(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0088967
(43) 공개일자 2023년06월20일

(51) 국제특허분류(Int. Cl.)
H02J 3/00 (2006.01) *G06Q 50/06* (2012.01)
(52) CPC특허분류
H02J 3/003 (2020.01)
G06Q 50/06 (2013.01)
(21) 출원번호 10-2021-0177250
(22) 출원일자 2021년12월13일
심사청구일자 2021년12월13일

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
황인준
서울특별시 성동구 매봉길 15(옥수동, 래미안 옥수 리버젠)
박성우
서울특별시 성동구 상원12길 15(성수동1가, 성수동동아그린아파트)
(뒷면에 계속)
(74) 대리인
특허법인주연케이알피

전체 청구항 수 : 총 14 항

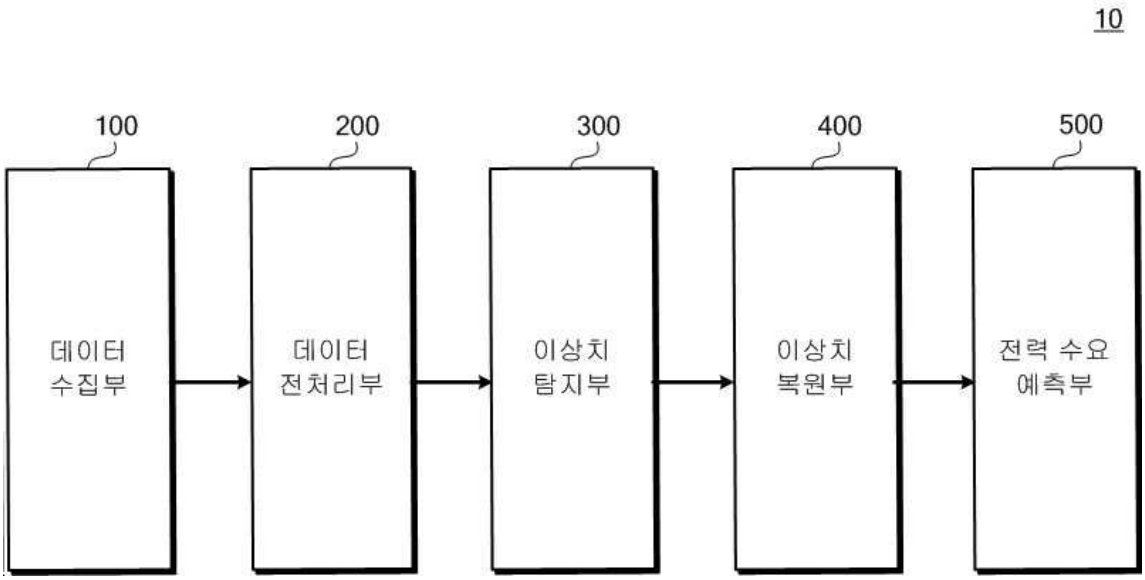
(54) 발명의 명칭 **이상 탐지 및 복원 기반의 전력 수요량 예측 장치 및 방법**

(57) 요 약

본 발명은 전력 수요량을 예측할 수 있는 장치 및 방법에 관한 것으로, 좀 더 자세하게는 기계학습을 통하여 미래의 전력 수요를 정확하게 예측할 수 있는 장치 및 방법에 관한 것이다. 본 출원의 실시 예에 따른 전력 수요량 예측 장치는 기상 데이터, 전력 데이터 및 시간 데이터를 수집하는 데이터 수집부; 상기 기상 데이터, 상기 전력

(뒷면에 계속)

대 표 도 - 도1



데이터 및 상기 시간 데이터에 대한 전처리 동작을 수행하는 데이터 전처리부; 변분 오토인코더를 통하여 구현되며, 상기 전처리된 전력 데이터를 이상 데이터와 정상 데이터로 분류하는 이상치 탐지부; 랜덤 포레스트 모델을 통하여 구현되며, 상기 이상 데이터를 복원하여 복원 데이터를 생성하는 이상치 복원부; 및 슬라이딩 윈도우 기반의 LightGBM 모델을 통하여 구현되며, 상기 정상 데이터와 상기 복원 데이터에 기초하여 예측 모델을 학습하고 전력 수요량을 예측하는 전력 수요 예측부를 포함한다. 본 출원의 실시 예에 따른 전력 수요량 예측 장치는 이상치를 탐지하고 이를 복원하여 학습 데이터로 사용한다. 데이터 수가 충분하지 않은 상황에서 이상치를 복원하여 학습 데이터로 사용함으로써, 본 출원의 실시 예에 따른 전략 수요량 예측 장치는 오버 피팅 없이 향상된 예측 성능을 제공할 수 있다.

(52) CPC특허분류
Y04S 10/50 (2020.08)
Y04S 40/20 (2020.08)

(72) 발명자
정승민
서울특별시 종로구 지봉로 81(창신동, 브라운스톤 창신)

정승원
서울특별시 노원구 우이천로 102(월계동, 초안아파트)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711127169
과제번호	2019M3F2A1073184
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	에너지클라우드기술개발(과기정통부)(R&D)
연구과제명	개방형 마이크로그리드 플랫폼을 위한 딥러닝 기반의 확률론적 에너지 사용량 예측 모델 구성 및 시뮬레이션 개발
기 여 율	1/1
과제수행기관명	고려대학교
연구기간	2021.01.01 ~ 2021.12.31

명 세 서

청구범위

청구항 1

기상 데이터, 전력 데이터 및 시간 데이터를 수집하는 데이터 수집부;

상기 기상 데이터, 상기 전력 데이터 및 상기 시간 데이터에 대한 전처리 동작을 수행하는 데이터 전처리부;

변분 오토인코더를 통하여 구현되며, 상기 전처리된 전력 데이터를 이상 데이터와 정상 데이터로 분류하는 이상치 탐지부;

랜덤 포레스트 모델을 통하여 구현되며, 상기 이상 데이터를 복원하여 복원 데이터를 생성하는 이상치 복원부; 및

슬라이딩 윈도우 기반의 LightGBM 모델을 통하여 구현되며, 상기 정상 데이터와 상기 복원 데이터에 기초하여 예측 모델을 학습하고 전력 수요량을 예측하는 전력 수요 예측부를 포함하는, 전력 수요량 예측 장치.

청구항 2

제1 항에 있어서,

상기 데이터 전처리부는

상기 기상 데이터에 대한 정규화 동작을 수행하는 기상 데이터 전처리부;

상기 전력 데이터에 대한 정규화 동작을 수행하는 전력 데이터 전처리부; 및

상기 시간 데이터를 2차원의 시간 데이터로 변환하는 시간 데이터 전처리부를 포함하는, 전력 수요량 예측 장치.

청구항 3

제2 항에 있어서,

상기 기상 데이터 전처리부는 상기 기상 데이터 중 기온, 습도, 풍속 데이터에 기초하여 체감 온도 데이터와 불쾌지수 데이터를 생성하고, 생성된 체감 온도 데이터와 불쾌지수 데이터에 대한 정규화 동작을 수행하는, 전력 수요량 예측 장치.

청구항 4

제2 항에 있어서,

상기 시간 데이터 전처리부는 주기 함수를 이용하여 상기 시간 데이터를 서로 다른 두 개의 2차원 시간 데이터로 변환하는, 전력 수요량 예측 장치.

청구항 5

제1 항에 있어서,

상기 데이터 수집부가 수집하는 기상 데이터는 기상 예보 데이터 및 측정 기상 데이터 중 어느 하나인, 전략 수요량 예측 장치.

청구항 6

제1 항에 있어서,

상기 전력 수요 예측부는 7의 윈도우 사이즈를 갖는, 전력 수요량 예측 장치.

청구항 7

데이터 수집부에서, 기상 데이터, 시간 데이터 및 전력 데이터를 수집하는 단계;

데이터 전처리부에서, 상기 기상 데이터, 상기 시간 데이터 및 상기 전력 데이터에 대한 전처리 동작을 수행하는 단계;

이상치 탐지부에서, 상기 전처리된 전력 데이터 중 이상치를 탐지하는 단계;

이상치 복원부에서, 상기 이상치를 복원하여 복원 데이터를 생성하는 단계; 및

전력 수요 예측부에서, 상기 복원 데이터를 학습 데이터로 하여 예측 모델을 학습하고, 상기 학습된 예측 모델에 기초하여 전력 수요량을 예측하는 단계를 포함하는, 전력 수요량 예측 방법.

청구항 8

제7 항에 있어서,

상기 전처리 동작을 수행하는 단계는

상기 기상 데이터 및 상기 전력 데이터를 정규화하는 단계; 및

상기 시간 데이터를 주기함수를 이용하여 서로 다른 두 개의 시간 데이터로 변환하는 단계를 포함하는, 전력 수요량 예측 방법.

청구항 9

제8 항에 있어서,

상기 기상 데이터를 정규화하는 단계는

상기 기상 데이터 중 기온, 습도, 풍속 데이터에 기초하여 체감 온도 데이터와 불쾌지수 데이터를 생성하는 단계; 및

상기 생성된 체감 온도 데이터와 상기 생성된 불쾌지수 데이터에 대한 정규화 동작을 수행하는 단계를 포함하는, 전력 수요량 예측 방법.

청구항 10

제9 항에 있어서,

상기 기상 데이터는 기상 예보 데이터 및 측정 기상 데이터 중 어느 하나인, 전력 수요량 예측 방법.

청구항 11

제7 항에 있어서,

상기 이상치 탐지부는 변분 오토인코더를 통하여 구현되는, 전력 수요량 예측 방법.

청구항 12

제7 항에 있어서,

상기 이상치 복원부는 랜덤 포레스트 모델을 통하여 구현되는, 전력 수요량 예측 방법.

청구항 13

제7 항에 있어서,

상기 전력 수요량 예측부는 슬라이딩 윈도우 기반의 LightGBM을 통하여 구현되는, 전력 수요량 예측 방법.

청구항 14

제13 항에 있어서,

상기 전력 수요량 예측부는 7의 윈도우 사이즈를 갖는, 전력 수요량 예측 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 전략 수요량을 예측할 수 있는 장치 및 방법에 관한 것으로, 좀 더 자세하게는 기계학습을 통하여 미래의 전략 수요를 정확하게 예측할 수 있는 장치 및 방법에 관한 것이다.

배 경 기 술

[0002] 스마트 그리드는 전 세계적으로 발생하는 환경문제 및 자원고갈 문제에 대한 실현 가능한 솔루션으로 많은 주목을 받고 있으며, 스마트 그리드를 구성하는 에너지 저장 시스템, 에너지 관리 시스템, 신재생 에너지 시스템 등 다양한 시스템들을 효율적으로 사용하기 위해서는 정확도 높은 전력 수요 예측이 필요하다.

[0003] 최근 컴퓨터 기술이 발전함에 따라 기계학습 및 딥러닝 기반의 예측 모델들에 대한 연구가 활발하게 진행되고 있으며 좋은 성능을 보여주고 있다. 이러한 기계학습 및 딥러닝 기반의 예측 모델들은 데이터의 양과 질의 영향을 많이 받는다.

[0004] 예측 모델의 학습에 사용되는 데이터에 이상치 혹은 결측치가 많이 존재할 경우 모델의 학습에 방해가 되어 예측 정확도가 낮아질 수 있다. 학습하기에 충분한 양의 데이터가 수집된 경우에는 이상치와 결측치를 제거한 후에 모델을 학습하여도 문제가 되지 않지만, 데이터 수가 충분하지 않을 때 이상치와 결측치를 제거하게 되면 오버 피팅으로 인하여 모델의 학습 자체가 어려워질 수 있다.

발명의 내용

해결하려는 과제

[0005] 본 발명은 데이터 수가 충분하지 않은 상황에서 이상치 탐지 및 복원을 통하여 오버 피팅 없이 예측 모델의 성능을 향상시킬 수 있는 전략 수요량 예측 장치를 제공하는데 있다.

과제의 해결 수단

[0006] 본 출원의 실시 예에 따른 전력 수요량 예측 장치는 기상 데이터, 전력 데이터 및 시간 데이터를 수집하는 데이터 수집부; 상기 기상 데이터, 상기 전력 데이터 및 상기 시간 데이터에 대한 전처리 동작을 수행하는 데이터 전처리부; 변분 오토인코더를 통하여 구현되며, 상기 전처리된 전력 데이터를 이상 데이터와 정상 데이터로 분류하는 이상치 탐지부; 랜덤 포레스트 모델을 통하여 구현되며, 상기 이상 데이터를 복원하여 복원 데이터를 생성하는 이상치 복원부; 및 슬라이딩 윈도우 기반의 LightGBM 모델을 통하여 구현되며, 상기 정상 데이터와 상기 복원 데이터에 기초하여 예측 모델을 학습하고 전력 수요량을 예측하는 전력 수요 예측부를 포함한다.

[0007] 실시 예에 있어서, 상기 데이터 전처리부는 상기 기상 데이터에 대한 정규화 동작을 수행하는 기상 데이터 전처리부; 상기 전력 데이터에 대한 정규화 동작을 수행하는 전력 데이터 전처리부; 및 상기 시간 데이터를 2차원의 시간 데이터로 변환하는 시간 데이터 전처리부를 포함한다.

[0008] 실시 예에 있어서, 상기 기상 데이터 전처리부는 상기 기상 데이터 중 기온, 습도, 풍속 데이터에 기초하여 체감 온도 데이터와 불쾌지수 데이터를 생성하고, 생성된 체감 온도 데이터와 불쾌지수 데이터에 대한 정규화 동작을 수행한다.

[0009] 실시 예에 있어서, 상기 시간 데이터 전처리부는 주기 함수를 이용하여 상기 시간 데이터를 서로 다른 두 개의 2차원 시간 데이터로 변환한다.

[0010] 실시 예에 있어서, 상기 데이터 수집부가 수집하는 기상 데이터는 기상 예보 데이터 및 측정 기상 데이터 중 어느 하나이다.

[0011] 실시 예에 있어서, 상기 전력 수요 예측부는 7의 윈도우 사이즈를 갖는다.

[0012] 본 출원의 실시 예에 따른 전력 수요량 예측 방법은 데이터 수집부에서, 기상 데이터, 시간 데이터 및 전력 데이터를 수집하는 단계; 데이터 전처리부에서, 상기 기상 데이터, 상기 시간 데이터 및 상기 전력 데이터에 대한 전처리 동작을 수행하는 단계; 이상치 탐지부에서, 상기 전처리된 전력 데이터 중 이상치를 탐지하는 단계; 이상치 복원부에서, 상기 이상치를 복원하여 복원 데이터를 생성하는 단계; 및 전력 수요 예측부에서, 상기 복원

데이터를 학습 데이터로 하여 예측 모델을 학습하고, 상기 학습된 예측 모델에 기초하여 전력 수요량을 예측하는 단계를 포함한다.

[0013] 실시 예에 있어서, 상기 전처리 동작을 수행하는 단계는 상기 기상 데이터 및 상기 전력 데이터를 정규화하는 단계; 및 상기 시간 데이터를 주기함수를 이용하여 서로 다른 두 개의 시간 데이터로 변환하는 단계를 포함한다.

[0014] 실시 예에 있어서, 상기 기상 데이터를 정규화하는 단계는 상기 기상 데이터 중 기온, 습도, 풍속 데이터에 기초하여 체감 온도 데이터와 불쾌지수 데이터를 생성하는 단계; 및 상기 생성된 체감 온도 데이터와 상기 생성된 불쾌지수 데이터에 대한 정규화 동작을 수행하는 단계를 포함한다.

[0015] 실시 예에 있어서, 상기 기상 데이터는 기상 예보 데이터 및 측정 기상 데이터 중 어느 하나이다.

[0016] 실시 예에 있어서, 상기 이상치 탐지부는 변분 오토인코더를 통하여 구현된다.

[0017] 실시 예에 있어서, 상기 이상치 복원부는 랜덤 포레스트 모델을 통하여 구현된다.

[0018] 실시 예에 있어서, 상기 전력 수요량 예측부는 슬라이딩 윈도우 기반의 LightGBM을 통하여 구현된다.

[0019] 실시 예에 있어서, 상기 전력 수요량 예측부는 7의 윈도우 사이즈를 갖는다.

발명의 효과

[0020] 본 출원의 실시 예에 따른 전력 수요량 예측 장치는 이상치를 탐지하고 이를 복원하여 학습 데이터로 사용한다. 데이터 수가 충분하지 않은 상황에서 이상치를 복원하여 학습 데이터로 사용함으로써, 본 출원의 실시 예에 따른 전략 수요량 예측 장치는 오버 피팅 없이 향상된 예측 성능을 제공할 수 있다.

도면의 간단한 설명

[0021] 도 1은 본 출원의 실시 예에 따른 전략 수요량 예측 장치(10)를 보여주는 블록도이다.

도 2는 도 1의 데이터 전처리부(200)를 좀 더 자세히 보여주는 블록도이다.

도 3a는 범주형 시간 데이터(Data_T)의 일 예를 보여주는 도면이다.

도 3b는 2차원 시간 데이터(Data_T1, Data_T2)의 일 예를 보여주는 도면이다.

도 4는 도 1의 이상치 탐지부(300)를 좀 더 자세히 보여주는 도면이다.

도 5는 이상치 탐지부(300)의 이상치 탐지 실험 결과를 보여주는 그래프이다.

도 6은 도 1의 이상치 복원부(400)를 좀 더 자세히 보여주는 도면이다.

도 7은 이상치 복원부(400)의 복원 실험 결과를 보여주는 그래프이다.

도 8은 도 1의 전력 수요 예측부(500)를 좀 더 자세히 보여주는 도면이다.

도 9는 전력 수요 예측부(500)의 윈도우 사이즈 결정을 위한 실험 결과를 보여주는 그래프이다.

도 10은 본 출원의 실시 예에 따른 전력 수요량 예측 장치(10)의 예측 결과를 다른 모델과 비교하는 그래프이다.

도 11은 도 1의 전력 수요량 예측 장치(10)의 동작을 보여주는 순서도이다.

발명을 실시하기 위한 구체적인 내용

[0022] 이하에서는, 본 출원의 기술 분야에서 통상의 지식을 가진 자가 본 출원의 기술적 사상을 용이하게 실시할 수 있을 정도로 상세히 설명하기 위하여, 본 출원의 실시 예들이 첨부된 도면을 참조하여 좀 더 자세히 설명될 것이다.

[0023] 도 1은 본 출원의 실시 예에 따른 전략 수요량 예측 장치(10)를 보여주는 블록도이다.

[0024] 도 1을 참조하면, 전략 수요량 예측 장치(10)는 데이터 수집부(100), 데이터 전처리부(200), 이상치 탐지부(300), 이상치 복원부(400) 및 전력수요 예측부(500)를 포함한다.

[0025] 데이터 수집부(100)는 전력 수요량 예측에 필요한 각종 데이터를 수집한다. 예를 들어, 데이터 수집부(100)는 기상 데이터, 전력 데이터 및 시간 데이터를 수집할 수 있다.

[0026] 예를 들어, 데이터 수집부(100)는 외부로부터 각종 기상에 대한 데이터를 수집할 수 있다. 데이터 수집부(100)가 수집하는 기상 데이터는, 예를 들어, 일 최고 기온, 일 최저 기온, 기온, 습도, 풍속, 전운량, 강수량에 대한 데이터일 수 있다. 데이터 수집부(100)는, 예를 들어, 기상청의 기상자료개방포털 등을 통하여 기상 데이터를 수집할 수 있다.

[0027] 예를 들어, 데이터 수집부(100)는 적어도 하나의 클러스터로부터 소모한 전력량에 대한 전력 데이터를 수집할 수 있다. 이 경우, 클러스터는 데이터의 편향을 방지하기 위하여 서로 다른 용도의 건물일 수 있다. 예를 들어, 데이터 수집부(100)는 교육용 건물들로 이루어진 클러스터 A, 기숙사들로 이루어진 클러스터 B, 공과대학 연구실로 이루어진 클러스터 C 및 이과대학 연구실로 이루어진 클러스터 D로부터 전력 데이터를 수집할 수 있다. 다만, 이는 예시적인 것이며, 클러스터의 개수 및 종류는 다양하게 설정될 수 있다.

[0028] 예를 들어, 데이터 수집부(100)는 각종 기상 및 전력 데이터를 수집할 때에, 이에 대응하는 시간 데이터를 함께 수집할 수 있다. 예를 들어, 데이터 수집부(100)는 기상 데이터를 수집할 때에 월(month), 일(day), 시(hour), 분(minute)에 대한 시간 데이터를 함께 수집할 수 있다. 또한, 데이터 수집부(100)는 전력 데이터를 수집할 때에 월(month), 일(day), 시(hour), 분(minute)에 대한 시간 데이터를 함께 수집할 수 있다.

[0029] 본 출원의 일 실시 예에 있어서, 데이터 수집부(100)는 실제로 측정된 기상 데이터 뿐만 아니라, 과거 시점으로부터 소정 기간 전에 예보된 기상 예보 데이터를 함께 수집할 수 있다. 예를 들어, 2018년 5월 5일에 소모된 전력 데이터 및 기상 데이터를 수집할 때에, 해당 2018년 5월 5일에 대응하는 기상 데이터는 하루 전인 2018년 5월 4일에 예보된 일 최소 기온, 일 최고 기온, 일 평균기온, 기온, 습도, 풍속, 전운량, 강수량 등에 대한 기상 예보 데이터일 수 있다. 이와 같이, 실제 측정된 기상 데이터 뿐만 아니라 기상 예보 데이터를 함께 수집하고, 이를 학습 모델 구현을 위한 학습 데이터로 제공함으로써, 본 출원의 실시 예에 따른 전력 수요량 예측 장치(10)는 기상 예보 데이터와 실제 기상 데이터 사이의 오차까지 함께 고려하여 학습 동작을 수행할 수 있다.

[0030] 다만, 이는 예시적인 것이며, 데이터 수집부(100)는 실제로 측정된 기상 데이터만을 수집할 수도 있으며, 이 경우에 학습 동작은 실제 측정된 기상 데이터만을 이용하여 수행될 수 있다. 다른 예로, 데이터 수집부(100)는 기상 예보 데이터만을 수집할 수 있으며, 이 경우에 학습 동작은 기상 예보 데이터만을 이용하여 수행될 수도 있다.

[0031] 데이터 전처리부(200)는 데이터 수집부(100)로부터 기상 데이터, 시간 데이터 및 전력 데이터를 수신한다. 데이터 전처리부(200)는 이상 탐지부(300)에서 활용될 수 있도록 수신한 기상 데이터, 전력 데이터 및 시간 데이터에 대한 전처리 동작을 수행한다.

[0032] 예를 들어, 데이터 전처리부(200)는 데이터 수집부(100)로부터 기상 데이터 및 전력 데이터를 수신하고, 수신한 기상 데이터 및 전력 데이터에 대한 정규화 동작을 수행할 수 있다. 또한, 데이터 전처리부(200)는 데이터 수집부(100)로부터 1차원의 시간 데이터를 수신하고, 수신한 1차원의 시간 데이터를 2차원의 시간 데이터로 변환할 수 있다. 데이터 전처리부(200)의 구성 및 동작은 이하의 도 2 및 도 3에서 좀 더 자세히 설명될 것이다.

[0033] 이상치 탐지부(300)는 데이터 전처리부(200)로부터 전처리된 전력 데이터를 수신한다. 이상치 탐지부(300)는 전처리된 전력 데이터에서 이상치를 탐지할 수 있다.

[0034] 본 출원의 일 실시 예에 있어서, 이상치 탐지부(300)는 변분 오토인코더(Variational Autoencoder, VAE)를 통하여 이상치를 탐지할 수 있다. 변분 오토인코더는 입력 값의 분포를 학습하여 출력 값을 생성하기 때문에, 이상치 탐지부(300)는 일반적인 전력 수요 분포에서 벗어난 이상치를 좀 더 잘 탐지할 수 있다. 이상치 탐지부(300)는 이하의 도 4 및 도 5에서 좀 더 자세히 설명될 것이다.

[0035] 이상치 복원부(400)는 이상치 탐지부(300)로부터 이상 데이터를 수신한다. 이상치 복원부(400)는 이상 데이터에 대응하는 시점의 다른 입력 변수를 기반으로 하여, 이상 데이터를 복원하여 복원 데이터를 생성한다.

[0036] 본 출원의 일 실시 예에 있어서, 이상치 복원부(400)는 랜덤 포레스트(RF) 모델을 통하여 이상치 데이터로부터 복원 데이터를 복원할 수 있다. 이상치 복원부(400)는 이하의 도 6 및 도 7에서 좀 더 자세히 설명될 것이다.

[0037] 전력수요 예측부(500)는 이상치 복원부(500)로부터 복원 데이터를 포함하는 입력 데이터를 수신한다. 여기서, 입력 데이터는 데이터 전처리부(200)에서 전처리된 데이터 및 이상치 복원부(400)에서 복원된 데이터를 포함한

다. 전력수요 예측부(500)는 입력 데이터를 사용하여 전력수요량 예측을 위한 모델을 학습시킬 수 있다.

[0038] 본 출원의 일 실시 예에 있어서, 전력수요 예측부(500)는 슬라이딩 윈도우(sliding window) 기반의 Light GBM 모델을 통하여 구현될 수 있다. 전력수요 예측부(500)는 이하의 도 8 및 도 9에서 좀 더 자세히 설명될 것이다.

[0039] 상술한 바와 같이, 본 출원의 실시 예에 따른 전력 수요량 예측 장치(10)는 변분 오토인코더를 통하여 이상치를 탐지하고, 랜덤 포레스트 모델을 통하여 이상치를 복원한다. 따라서, 데이터 수가 충분하지 않은 상황에서도 안정적인 학습 데이터 제공이 가능하며, 이에 따라 전력 수요량 예측 장치(10)는 오버 피팅 없이 향상된 예측 성능을 제공할 수 있다. 아울러, 전력 수요량 예측 장치(10)는 슬라이딩 윈도우 기반의 Light GBM 모델을 통하여 예측 모델을 구현하며, 이에 따라 예측 시점과 가까운 최신 데이터 패턴을 적절히 반영할 수 있어서 예측 성능을 더욱 향상시킬 수 있다.

[0040] 도 2는 도 1의 데이터 전처리부(200)를 좀 더 자세히 보여주는 블록도이다.

[0041] 도 2를 참조하면, 데이터 전처리부(200)는 기상 데이터 전처리부(210), 전력 데이터 전처리부(220) 및 시간 데이터 전처리부(230)를 포함한다.

[0042] 기상 데이터 전처리부(210)는 데이터 수집부(100)로부터 기상 데이터(Data_W)를 수신한다. 기상 데이터 전처리부(210)는 수신된 기상 데이터(Data_W)에 대한 전처리 동작을 수행하여 제1 기상 데이터(Data_W1)를 생성한다.

[0043] 예를 들어, 기상 데이터 전처리부(210)는 데이터 수집부(100)로부터 일 최고 기온, 일 최저 기온, 기온, 습도, 풍속, 전운량, 강수량 등의 데이터를 수신할 수 있다. 기상 데이터 전처리부(210)는 수신된 기상 데이터들에 대한 정규화 동작을 수행하는 것을 통해 제1 기상 데이터(Data_W1)를 생성할 수 있다.

[0044] 또한, 기상 데이터 전처리부(210)는 수신된 기온, 습도, 풍속 데이터에 기초하여 체감 온도 데이터와 불쾌지수 데이터를 생성할 수 있다. 이후, 기상 데이터 전처리부(210)는 체감 온도 데이터와 불쾌지수 데이터를 정규화하여 제2 기상 데이터(Data_W2)를 생성할 수 있다.

[0045] 전력 데이터 전처리부(220)는 데이터 수집부(100)로부터 적어도 하나의 클러스터에서 소모된 전력량에 대한 전력 데이터를 수신할 수 있다. 전력 데이터 전처리부(220)는 수신된 전력 데이터에 대한 전처리 동작을 수행하여 전처리된 전력 데이터(Data_Pp)를 생성할 수 있다.

[0046] 시간 데이터 전처리부(230)는 데이터 수집부(100)로부터 시간 데이터(Data_T)를 수신한다. 여기서, 시간 데이터는 범주형 데이터를 반영하는데 좋은 1차원 데이터일 수 있다. 시간 데이터 전처리부(230)는 1차원의 시간 데이터(Data_T)를 그대로 출력하거나, 1차원의 시간 데이터(Data_T)를 주기성 정보를 반영할 수 있는 2차원의 시간 데이터((Data_T1, Data_T2)로 변환할 수 있다.

[0047] 좀 더 자세히 설명하면, 1차원 시간 데이터는 1월, 2월과 같은 월(month)을 나타내는 범주형 정보, 1일, 2일과 같이 일(day)을 나타내는 범주형 정보 및 1시, 2시와 같이 시(hour)를 나타내는 범주형 정보는 잘 반영한다. 그러나, 1차원 시간 데이터는 주기성 정보는 잘 반영하지 못한다는 문제가 있다. 예를 들어, 23시와 0시는 연속적인 시간임에도 불구하고, 1차원 데이터 상으로는 23의 차이가 발생하게 된다.

[0048] 따라서, 시간의 주기성 정보가 잘 반영될 수 있도록, 시간 데이터 전처리부(230)는 1차원의 시간 데이터(Data_T)를 2차원의 시간 데이터((Data_T1, Data_T2)로 변환할 수 있다. 이때, 시간 데이터 전처리부(230)는 사인(sin) 함수와 코사인 함수(cos)와 같은 주기 함수를 통하여 변환 동작을 수행할 수 있다. 예를 들어, 시간 데이터 전처리부(230)는 다음의 수식을 통하여 1차원의 시간 데이터(Data_T)를 2차원의 시간 데이터(Data_T1, Data_T2)로 변환할 수 있다.

$$time_x = \sin(time \times (2\pi/cycle))$$
$$time_y = \cos(time \times (2\pi/cycle))$$

[0049]

[0050] 여기서, "cycle"은 시간 데이터의 주기를 나타낸다. 예를 들어, "time"이 월(month) 데이터인 경우에 "cycle"은 "12"일 수 있고, "time"이 일(day) 데이터인 경우에 "cycle"은 해달 월의 일수(Day of the Month)일 수 있으며, "time"이 시(hour) 데이터인 경우에 "cycle"은 "24"일 수 있다. 또한, 여기서, "$time_x$"와 "$time_y$"는 각각 도 2의 Data_T1과 Data_T2에 대응할 수 있다.

[0051] 한편, 시간 데이터 전처리부(230)가 2개의 삼각함수 값을 통해 2차원으로 표현하는 이유는, 예를 들어 주기가 12인 하나의 삼각함수 값을 통해 표현할 경우에는 두 개의 x값에 대해서 같은 y값이 결정되며, 이 경우에 y값만으로 시기를 특정하기 어렵기 때문이다. 따라서, 동일한 x값이라도 서로 다른 y값을 갖는 두 개의 삼각함수를 사용하여 이러한 문제를 해결할 수 있도록, 시간 데이터 전처리부(230)는 2개의 삼각함수를 이용하여 2차원 데이터(Data_T1, Data_T2)를 생성한다.

[0052] 상술한 바와 같이, 데이터 전처리부(200)는 기상 데이터(Data_W), 전력 데이터(Data_P) 및 시간 데이터(Data_T)에 대한 전처리 동작을 수행할 수 있다. 특히, 데이터 전처리부(200)는 특히 범주형 정보를 잘 반영하는 1차원의 시간 데이터(Data_T)와 주기성 정보를 잘 반영하는 2차원의 시간 데이터(Data_T1, Data_T2)를 함께 출력할 수 있다.

[0053] 도 3은 도 2의 시간 데이터 전처리부(230)에 의하여 출력되는 시간 데이터의 일 예를 보여주는 도면이다. 구체적으로, 도 3a는 범주형 시간 데이터(Data_T)의 일 예를 보여주며, 도 3b는 2차원 시간 데이터(Data_T1, Data_T2)의 일 예를 보여준다.

[0054] 도 3에 도시된 바와 같이, 시간 데이터 전처리부(230)는 시간 데이터(Data_T)에 대한 전처리 동작을 수행하여 2차원 시간 데이터(Data_T1, Data_T2)를 생성할 수 있다.

[0055] 도 4는 도 1의 이상치 탐지부(300)를 좀 더 자세히 보여주는 도면이고, 도 5는 이상치 탐지부(300)의 이상치 탐지 실험 결과를 보여주는 그래프이다.

[0056] 도 4를 참조하면, 이상치 탐지부(300)는 데이터 전처리부(200)로부터 전처리된 전력 데이터(Data_Pp)를 수신한다. 이상치 탐지부(300)는 전처리된 전력 데이터(Data_Pp)에서 이상치를 탐지하여, 정상 데이터(Normal Data)와 이상 데이터(Abnormal Data)로 구분할 수 있다.

[0057] 본 출원의 실시 예에 있어서, 이상치 탐지부(300)는 변분 오토인코더를 통하여 구현될 수 있다. 즉, 이상치 탐지부(300)는 전처리된 전력 데이터를 이용하여 변분 인코더를 학습하고, 학습된 변분 오토인코더를 통해 재구성된 출력 값과 입력 값의 차이를 이용하여 이상치를 탐지할 수 있다.

[0058] 만약 이상치 탐지부가 일반 오토인코더(AE)를 통하여 구현한다면, 오토인코더 기반의 이상 탐지부는 모든 입력 변수를 재구성한 후에 각 입력 변수들의 재구성 오류를 모두 더한다. 이후, 오토인코더 기반의 이상 탐지부는 재구성 오류의 합이 일정 값 이상이 될 경우에 이상치라고 판단한다. 그러나, 이 경우, 실제 전력 소모량에 이상치가 발생한 경우가 아닐 때에도, 오토인코더 기반의 이상치 탐지부는 이상치라고 판단할 수 있는 위험이 있다.

[0059] 예를 들어, 여름철에 장마로 인하여 갑작스럽게 어느 한 시점에 비가 많이 내렸다고 가정하자. 이 경우, 오토인코더에 기반의 이상 탐지부는 모든 입력 변수들의 재구성 오류를 고려하기 때문에, 전력 사용량을 비롯하여 재구성된 입력 변수들이 정상임에도 불구하고 재구성된 강수량 관련 변수에서 차이가 많이 발생하여 결과적으로 이상치라고 잘못 판단할 수 있다.

[0060] 이러한 오류 위험성을 방지하기 위하여, 본 출원의 실시 예에 따른 이상치 탐지부(300)는 변분 오토인코더를 통하여 구현될 수 있다. 이 경우, 모든 입력 변수의 재구성 오류를 더하는 것을 통해 이상치를 판단하는 것이 아니라, 전처리된 전력 데이터(Data_Pp)의 재구성 오류만을 이용하여 이상치가 판단되기 때문에, 본 출원의 실시 예에 따른 이상치 탐지부(300)는 좀 더 향상된 이상치 탐지 성능을 가질 수 있다.

[0061] 도 5를 참조하면, 본 출원의 실시 예에 따른 변분 오토인코더 기반의 이상치 탐지부(300)가 다른 모델에 비하여 좀 더 향상된 성능을 갖는 것을 확인할 수 있다. 도 5에서, IQR, IForest, LOF는 각각 사분위수 범위(interquartile range, IQR)를 활용한 이상치 탐지 모델, Isolation Forest 이상치 탐지 모델 및 Local Outlier Factor 이상치 탐지 모델을 의미한다.

[0062] 도 6은 도 1의 이상치 복원부(400)를 좀 더 자세히 보여주는 도면이고, 도 7은 이상치 복원부(400)의 복원 실험 결과를 보여주는 그래프이다.

[0063] 도 6을 참조하면, 이상치 복원부(400)는 이상치 탐지부(300)로부터 정상 데이터(Normal Data)와 이상 데이터(Abnormal Data)를 수신한다. 이상치 복원부(400)는 이상 데이터가 탐지된 시점의 다른 입력 변수를 기반으로 하여 이상 데이터를 복원 데이터(Repair Data)로 복원할 수 있다.

[0064] 본 출원의 일 실시 예에 있어서, 이상치 복원부(400)는 랜덤 포레스트(RF) 모델을 사용하여 구현될 수 있다.

즉, 이상치 복원부(400)는 정상 데이터를 사용하여 랜덤 포레스트 모델을 학습하고, 학습된 랜덤 포레스트 모델에 이상치가 발생한 시점의 입력 변수들을 넣었을 때에 도출된 값으로 복원 데이터를 생성할 수 있다. 랜덤 포레스트 모델은 부분 데이터 셋을 사용하여 각각의 의사 결정 트리를 학습한다. 따라서, 이상치 복원부(400)는 오버 피팅이 발생하지 않을 뿐 아니라, 입력 변수가 많음에도 불구하고 향상된 성능을 가질 수 있다.

[0065] 도 7을 참조하면, 본 출원의 실시 예에 따른 랜덤 포레스트 기반의 이상치 복원부(400)가 다른 모델에 비하여 좀 더 향상된 성능을 갖는 것을 확인할 수 있다. 도 7에서, Zero, Linear, RF 는 각각 zero interpolation을 사용하여 이상치를 복원하였을 때, Linear interpolation을 사용하여 이상치를 복원하였을 때 및 Random Forest를 사용하여 이상치를 복원하였을 때를 나타낸다.

[0066] 도 8은 도 1의 전력 수요 예측부(500)를 좀 더 자세히 보여주는 도면이고, 도 9는 전력 수요 예측부(500)의 윈도우 사이즈 결정을 위한 실험 결과를 보여주는 그래프이다.

[0067] 도 8을 참조하면, 전력 수요 예측부(500)는 이상치 복원부(400)부로부터 전력 소모량에 대한 정상 데이터 및 복원 데이터를 수신하고, 데이터 전처리부(200)로부터 제1 및 제2 기상 데이터(Data_W1, Data_W2), 범주형 시간 데이터(Data_T) 및 주기성 시간 데이터(Data_T1, Data_T2)를 수신할 수 있다. 전력 수요 예측부(500)는 수신된 입력 데이터들에 기초하여, 전력 수요량 예측 모델을 학습하고, 학습된 전력 수요량 예측 모델을 통하여 미래의 전력 수요량을 예측할 수 있다.

[0068] 본 출원의 일 실시 예에 있어서, 전력 수요 예측부(500)는 슬라이딩 윈도우(sliding window) 기반의 Light GBM 모델을 통하여 구현될 수 있다. 슬라이딩 윈도우 기법을 적용함으로써, 전력 수요 예측부(500)는 최신 추세 및 패턴을 적절하게 반영할 수 있다.

[0069] 예를 들어, 도 8에서, 점들은 각각 하루치의 입력 데이터를 의미할 수 있다. 이 경우, 전력 수요 예측부(500)는 예측하고자 하는 시점으로부터 이전 일주일의 입력 데이터를 사용하기 때문에, 최신 데이터를 반영할 수 있어 좋은 성능을 가질 수 있다.

[0070] 본 출원의 일 실시 예에 있어서, 전력 수요 예측부(500)는 윈도우 사이즈(window size)로 '7'을 설정할 수 있다. 도 9를 참조하면, 윈도우 사이즈가 '1'에서부터 '7'로 증가하는 동안에는 그 성능이 향상되지만, 윈도우 사이즈가 '7'을 초과하는 경우부터는 성능 향상은 미미한 반면에 학습 시간만 늘어나는 것을 확인할 수 있다. 이는 전력 수요량에 대한 데이터가 일정 부분 주기적 패턴으로 반복되는 경향을 보이기 때문이다. 즉, 예측하고자 하는 시점이 월요일인 경우, 입력 변수로 지난주 월요일의 데이터를 입력하면 주기적 패턴을 반영할 수 있어서 성능이 향상될 수 있다. 본 출원의 실시 예에 따른 전력 수요 예측부(500)는 전력 수요량에 대한 이러한 주기적 패턴을 반영하기 위하여, 윈도우 사이즈를 '7'로 설정한다.

[0071] 한편, 슬라이딩 윈도우 기법을 적용하는 경우, 예측하고자 하는 시점마다 모델을 새롭게 구성해야 하는 단점이 존재한다. 본 출원의 실시 예에 따른 전력 수요 예측부(500)는 이러한 단점을 보완하기 위하여 모델 구성 속도가 빠르면서도 예측 성능이 뛰어난 Light GBM 모델을 사용하여 구현될 수 있다. 이 경우, 전력 수요 예측부(500)는 데이터 중 기울기가 큰 부분만을 사용하여 정보를 얻는 GOSS(Gradient-based One Side Sampling) 기법과 상호 배타적 변수들을 묶어서 처리하는 EFB(Exclusive Featuree Bundling) 기법을 사용하여 모델의 구성 속도를 좀 더 빠르게 할 수 있다.

[0072] 도 10은 본 출원의 실시 예에 따른 전력 수요량 예측 장치(10)의 예측 결과를 다른 모델과 비교하는 그래프이다.

[0073] 도 10에 도시된 바와 같이, 서로 다른 클러스터 4개에 대한 전력 수요를 예측한 결과, 전력 수요량 예측 장치(10)의 MAPE(Mean Absolute Percentage Error)는 각각 4.545%, 3.755%, 2.7%, 2.144%를 기록하였으며, 다른 모델에 비하여 뛰어난 예측 성능을 갖는 것을 확인할 수 있다.

[0074] 도 11은 도 1의 전력 수요량 예측 장치(10)의 동작을 보여주는 순서도이다.

[0075] S110 단계에서, 데이터 수집부(100)는 기상 데이터, 시간 데이터 및 전력 데이터를 수집할 수 있다.

[0076] S120 단계에서, 데이터 전처리부(200)는 수집된 기상 데이터, 시간 데이터 및 전력 데이터에 대한 전처리 동작을 수행할 수 있다. 예를 들어, 데이터 전처리부(200)는 기상 데이터 및 전력 데이터에 대한 정규화 동작을 수행할 수 있다. 예를 들어, 데이터 전처리부(200)는 1차원의 시간 데이터를 2차원의 시간 데이터로 변환할 수 있다.

[0077] S130 단계에서, 이상치 탐지부(300)는 전처리된 전력 데이터 중 이상치를 탐지하는 동작을 수행할 수 있다. 예를 들어, 이상치 탐지부(300)는 전력 데이터를 이상 데이터와 정상 데이터로 분류할 수 있으며, 이상치 탐지부(300)는 변분 오토인코더(VAE)를 통하여 구현될 수 있다.

[0078] S140 단계에서, 이상치 복원부(400)는 이상 데이터를 복원할 수 있다. 예를 들어, 이상치 복원부(400)는 랜덤 포레스트 모델을 통하여 이상 데이터를 복원하여 복원 데이터를 생성할 수 있다.

[0079] S150 단계에서, 전력 수요 예측부(500)는 복원 데이터를 포함하는 입력 데이터에 기초하여 예측 모델을 학습할 수 있다. 예를 들어, 전력 수요 예측부(500)는 슬라이딩 윈도우 기반 LightGBM을 통하여 구현될 수 있다. 학습된 예측 모델을 통하여, 전력 수요 예측부(500)는 미래 시점의 전력 수요량을 정확하게 예측할 수 있다.

[0080] 이상에서는 본 발명에 따른 바람직한 실시 예들에 대하여 도시하고 또한 설명하였다. 그러나 본 발명은 상술한 실시 예에 한정되지 아니하며, 특허 청구의 범위에서 첨부하는 본 발명의 요지를 벗어남이 없이 당해 발명이 속하는 기술 분야에서 통상의 지식을 가진 자라면 누구든지 다양한 변형 실시가 가능할 것이다.

부호의 설명

[0081] 100: 데이터 수집부

200: 데이터 전처리부

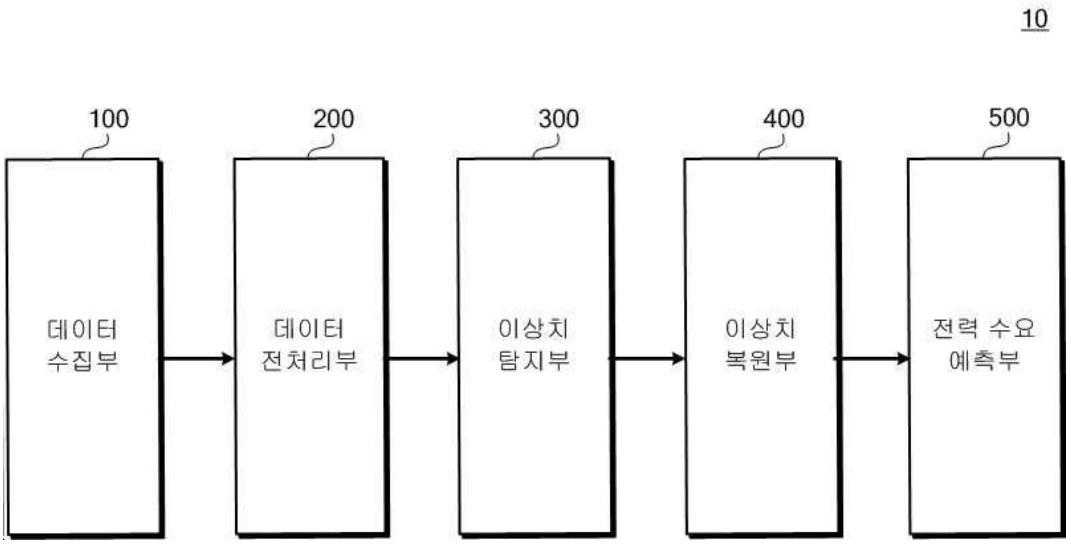
300: 이상치 탐지부

400: 이상치 복원부

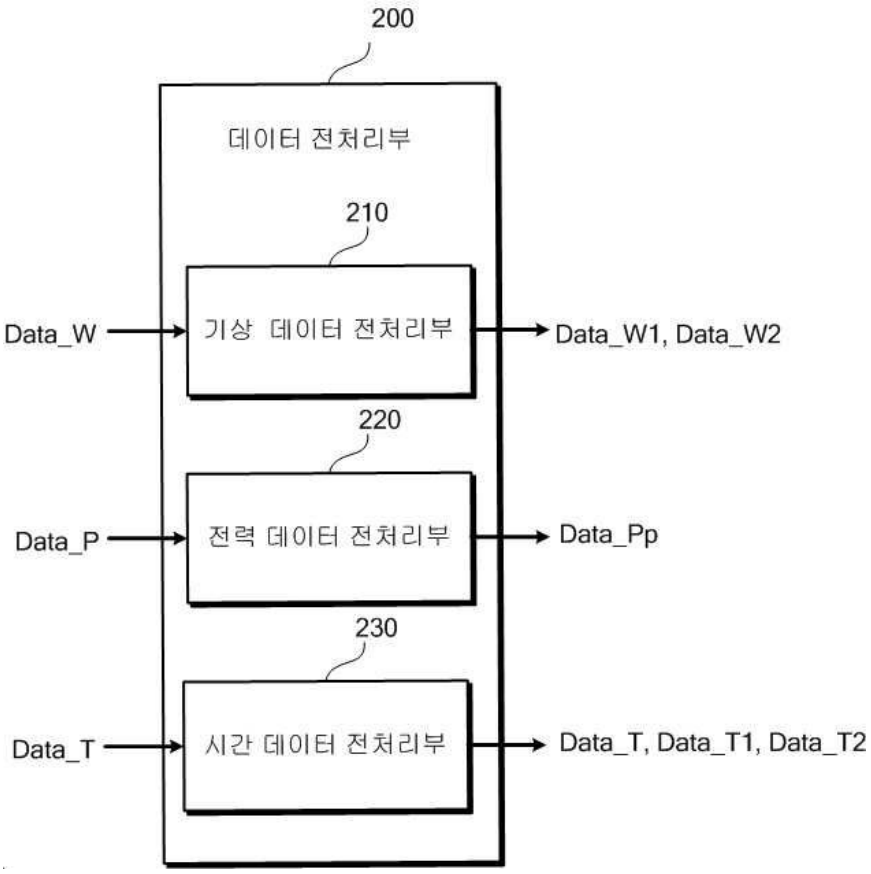
500: 전력 수요 예측부

도면

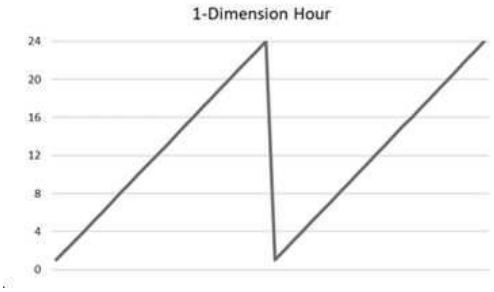
도면1



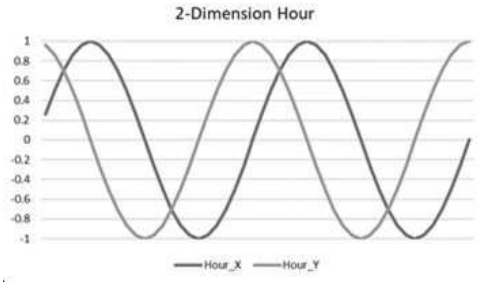
도면2



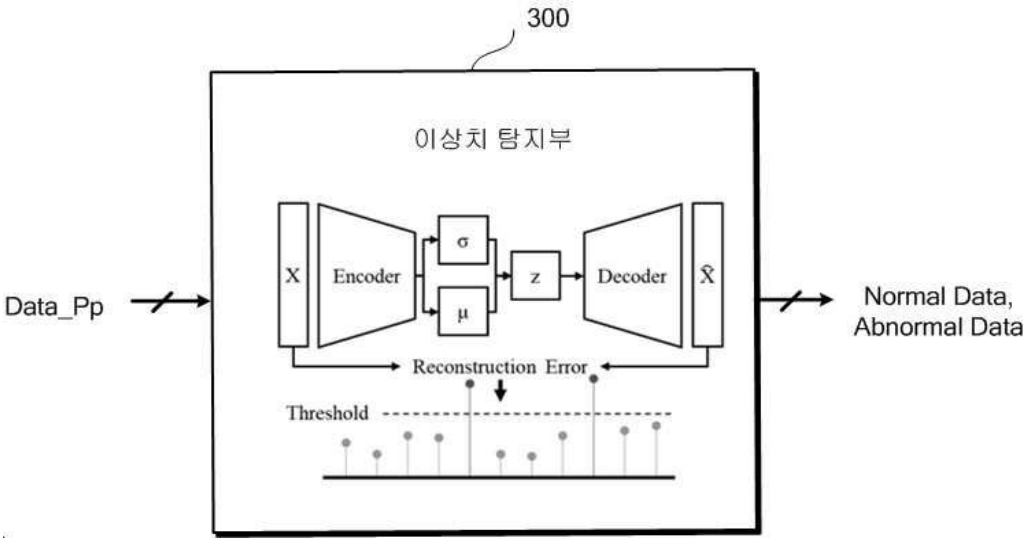
도면3a



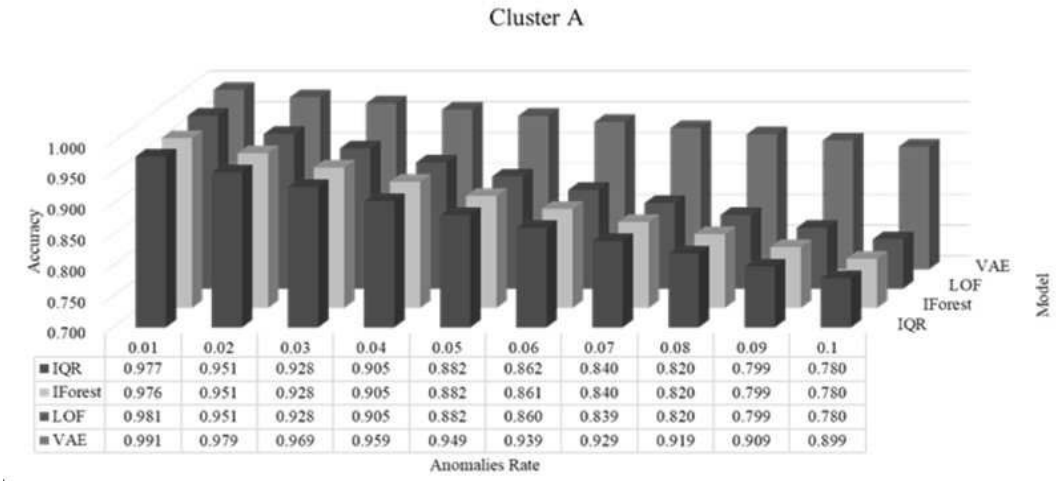
도면3b



도면4

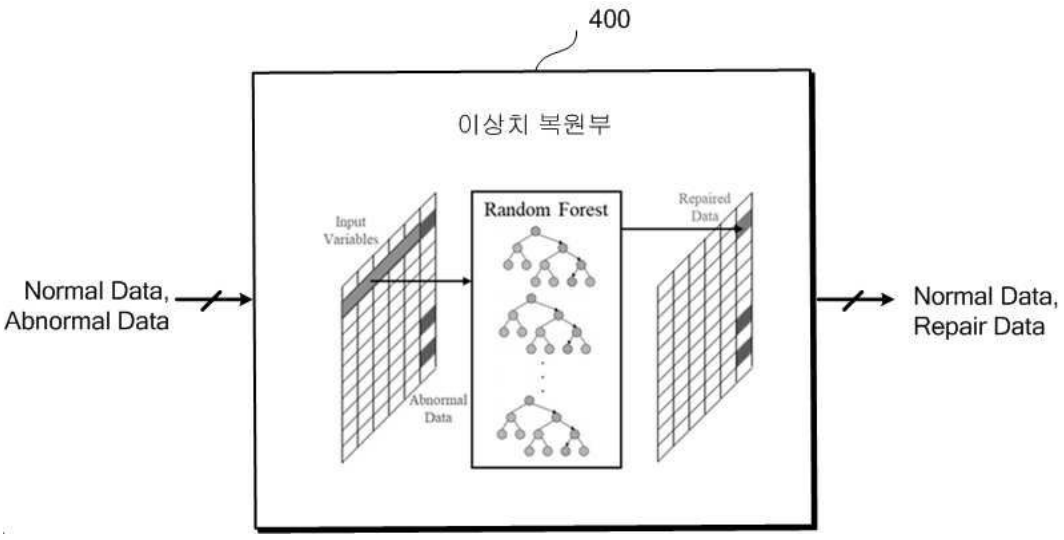


도면5

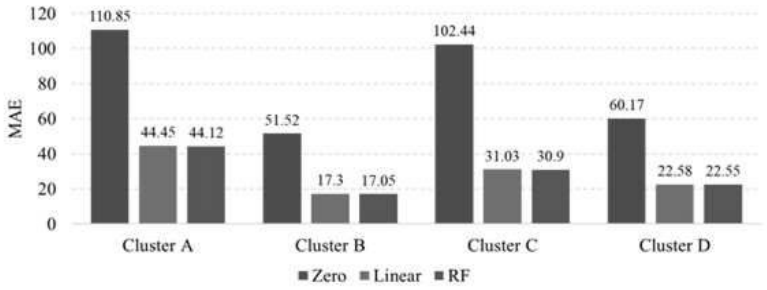


	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
IQR	0.977	0.951	0.928	0.905	0.882	0.862	0.840	0.820	0.799	0.780
IForest	0.976	0.951	0.928	0.905	0.882	0.861	0.840	0.820	0.799	0.780
LOF	0.981	0.951	0.928	0.905	0.882	0.860	0.839	0.820	0.799	0.780
VAE	0.991	0.979	0.969	0.959	0.949	0.939	0.929	0.919	0.909	0.899

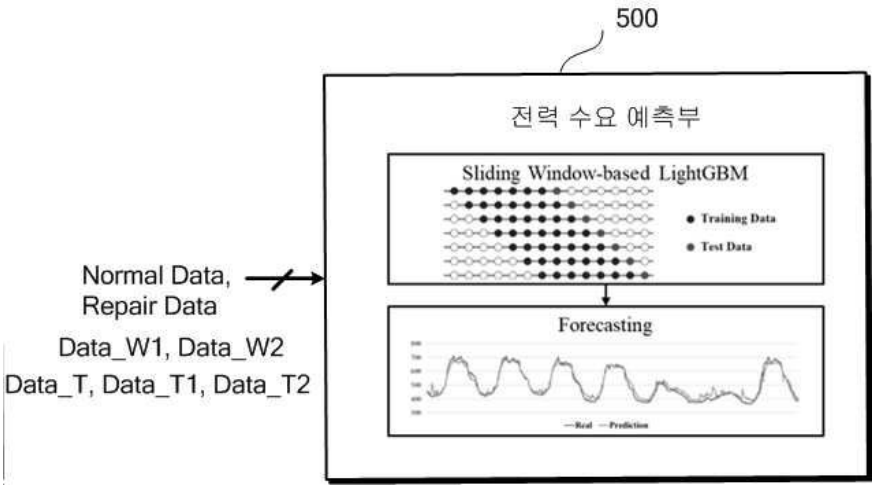
도면6



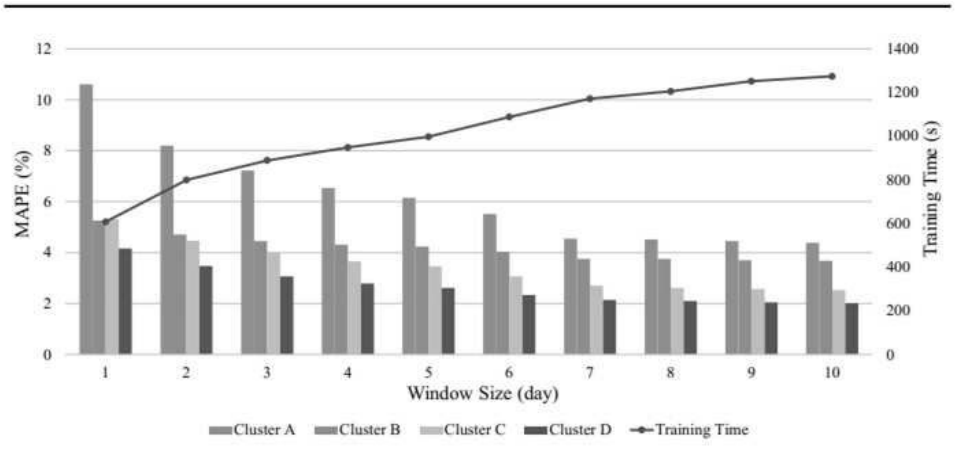
도면7


120
100
80
60
40
20
0
MAE
110.85
44.45
44.12
51.52
17.3
17.05
102.44
31.03
30.9
60.17
22.58
22.55
Cluster A
Cluster B
Cluster C
Cluster D
Zero
Linear
RF

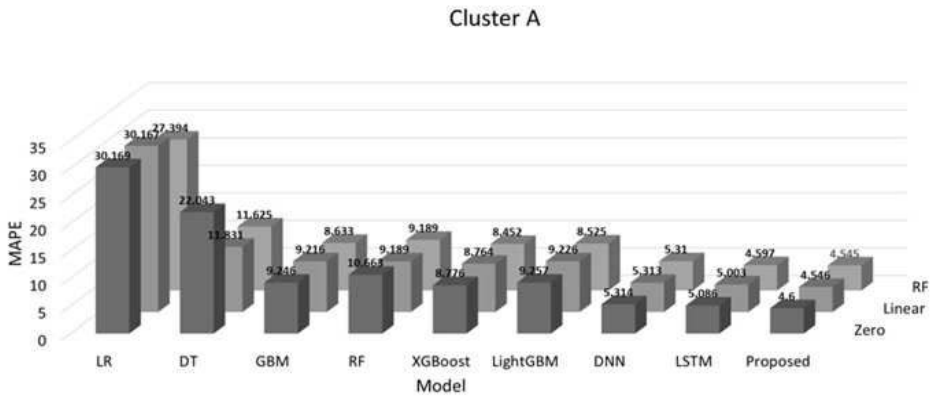
도면8


500
전력 수요 예측부
Sliding Window-based LightGBM
Training Data
Test Data
Forecasting
Real
Prediction
Normal Data,
Repair Data
Data_W1, Data_W2
Data_T, Data_T1, Data_T2

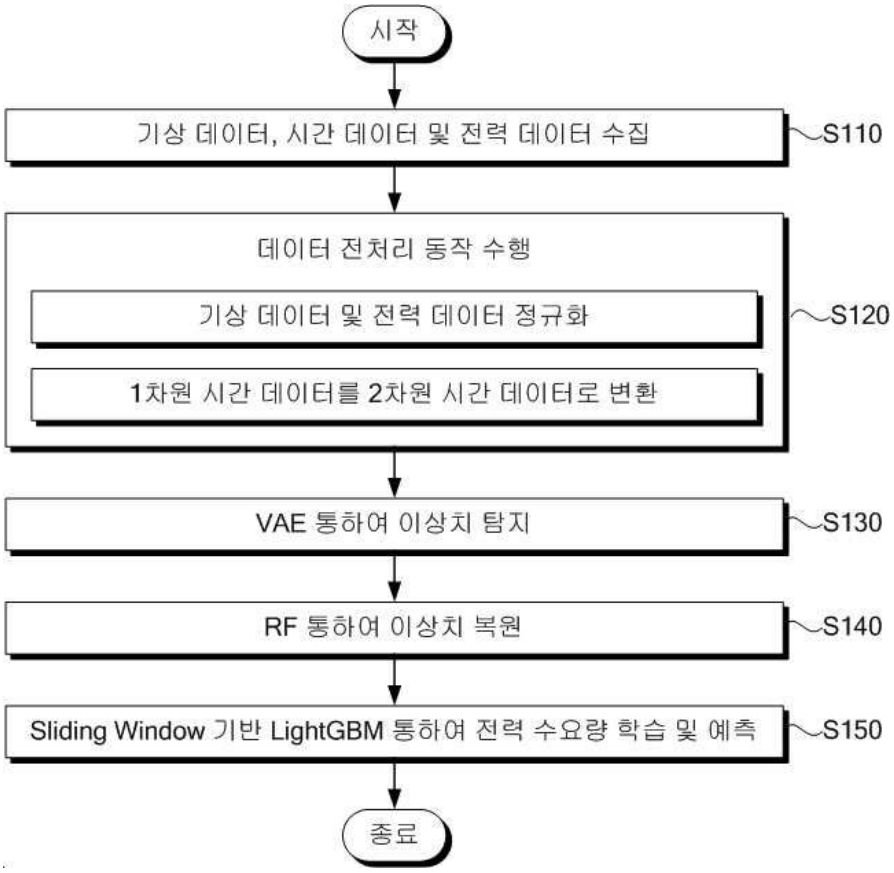
도면9


12
10
8
6
4
2
0
MAPE (%)
1400
1200
1000
800
600
400
200
0
Training Time (s)
1
2
3
4
5
6
7
8
9
10
Window Size (day)
Cluster A
Cluster B
Cluster C
Cluster D
Training Time

도면10


Cluster A
MAPE
35
30
25
20
15
10
5
0
30.169
30.167
27.394
22.043
13.231
11.625
9.216
8.633
9.216
10.563
9.189
8.776
9.189
8.764
8.452
9.257
9.226
8.525
5.313
5.311
5.31
5.086
5.003
4.6
4.597
4.546
4.545
RF
Linear
Zero
LR
DT
GBM
RF
XGBoost
LightGBM
DNN
LSTM
Proposed
Model

도면11


시작
기상 데이터, 시간 데이터 및 전력 데이터 수집
S110
데이터 전처리 동작 수행
기상 데이터 및 전력 데이터 정규화
1차원 시간 데이터를 2차원 시간 데이터로 변환
S120
VAE 통하여 이상치 탐지
S130
RF 통하여 이상치 복원
S140
Sliding Window 기반 LightGBM 통하여 전력 수요량 학습 및 예측
S150
종료