

기상정보를 활용한 LSTM 기반 태양광 발전량 예측 기법

2021 winter team
안혜림

기상정보를 활용한 LSTM 기반 태양광 발전량 예측 기법

요약

태양광 발전 관리(효율적 운영, 경제성 향상)를 위한 정확한 태양광 발전량 예측 필요



시계열 데이터 분석 및 LSTM(Long Short-Term Memory)과
기온, 습도, 전운량, 자외선 지수 등의 기상정보를 통해
시간 단위의 태양광 발전량을 예측하는 기법 제안



정해진 시점이 아닌 원하는 구간의 발전량 예측 -> **가용성** 높음
태양광 발전량과 밀접한 관계 있는 **일사량** 먼저 예측 -> 발전량 예측 **정확도** ↑



LSTM 기반 예측 모델의 발전량 예측 결과
오차율인 MAE : 1.5424 / NMAE : 0.0454이며,

DNN 기반 예측 모델의 MAE : 1.9374 / NMAE : 0.0569이므로
LSTM 기반이 DNN에 비해 오차율이 낮은 것을 확인

기상정보를 활용한 LSTM 기반 태양광 발전량 예측 기법

서론

태양광 발전은 외부 조건에 따라 변동이 심함

-> 기존의 전력 발전과 태양광 발전 사이의 전력 생산 균형을 맞추고 전력 공급 계획을 효율적으로 세우기 위해
정확한 태양광 발전량 예측이 필요

[진행중인 예측 연구]

1. Sharma 등 - SVM(서포트 벡터 머신)

▶ 단일 차원 벡터에 비해 약 27% 성능 향상했으나, 3시간 보다 이후의 시점이나 어떤 구간에 대한 발전량 예측X

2. 이승민 등 - 신경망, SVM, 딥러닝

▶ 일별 발전량과 기상정보 데이터 학습하여 하루 단위 발전량 예측, 여러 개의 은닉층 가지는 심층 신경망이 가장 높은 예측 정확도 보임

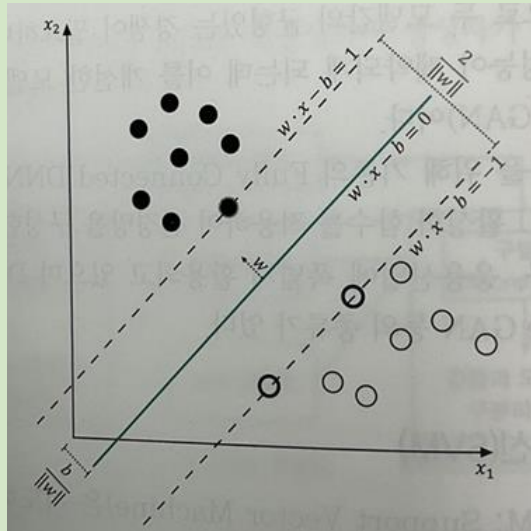
서포트벡터머신(SVM: Support Vector Machine)

: 지도학습 기법으로 고차원 또는 무한 차원의 공간에서 초평면(의 집합)을 찾아 이를 이용하여 분류와 회귀를 수행

'여백(마진) 최대화'로 일반화 능력의 극대화 추구

* SVM의 주요 요소

- ① **벡터(Vector)**
점들 간 클래스(class)
- ② **결정영역(Decision Boundary)**
클래스들을 잘 분류하는 선
- ③ **초평면(Hyperplane)**
서로 다른 분류에 속한 데이터들 간 거리를 가장 크게 하는 분류 선
- ④ **서포트벡터(Support Vector)**
두 클래스를 구분하는 경계선
- ⑤ **마진(Margin)**
서포트벡터를 지나는 초평면 사이의 거리



* 마진이 가장 큰 초평면을 분류기로 사용할 때, 새로운 자료에 대한 오분류가 가장 낮아짐

* 그림에서 가운데 직선이 초평면이면 양쪽 점선이 서포트 직선. 직선을 두개 정의하는데 두 직선 다 초평면과 같은 법선 벡터를 가지고 있고 1만큼 거리를 둔 식으로 $w \cdot x - b = 1$, $w \cdot x - b = -1$ 로 나타냄.

* 초평면의 마진은 각 서포트 벡터를 지나는 초평면 사이의 거리를 의미. 기하학적 의미는 두 초평면 사이의 거리, 즉 $2/||w||$ 라는 것을 알 수 있으며 마진을 최대화해야 하므로 w 의 크기가 최소가 되어야 함

기상정보를 활용한 LSTM 기반 태양광 발전량 예측 기법

서론

3. Jidong 등 - FOS-ELM 알고리즘

- ▶ 단기간 태양광 발전량 예측, 과거 기상정보와 과거 발전량 데이터 이용하여 미래 발전량 예측 프로세스 적용

4. 배국열 등 - ANN, SVM

- ▶ 데이터 클러스터링, 학습 및 테스트 단계를 통해 발전량 예측, 날씨 예측오차의 태양광 출력 예측 정확도에 끼치는 영향 분석

인공신경망(ANN : Artificial Neural Network)

: 인간의 두뇌 신경세포인 뉴런을 기본으로 한 기계학습 기법으로 하나의 뉴런이 다른 뉴런들과 연결되어 신호를 전달, 처리하는 구조를 본뜬
입력데이터가 들어가면서 신호의 강도에 따라 가중치 처리되고 활성화 함수를 통해 출력이 계산되는데 학습을 거쳐 원하는 결과가 나오게끔 가중치가 조정됨

* 인공신경망 원리

지도학습의 경우 하나의 뉴런은 입력 값(X)과 목표 출력 값(Y)이 있을 때 다음 뉴런으로 전달하는데 적절한 출력 값을 생성하기 위해 가중치 W를 곱한 값에 편향(bias)을 더하여 이를 조정하면서 학습 최적화 과정을 거치게 되며 최종적으로 활성화 함수를 활용함.

지도학습	학습데이터로 입력벡터와 함께 기대되는 출력벡터, 즉 답을 제시한다. 신경망에서 출력된 결과가 기대되는 출력과 다르면, 그 차이를 줄이는 방향으로 연결가중치를 조절한다.
비지도학습	학습 벡터에 목표가 없을 때, 학습 데이터의 관계를 추론하여 학습을 진행하는 방식으로 입력벡터들을 집단으로 그룹핑하여 해당 집단을 대표하는 데이터를 선정한다.
강화학습	특정 환경 안에서 에이전트가 현재 상태를 인식하여 보상을 최대화하는 방향으로 동작을 선택하는 방법이다.

* 인공신경망의 주요 요소

노드 : 신경계 뉴런, 가중치와 입력값으로 활성화함수를 통해 다음 노드로 전달

가중치 : 신경계 시냅스, 노드와의 연결계수 · 활성화함수 : 임계값을 이용, 노드

의 활성화 여부를 결정

입력층 : 학습 위한 데이터 입력

은닉층 : 다층 네트워크에서 입력층과 출력층 사이, 데이터를 전파학습

출력층 : 결과값 출력

* 뉴런 간의 연결 방법

층간 연결 : 서로 다른 층에 존재하는 뉴런과 연결

층내 연결 : 동일 층 내의 뉴런과의 연결

순환 연결 : 어떠한 뉴런의 출력이 자기 자신에게 입력되는 연결

기상정보를 활용한 LSTM 기반 태양광 발전량 예측 기법

서론

5. Pedro 등 - **ARIMA** 및 KNN, ANN

▶ 태양광 발전량 예측 진행 후 비교 및 평가

자동회귀누적이동평균(ARIMA : Auto Regressive Integrated Moving Average)

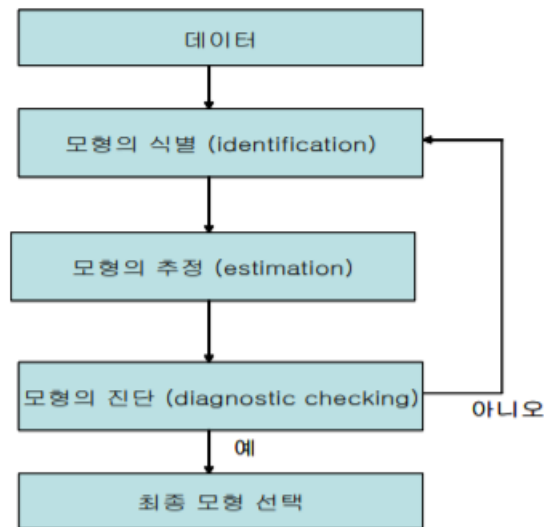
: 자동회귀이동평균(ARMA) 모델의 일반화로 미래 지점을 예측하기에 적합. 데이터가 비정상성이 아닌 증거를 나타내는 경우에 적용되며, 초기 차분 단계를 한번 이상 적용하여 비정상성을 제거할 수 있음.

* **AR** - 진화하는 관심 변수가 시차(이전)값으로 회귀됨

I(누적) - 데이터 값이 해당 값과 이전 값의 차이로 바뀌었음

MA - 회귀 오류가 실제로 과거 여러 시간에 동시에 발생한 오류 항의 선형 조합

8.2 ARIMA modeling 과정



① 모형의 식별

i) 정상시계열과 비정상시계열의 특징

정상시계열	비정상시계열
뚜렷한 추세가 없다. 진폭이 시간의 흐름에 따라 일정하다.	평균수준이 시간대에 따라 다르다. 추세를 가진다. 계절성을 가진다. 분산이 변한다.

ii) 비정상시계열의 정상화

ㄱ. 분산이 일정하지 않은 경우

: 분산안정화변환 (로그변환, 제곱근변환, Box-Cox 변환을 시도)

ㄴ. 추세를 가지는 경우

결정적 추세 -> 분해법 또는 추세항 모형에 포함

확률적 추세 (Dickey-Fuller의 단위근 검정) -> 차분

ㄷ. 계절성을 가지는 경우

결정적 계절추세 -> 계절 추세항 모형에 포함

확률적 계절추세 (계절형 단위근 검정) -> 계절차분

기상정보를 활용한 LSTM 기반 태양광 발전량 예측 기법

서론

5. Pedro 등 - ARIMA 및 KNN, ANN

▶ 태양광 발전량 예측 진행 후 비교 및 평가

K-최근접 이웃법(K-Nearest Neighbor)

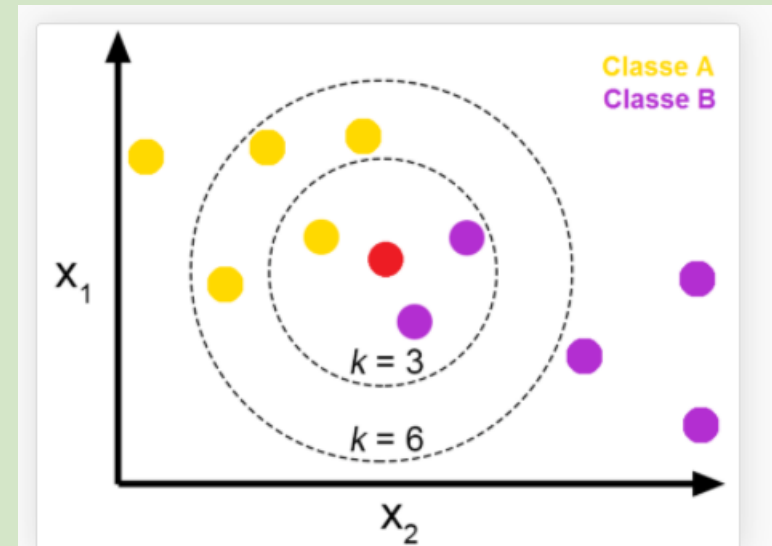
: 새로운 데이터를 입력 받았을 때 가장 가까이 있는 것이 무엇이나를 중심으로 새로운 데이터의 종류를 정해주는 알고리즘

* 특징

- n개의 특성(feature)을 가진 데이터는 n차원의 공간에 점으로 개념화 할 수 있다.
- 유사한 특성을 가진 데이터들끼리는 거리가 가깝다. 그리고 거리 공식을 사용하여 데이터 사이의 거리를 구할 수 있다.
- 분류를 알 수 없는 데이터에 대해 가장 가까운 이웃 k개의 분류를 확인하여 다수결을 할 수 있다.
- 분류기의 효과를 높이기 위해 파라미터를 조정할 수 있다.
- K-Nearest Neighbors의 경우 k 값을 변경할 수 있다.
- 분류기가 부적절하게 학습되면 overfitting 또는 underfitting이 나타날 수 있다.
- K-Nearest Neighbors의 경우 너무 작은 k는 overfitting, 너무 큰 k는 underfitting을 야기한다.

새로운 데이터가 주어졌을 때 (빨간 점) 이를 Class A로 분류할지, Class B로 분류할지 판단하는 문제.
k=3일 때, 즉 안 쪽 원을 먼저 살펴보면, k가 3이라는 것은 가장 가까운 주변의 3개 데이터를 본 뒤, 3개의 주변 데이터가 더 많이 포함되어 있는 범주로 분류하겠다는 것.
빨간 점 주변에 노란색 점(Class A) 1개와 보라색 점(Class B) 2개가 있음. 따라서 k=3 일 때는 해당 데이터가 Class B (보라색 점)으로 분류됨.
k=6일 때, 원이 더 커졌음. 이제 원 안에 노란색 점 4개와 보라색 점 2개가 있음. 따라서 k=6일 때는 노란색 점으로 분류.

KNN은 k를 어떻게 정하냐에 따라 결과 값이 바뀔 수 있음. k가 너무 작아서도 안 되고, 너무 커서도 안 됨. k의 default 값은 5. 가장 가까운 주변 5개 데이터를 기반으로 분류한다는 것. 일반적으로 k는 홀수를 사용. 짝수일 경우 동점이 되어 하나의 결과를 도출할 수 없기 때문.



출처: towardsdatascience

기상정보를 활용한 LSTM 기반 태양광 발전량 예측 기법

서론

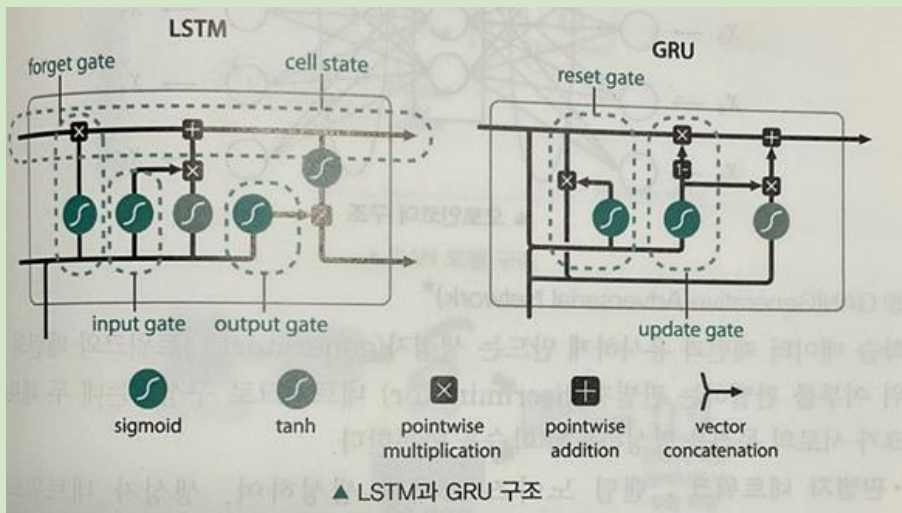
■ 본 연구 - LSTM

▶ 기온, 습도, 전운량, 자외선 지수 등의 기상예보 데이터 통해 시간 단위의 태양광 발전량을 높은 정확도로 예측하는 기법 제안

LSTM(Long Short-Term Memory Network)

: RNN*의 단점을 보완하기 위해 변형된 알고리즘으로 보통 신경망 대비 4배 이상 파라미터를 보유하여 많은 단계를 거치더라도 오랜 시간동안 데이터를 잘 기억함

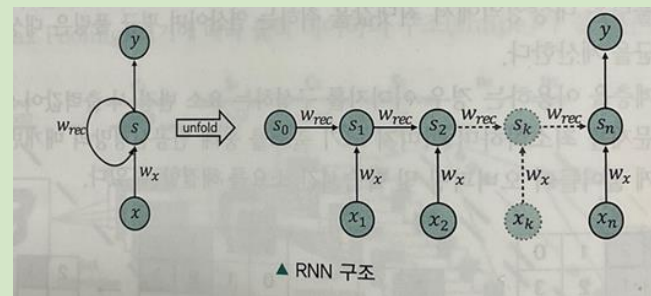
LSTM은 다음의 3가지 게이트(입력 게이트(Input Gate), 출력 게이트(Output Gate), 망각 게이트(Forget Gate))로 보완된 구조를 통해 가중치를 곱한 후 활성화 함수를 거치지 않고 컨트롤 게이트를 통해 상황에 맞게 값을 조절함으로써 문제를 해결함.



LSTM은 은닉층 이외 셀(Cell)이라는 층을 구성하는데 셀은 장기(longterm) 메모리를 기억하는 셀로 망각 게이트와 입력 게이트를 과거와 현재 상태의 셀로 조합, 과거정보를 얼마나 망각할지 현재 정보를 얼마나 반영할지를 결정함.
해당 메모리값이 활성화 함수를 거치고 출력 게이트를 통해 얼마나 밖으로 표현될지가 결정되면 현재의 은닉층 값이 정해지게 됨.

RNN(Recurrent Neural Network)*

- 순서를 가진 데이터를 입력하여 단위 간 연결이 시퀀스를 따라 방향성 그래프를 형성하는 신경네트워크 모델로 내부 상태(메모리)를 이용하여 입력 시퀀스를 처리함.
- 중간층(은닉층)이 순환구조로 동일한 가중치를 공유. 가중치(weights)와 편향(bias)에 대한 오차함수의 미분을 계산하기 위해 확률적 경사하강법(SGD: Stochastic Gradient Descent)을 이용.
- RNN은 가중치 업데이트를 위해 과거시점까지 역전파하는 BPTT(Back Propagation Through Time)를 활용. 입력 데이터의 순서(sequence)로 모두 동일 연산을 수행하며, 입력 시점마다 가중치가 공유. 계산 기울기(gradient)는 현재 상태와 이전 상태에 대해 의존적이므로, 순차적 데이터 처리에 유용.



- RNN은 점차 데이터가 소멸해 가는 문제를 발생하는데, 관련 정보와 그 정보를 사용하는 지정사이 거리가 멀 경우 역전파 시 기울기가 점차 줄어들어 학습 능력이 떨어짐 -> LSTM 이용

기상정보를 활용한 LSTM 기반 태양광 발전량 예측 기법

본론 : 발전량과 기상정보 간 상관관계 분석

■ 활용 데이터

- ① Dark Sky API : 원하는 지역의 위도 및 경도를 입력하여 결측값이 없는 완전한 기상정보 데이터 제공
-> 과거 기상정보와 기상예보 데이터 종류 같음 -> 발전량 예측에 효과적
- ② 충남 지역 2곳의 발전소, 제주 지역 1곳의 발전소 데이터 : 태양광 발전량 데이터뿐만 아니라 발전 설비가 받아들이는 일사량까지 포함(2019년2월9일 ~ 2019년 9월 30일 데이터)

■ 발전량과 기상정보 간 상관관계 분석 ; 발전량에 영향을 가장 많이 미치는 변수 파악 위함

표 1. 발전량 및 일사량과 기상정보 간 상관관계
Table 1. A correlation between solar photovoltaic power and weather factors including insolation.

Weather factor	Correlation coefficient	
	Power Generation	Insolation
Insolation	0.9865	1
UV index	0.8452	0.8463
Humidity	-0.3223	-0.3375
Temperature	0.2783	0.2871
Cloud cover	-0.1825	-0.1775
Wind speed	-0.0423	-0.0517

$$\rho_{X,Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \quad (1)$$

식 (1)에서 $\rho_{X,Y}$ 는 변수 X 와 Y 의 상관 계수, σ_X 와 σ_Y 는 각각 X 와 Y 의 표준편차, μ_X 와 μ_Y 는 각각 X 와 Y 의 평균, E 는 기댓값을 의미한다. 상관 계수의 절댓값이 1에 가까울수록 두 변수가 밀접한 관계에 있음을 나타낸다.

일사량
자외선 지수
습도
기온
전운량
풍속

기상정보를 활용한 LSTM 기반 태양광 발전량 예측 기법

본론 : 데이터 구성 및 전처리

■ 입력 데이터 정규화

- ① 입력 변수 추가 : 시간 정보 나타내기 위한 월(month), 시간(hour) 선정
- ② 데이터 값이 0~1의 분포를 갖도록 변형하는 min-max 정규화* 사용하여 입력 데이터 정규화

Min-Max 정규화(최소-최대 정규화)*

: 모든 featur에 대해 각각의 최소값 0, 최대값 1로, 그리고 다른 값들은 0과 1 사이의 값으로 변환하는 것.

ex. 어떤 특성의 최소값이 20이고 최대값이 40인 경우 30은 딱 중간이므로 0.5로 변환

머신러닝 알고리즘은 데이터가 가진 feature(특성)들을 비교하여 데이터의 패턴을 찾는 것. 그런데 데이터가 가진 feature의 스케일이 심하게 차이가 나는 경우 문제가 됨. 그래서 모든 데이터 포인트가 동일한 정도의 스케일(중요도)로 반영되도록 해주는 게 정규화(Normalization) 해야함. Min-Max 정규화는 데이터를 정규화하는 가장 일반적인 방법

$$z = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (2)$$

■ 사용 데이터 구성

- ① 발전소마다 10분 간격으로 태양광 발전량 수신, 일사량 정보 기록
- ② Dark Sky API를 통해 발전소 지역마다 시간 단위의 기온, 습도, 전운량, 자외선 지수 등 과거 기상정보와 기상예보 데이터 이용 (최대 168시간 이내의 시간 단위 기상예보 제공)

기상정보를 활용한 LSTM 기반 태양광 발전량 예측 기법

본론 : 예측 모델 구성

■ LSTM(Long short-Term Memory) 사용

- ① 과거 학습결과를 현재 학습에 사용하는 딥러닝 네트워크인 RNN 일종
- ② Cell state를 통해 학습이 반복됨에 따라 과거 학습 정보가 사라지는 장기 의존성 문제 해결
- ③ neural network 모듈을 반복시키는 체인과 같은 형태

1. 시그모이드 함수 (Sigmoid)

시그모이드 함수는 Logistic 함수라 불리기도한다. 선형인 멀티퍼셉트론에서 비선형 값을 얻기 위해 사용하기 시작했다. 함수는 아래와 같이 구성된다.

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$
$$\sigma'(x) = \sigma(x)(1 - \sigma(x))$$

sigmoid에 대해 특징을 살펴보자.

- 우선 함수값이 (0, 1)로 제한된다.
- 중간 값은 $\frac{1}{2}$ 이다.
- 매우 큰 값을 가지면 함수값은 거의 1이며, 매우 작은 값을 가지면 거의 0이다.

2. tanh 함수, (Hyperbolic tangent function)

하이퍼볼릭탄젠트란 쌍곡선 함수중 하나이다.

쌍곡선 함수 : 쌍곡선 함수란 삼각함수와 유사한 성질을 가지고, 표준 쌍곡선을 매개변수로 표시할 때 나오는 함수이다.

하이퍼볼릭탄젠트 함수는 시그모이드 함수를 transformation해서 얻을 수 있다.

-1 vertical shift & 1/2 horizontal squeeze & 2 vertical stretch

함수는 다음과 같이 정의된다.

$$\tanh(x) = 2\sigma(2x) - 1$$
$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$
$$\tanh'(x) = 1 - \tanh^2(x)$$

- tanh 함수는 함수의 중심값을 0으로 옮겨 sigmoid의 최적화 과정이 느려지는 문제를 해결했다.
- 하지만 미분함수에 대해 일정값 이상 커질시 미분값이 소실되는 **gradient vanishing** 문제는 여전히 남아있다.

sigmoid 함수를 통해 어떤 정보를 버릴지 결정

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

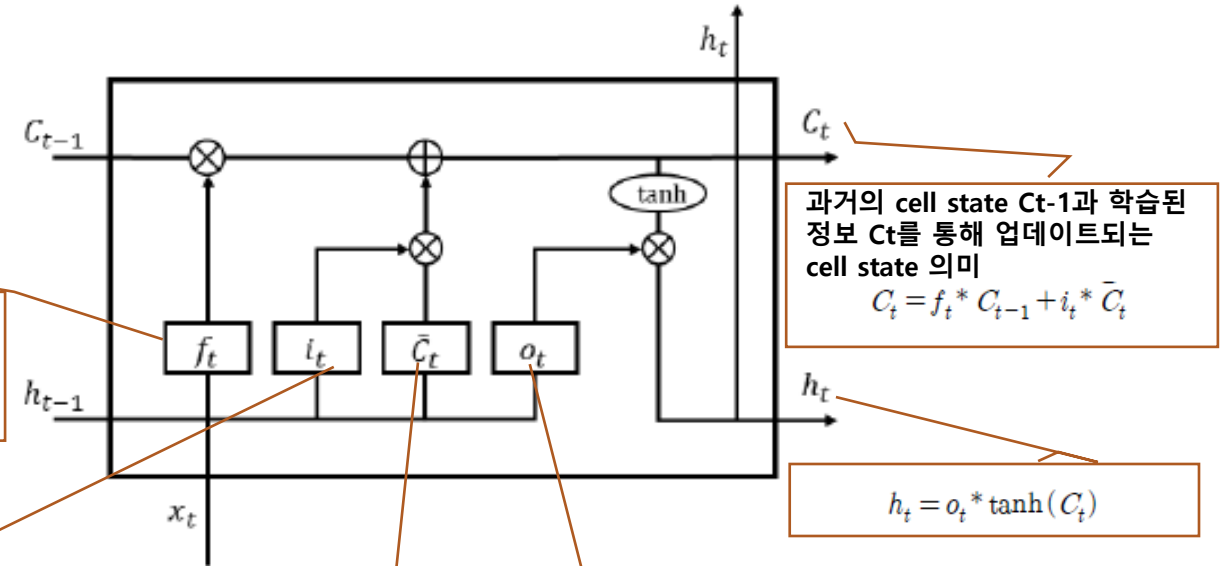
sigmoid* 및 tanh함수*를 통해 cell state에 어떤 정보를 저장할지 결정

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\bar{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

업데이트된 cell state를 바탕으로 어떤 정보를 내보낼지 결정

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$



과거의 cell state C_{t-1} 과 학습된 정보 C_t 를 통해 업데이트되는 cell state 의미

$$C_t = f_t * C_{t-1} + i_t * \bar{C}_t$$

$$h_t = o_t * \tanh(C_t)$$

그림 1. LSTM 반복 모듈
Fig. 1. LSTM iterative module

기상정보를 활용한 LSTM 기반 태양광 발전량 예측 기법

본론 : 일사량 및 발전량 예측 모델 구조

■ 일사량 예측 모델

기상예보 데이터에 일사량 존재 X

-> 기상정보 통해 일사량 예측 모델 구성 후 예측한 일사량과 기상정보 통해 최종 발전량 예측 모델 분리

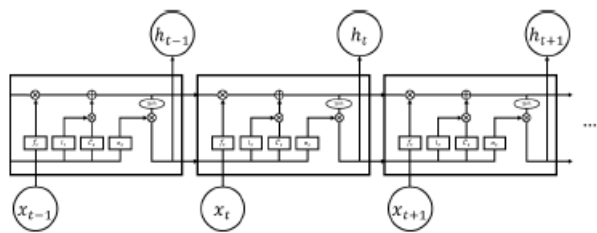


그림 2. LSTM 기반 예측 모델 구조

- 그림1의 LSTM 반복 모듈이 체인 형태로 연결되어 있음

- x: 예측 모델 입력 변수

h: 예측 모델의 출력 데이터

- 장점: 시간 단위의 일사량 및 발전량을 구간 제한없이 자유롭게 학습 및 예측 가능

■ 태양광 발전량 예측 프로세스

72시간 동안의 일사량 및 발전량 예측 진행

-> 일사량 예측 모델을 추가적으로 사용하는 것이 발전량 예측 모델만 사용하는 것보다 더 높은 예측 정확도 보임

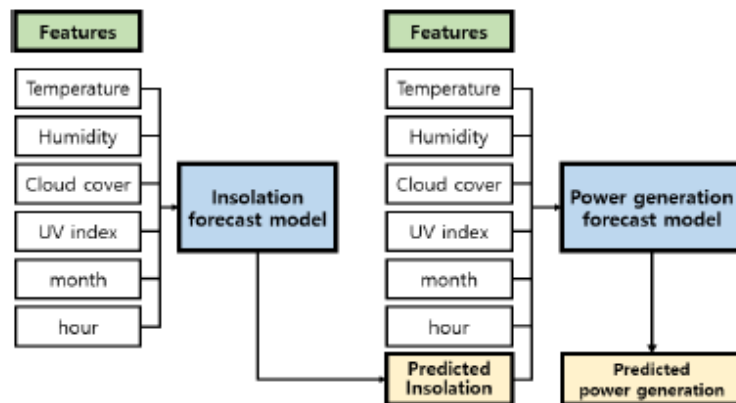


그림 3. 태양광 발전량 예측 프로세스 예측한 일사량 추가

기상정보를 활용한 LSTM 기반 태양광 발전량 예측 기법

실험

■ 실험 조건

- ① 사용 라이브러리 : TensorFlow 및 Keras 라이브러리
- ② 데이터 : 시간 단위의 일사량 및 발전량, 기상 정보 데이터(시간 단위의 기온, 습도, 전운량, 자외선 지수 포함)
- ③ 데이터 날짜 : 2019년 2월 9일~2019년 9월 30일 -> 학습 데이터 기간 1년 미만으로 샘플링 편향 문제 발생할 수 있는 한계점 존재

■ 학습 데이터

- ① train set : 2개의 인버터를 통해 수신한 발전량 데이터
- ② validation set : 1개의 인버터를 통해 수신한 발전량 데이터

인버터

파워 인버터 또는 인버터는 직류 성분을 교류 성분으로 바꾸기 위한 전기변환장치를 호칭한다. 적절한 변환 방법이나 스위칭 소자, 제어 회로를 통해 원하는 전압과 주파수 출력값을 얻는다. 위키백과

■ 예측 모델의 정확도 평가

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - x_i|$$

$$NMAE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - x_i|}{capacity}$$

(4)

- n: 데이터 총 개수 / y: 실제 데이터 값 / x: 예측한 데이터 값 / Capacity: 인버터 발전 용량
- NMAE(Normalized MAE) : MAE(Mean Absolute Error)에 발전 용량 나눈 값
 - ▶ 단위 발전 용량에 대한 오차율을 나타냄으로써 MAE보다 객관적 평가 가능

MAE - mean absolute error - 평균절대오차

: 통계에서 평균 절대 오차는 동일한 현상을 표현하는 쌍을 이루는 관측치 간의 오차 측정 값

AE - absolute error - 절대오차 -

실제 값과 측정(예측) 값과의 차이

$$\Delta x = x_i - x$$

x_i - 측정값 / x - 실제값

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - x|$$

1. 각 측정값과 실제값 사이의 절대 오차를 구하기
2. 구한 절대 오차들을 모두 더하라
3. 이를 절대 오차의 평균 (개수값)을 구하기

평균 절대 오차 (Mean Absolute Error)

평균 절대 오차(MAE)는 모든 절대 오차의 평균이다.

n = 오차의 갯수

Σ = 합을 나타내는 기호

$|x_i - x|$ = 절대 오차

기상정보를 활용한 LSTM 기반 태양광 발전량 예측 기법

실험

■ LSTM 기반 발전량 예측 모델 학습 결과

LSTM 반복 모듈 개수 10개 / 손실 함수-오차율 산정 방식과 같은 MAE / 최적화 기법-ADAM(Adaptive Moment Estimation)* / 학습 epoch 횟수-500회

아담(Adaptive Moment Estimation, Adam)은 모멘텀(Momentum)과 RMSProp의 장점을 조합한 방식으로 현재 가장 많이 사용되는 최적화 기법이다. 기울기의 과거 변화를 어느 정도 유지함으로써 경로의 효율성을 주는 모멘텀의 $V(\theta)$ 값과 최신의 정보를 과거 보다 크게 반영하는 RMSProp의 $G(\theta)$ 값을 모두 사용한다.

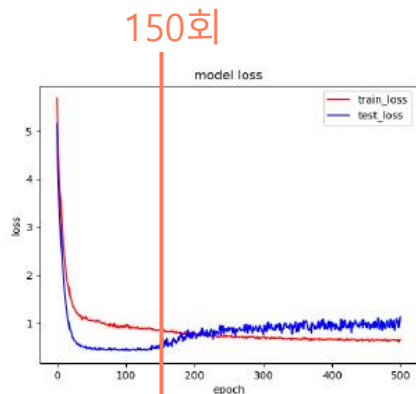


그림 4. LSTM 기반 발전량 예측 모델 학습 결과
Fig. 4. LSTM based Power generation forecast model training result

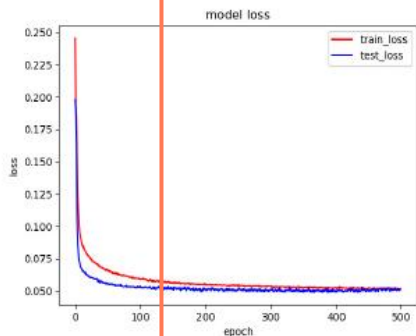


그림 5. LSTM 기반 일사량 예측 모델 학습 결과
Fig. 5. LSTM based Insolation forecast model training result

- **train_loss** : train set 학습시의 loss
- **test_loss** : validation set 학습 시의 loss

epoch* 150회 이후부터 test_loss가 낮아지지 않아 학습이 제대로 이뤄지지 X

발전량 예측 모델(그림4)의 경우 test_loss가 상승하여 overfitting(과적합) qkftod

- epoch

One Epoch is when an ENTIRE dataset is passed forward and backward through the neural network only ONCE

(한 번의 epoch는 인공 신경망에서 전체 데이터 셋에 대해 forward pass/backward pass 과정을 거친 것을 말함. 즉, 전체 데이터 셋에 대해 한 번 학습을 완료한 상태)

▶ 신경망에서 사용되는 역전파 알고리즘(backpropagation algorithm)은 파라미터를 사용하여 입력부터 출력까지의 각 계층의 weight를 계산하는 과정을 거치는 순방향 패스(forward pass), forward pass를 반대로 거슬러 올라가며 다시 한 번 계산 과정을 거쳐 기존의 weight를 수정하는 역방향 패스(backward pass)로 나뉩니다. 이 전체 데이터 셋에 대해 해당 과정(forward pass + backward pass)이 완료되면 한 번의 epoch가 진행됐다고 볼 수 있습니다.

역전파 알고리즘이 무엇인지 잘 모른다고 하더라도 epoch를 전체 데이터 셋에 대해 한 번의 학습 과정이 완료됐다고 단편적으로 이해하셔도 모델을 학습 시키는 데는 무리가 없습니다.

epochs = 40이라면 전체 데이터를 40번 사용해서 학습을 거치는 것입니다.

기상정보를 활용한 LSTM 기반 태양광 발전량 예측 기법

실험

■ DNN 오차율과 비교

LSTM 기반의 예측 모델 효율성 평가를 위함

하이퍼파라미터 설정-5개의 layer 구성 / 손실함수-MAE / 최적화 기법-ADAM / 활성화 함수 -ReLU

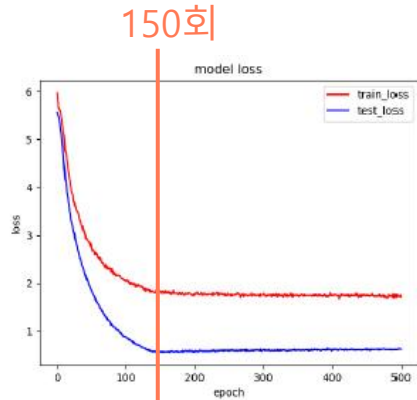


그림 6. DNN 기반 발전량 예측 모델 학습 결과
Fig. 6. DNN based Power generation forecast model training result

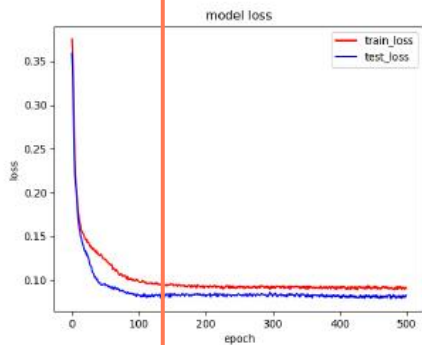


그림 7. DNN 기반 일사량 예측 모델 학습 결과
Fig. 7. DNN based Insolation forecast model training result

- **train_loss** : train set 학습시의 loss
- **test_loss** : validation set 학습 시의 loss

epoch* 150회 이후부터 test_loss가 감소X
가장 낮은 test_loss가 LSTM에 비해 높음

▶ LSTM이 DNN보다 발전량 및 일사량 데이터의 특성을 잘 학습하였음을 의미

3. ReLU 함수 (Rectified Linear Unit)

ReLU함수는 최근 가장 많이 사용되는 활성화 함수이다. 함수는 아래와 같이 정의된다.

$$f(x) = \max(0, x)$$

- $x > 0$ 이면 기울기가 1인 직선이고, $x < 0$ 이면 함수값이 0이된다.
- sigmoid, tanh 함수와 비교시 학습이 훨씬 빨라진다.
- 연산 비용이 크지않고, 구현이 매우 간단하다.
- $x < 0$ 인 값들에 대해서는 기울기가 0이기 때문에 뉴런이 죽을 수 있는 단점이 존재한다.

표 2. DNN과 LSTM 기반 예측 모델의 오차율 비교

Table 2. Comparison of error rate with DNN and LSTM based forecast model

Error	LSTM	DNN
MAE	1.5424	1.9347
NMAE	0.0454	0.0569

기상정보를 활용한 LSTM 기반 태양광 발전량 예측 기법

실험

■ 일사량 예측 모델 사용 vs 사용X

일사량 예측 모델 사용 : 기상정보+예측한 일사량 통한 최종 발전량 예측

일사량 예측 모델 사용X : 기상정보(기온, 습도, 전운량, 자외선 지수)만으로 발전량 예측

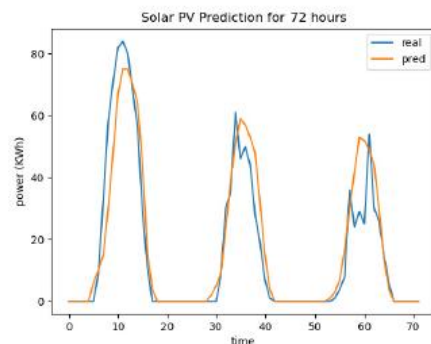
표 3. 예측 일사량을 사용한 경우와 사용하지 않은 경우의 오차율 비교

Table 3. Comparison of error rate with and without forecasted insolation

Error	With forecasted insolation	Without forecasted insolation
MAE	1.5424	1.9347
NMAE	0.0454	0.0569

▶ 일사량 예측 모델을 사용한 경우가 사용X보다 발전량 예측에 있어 더 낮은 오차율 보임

■ 특정 시점의 72시간 기상예보를 통한 예측 발전량과 해당 구간의 실제 발전량 비교



- real: 실제 발전량
- pred: 예측 발전량

비교적 기상예보 시작 시점에서 가까운 구간의 예측 정확도가 높음
기상예보 정확도가 가까운 시점일수록 높기 때문에 나타나는 현상

그림 8. 72시간 동안 실제 발전량 및 예측 발전량 비교
Fig. 8. Comparison of real power generation and predicted power generation for 72 hours

기상정보를 활용한 LSTM 기반 태양광 발전량 예측 기법

결론

LSTM 기반 일사량 예측 모델과 발전량 예측 모델을 통해
시간 단위의 태양광 발전량을 구간 제한 없이 예측하는 기법 제안



일사량 예측한 경우가 그렇지 않은 경우보다 높은 예측 정확도
DNN보다 LSTM(시계열 특성 잘 학습) 기반 예측 모델의 정확도가 더 우수
(NMAE 0.0454의 낮은 오차율 기록)



기존 연구에 비해 시간 단위의 세밀한 발전량 예측 가능,
정해진 시점이 아닌 원하는 구간의 발전량 예측을 하여 가용성 높임.

정확도 높은 태양광 발전량 예측 제공으로
태양광 발전 및 기존 화석 연료 발전의 전력 생산 계획을 효율적으로 세울 수 있고
예상 발전량 통해 합리적인 전력 거래 가격을 결정할 수 있으며
전력 계통의 안정성을 확보하여 경제성을 향상시킬 수 있을 것으로 기대함

감사합니다.