

다양한 feature를 이용하여 태양광 발전량 예측

유재상^o 장유진^o 이영구

경희대학교 컴퓨터공학과

jsyoo1996@khu.ac.kr 2015104214@khu.ac.kr yklee@khu.ac.kr

Photovoltaic prediction by using various features

Jaesang Yoo^o Yujin Jang^o, Youngkoo Lee

Department of Computer Science and Engineering, Kyung Hee University

요약

최근 신재생 에너지의 중요성이 커짐에 따라 정부도 이에 대한 지원이 커지고 있다. 그 중에서 태양광 발전은 다른 신재생 에너지와 다르게 설치하기 용이하여 다양한 공기업, 사기업, 공장 등의 건물 옥상에 설치되고 있다. 현재 태양광 발전을 하는 고객과 앞으로 참여할 잠재 고객의 태양광 발전량을 예측하는 딥러닝 모델을 구현한다. 예측된 태양광 발전량은 주식회사 시너지의 태양광발전 관리 시스템에 적용할 수 있고, 이는 소규모 분산 자원 거래 시장에서 입찰 정확성을 높일 수 있고 이를 이용하여 추가 인센티브를 확보할 수 있다.

1. 서론

1.1. 연구배경

화석연료의 고갈로 인한 자원확보 경쟁 및 고유가의 지속으로 인해 신재생 에너지로의 전환은 이제 필수에 준하는 상황에 왔다. 한국을 포함해 세계적인 추세는 기존의 석탄, 석유와 더불어 원자력 발전까지 신재생 에너지로의 전환이 필요한 자원으로 보고 있으며, 각종 에너지 전환 정책들과 함께 친환경 발전을 장려하고 있다. 태양광, 태양열, 풍력 등 많은 종류의 신재생 에너지가 있지만, 현재로서는 대부분이 설치 과정에 까다로운 조건이 붙거나, 설치 비용 대비 발전효율이 떨어지고 소규모의 작업장에는 적용하기 힘든 것이 대부분이다. 태양광 발전은 다른 발전 방법에 비해 설치에 제한 사항이 적으며, 기업을 포함해 규모가 작은 일반 가정에서도 쉽게 접할 수 있다.

태양광 발전이 다른 신재생에너지에 비해 설치 조건이 까다롭지 않은 것은 맞지만, 발전의 효율을 높이기 위해서는 여러가지 요소들을 만족해야 한다. 일조량이 곧 전력생산량과 비례하기 때문에 가장 우선적으로 고려해야 하는 것이며, 패널의 온도가 너무 높거나 낮아도 발전의 효율이 떨어지며, 초기 투자비가 많이 소요된다. 이러한 요소들로 인해, 이미 운용 중인 고객 층과 잠재 고객 층 모두에게 특정 부지 (설치 후보지) 및 조건에서의 발전량 예측은 투자 대비 효율을 높일 수 있는 효과적인 방법이 될 수 있다.

현재까지 발전량 데이터와 해당 기간의 기상 정보를 딥러닝 모델에 학습시킨 후, 차후의 기상 정보가 있다면 발전량을 예측할 수 있다. 또한, 태양광 발전에 영향을 미치는 요소들 각각의 가중치를 책정할 수 있고 이를 관리 시스템에 적용해 발전 효율을 높일 수 있다. 아직 설치하지 않은 잠재고객의 경우, 설치 후보지에 일정 수준 이상의 데이터가 존재한다면 유의미한 수준의 예측이 가능할 것으로 예상되며, 초기 설치 비용의 감소와 투자 대비 효율을 높이는 파생효과를 기대할 수 있을 것이다.

1.2. 연구목표

LSTM, k-nn, SVR의 모델을 이용하여 진행한다. 기상데이터포털에서 제공하는 기상 정보 중, 발전량과 상관계수가 높은 항목들을 구별하여 학습 데이터로 적용하고, 실제 기상 예보에서 얻을 수 있는 항목들로 재검토를 진행한다. 지역별로 상이할 수 있으므로 여러 지역을 검토한다.

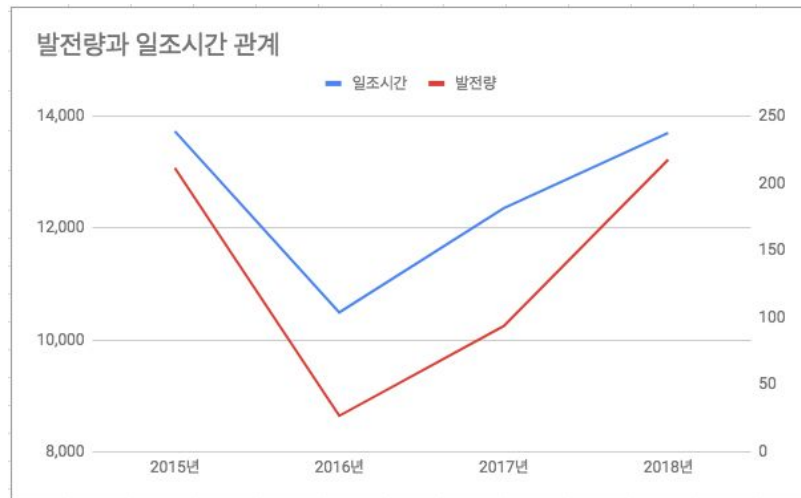
또한, 한 지역의 학습모델로 다른 지역의 발전량 예측이 가능한지 각 모델로 확인한다. 확실한 비교를 위해, 해당 지역 학습모델로 예측도 동시에 진행해 그 차이를 비교하고, 어느정도의 효용성이 있는지 판단한다.

2. 관련연구

2.1. 태양광 발전과 날씨의 관계

2.1.1. 일조시간

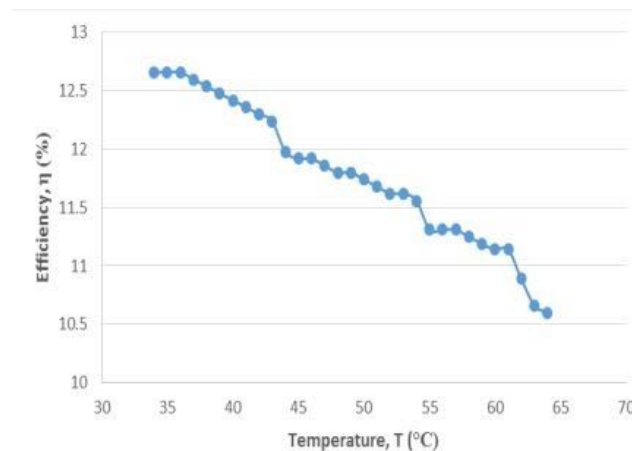
일조시간이란 태양의 빛이 일정 수준(120w/m^2) 이상일 때 1, 그 이하일 때 0으로 분당 한번씩 체크하여 1시간 동안 1인 값이 얼마나 기록되었는지를 기준으로 측정된다. 일사량은 일조시간과 다른 개념인데, 일사량은 태양으로 부터 온 태양 복사 에너지가 지표면에 닿는 양을 나타낸다. 따라서 태양광 발전은 이름에서 알 수 있듯이 일사량이 아니라 일조시간과 관련이 있고, 실제로 연간 일조시간과 태양광 발전량의 상관 그래프는 [그림 1]과 같다.



[그림 1] 일조시간과 태양광 발전량의 관계 그래프[4]

2.1.2. 온도

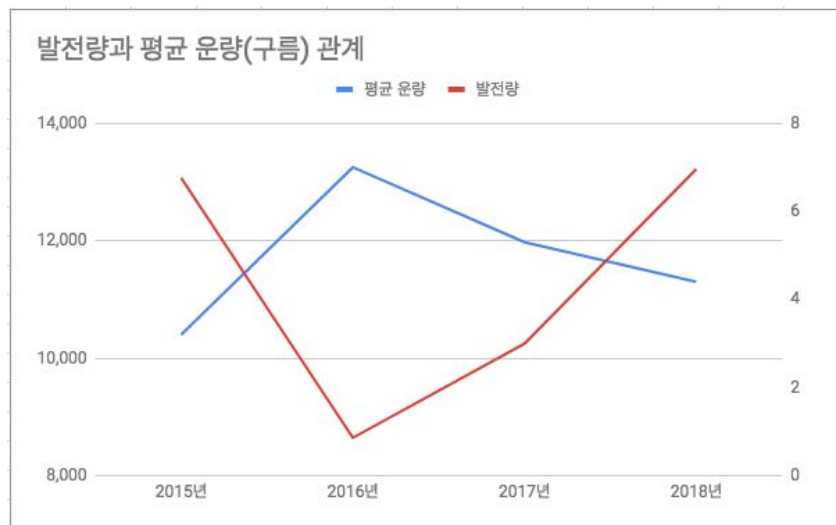
일조시간이 길어지면 일조 표면의 온도는 점진적으로 상승하는데, 태양광 패널 역시 이 영향을 받는다. 패널의 온도에 따라 발전 효율이 변화하는데, 일정 수준을 초과하게 될 경우 [그림 2]와 같이 효율이 떨어지는 모습을 보인다. 1°C 상승마다 0.27-0.77% 정도 효율이 점점 떨어지게 되며, 이는 태양광 패널의 소재나 모듈의 설계에 따라 편차가 있다[1]. 때문에 여름이 일사량이 가장 많지만, 평균 발전량은 봄, 가을, 여름, 겨울 순으로 높다. 이를 위한 해결 방안으로 온도 조절을 위해 모듈의 배치를 조정하거나 냉각수를 분사하는 등의 방법을 통해 패널의 온도를 낮추는 시도를 하고 있으며, IBM 사의 sunflower 같은 다양한 방법들이 연구되고 있다.



[그림 2] 온도와 태양광 발전 효율의 관계 그래프

2.1.3. 평균 운량

기상데이터에서 제공하는 전운량은 구름의 수치에 따라 0~10까지의 숫자로 나타낸다. 10일 경우 구름이 하늘의 10할을 차지하고 있는 경우이다. 기상청에서는 0~5까지를 맑음, 6~8까지를 구름많음, 9~10을 흐림으로 구분하고 있다. 전운량의 증가는 곧 일조량의 저하로 이어지며, 결과적으로 발전량 저하의 주요 요인 중 하나로 작용한다. [그림3]과 같이 연간 평균운량이 7에 육박할 정도로 높았던 2016년에는 평균 발전량이 2015~2018년 중 최저치였고, 대체적으로 운량이 적었던 2015년과 2018년에는 높은 발전량을 보인다.

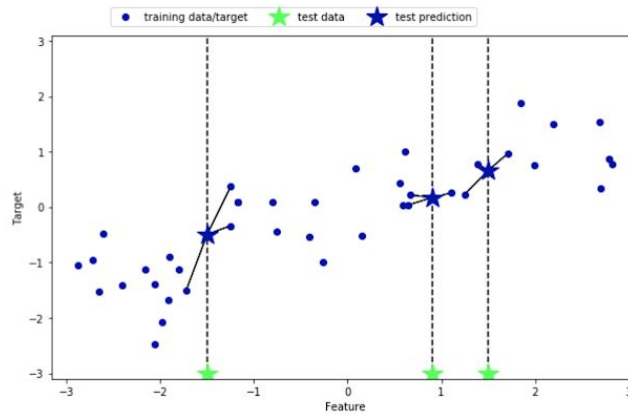


[그림 3] 평균 운량과 태양광 발전 효율의 관계 그래프[4]

2.2 Regression Model

2.2.1. K-NN Regression

K-최근접 이웃(K-Nearest Neighbor, KNN)은 지도 학습 알고리즘 중 하나이다. 어떤 데이터가 주어지면 그 주변의 데이터를 살펴본 뒤 더 많은 데이터가 포함되어 있는 범주로 분류하는 방식이다. K-NN의 특징은 훈련이 따로 필요 없다는 것이다. 훈련 데이터를 기반으로 모델을 만들고 테스트 데이터로 테스트를 하는 방식이다. K-NN은 분류가 아니라 회귀로도 사용할 수 있다. K-NN은 단순히 인접한 K개의 데이터 중 가장 많은 데이터로 분류하지만, K-NN Regression의 경우, 인접 데이터의 종속변수의 평균값이 예측값이 된다. 수백 개 이상의 특성을 갖거나, 특성값에 0이 많은 데이터 셋의 경우 제대로 동작하지 않는다.



[그림 4] K-NN Regression을 이용한 예측

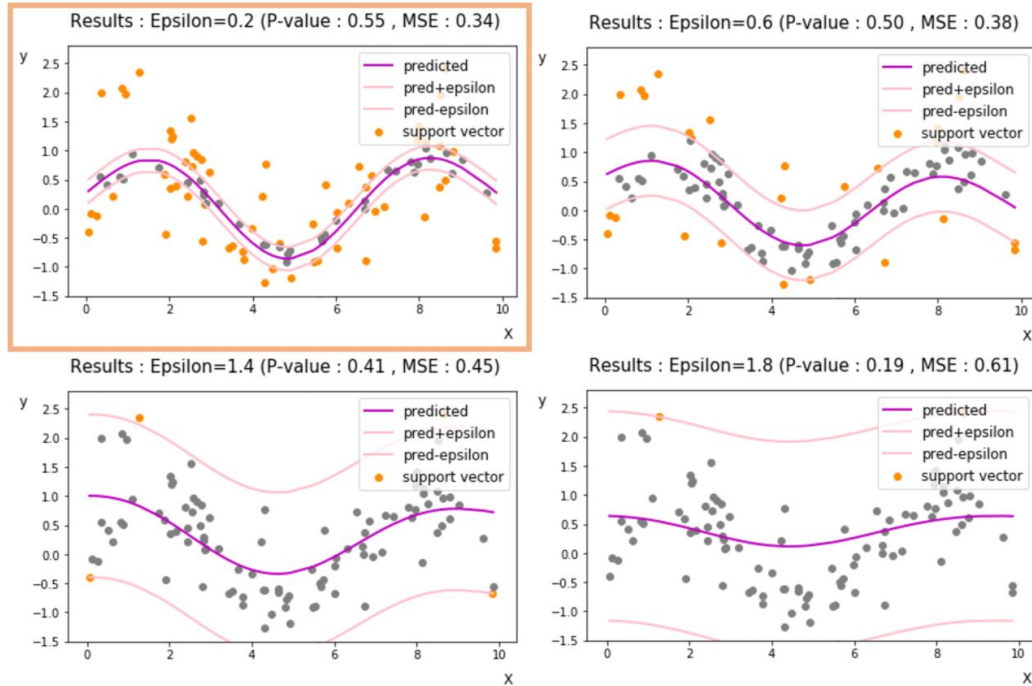
[그림 4]을 보면 분류할 때와 마찬가지로 주변의 데이터를 참고하여 그 주변 데이터의 값과 거리 등을 고려하여 예측한다.

2.2.2 SVM-regression

서포트 벡터 머신(support vector machine SVM)은 기계학습의 분야 중 하나로 패턴 인식, 자료분석을 위한 지도 학습 모델이며, 주로 분류와 회귀 분석을 위해 사용한다. 이를 통해 만들어진 모델은 초평면에서 경계로 표현되는데 그 중 가장 큰 폭(마진)을 가진 경계를 찾는 알고리즘이다. 회귀를 위한 SVM은 SVR(support vector regression)이라고 불린다. SVR에 의해 생성된 모델은 손실 함수가 모델 예측에 가까운 모든 학습 데이터를 무시하기 때문에 학습 데이터의 하위 집합에만 의존한다. SVM과는 다르게 제한된 마진 오류 내에서 경계 안에 가능한 많은 샘플이 들어가도록 학습한다.

아래의 [그림 5]에서 확인할 수 있듯이, epsilon 값을 통해 마진 오류의 제한 값을 조정할 수 있으며, epsilon의 값이 증가할수록 더 넓은 공간에서 데이터에 penalty를 부여하지 않게되고 결과적으로 점점 더 평평한 경계면을 갖게되기 때문에, 정확한 예측을 위해서는 적당한 epsilon값을 부여해야한다.

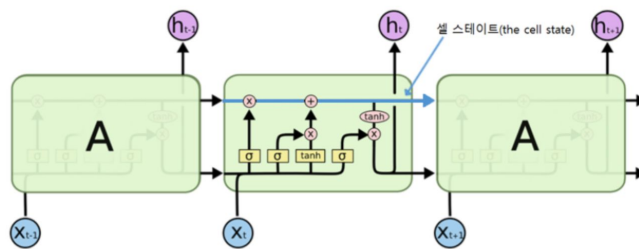
- Epsilon 변경 [0.2, 0.6, 1.4, 1.8] (Cost = 1 , gamma = 0.1)



[그림 5] Epsilon 값에 따른 SVR 경계의 변화

2.2.3 LSTM

RNN의 경우, 데이터간의 거리가 멀어질수록 데이터의 상관관계가 불명확해지는데 이를 개선한 모델이 바로 LSTM(Long Short Term Memory Network)이다. 가장 기본적인 RNN 모델의 경우, 학습이 거듭될수록 초기의 weight값이 유지되지 않아 학습률이 저하되는 Long-Term Dependency 현상이 생긴다.



[그림 6] LSTM 구조와 셀 스테이트

[그림 6]과 같이 LSTM에서는 Cell State라는 layer를 추가해서 weight를 계속 기억할 것인지 결정하여 이 문제를 개선했다.[2] 이를 통해 직전의 데이터만 참고하는 Vanilla RNN과 달리 LSTM은 cell state에서 다른 weight를 참고할 수 있게 되었다.

3. 프로젝트 내용

3.1. 태양광 발전량과 다양한 feature들과의 상관관계 분석

기상청에서 받을 수 있는 1시간 단위의 모든 날씨 데이터를 받아서 태양광 발전량과의 상관관계를 분석했다. 태양광 발전소는 ‘영암에프원’으로 위경도는 (34.745356, 126.411182)이고, 날씨 데이터는 직선 거리로 약 8km 떨어져있는 가장 가까운 목포 기상대(34.817064, 126.381180)의 관측 데이터를 이용했다.

상관관계를 분석한 모든 날씨 데이터는 [표 1]과 같다.

기온(°C)	기온 QC 플래그	강수량(mm)	강수량 QC 플래그	풍속(m/s)
풍속 QC 플래그	풍향(16방위)	풍향 QC 플래그	습도(%)	습도 QC 플래그
증기압(hPa)	이슬점온도(°C)	현지기압(hPa)	해면기압(hPa)	일조(hr)
일조 QC 플래그	일사(MJ/m2)	적설(cm)	3시간신적설(cm)	전운량(10분위)
중하층운량(10분위)	최저운고(100m)	시정(10m)	지면온도(°C)	5cm 지중온도(°C)
10cm 지중온도(°C)	20cm 지중온도(°C)	30cm 지중온도(°C)		

[표 1] 날씨 데이터 종류

[표 1]에 있는 feature들과 태양광 발전량과의 상관계수는 아래의 [표 2]와 같다.

기온(°C)	0.27	풍향 QC 플래그	0.03	일조(hr)	0.85	최저운고(100m)	-0.12
기온 QC 플래그	-0.01	습도(%)	-0.57	일조 QC 플래그	-0.60	시정(10m)	0.25
강수량(mm)	-0.07	습도 QC 플래그	-0.01	일사(MJ/m2)	0.96	지면온도(°C)	0.50
강수량 QC 플래그	-0.05	중기압(hPa)	-0.01	적설(cm)	-0.06	5cm 지중온도(°C)	0.18
풍속(m/s)	0.24	이슬점온도(°C)	-0.01	3시간신적설(cm)	-0.03	10cm 지중온도(°C)	0.09
풍속 QC 플래그	0.03	현지기압(hPa)	-0.02	전운량(10분위)	-0.18	20cm 지중온도(°C)	0.03
풍향(16방위)	0.23	해면기압(hPa)	-0.02	중하층운량(10분위)	-0.24	30cm 지중온도(°C)	0.03

[표 2] 날씨 feature와 태양광 발전량의 상관관계

상관관계 분석 결과 상관계수의 절댓값이 0.1 이상인 feature만 사용했다. [표 2]에 표시된 12개 값이 사용한 feature들 이다. 그리고 hour 값도 0~23으로 하여 학습 feature로 사용했다.

3.2. 예측 모델

3.2.1. 시계열 데이터를 통한 근미래의 태양광 발전량 예측 모델

수집한 날씨 데이터 중 사용할 feature만 남도록 전처리한다. 그리고 태양광 발전량 데이터와 날짜 및 시간이 맞게 인덱싱을 하여 하나의 데이터 프레임으로 합친다. 데이터의 완성도를 높이기 위해 누락된 부분이 있는지 확인하고, 값이 없는 부분은 0으로 대체한다. 그 후 모든 값에 대해서 각 feature별로 0~1의 값으로 정규화를 진행한다. 그리고 다양한 feature를 이용하여 하나의 feature의 이후 값을 예측하는 lstm 모델을 설계한다.


```

with tf.device('/device:GPU:0'):
    multi_step_model = tf.keras.models.Sequential()

    multi_step_model.add(tf.keras.layers.LSTM(32, return_sequences=True,
                                                input_shape=x_train_multi.shape[-2:]))
    multi_step_model.add(tf.keras.layers.LSTM(16))
    multi_step_model.add(tf.keras.layers.Dense(24))

    multi_step_model.compile(optimizer=tf.keras.optimizers.RMSprop(clipvalue=1.0),
                             loss='mae', metrics=['accuracy'])

    callback = tf.keras.callbacks.EarlyStopping(monitor='val_loss', patience=20)

```

[그림 7] LSTM 모델

[그림 7]은 전처리한 시계열 데이터를 학습할 LSTM 모델이다. 2개의 LSTM층으로 구성되어 있고, input_shape는 (720, 14)로 날씨 feature 12개, 발전량, 시간(hour)으로 이루어진 14개의 feature를 열(column)로 가지고 24시간 * 30일인 720 개의 행(row)으로 가진다. 그리고 출력값은 다음날 값인 24시간으로 나오는데 그중에서 generation 값만 사용한다. optimizer는 RMSprop를 사용하고, 손실 함수는 MAE를 사용한다. 그리고 모델의 오버피팅을 막기 위해 조기종료를 위한 인내 횟수는 20번으로 설정한다.

3.2.2 실제 발전소의 데이터를 바탕으로 가상 발전소의 발전량 예측 모델

실제 발전소의 발전량 데이터와 발전소 주변의 기상 데이터 등을 기반으로 그 features에 기반한 초평면(hyperplane)을 생성한다.

SVM 기반의 회귀모델의 경우 결정 경계(decision boundary)를 생성하게 되는데, 최적의 결정 경계를 위해서 마진(margin)의 조정이 필요하다. 마진이 최대일 경우가 최적이라고 해서 너무 큰 마진을 주는 soft margin의 경우 너무 많은 데이터를 페널티 없이 참고해 정상적인 예측이 불가능하고, 반대의 hard margin의 경우에는 너무 적은 데이터만을 참고해 특이점이 발생할 수 있다. 때문에 어디까지 penalty를 받지않고 데이터를 받아들일지 정하는 epsilon값과 해당 범위 밖의 데이터에 얼마나 penalty를 줄지 결정하는 C값의 완급을 통해 적절한 예측값을 도출해낸다.

```

#C=300 epsilon=0.05인 poly type svr
svm_poly_reg_C300_e05=SVR(kernel='poly', C=300, gamma='auto', degree=6, epsilon=0.05, coef0=1)
svm_poly_reg_C300_e05.fit(X_train,y_train)

```

[그림 8] SVR 모델

KNN 기반의 회귀모델의 경우 실제 발전소의 데이터 집단에 가상 발전소의 데이터를 삽입하여 발전량을 예측하게 된다. 이때, 몇 개의 훈련 표본을 확인할 것인지 정하는 상수 K값을 지정하고, 가상 발전소의 데이터로부터 가까운 K개의 데이터에서 발전량을 도출해낸다. K의 값이 너무 크거나 작을 경우, 적절한 예측이 불가능하기 때문에, 다양한 K값을 시도해보아야 한다.

```
#k=50 knn 모델
regressor_k50 = KNeighborsRegressor(n_neighbors=50, weights="distance")
regressor_k50.fit(X_train, y_train)
```

[그림 9] KNN 모델

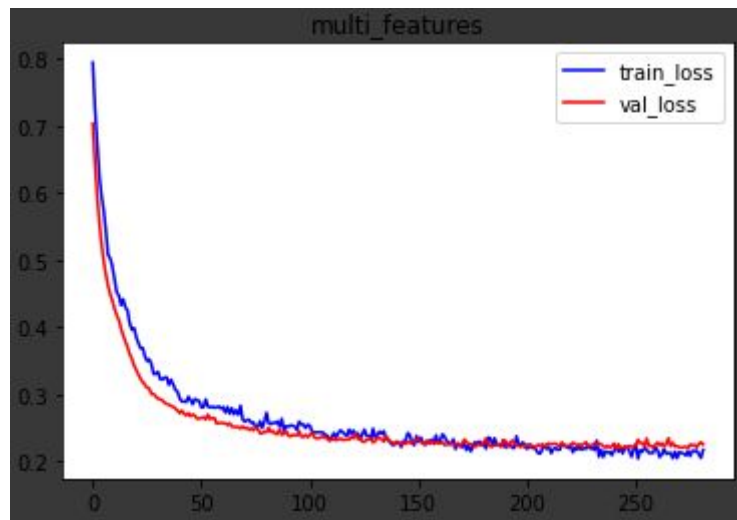
위의 두 모델의 경우 특성값에 0이 존재하는 것이 치명적이기 때문에 결측값이 많은 구간은 사전에 배제해 주는 작업이 필요하다. 때문에 초기에는 발전량이 0인 시간을 모두 제외했었고, 최종적으로는 오후 9시에서 익일 오전 6시까지의 데이터는 배제했다. 또한, 시간의 누락 역시 예측에 악영향을 미쳐 따로 추가해주었다.

4. 프로젝트 결과

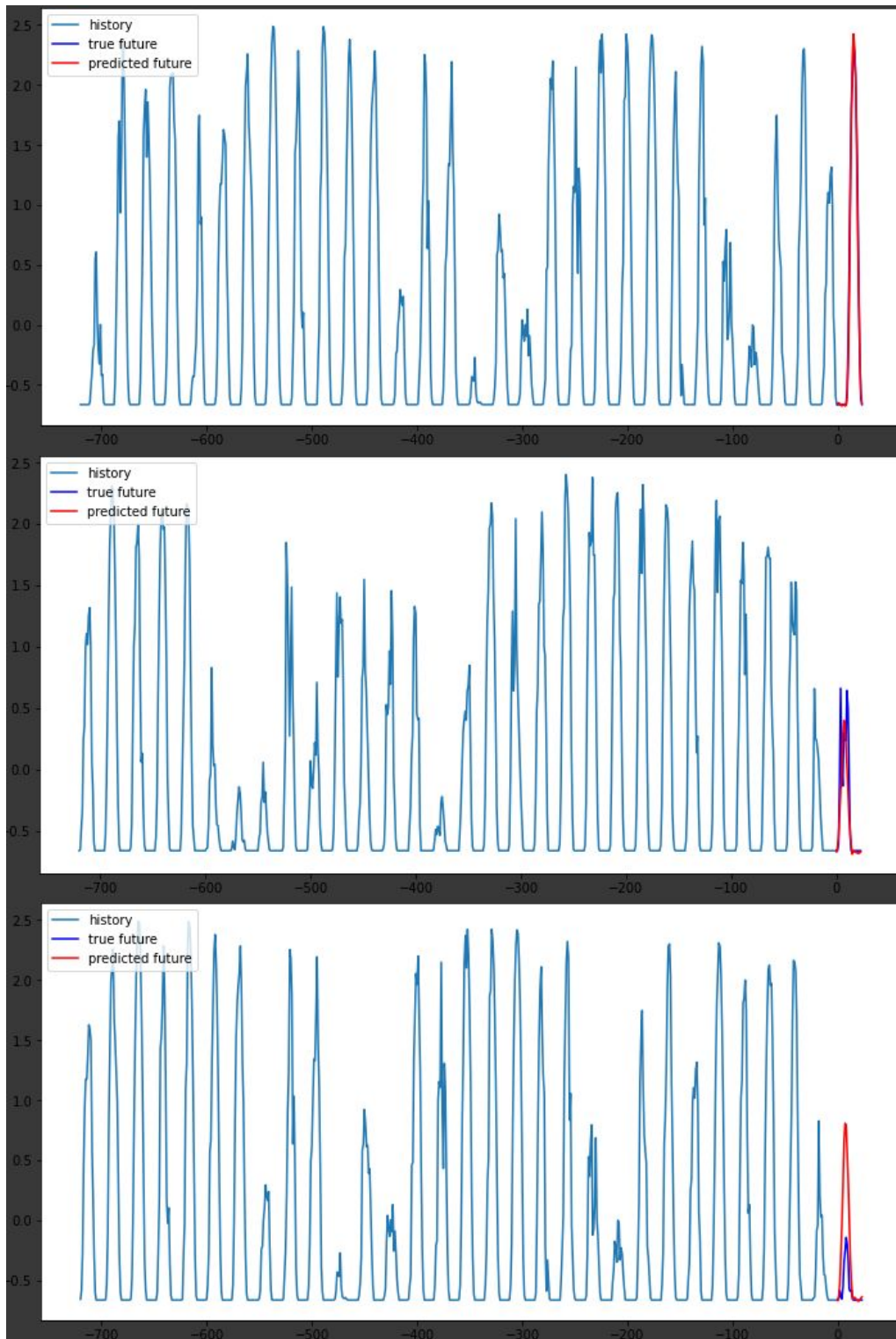
4.1. 연구 결과

4.1.1. 모든 날씨 feature를 사용한 태양광 발전량 예측

12 종류의 모든 날씨 feature 및 발전량과 시간을 사용하여 가상의 발전소에 대해 태양광 발전량을 각 모델별로 예측했다. 우선 LSTM 모델에서 학습과정에서 train 손실 값과 validation 손실 값은 [그림 10]과 같이 안정적으로 점차 감소했다.



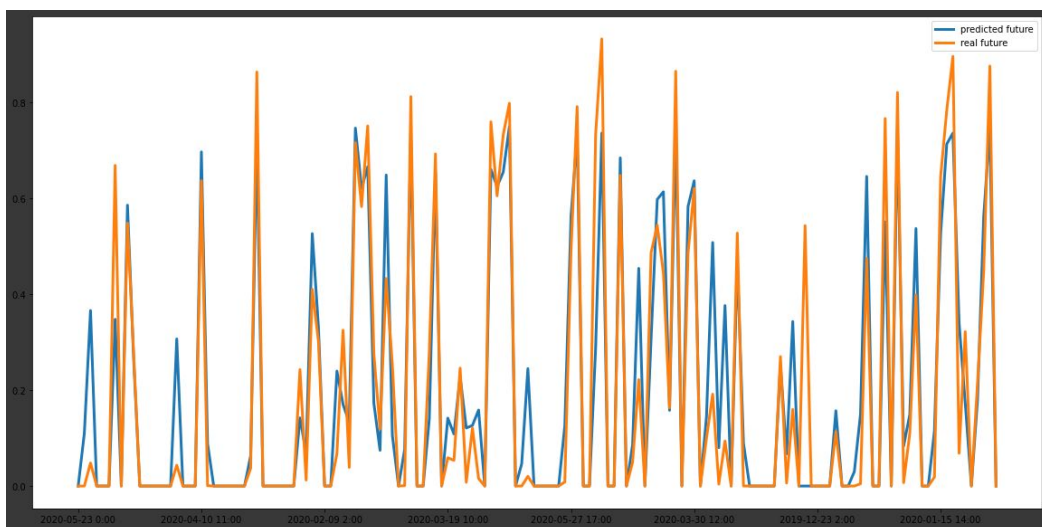
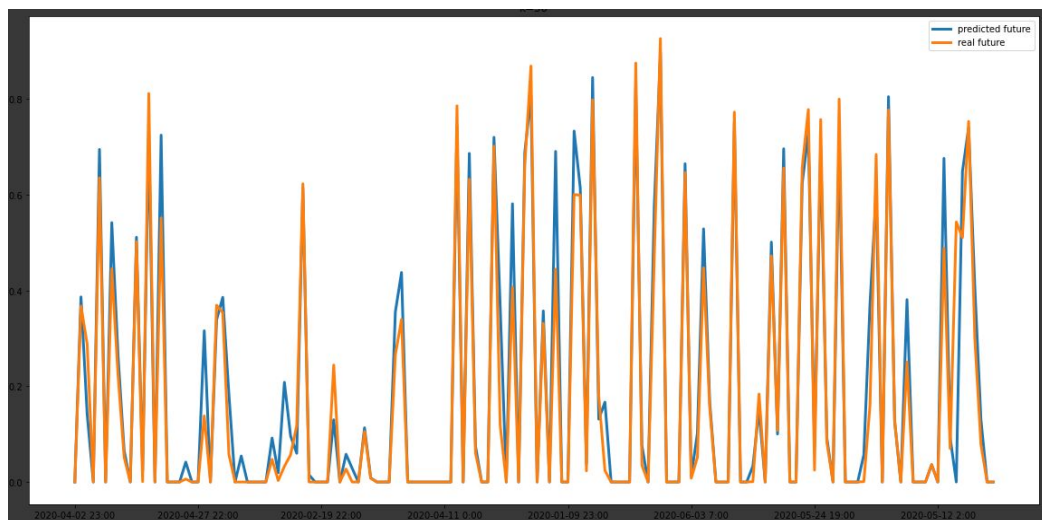
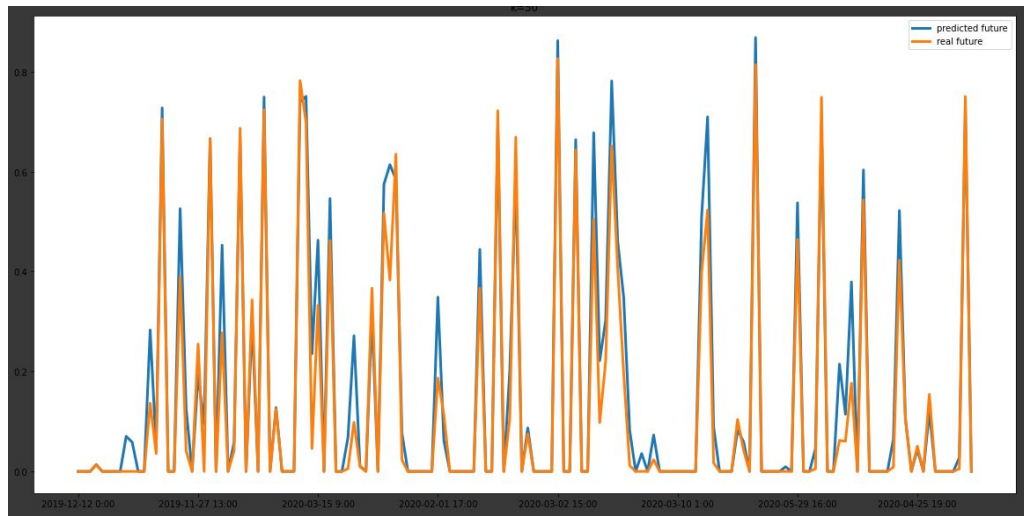
[그림 10] LSTM 모델의 훈련 중 loss 값 변화



[그림 11] 과거 데이터를 통한 예측 값(red)과 실제 값(blue)

하지만 [그림 11]을 보면 알 수 있듯이 날씨가 맑은 날에는 매우 높은 정확도를 보이지만, 흐린 날인 경우 일정 값 이하를 잘 예측하지 못하는 것을 볼 수 있다.

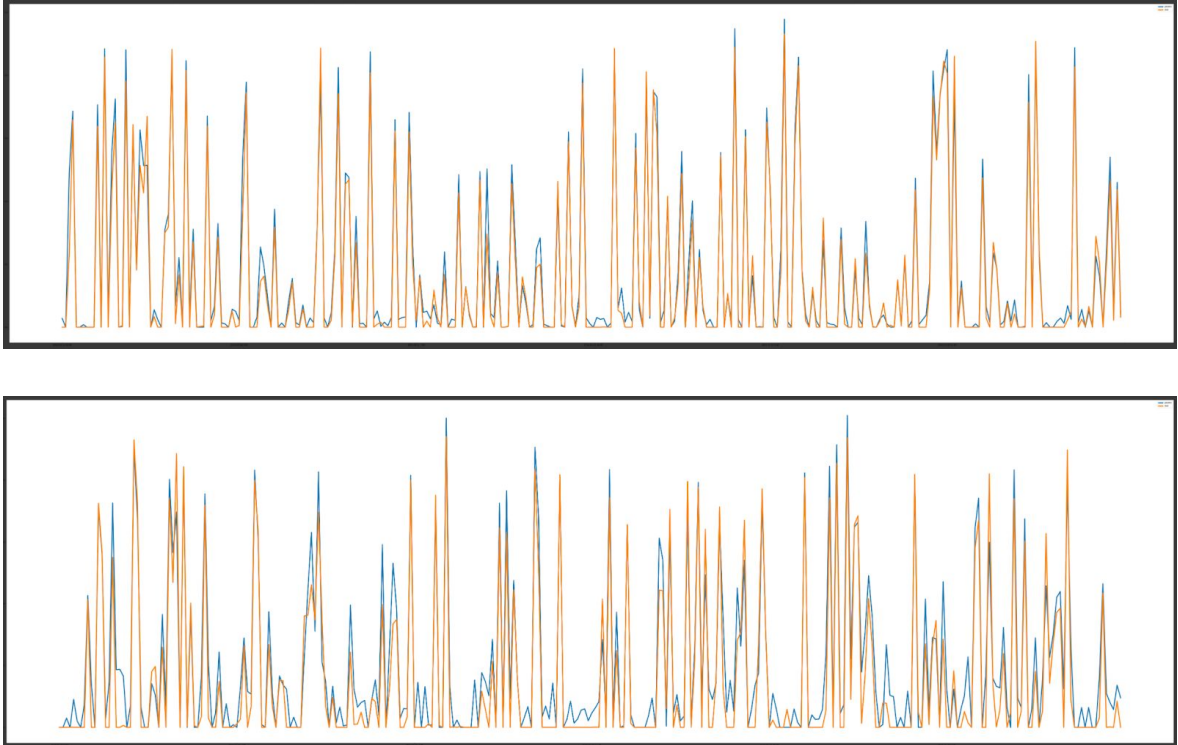
KNN 회귀모델의 경우 데이터셋 크기의 제공근 값인 156과 작은 값 50, 큰값 900으로 세 번에 걸쳐 진행했다.



[그림12] 순서대로 K=156, 50, 900 의 KNN 예측 sample(orange=실제발전량, blue=예측발전량)

육안상으로 보았을 때, 156, 50 모두 낮은 오차를 보였으나, 900의 경우 너무 많은 데이터를 인접 데이터로 인식해 실제로 낮은 발전량에서 높은 발전량을 예측하는 경향을 보였다.

이어서 SVR을 진행했으며 epsilon 값을 0.05와 0.2로 비교했고, 경계 외의 데이터에는 충분한 penalty를 부여하기 위해 C=300으로 높게 부여했다.



[그림13] 순서대로 epsilon=0.05, 0.2의 SVR 예측 sample(orange=실제발전량, blue=예측발전량)

0.05의 경우 예측값인 파란색 선이 거의 가려질 정도로 낮은 오차율을 보인다. 반면에 0.2의 경우 knn의 900과 비슷한 양상으로 실제로 낮은 발전량에서 높은 발전량을 예측하는 경향을 보인다.

각 모델의 정확도 측정하기 위해 R2 score를 사용했다. R2 score는 결정계수라고도 하며 0~1의 값을 가지고 실제값과 일치하면 1, 전혀 다르면 0을 나타낸다. 공식은 $R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$ 이다. SSE는 Sum of Square Error의 약자로 예측값과 실제값의 차이를 의미한다. SSR는 Sum of Square Regression의 약자로 예측값과 평균값의 차이를 의미한다. SST는 Sum of Square Total의 약자로 편차의 제곱합 이다. SST는 실제값과 예측값, 평균값 사이에 발생한 차이를 더한 것이다. 다른 성능 지표인 MAE 등은 데이터의 scale에 따라 값이 달라지는데 R2(결정계수)의 경우 상대적인 성능이기 때문에 판단하기 쉽다.

각 모델의 R2 score를 보면 [표 3]과 같다.

모델	R2 score
LSTM	0.78
KNN(K=50)	0.93
KNN(K=156)	0.91
KNN(K=900)	0.87
SVR(epsilon=0.05)	0.96
SVR(epsilon=0.2)	0.88

[표 3] 모든 날씨 feature를 사용한 모델별 R2 score

LSTM 모델의 경우 과거 데이터를 기준으로 학습하여 미래의 발전량을 예측한다. 즉, 각 feature들과 발전량의 관계가 아니라 과거부터 현재까지의 시간순에 따라 feature들의 변화를 보고 발전량을 예측하기 때문에 가장 최근에 나온 딥러닝 모델임에도 불구하고 정확도가 제일 낮다고 판단된다. 발전량의 가장 영향을 크게 미치는게 날씨인데 다음날 날씨는 이 연구보다 더 많은 데이터를 슈퍼컴퓨터로 분석하는 기상청도 정확하게 맞추지 못하기 때문에 이 LSTM 모델 역시 다음날의 날씨를 예상하기 어렵다. 따라서 과거의 날씨 데이터 및 발전량 데이터의 변화 추이를 보고 다음날 발전량을 예측하는 LSTM은 정확도가 제일 낮다.

날씨가 맑으면 발전량이 잘 나올 것이라는 것을 직관적으로 알 수 있듯이, 오히려 LSTM과 같은 딥러닝 모델보다 단순 계산에 가까운 KNN-regression이나 SVR이 오히려 높은 R2 score를 보일 수 있고, 동일조건 하에서의 발전량은 동일한 값이 나오기때문에 실제로 뒤의 두 모델들이 더 정확한 예측값을 보였다. 여기서 KNN보다 SVR이 높은 정확도를 보인 것은 시간을 포함한 13개의 feature를 사용한다는 점에 있을 것이다. 이는 뒤의 예보 feature를 통한 예측에서도 보이는데, KNN이 단순 거리 계산을 통한 도출인 반면에, SVR은 Kernal trick을 통해 데이터를 고차원 공간으로 매핑하고 학습하기에 비선형성을 띄는 경계를 구축하는데 용이하다. 13개의 feature들이 이와 상호작용해 더욱 높은 정확도를 보일 수 있는 것이다.

4.1.2. 예보 날씨 feature를 사용한 태양광 발전량 예측

위에서 사용한 날씨 데이터들은 측정하기 전까지 알 수 없는 feature들이 많다. 실제로 기상청에서 공개하는 예보 데이터는 온도, 풍속, 풍향, 습도, 강수확률 뿐이다. 이중에서 강수확률은 예보 데이터에만 존재하고 측정 데이터에는 없는 feature이므로 제외한 온도, 풍속, 풍향, 습도, 발전량,

시간 6개의 feature만 사용해서 근미래의 발전량을 예측하는 모델을 설계했다. 학습에 사용한 모델의 구조는 위의 4.1.1.에 사용한 모델과 같다.

각 모델별로 R2 score를 보면 [표 4]와 같다.

모델	R2 score
LSTM	0.75
KNN(K=50)	0.85
KNN(K=156)	0.84
KNN(K=900)	0.78
SVR(epsilon=0.05)	0.82
SVR(epsilon=0.2)	0.80

[표 4] 예보 날씨 feature만 사용한 모델별 R2 score

LSTM 모델은 예보 feature만 사용해서 학습을 했을 때 모든 모델들 중에 정확도가 가장 적게 감소했다. 하지만 역시 다른 모델의 정확도가 더 높기 때문에 사용하지 않는 것이 좋을 것이라 판단된다.

이전의 모든 날씨 feature를 사용한 것과 다르게 예보 feature만 사용한 모델에서는 KNN이 SVR보다 더 높은 정확도를 보이고 있다. 이는 SVR은 여러 feature들을 사용하여 태양광 발전량 값이 예상되는 곳에 초평면을 그리는 식인데, 상관계수가 높은 feature 수가 적어지면서 정확도가 떨어진 것으로 판단된다. 그리고 KNN의 경우 여러 feature들에 대응되는 발전량이 초평면에 찍히고, 주어진 예보 데이터의 경우에 최근접 발전량(K개)들의 평균값을 가지기 때문에 feature가 적은 경우엔 오히려 SVR보다 높은 정확도를 보인 것으로 판단된다.

5. 결론

5.1. 기대효과

본 논문에서는 태양광 발전량 예측을 통해 시너지의 태양광 관리 시스템에 접목시켜 태양광 발전 효율의 증대를 도모한다. 이를 통해 전력거래 시장에서의 입찰 정확성을 올릴 수 있으며, 추가적인 이익으로 귀결된다. 태양광 설치시에 정부가 많은 지원을 해주고 있지만, 그럼에도 여전히 부담되는 가격인 것은 사실이며, 적합한 지역에 설치하지 않으면 기대만큼의 효과를 볼 수 없다. 그렇기에

부지 선정에 만전을 기해야 하는데 잠재고객에게 이 정보를 토대로 구체적이고 가시적인 발전량 예측을 제공하면 부지 선정의 정확성을 높일 수 있으며 신규 고객 유치가 용이하다.

5.2. 추후 연구 방향

이 연구는 발전량 데이터 및 날씨 데이터가 누락된 날짜가 없는 곳을 기준으로 진행하였다. 그리고 그 장소들 중에서 날씨 데이터의 종류가 가장 많은 곳을 선택하였다. 즉, 이 연구를 위한 자료조사 과정에서 발전량 데이터가 누락된 날짜가 있는 것이 많았고, 지역별로 측정하는 날씨 데이터의 종류가 달라서 다른 지역의 발전소에 적용하기는 어려울 것으로 판단된다.

발전량과 상관계수값이 높게 측정되는 날씨 feature가 누락되는 지역이 많았고, 이러한 지역에 본 연구에 사용한 모델을 적용해본 결과 R2 score가 0.7 이하로 떨어지는 결과를 관측할 수 있었다. 이러한 문제점을 해결하는 가장 좋은 방안은 정부에서 각 기상대가 취득하는 날씨 데이터를 통일시키는 것인데, 이는 본 연구 진행자들이 관여할 수 없는 부분이다.

따라서 이러한 지역을 위해 태양광 발전량과 높은 상관관계를 가지는 새로운 feature를 발견하여 사용하고, 새로운 딥러닝 모델을 설계해야 할 것이라고 판단된다.

참고문헌

[1] Li Wah Thong, Sharmeeni Murugan, Poh Kiat Ng and Cha Chee Sun, Analysis of Photovoltaic Panel Temperature Effects on its Efficiency, 2016.

[2] Christopher Olah, Understanding LSTM Networks, 2015.

<https://colah.github.io/posts/2015-08-Understanding-LSTMs/>

[3] Andreas C.Muller and Sarah Guido, Introduction to machine learning with python, 2017.

[4] Seul-Gi Kim, Jae-Yoon Jung, Min Kyu Sim, A Two-Step Approach to Solar Power Generation Prediction Based on Weather Data Using Machine Learning, 2019

[5] (주)날씨맑음, 실제 운영중인 100kW 태양광 발전소 2018년 10월 발전량 - 날씨와 발전량 이야기, 2018.