는 과정은 EM알고리즘의 E단계에 해당한다.

[0124] 이후 행해지는 M스텝에서는 잠재변수 z 에 대한 기댓값을 최대화하는 π, μ, Σ 를 계산한다. 미분 등을 통해 계산한 최적 π, μ 는 다음과 같다.

[0125]
$$\pi_k = \frac{1}{N} \sum_i \gamma_{i,k}, \quad \mu_k = \frac{\sum_i \gamma_{i,k} x}{\sum_i \gamma_{i,k}}$$

[0126] 특히, M스텝에서 공분산 Σ 를 계산하는 경우, 파라미터가 많아 과적합의 위험성이 있다. 이에, sparse covariance estimation 방법론 중 하나인 공분산 라쏘 알고리즘(Covariance Lasso algorithm)을 이용한다. 공분산 라쏘 알고리즘은 예측을 수행하는데 사용되는 변수를 임베디드 기법에 기초하여 선택하는 알고리즘으로서, 모델의 정확도에 기여하는 변수를 학습하여 좀 더 적은 계수를 가지는 회귀식을 찾는 방향으로 공분산 Σ 를 추정할 수 있다.

[0127] 공분산 행렬 Σ 를 추정하기 위해, Σ 의 추정 행렬(이하, '근사 행렬'이라고도 함)을 W 로 정의하고 수치적 근사 과정에서 실험적 공분산 행렬을 S 라 정의한다. 또한, 공분산 행렬 Σ 를 추정하기 위해 공분산 행렬의 역행렬인 정확도 행렬(Sparse Precision matrix) $\Theta = \Sigma^{-1}$ 을 추정할 수 있고, 이 문제는 아래 (수학식 10a)을 최대화 하는 문제로 귀결된다. 그러나 최대화 문제는 수치적 알고리즘인 공분산 라쏘 알고리즘을 통해 해결할 수 있다.

[0128]
$$\log \det \Theta - \text{tr}(S\Theta) - \rho \|\Theta\|_1 \quad \dots \text{ (수학식 10a)}$$

[0129] 여기서, tr 은 trace를 의미하며, 행렬의 대각선 성분의 합을 의미한다. 또한, $\|\cdot\|_1$ 는 L_1 노름(norm)을 의미하며, 행렬의 열별로 각 성분의 절댓값을 취한 값의 합산의 최댓값으로 정의한다. 또한, ρ 는 Lasso Problem의 하이퍼파라미터를 의미한다.

[0130] 보다 상세하게, 최대화 문제를 해결하기 위해 수행되는 공분산 라쏘 알고리즘은 아래 3단계를 수행한다.

[0131] 먼저, 행렬 W 를 추정하기 위해 $W = S + \rho I$ 로 초기화한다(S503-1). 이때, 행렬 W 의 대각선 성분(diagonal entry, $w_{ii} = s_{ii} + \rho, i = 1, 2, \dots, N$)은 고정한다.

[0132] 이어서, Coordinate Descent 알고리즘을 적용하기 위해 두 행렬 W 와 S 를 아래와 같이 분할한다(S503-2).

[0133]
$$W = \begin{pmatrix} W_{11} & w_{12} \\ w_{12}^T & w_{22} \end{pmatrix}, \quad S = \begin{pmatrix} S_{11} & s_{12} \\ s_{12}^T & s_{22} \end{pmatrix} \quad \dots \text{ (수학식 10b)}$$

[0134] 여기서, Coordinate Descent 알고리즘은 행렬의 각 좌표축을 따라 반복적으로 변경하며 목적함수의 최솟값을 찾는 최적화 알고리즘이다.

[0135] 아래의 (수학식 10c)를 목적함수로 하여 Coordinate Descent 알고리즘을 수행한다.

[0136]
$$\min_{\beta} \left\{ \frac{1}{2} \|W_{11}^{1/2} \beta - b\|^2 + \rho \|\beta\|_1 \right\}, \quad b = W_{11}^{-1/2} s_{12} \quad \dots \text{ (수학식 10c)}$$

[0137] 즉, (식 10c)를 만족하는 (N-1) 차원 벡터 $\hat{\beta}$ 은 아래 식을 만족한다.

[0138]
$$w_{12} = W_{11} \hat{\beta} \quad \dots \text{ (수학식 10d)}$$

[0139] 행렬 W 에 대응되는 행과 열을 벡터 $\hat{\beta}$ 로 치환한다.

[0140] 마지막으로, 하이퍼파라미터인 오차범위(tolerance) ϵ_{cov} 에 따라 $\Delta \|\Sigma\|_1 < \epsilon_{cov}$ 로 수렴할 때까지 2번 과정을 반복한다(S503-3).

[0142] 결과적으로, 모델 생성부(140)는 EM스텝을 $\Delta L < \epsilon_{EM}$ 일 때까지 반복한다. 본 발명에서의 EM스텝은 날씨 데이터로 GMM모델 파라미터를 고도화하는 과정을 의미한다.

[0144] 이후, 상술한 3단계에 거친 S500 단계를 통해 산출되어 학습된 GMM 파라미터를 이용하여 기상예측 데이터가 주어졌을 때 발전량의 분포를 산출하여 예측 발전량의 기댓값과 표준편차를 예측한다. 보다 상세하게, 도 4를 참조하면, 발전량 예측부(150)는 기댓값과 표준편차를 통해 클러스터(Cluster)의 개수와 기상예측 데이터의 PCA 컴포넌트 개수(Comp)에 따른 예측값과 예측구간(Prediction Interval)을 산출한다.

[0145] 이어서, 모델 생성부(140)는 확률 모델(즉, 혼합기상모델)의 평균과 표준편차를 이용하여 하이퍼파라미터별로 태양광 발전량 시나리오를 구성한다. 이때, 태양광 발전량 시나리오는 도 5에 도시된 바와 같이 확률 분포를 가장 잘 표현할 수 있는 트리(tree) 형태의 시나리오를 생성할 수 있다. 이에 대한 보다 상세한 설명은 후술하기로 한다.

[0146]

[0147] 4. 발전량의 분포 예측 과정

[0148] S500 단계 이후에, 발전 예측부(150)는 기상 데이터 조건부 발전량의 분포 예측을 수행한다(S600 단계).

[0149] S600 단계에서, 발전 예측부(150)는 모델 생성부(140)에서 학습된 하이퍼파라미터(PCA차원(p), Region(r), 라쏘 정규화 하이퍼파라미터(Lasso regularization hyperparameter)(lambda, λ), GMM클러스터 개수(c))를 이용하여 기상예측 데이터가 주어졌을 때, 발전량의 분포를 산출하여 예측 발전량의 기댓값과 표준편차 예측을 수행한다.

[0150] 이때, 발전량 분포 예측은 다음과 같다.

$$p(Y = y|W = w) = \frac{\sum_k p(Z_k)p(Z_k)}{p(W=w)}$$

$$= \frac{\sum_k p(Z_k)p(Z_k)}{\sum_k p(Z_k)p(Z_k)}$$

[0151]

[0152] 기존 데이터의 평균과 공분산을 날씨와 발전량에 따라 다음과 같이 $\mu = [\mu_w \mu_y]$, $\Sigma = \begin{bmatrix} \Sigma_w & \Sigma_c^T \Sigma_y \end{bmatrix}$ 라고 하면, 발전량의 분포의 평균과 분산은 각각 다음과 같이 나타낼 수 있다.

$$\hat{\mu} = \sum_y \hat{\mu}_y * \frac{\pi_k N(\mu_k, \Sigma_k)}{\sum_k \pi_k N(\mu_k, \Sigma_k)}, \hat{\mu}_y = \mu_y + \Sigma_c^T \Sigma_w^{-1} (W - \mu_w)$$

$$\hat{\Sigma} = \sum_w \hat{\Sigma}_w * \frac{\pi_k N(\mu_k, \Sigma_k)}{\sum_k \pi_k N(\mu_k, \Sigma_k)}, \hat{\Sigma}_y = \Sigma_y - \Sigma_c^T \Sigma_w^{-1}$$

[0153]

[0155] 도 9 내지 도 10은 본 발명의 비교예에 따른 에러지표를 이용하여 하이퍼파라미터 별로 생성된 시나리오의 편차를 나타낸 예시도이다. 도 11 내지 도 12는 본 발명의 실시예에 따른 에러지표를 이용하여 하이퍼파라미터 별로 생성된 시나리오의 편차를 나타낸 예시도이다.

[0157] 발전량 분포 예측의 정확도를 측정하는 KPI

[0158] 이때, 산출된 발전량 분포 예측값이 정확한지 판단하기 위해 발전 예측부(150)는 도 11 내지 도 12와 같이 L1 error KPI(이하, 'L1 에러지표'라고도 함) 및 L2 error KPI(이하, 'L2 에러지표'라고도 함)를 에러지표로서 사용한다. 일반적으로는 예측의 정확도를 판단하기 위해 도 9 내지 도 10과 같이 Mean Absolute Error(MAE), Root Mean Squared Percentage Error(RMSE), Root Mean Squared Percentage Error(RMSPE) 등과 같은 에러지표를 사용하는데, 이러한 지표로는 분포 예측의 정확도를 판단하는데 한계가 있다.

[0159] 이하에서 설명하는 에러지표에서, *Real* 은 실측값을 의미하고, *Pred* 는 예측분포의 평균값을 의미하고, σ 는 예측분포의 표준편차를 의미하고, Capacity(이하, 'Cap'이라고도 함)는 총 발전최대용량을 의미하고, N은 데이터 개수를 의미한다.

[0160] 예컨대, RMSPE 에러지표 $\left(\frac{\sum (Real-Pred)^2}{N} \right)^{\frac{1}{2}}$ 는 단순히 예측의 정확도를 보여줄 수 있다. 그러나, Squared Sum(제곱합) 에러지표 $\left(\frac{\sum \left(\frac{Real-Pred-\alpha}{Capacity} \right)^2 + \left(\frac{Real-Pred+\alpha}{Capacity} \right)^2}{N} \right)^{\frac{1}{2}}$ 를 이용하면 양쪽의 예측 구간과 실제 값의 정확도를 알 수 있다.

[0161] L1 에러지표 $\frac{|Real-Pred| + \min\left(\frac{Real-Pred-\alpha}{Capacity}, \frac{Real-Pred+\alpha}{Capacity}\right)}{N}$ 는 예측의 정확도 $\left(\frac{Real-Pred-\alpha}{Capacity} \right)$ 와 예측 구간의 정확도 $\left(\frac{Real-Pred+\alpha}{Capacity} \right)$ 의 최소값을 Absolute Sum(절대합) $|Real-Pred|$ 과 합하여 분포 예측의 정확도를 측정한다.

[0162] L2 에러지표 $\left(\frac{|Real-Pred|^2 + \left(\left(\frac{Real-Pred-\alpha}{Capacity} \right)^2 + \left(\frac{Real-Pred+\alpha}{Capacity} \right)^2 \right)}{N} \right)^{\frac{1}{2}}$ 는 예측의 정확도 $\left(\frac{Real-Pred-\alpha}{Capacity} \right)$ 와 예측 구간의 정확도 $\left(\frac{Real-Pred+\alpha}{Capacity} \right)$ 를 square sum(제곱합) $|Real-Pred|^2$ 과 합하여 분포 예측의 정확도를 측정하되, 둘 중 하나의 값(즉, 예측의 정확도에 $|Real-Pred|^2$ 을 합한 값과 예측 구간의 정확도에 $|Real-Pred|^2$ 을 합한 값 중 하나)이 되는 것을 L1 에러지표 보다 더 엄격하게 평가한다. 다시 말해, L2 에러지표의 두 값(발전량 분포 예측의 정확도 또는 발전량 예측 구간의 정확도) 중 L1 에러지표의 두 값과 오차가 큰 값이 정확도가 낮다고 판단할 수 있다.

[0163] 따라서, 발전 예측부(150)는 L1 에러지표와 L2 에러지표를 활용함으로써 발전량 예측 구간의 정확도를 고려하여 혼합기상모델을 평가할 수 있고, 하이퍼파라미터별로 생성된 시나리오의 편차를 낮출 수 있다. 발전 예측부(150)는 L1 에러지표와 L2 에러지표가 가장 낮을 때에 분포 예측의 정확도가 높다고 판단하고, 이를 기반으로 예측을 수행함에 따라 최적의 입찰량을 결정할 수 있는 효과가 있다.

[0164] 또한, 분포 예측의 정확도에 변화가 생길 수 있기 때문에 적절한 시간 간격으로 예측 오차 분포를 수정하고, 태양광 발전량 시나리오 트리(PVST)를 생성한다.

[0165] 이처럼, 본 발명의 실시예에 따르면 발전 예측부(150)가 발전량의 분포 예측의 정확도를 판단하여 예측을 수행함에 따라 예측 모델의 성능이 고도화되면, 사용자에게 더 많은 입찰 수익을 가져다줄 수 있다.

[0167] 5. 최적 입찰 결정 수행 과정

[0168] S600 단계 이후에, 입찰부(160)는 최적 입찰 결정을 수행한다(S700 단계).

[0169] 구체적으로, 입찰부(160)는 앞서 생성된 태양광 발전량 시나리오 트리(PVST)를 기반으로 최적 입찰량을 산출한다.

[0170] 도 4를 참조하면, 태양광 발전량 시나리오 트리(PVST)는 최초 입찰 시간대(6시~9시)와 마지막 입찰 시간대(15시~18시)에는 두 개의 노드(n1, n2)가 존재한다.

[0171] 이때, 최상위 노드(n1)의 대표값은 평균(μ)과 표준편차(σ with 0.5)의 오차(즉, $\mu - \sigma$ with 0.5)이고, 50%의 전이확률을 갖는다. 최하위 노드(n2)의 대표값은 평균(μ)과 표준편차(σ with 0.5)의 합(즉, $\mu + \sigma$ with 0.5)이고, 50%의 전이확률을 갖는다.

[0172] 그리고, 중간 시간대(9시~12시/12시~15시)에는 두 개 또는 세 개의 노드가 존재한다.

[0173] 먼저, 중간 시간대의 노드가 두 개인 경우는 (평균 + (표준편차*sqrt(6)/2) > 발전용량) 또는 (평균 - (표준편차 *sqrt(6) / 2) < 0)인 경우가 3시간 범위(batch) 안에 존재할 때이다.

[0174] 이때, 최상위 노드와 최하위 노드는 상술한 최초 입찰 시간대의 노드와 같다. 즉, 최상위 노드와 최하위 노드는 각각 $\mu - \sigma$ with 0.5, $\mu + \sigma$ with 0.5 의 대표값을 갖는다.

[0175] 중간 시간대의 노드가 세 개인 경우는 (평균 + (표준편차*sqrt(6)/2) ≤ 발전용량) 및 (평균 - (표준편차 *sqrt(6)/2) ≥ 0)일 때이다.

[0176] 이때, 최상위 노드의 대표값은 평균(μ)과 표준편차(σ)*sqrt(6)/2의 오차(즉, $\mu - \sigma\sqrt{6}/2$)이고, 33%의 전이확률을 갖는다. 그리고, 중간 노드의 대표값은 평균(μ)이고, 33%의 전이확률을 갖는다. 마지막으로, 최하위 노드의 대표값은 평균(μ)과 표준편차(σ)*sqrt(6)/2의 합(즉, $\mu + \sigma\sqrt{6}/2$)이고, 33%의 전이확률을 갖는다.

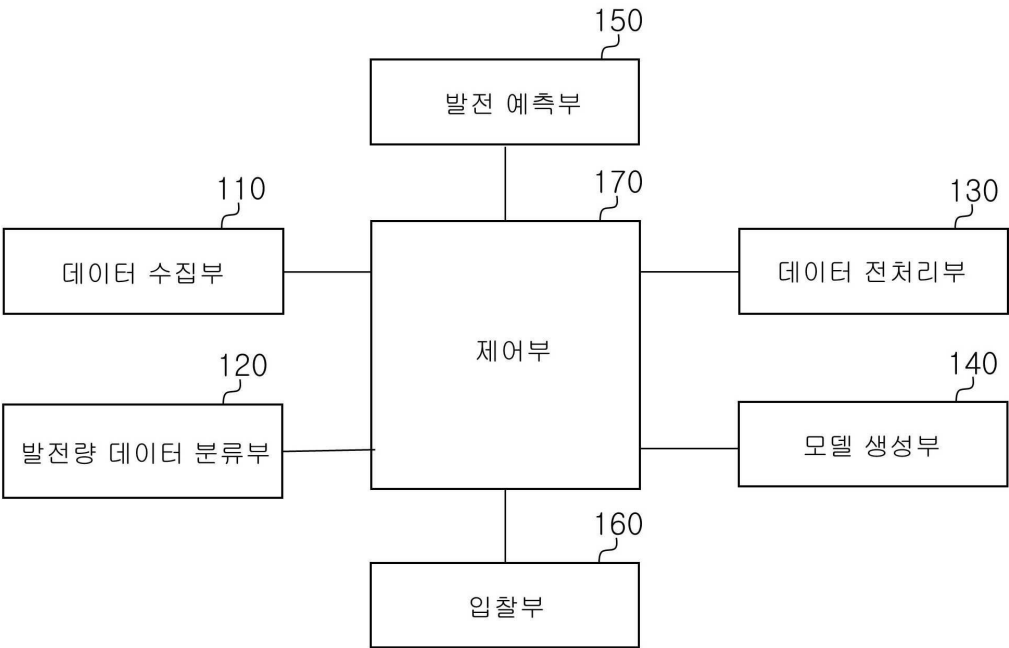
- [0177] 태양광 발전량 시나리오 트리(PVST)는 도 4에 도시된 바와 같이 Lattice 형태를 갖고, 총 시나리오 개수는 36개가 생성된다.
- [0178] 이와 같이, 본 발명의 실시예에 따른 최적 입찰량 결정 시스템은, 다양한 관측소로부터 수집되는 방대한 양의 기상예측 데이터를 리스케일링하여 단위를 균일하게 하고 차원 축소 알고리즘을 통해 최적 입찰 시나리오 생성을 위한 학습 모델에 사용되는 피처의 개수를 감소시켜 모델 학습의 복잡성 또는 과적합 등의 문제를 해결할 수 있다.
- [0179] 또한, 본 발명의 실시예에 따른 최적 입찰량 결정 시스템은, 리스케일링하고 차원 축소된 기상예측 데이터에 태양광 발전량을 합하여 추가적인 리스케일링 과정을 진행함으로써 모델 파라미터를 최소화할 수 있다.
- [0180] 다시 말해, 리스케일링된 태양광 발전량으로 모델의 하이퍼파라미터와 관련있는 인덱스 값을 찾은 뒤, 각 지역별로 리스케일링 되고 차원 축소된 기상예측 데이터를상기 리스케일링된 태양광 발전량 데이터에 합한 데이터의 인덱스를 하이퍼파라미터 탐색에 사용함으로써 발전 예측 값의 불확실성을 최소화할 수 있다.
- [0181] 또한, 본 발명의 실시예에 따른 최적 입찰량 결정 시스템은, 예측 구간의 정확도까지 판단할 수 있는 에러지표를 적용함으로써 학습 모델을 이용한 발전량 분포 예측 시 예측의 정확도뿐만 아니라 발전량 예측 구간의 정확도까지 높일 수 있다.
- [0182] 따라서, 정확도 높은 태양광 발전량 예측값 및 예측 구간을 기반으로 예측을 수행하여 최적의 입찰량을 산출함에 따라 최적 입찰량 결정 시스템의 성능을 고도화하여 예측의 신뢰도를 향상시킬 수 있다.
- [0183] 뿐만 아니라, 본 발명은 발전량 예측값 및 예측 구간의 정확도가 향상된 시나리오를 기반으로 발전량을 예측하는 사용자들의 이익을 극대화시킬 수 있다.
- [0184] 이상 첨부된 도면을 참조하여 본 발명의 실시예들을 더욱 상세하게 설명하였으나, 본 발명은 반드시 이러한 실시예로 국한되는 것은 아니고, 본 발명의 기술사상을 벗어나지 않는 범위 내에서 다양하게 변형 실시될 수 있다. 따라서, 본 발명에 개시된 실시예들은 본 발명의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시예에 의하여 본 발명의 기술 사상의 범위가 한정되는 것은 아니다. 그러므로, 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다. 본 발명의 보호 범위는 아래의 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 발명의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

부호의 설명

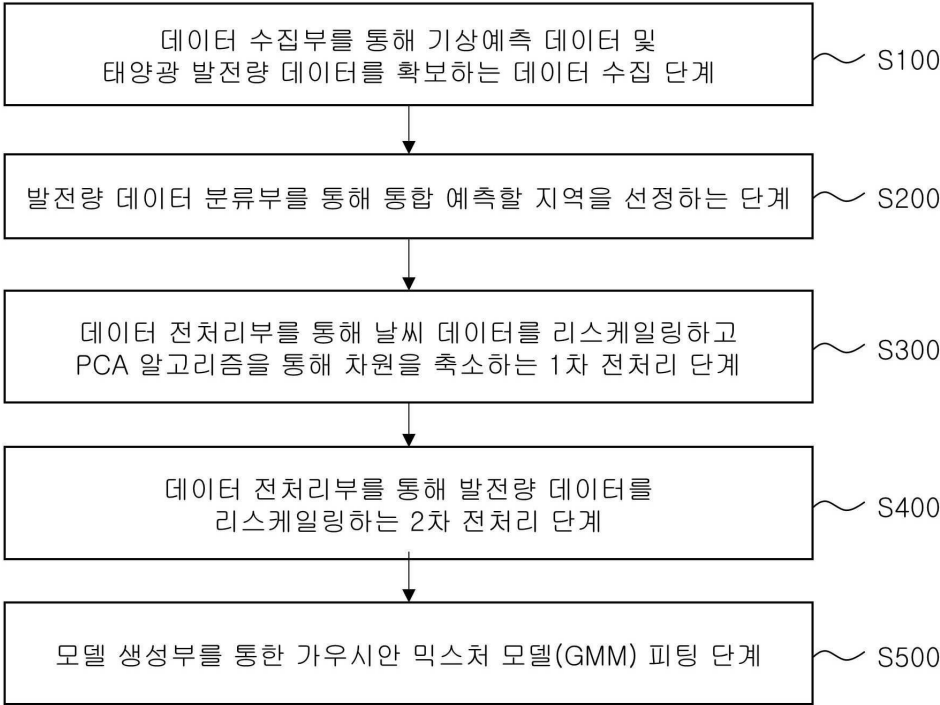
- [0185]
- | | |
|---------------|------------------|
| 110: 데이터 수집부 | 120: 발전량 데이터 분류부 |
| 130: 데이터 전처리부 | 140: 모델 생성부 |
| 150: 발전 예측부 | 160: 입찰부 |
| 170: 제어부 | |

도면

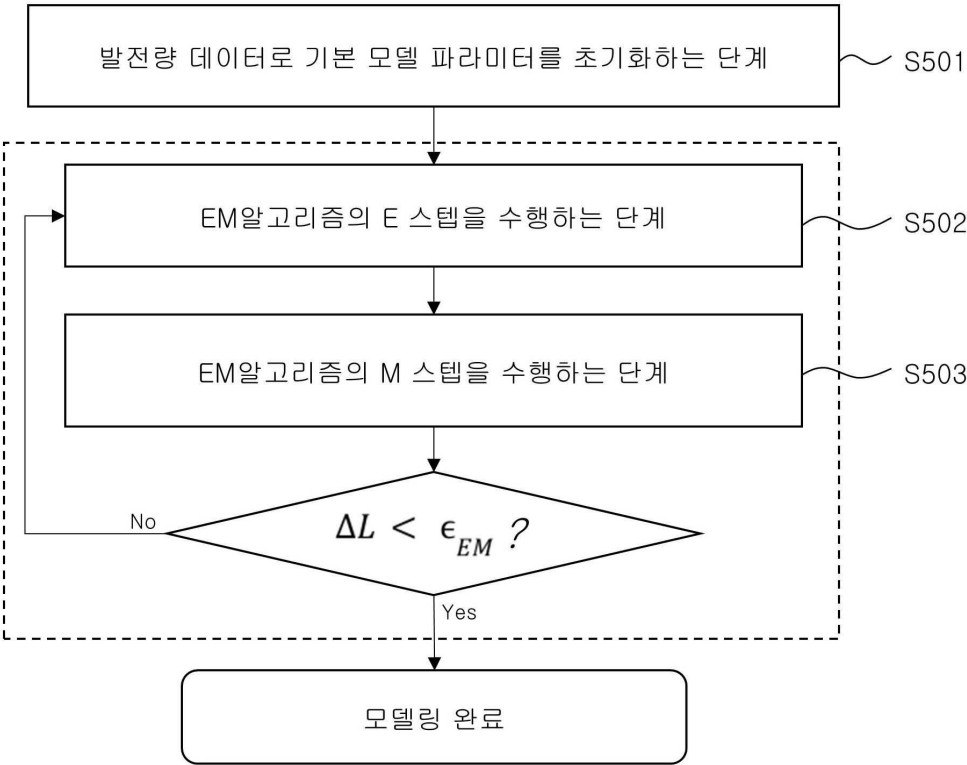
도면1



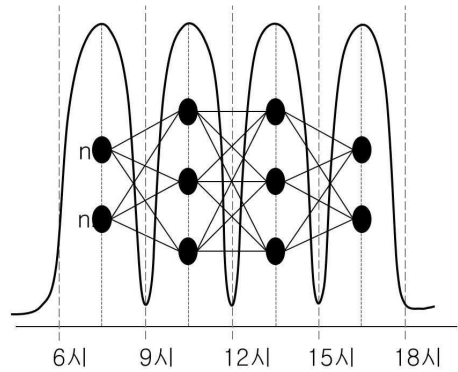
도면2



도면3



도면4



도면5

관측소(20)

26개의 기상예측 데이터(10)

idx	observ_id	ts	recv_ts	cloud	temp	temp_max	temp_min	humidity	ground_pr	wind_spe	wind_dir	rain	snow	real_feel	level_bulb	dew_point	wind_spd	wind_gust	rel_hum	vis	ceiling	uv_idx	precip	protrain_prob	snow_prob	ice_prob	total_q	ice	old_cvr	
0	1832501	2020-08-12	2020-07-1	97	22.9	22.9	22.9	92	1013	0.91	160	0	0																	
1	1832501	2020-08-12	2020-07-1	96	22.78	22.78	22.78	92	1013	1.26	167	0	0																	
2	1832501	2020-08-12	2020-07-1	97	22.58	22.58	22.58	93	1012	1.26	155	0	0																	
⋮																														
961212	3430522	2021-09-20	2021-09-21 21:11:56	24.4444								243	4.572	0	26.1111	23.3333	22.7778	5.6327	7.40239	90	4.82803	467.68	0	60	60	0	0	4.572	0	100
961213	3430522	2021-09-20	2021-09-21 22:11:56	24.4444								202	2.286	0	26.1111	23.3333	22.7778	7.40298	9.3342	90	4.82803	467.68	0	55	55	0	0	2.286	0	100
961214	3430522	2021-09-20	2021-09-21 23:11:56	23.8889								197	0	0	27.2222	23.3333	22.7778	7.40298	9.3342	92	4.82803	467.68	0	49	49	0	0	0	0	100

(i)

도면6

시간 지역 관측속
(30) (40) (20)

26개의 기상예측 데이터(10)

리스크에어링1

리스크에어링2

리스크에어링3

도면7

PCA차원1 PCA차원2 PCA차원3 PCA차원4 PCA차원5 PCA차원6

시간 지역 관측속
(30) (40) (20)

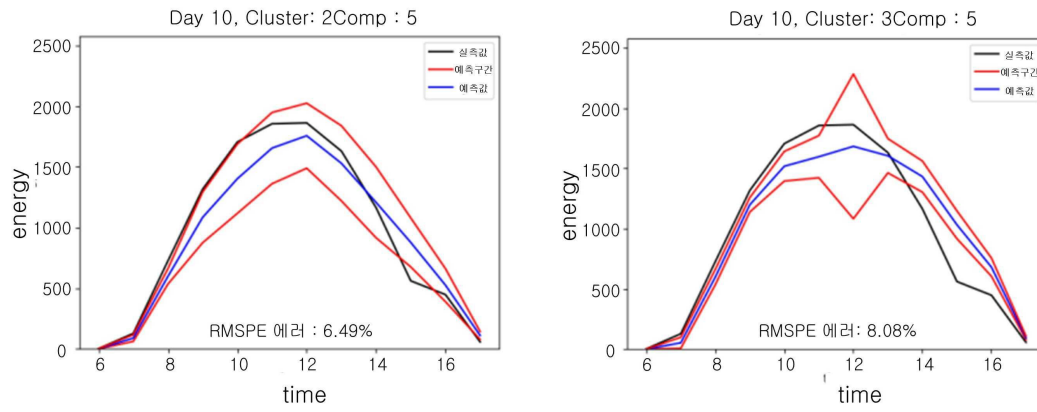
26개의 기상예측 데이터(10)

리스크에어링1

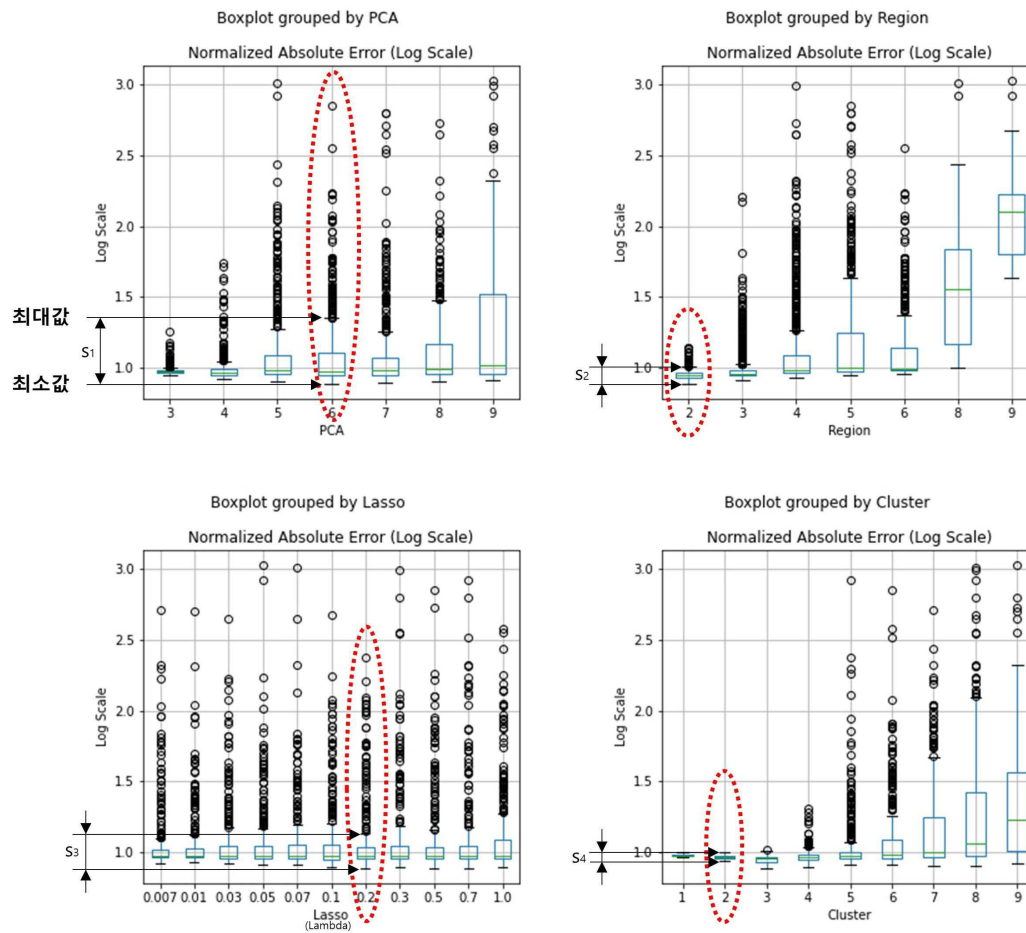
리스크에어링2

리스크에어링3

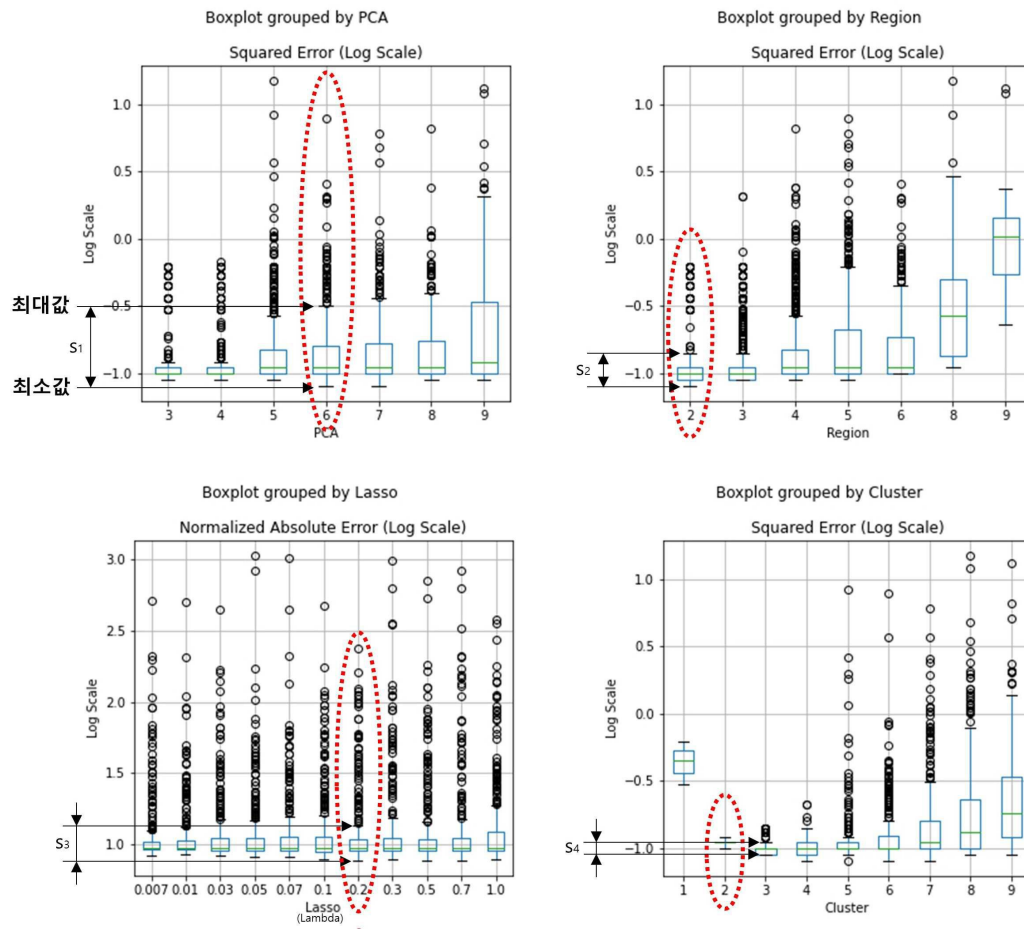
도면8



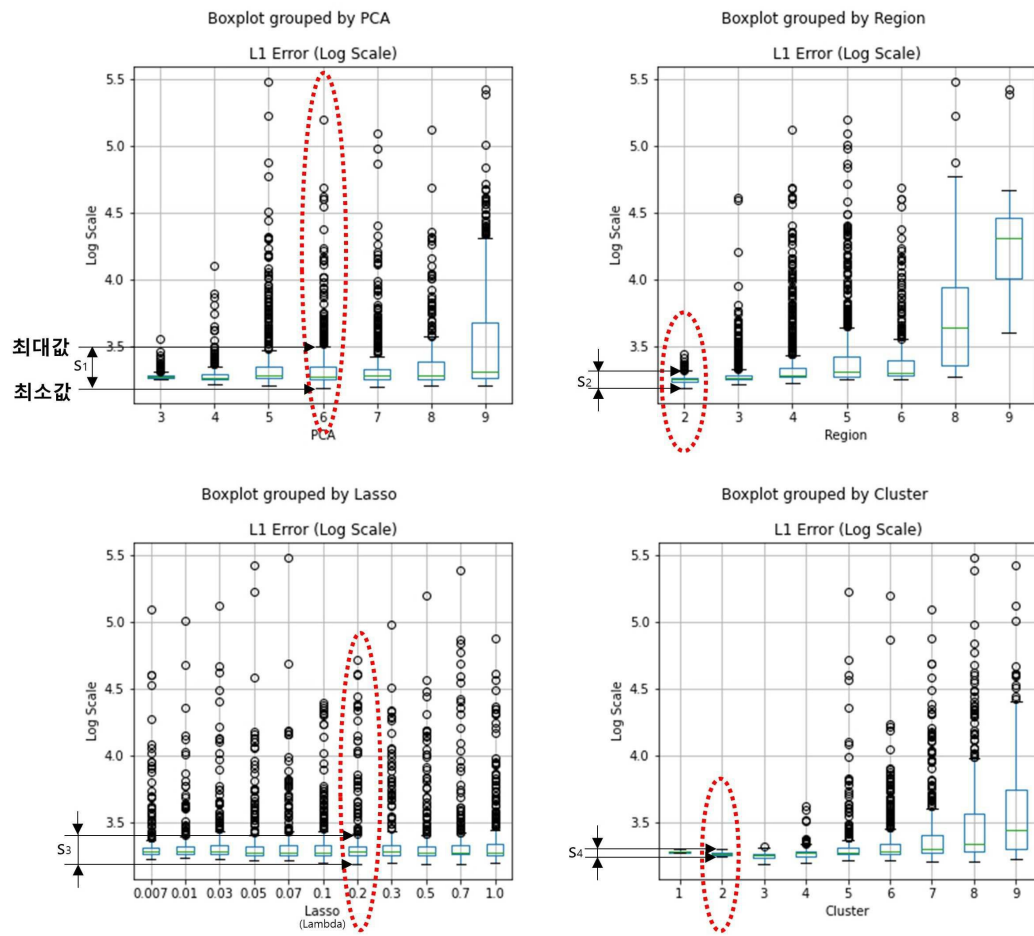
도면9



도면10



도면11



도면12

