

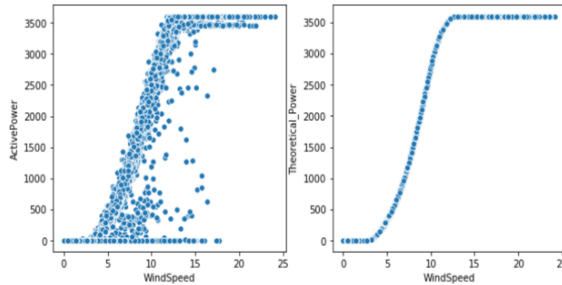
「재생에너지 발전량 예측 경진대회」 참가신청서

성명	전민규, 유현우	소속 구분	<input type="checkbox"/> 일반 <input checked="" type="checkbox"/> 기업																																																																																																																																									
연락처	+82-10-5294-4884 +82-10-2720-7948	이메일	jskstar12@makinarocks.ai hyeon95y@makinarocks.ai																																																																																																																																									
세부 주제	<input checked="" type="checkbox"/> 풍력발전량 예측 <input checked="" type="checkbox"/> 태양광발전량 예측																																																																																																																																											
분석 데이터 (해당사항 모두 체크)	요청 데이터	<input checked="" type="checkbox"/> 시간별 풍력발전량 <input checked="" type="checkbox"/> 시간별 태양광발전량 <input checked="" type="checkbox"/> 기상데이터(시간별, 지역별 예보 데이터)																																																																																																																																										
	활용 데이터	한국 기상예보 과거/실시간 데이터 (기상자료개방포털, 기상청 날씨누리) 한국/유럽 기상예보, 기상관측치, 위성사진 과거/실시간 데이터 (SK플래닛 웨더폼) 미세먼지예보 과거/실시간 데이터 (에어코리아) 한국 풍속/풍향/일사량 과거/실시간 데이터 (국립기상과학원, 신재생에너지 데이터센터) 유럽 일사량/태양광/풍력 발전량데이터 다수(Kaggle, NOAA, Open Power System Data)																																																																																																																																										
분석도구 (해당사항 모두 체크)	<input type="checkbox"/> R <input checked="" type="checkbox"/> PYTHON <input type="checkbox"/> SPSS <input type="checkbox"/> SAS <input type="checkbox"/> STATA <input type="checkbox"/> EXCEL <input type="checkbox"/> 기타()																																																																																																																																											
과제1	풍력발전량 예측																																																																																																																																											
내용	<p>문제정의</p> <p>데이터의 양에 따라 어떤 알고리즘은 잘 작동하고, 어떤 알고리즘은 잘 작동하지 않을수 있으므로 여러가지 시나리오를 대비해 아래와같이 문제를 정의하였음.</p> <table border="1"> <thead> <tr> <th>예보 사용여부</th><th>문제정의</th><th>Input</th><th>Output</th><th>비고</th></tr> </thead> <tbody> <tr> <td rowspan="2">기상예보 사용</td><td>문제1</td><td>기상예보 (1시간)</td><td>예측 발전량 (1시간)</td><td>- 기상예보 데이터로 트레이닝 - 기상예보에 종속적인 모델 - 각 시간의 연속성을 반영하지 못하는 모델</td></tr> <tr> <td>문제2</td><td>기상예보 (24시간)</td><td>예측 발전량 (24시간)</td><td>- 기상예보 데이터로 트레이닝 - 많은 수의 데이터 필요 - 데이터수가 부족하면 트레이닝이 잘 되지 않음 - 각 시간에 대한 연속성을 반영하는 모델</td></tr> <tr> <td rowspan="2">기상예보 사용하지 않음</td><td>문제3</td><td>(n-1)일동안의 기상 관측치</td><td>n일(당일)의 예측 발전량(24시간)</td><td>- 기상 관측치 데이터로 트레이닝 - 해당 지역에 대한 정보만 사용</td></tr> <tr> <td>문제4</td><td>(n-1)일동안의 위성사진</td><td>n일(당일)의 예측 발전량(24시간)</td><td>- 위성사진 데이터로 트레이닝 - 주위 지역에 대한 정보도 반영</td></tr> </tbody> </table> <p>아래는 각 문제의 프로세스와 Input/Output을 시각화한 이미지</p> <div> <div> <p>Ensemble</p> <p>문제1</p> <p>기상예보 (1시간) → CatBoost, LightGBM → 발전량 예측 (1시간)</p> <table border="1"> <tr><td>days</td><td>t_1</td><td>...</td><td>t_{43}</td><td>...</td><td>t_{72}</td></tr> <tr><td>Day n 예보</td><td>0</td><td>...</td><td>546</td><td>...</td><td>0</td></tr> </table> <p>↓</p> <table border="1"> <tr><td>days</td><td>t_1</td><td>...</td><td>t_{43}</td><td>...</td><td>t_{72}</td></tr> <tr><td>Day n 발전량 예측</td><td>0</td><td>...</td><td>1021</td><td>...</td><td>0</td></tr> </table> </div> <div> <p>문제2</p> <p>기상예보 (24시간) → LSTM, Transformer → 발전량 예측 (24시간)</p> <table border="1"> <tr><td>days</td><td>t_1</td><td>...</td><td>t_{43}</td><td>...</td><td>t_{72}</td></tr> <tr><td>Day n 예보</td><td>0</td><td>...</td><td>546</td><td>...</td><td>0</td></tr> </table> <p>↓</p> <table border="1"> <tr><td>days</td><td>t_1</td><td>...</td><td>t_{43}</td><td>...</td><td>t_{72}</td></tr> <tr><td>Day n 발전량 예측</td><td>0</td><td>...</td><td>1022</td><td>...</td><td>0</td></tr> </table> </div> <div> <p>문제3</p> <p>기상관측치 (이전 n-1일분) → LSTM, Transformer → 발전량 예측 (1일)</p> <table border="1"> <tr><td>days</td><td>t_1</td><td>...</td><td>t_{43}</td><td>...</td><td>t_{72}</td></tr> <tr><td>Day 1 관측치</td><td>0</td><td>...</td><td>546</td><td>...</td><td>0</td></tr> <tr><td>Day 2 관측치</td><td>0</td><td>...</td><td>742</td><td>...</td><td>0</td></tr> <tr><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr> <tr><td>Day n-1 관측치</td><td>0</td><td>...</td><td>655</td><td>...</td><td>0</td></tr> </table> <p>↓</p> <table border="1"> <tr><td>days</td><td>t_1</td><td>...</td><td>t_{43}</td><td>...</td><td>t_{72}</td></tr> <tr><td>Day n 발전량</td><td>0</td><td>...</td><td>1023</td><td>...</td><td>0</td></tr> </table> </div> <div> <p>문제4</p> <p>위성사진 (이전 n-1일분) → 발전량 예측 (1일)</p> <p>Days 1 t_1 위성사진, Days n-1 t_{72} 위성사진</p> <p>↓</p> <table border="1"> <tr><td>days</td><td>t_1</td><td>...</td><td>t_{43}</td><td>...</td><td>t_{72}</td></tr> <tr><td>Day n 발전량</td><td>0</td><td>...</td><td>1023</td><td>...</td><td>0</td></tr> </table> </div> <div> <p>Ensemble된 최종 예측값</p> <table border="1"> <tr><td>days</td><td>t_1</td><td>...</td><td>t_{43}</td><td>...</td><td>t_{72}</td></tr> <tr><td>Day n 발전량</td><td>0</td><td>...</td><td>1022</td><td>...</td><td>0</td></tr> </table> <p>단일 모델에 비해 더 안정적이고 오차가 적은 결과를 출력함</p> </div> </div>			예보 사용여부	문제정의	Input	Output	비고	기상예보 사용	문제1	기상예보 (1시간)	예측 발전량 (1시간)	- 기상예보 데이터로 트레이닝 - 기상예보에 종속적인 모델 - 각 시간의 연속성을 반영하지 못하는 모델	문제2	기상예보 (24시간)	예측 발전량 (24시간)	- 기상예보 데이터로 트레이닝 - 많은 수의 데이터 필요 - 데이터수가 부족하면 트레이닝이 잘 되지 않음 - 각 시간에 대한 연속성을 반영하는 모델	기상예보 사용하지 않음	문제3	(n-1)일동안의 기상 관측치	n일(당일)의 예측 발전량(24시간)	- 기상 관측치 데이터로 트레이닝 - 해당 지역에 대한 정보만 사용	문제4	(n-1)일동안의 위성사진	n일(당일)의 예측 발전량(24시간)	- 위성사진 데이터로 트레이닝 - 주위 지역에 대한 정보도 반영	days	t_1	...	t_{43}	...	t_{72}	Day n 예보	0	...	546	...	0	days	t_1	...	t_{43}	...	t_{72}	Day n 발전량 예측	0	...	1021	...	0	days	t_1	...	t_{43}	...	t_{72}	Day n 예보	0	...	546	...	0	days	t_1	...	t_{43}	...	t_{72}	Day n 발전량 예측	0	...	1022	...	0	days	t_1	...	t_{43}	...	t_{72}	Day 1 관측치	0	...	546	...	0	Day 2 관측치	0	...	742	...	0	Day n-1 관측치	0	...	655	...	0	days	t_1	...	t_{43}	...	t_{72}	Day n 발전량	0	...	1023	...	0	days	t_1	...	t_{43}	...	t_{72}	Day n 발전량	0	...	1023	...	0	days	t_1	...	t_{43}	...	t_{72}	Day n 발전량	0	...	1022	...	0
예보 사용여부	문제정의	Input	Output	비고																																																																																																																																								
기상예보 사용	문제1	기상예보 (1시간)	예측 발전량 (1시간)	- 기상예보 데이터로 트레이닝 - 기상예보에 종속적인 모델 - 각 시간의 연속성을 반영하지 못하는 모델																																																																																																																																								
	문제2	기상예보 (24시간)	예측 발전량 (24시간)	- 기상예보 데이터로 트레이닝 - 많은 수의 데이터 필요 - 데이터수가 부족하면 트레이닝이 잘 되지 않음 - 각 시간에 대한 연속성을 반영하는 모델																																																																																																																																								
기상예보 사용하지 않음	문제3	(n-1)일동안의 기상 관측치	n일(당일)의 예측 발전량(24시간)	- 기상 관측치 데이터로 트레이닝 - 해당 지역에 대한 정보만 사용																																																																																																																																								
	문제4	(n-1)일동안의 위성사진	n일(당일)의 예측 발전량(24시간)	- 위성사진 데이터로 트레이닝 - 주위 지역에 대한 정보도 반영																																																																																																																																								
days	t_1	...	t_{43}	...	t_{72}																																																																																																																																							
Day n 예보	0	...	546	...	0																																																																																																																																							
days	t_1	...	t_{43}	...	t_{72}																																																																																																																																							
Day n 발전량 예측	0	...	1021	...	0																																																																																																																																							
days	t_1	...	t_{43}	...	t_{72}																																																																																																																																							
Day n 예보	0	...	546	...	0																																																																																																																																							
days	t_1	...	t_{43}	...	t_{72}																																																																																																																																							
Day n 발전량 예측	0	...	1022	...	0																																																																																																																																							
days	t_1	...	t_{43}	...	t_{72}																																																																																																																																							
Day 1 관측치	0	...	546	...	0																																																																																																																																							
Day 2 관측치	0	...	742	...	0																																																																																																																																							
...																																																																																																																																							
Day n-1 관측치	0	...	655	...	0																																																																																																																																							
days	t_1	...	t_{43}	...	t_{72}																																																																																																																																							
Day n 발전량	0	...	1023	...	0																																																																																																																																							
days	t_1	...	t_{43}	...	t_{72}																																																																																																																																							
Day n 발전량	0	...	1023	...	0																																																																																																																																							
days	t_1	...	t_{43}	...	t_{72}																																																																																																																																							
Day n 발전량	0	...	1022	...	0																																																																																																																																							

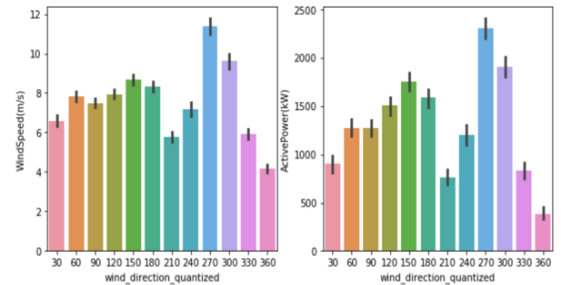
데이터 전처리

여러 발전소의 데이터를 사용하기 위해서 각 발전소의 데이터를 1시간 단위로 Interpolation하거나 합치고, 한국/해외의 기상예보 또한 동일한 포맷으로 전처리한다. 문제4¹의 경우에는 지역에 따라 모델이 학습하는 사진이 다르므로 다른 위치에 있는 발전소 데이터를 사용할 수 없다.

풍력 발전량에 영향을 주는 요소들



풍속이 발전량에 큰 영향을 미치지만 이론(오른)과 달리 실제 (왼) 다른 요소의 영향도 많이 받는다는 것을 알 수 있음



풍향이 풍속/발전량에 영향을 미치며 풍속에 미치는 것과 관련 있게 발전량에도 영향을 미침

2

풍속, 공기밀도, 출력계수, 회전면적 등 몇가지 값을 알면 풍력 터빈의 출력값을 계산할 수 있으므로 풍속을 예측하기 위한 방법에 초점을 맞추려 한다. 보통 10~3월까지의 풍력발전 이용률이 증가하지만, 바람이 잘 불지 않는 4~9월엔 풍속이 느려져 이용률이 떨어지기도 한다. 따라서 계절이 발전량에 영향을 미칠 것이다.

이외에도 풍속에 영향을 미치는 변수로 온도, 압력, 습도 등 여러 요인³이 있다. 특히 온난화에 따라 기온이 증가하면 공기 밀도가 저하되고, 따라서 풍속의 감소를 기인한다는 연구⁴ 결과가 있으므로 참고할 예정이다.

또한 수평형 풍력 발전기는 전혀 다른쪽으로 바람이 불면 블레이드가 돌지 않으므로 풍향도 발전량에 중요한 영향을 미칠 것이라 예상한다.

SK ICT 파트너 회사로서 데이터 활용 협력

SK ICT파트너사로서, 기상청 데이터보다 세세한 구역에 대한 정보를 제공하는 SK플래닛 웨더폰 기상 서비스 데이터도 활용하여 데이터 범위와 품질을 향상시키고자 함. 풍력 발전소를 건설하고 운영하는 SK D&D, SK E&S와의 협력관계를 통해 발전산업 전문지식을 협조받아 도메인 지식을 활용한 모델링으로 모델 성능 향상.

예측 기법

문제에 따른 모델 제시

문제	모델	설명
문제1	CatBoost, XGboost, LightGBM	Categorical Variable인 풍속을 다루는데에 CatBoost가 다른 두 모델보다 유리하나, 세 모델 모두 Regression을 시도하여 성능 비교
문제2	LSTM, Transformer	Sequence to Sequence를 다루는 LSTM과 Transformer를 활용. Transformer가 이론적으로 LSTM보다 학습이 빠르나, 두 모델 모두 학습하여 성능 비교
문제3	LSTM, Transformer	n일동안의 기상 관측치를 Input으로 활용하여 이후 +72시간동안의 발전량을 예측. 두 모델 모두 시도하여 성능 비교
문제4	CNN + LSTM with Attention	시간에 따라 변하는 위성 이미지의 Feature를 CNN으로 추출. Attention을 통해 어떤 Feature가 영향을 많이 미치는지 학습. 추출된 Feature를 LSTM의 Input으로 활용하여 이후 +72시간동안의 발전량을 예측

Feature Engineering Automation

가능한 모든 Feature Engineering Idea에 대해 Bayesian Optimization을 통해 최적의 성능을 내는 Feature Engineering Idea의 Combination을 탐색한다. 이는 각각의 방법이 적용 되었을 때의 즉각적인 성능 차이는 개별적인 실험으로 확인할수 있지만 방법간 혹은 방법과 모델간 어떤 Interaction이 있는지 알지 못하기 때문에 Optimization Method를 통해 Grid Search보다 빠르게 찾는 방법이다. 추가로 아래와 같은 아이디어들을 적용할 수 있다.

- 예보와 관측치를 비교하여, 관측치와 크게 차이나는 예보는 트레이닝 데이터에서 제거
- Polynomial Features 추가 ('Wind Turbine Scada' Kaggle data로 성능향상 확인)

¹ <http://www.greenmap.go.kr/inquiry.do> / 웨이폰 위성영상을 사용 예정이며 어느 위치의 이미지를 사용할지 한정 지어야 함

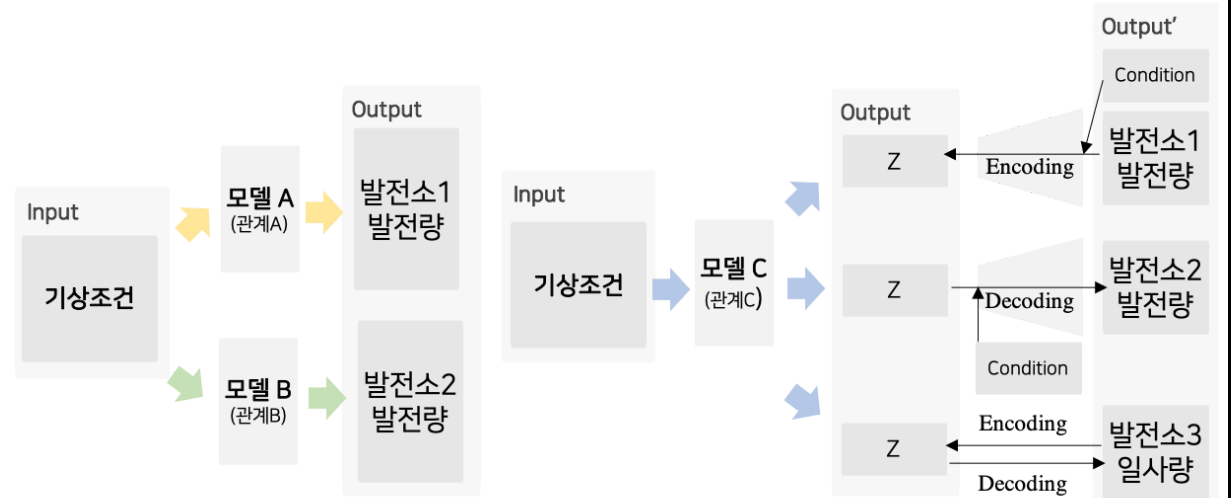
² Wind Turbine Scada Dataset, Kaggle 분석결과

³ '제주도 지형 조건에 의한 풍력자원의 공간적 특성(2017)', 김대영 외 3

⁴ '한반도 온난화가 풍력발전에 미치는 영향에 관한 연구(2017)', 서성구

타지역 풍력 발전소 데이터 활용

기상자원기술개발과 Kaggle에 공개된 한국/유럽의 일사량/태양광 발전량 데이터를 활용하여 데이터수 부족 문제를 해결하고자 한다. 동일한 기상조건에서도 풍력 발전소는 다른 출력의 발전량을 내게 된다.

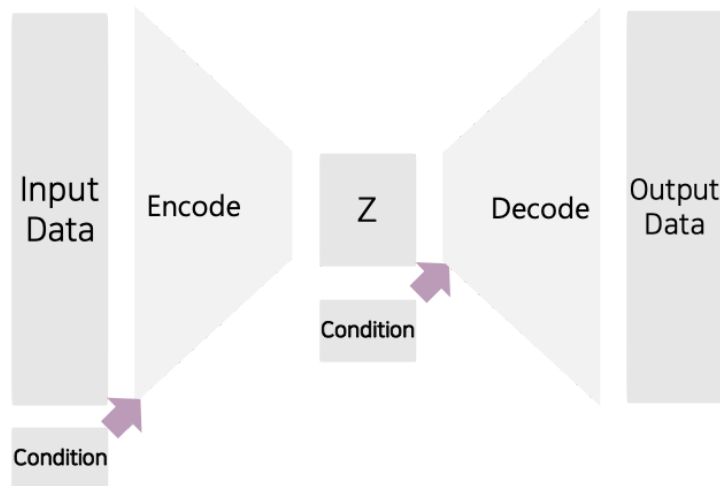


<기존 : 발전소별로 다른 데이터, 모델>

<CVAE 활용 : 여러 발전소 데이터 모두 활용, 하나의 모델>

기존의 방식으로서는 동일한 기상조건에 대해 각 발전소가 어떤 발전량을 출력하는지에 대한 개별적인 **관계A(모델A)**, **관계B(모델B)**만을 학습할수 있었다.

Conditional Variational Autoencoder (CVAE)



Input Data : 해당 지점의 풍속 또는 해당 지점의 풍력 발전량

Condition: 데이터에서 제거하고 싶은 내재된 요소 (위도 등)

Z : 데이터에서 Condition이 주는 영향이 제거된 Location Invariant한 압축된 데이터

Output Data : Z로부터 Condition을 통해 해당 지점의 풍력 발전량 복원

Input Data와 Output Data는 같다

Conditional Variational Autoencoder(CVAE)⁵를 활용하면 각 발전소의 특성을 Condition으로 두어 Encoding함으로써 각 발전소간의 차이가 제거된 발전량 Z(Output)를 얻을 수 있다.

모델C(문제1~문제4를 푸는 모델)는 주어진 기상조건에 대해서 항상 같은 Z(Output)를 예측하도록 트레이닝 된다. 마지막으로 Z와 각 발전소에 대한 Condition을 이용하여 Decoding함으로써, 각 발전소의 위도/설비가 반영된 최종적인 예측 발전량(Output)을 얻게 된다. 이러한 방법을 통해 위치와 설비가 다른 모든 발전소의 데이터(Output)를 이용하여 모델C와 CVAE를 학습할 수 있게 된다. 더 많은 기상조건-발전량 관계를 학습했으므로, 단일 발전소의 데이터만을 활용하여 학습시킨 모델A, B보다 성능이 높다.

활용 방안

다른 발전소 데이터로부터 **Transfer Learning**으로 신규 발전소도 높은 정확도로 예측

상대적으로 데이터수가 적은 신규발전소도 해당 발전소의 데이터로 CVAE에서 Condition만 학습하여 이미 Training된 모델을 통해 높은 정확도로 발전량을 예측할수 있다는 장점이 있다.

건설 예정인 신규 발전소에 대하여 일사량을 측정함으로써 **Transfer Learning**으로 연간 발전량 예측

건설 예정인 신규 발전소에 대하여 어느정도의 발전량이 나올지 예측하고 싶을때, 해당 위치에서의 풍속을 측정하여 Condition을 구하고, Transfer Learning을 통해 과거 1년간 기상예보로 Inference함으로써 연간 예측 발전량을 구할 수 있다.

검증된 예측 모델을 스케줄링 시스템 내 연동

AI 모델의 시스템 내 적용을 위해서는 변화하는 환경에 맞춰 지속적으로 모델을 학습시킬 수 있는 Continual Learning 기술과 다양한 모델을 효율적으로 관리할 수 있는 AI Model Management 기술이 필요. 팀이 보유하고 있는 검증된 AI Operation & Deployment 기술을 활용하여 실 시스템 적용을 타깃으로 함.

⁵ 'Learning structured output representation using deep conditional generative models(2015)', Sohn et al

과제2

태양광발전량 예측

내용

문제정의

데이터의 양에 따라 어떤 알고리즘은 잘 작동하고, 어떤 알고리즘은 잘 작동하지 않을수 있으므로 여러가지 시나리오를 고려하여 아래와 같이 문제를 정의하였습니다.

예보 사용여부	문제정의	Input	Output	비고
기상예보 사용	문제1	기상예보 (1시간)	예측 발전량 (1시간)	- 기상예보 데이터로 트레이닝 - 기상예보에 종속적인 모델 - 각 시간의 연속성을 반영하지 못하는 모델
	문제2	기상예보 (24시간)	예측 발전량 (24시간)	- 기상예보 데이터로 트레이닝 - 많은 수의 데이터 필요 - 데이터수가 부족하면 트레이닝이 잘 되지 않음 - 각 시간에 대한 연속성을 반영하는 모델
기상예보 사용하지 않음	문제3	(n-1)일동안의 기상 관측치	n일(당일)의 예측 발전량(24시간)	- 기상 관측치 데이터로 트레이닝 - 해당 지역에 대한 정보만 사용
	문제4	(n-1)일동안의 위성사진	n일(당일)의 예측 발전량(24시간)	- 위성사진 데이터로 트레이닝 - 주위 지역에 대한 정보도 반영

아래는 각 문제정의별로 프로세스와 Input/Output을 시각화한 이미지입니다.



각 문제에 해당하는 데이터를 사용하고 모델을 트레이닝한 뒤, Ensemble을 통해 각 데이터와 모델이 가진 장점은 활용하며 단점을 상쇄시켜 오차를 줄이고 안정적인 성능을 얻게 됩니다. 최종적으로는 문제1~문제4에 해당하는 데이터와 모델을 모두 Ensemble하여 오차가 적고 안정적인 성능을 가지는 모델을 수립하고자 합니다.

데이터 전처리

문제	전처리 포맷	공통
문제1	- 동일한 기상예보 포맷(Column명, 시간, 단위)	- 일출/일몰 시간 외에 일사량과 태양광 발전량은 0이기 때문에 일출/일몰 예보를 활용하여 일출/일몰 전후 발전량 절삭하여 Output Dimension을 줄임 - Singular Value Decomposition을 통해 Output Dimension을 줄임
문제2	- 동일한 발전량 포맷(시간, 단위)	
문제3	- 동일한 기상 관측치 포맷(Column명, 시간, 단위)	

여러 발전소의 데이터를 활용하기 위해서는 발전소별 데이터를 같은 포맷으로 전처리해주어야 합니다. 각 발전소의 데이터를 1시간 단위로 Interpolation하거나 합치고, 기상예보도 한국의 기상예보와 유럽의 기상예보를 동일한 포맷으로 전처리합니다. 문제4의 경우에는 지역에 따라 모델이 학습하는 사진이 다르므로 다른 위치에 있는 발전소 데이터를 사용할 수 없습니다.

일반적으로 Input/Output Dimension이 클수록 모델의 예측 성능이 떨어지기 때문에, 위에서 제시된 Output Dimension을 줄이는 방법들을 통해 모델의 Search Space를 줄입니다. 이를 통해 성능을 향상시킵니다.

태양광 발전량에 영향을 주는 변수들⁶⁷

기온
최고기온
최저기온
일사량
습도
강수형태
강수확률
강수량
하늘상태
운량
풍속
풍향
적설

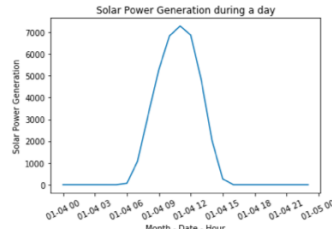
기온
최고기온
최저기온
일사량

도달하는 빛에너지의
양에 비례

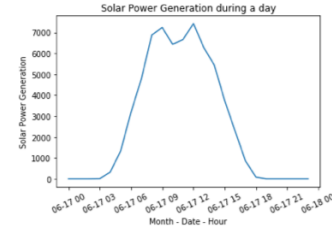
강수형태
강수확률
강수량
하늘상태
운량
적설

발전량에 영향을
주는 변수들

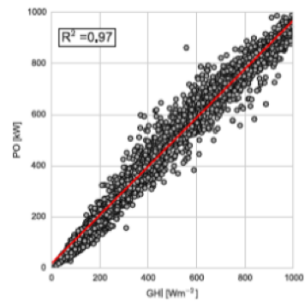
빛에너지의 도달을
방해하는 요소



하루중 태양광 발전량
(맑은날)



하루중 태양광 발전량
(흐린날)



X: 일사량, Y: 태양광 발전량

- 태양광 발전량은 일사량에 선형으로 비례
- 일사량은 $\frac{1}{(\text{발전소}-\text{태양간 거리})^2}$ 에 비례하여 종 모양을 보임
- 이 모양이 기후 영향에 따라 어떻게 변하는지를 찾아내는것이 과제

태양광 발전량은 하루중의 시간에 따라 종모양의 일정한 형태를 띄고 기후에 의해 그 모양과 스케일이 변합니다. 따라서 맑은날의 태양광 발전량으로부터 생성한 Empirical Distribution을 문제1의 인풋으로 추가하여, 모델이 기후에 의해 이 종 모양이 어떻게 변하는지만 찾는 데에만 집중할 수 있도록 Optimization 부담을 줄입니다.

SK ICT 파트너 회사로서 데이터 활용 협력

SK ICT파트너사로서, 기상청 데이터보다 훨씬 세세한 구역에 대한 정보를 제공하는 SK플래닛 웨더포그 기상 서비스 데이터도 활용하여 데이터 범위와 품질을 향상시키고자 함. 태양광발전소를 건설하고 운영하는 SK D&D, SK E&S와의 협력관계를 통해 발전산업 전문지식을 협조받아 도메인 지식을 활용한 모델링으로 모델 성능 향상.

예측 기법

문제정의에 따른 모델 제시

문제	모델	설명
문제1	CatBoost, XGBoost, LightGBM	Categorical Variable인 풍속을 다루는데에 CatBoost가 다른 두 모델보다 유리하나, 세 모델 모두 Regression을 시도하여 성능 비교
문제2	LSTM, Transformer	Sequence to Sequence를 다루는 LSTM과 Transformer를 활용. Transformer가 이론적으로 LSTM보다 학습이 빠르나, 두 모델 모두 학습하여 성능 비교
문제3	LSTM, Transformer	n일동안의 기상 관측치를 Input으로 활용하여 이후 +72시간동안의 발전량을 예측. 두 모델 모두 시도하여 성능 비교
문제4	CNN + LSTM with Attention	시간에 따라 변하는 위성 이미지의 Feature를 CNN으로 추출. Attention을 통해 어떤 Feature가 영향을 많이 미치는지 학습. 추출된 Feature를 LSTM의 Input으로 활용하여 이후 +72시간동안의 발전량을 예측

Feature Engineering Automation

가능한 모든 Feature Engineering Idea를 제시하고, Bayesian Optimization을 통해 최적의 성능을 내는 Feature Engineering Idea의 Combination을 탐색합니다. 각각의 방법이 적용되었을때의 즉각적인 성능 차이는 개별적인 실험으로 확인할 수 있지만 방법간/방법과 모델간 어떤 Interaction이 있는지 알지 못하기 때문에 Grid Search로 이를 찾아야 하는데, 이를 Optimization Method를 통해 Grid Search보다 빠르게 찾는 방법입니다. 아래는 적용할 수 있는 아이디어들의 예시이며, 이외에도 데이터 분석과 도메인 지식을 통해 다양한 방법을 적용할 수 있습니다.

- 맑은날 발전량의 Empirical Distribution을 Input으로 추가
- 예보와 관측치를 비교하여, 관측치와 크게 차이나는 예보는 트레이닝 데이터에서 제거
- Singular Value Decomposition으로 Output Dimension 축소
- Polynomial Features 추가 (Kaggle 데이터⁸를 활용하여 분석했을때 성능 향상 확인)

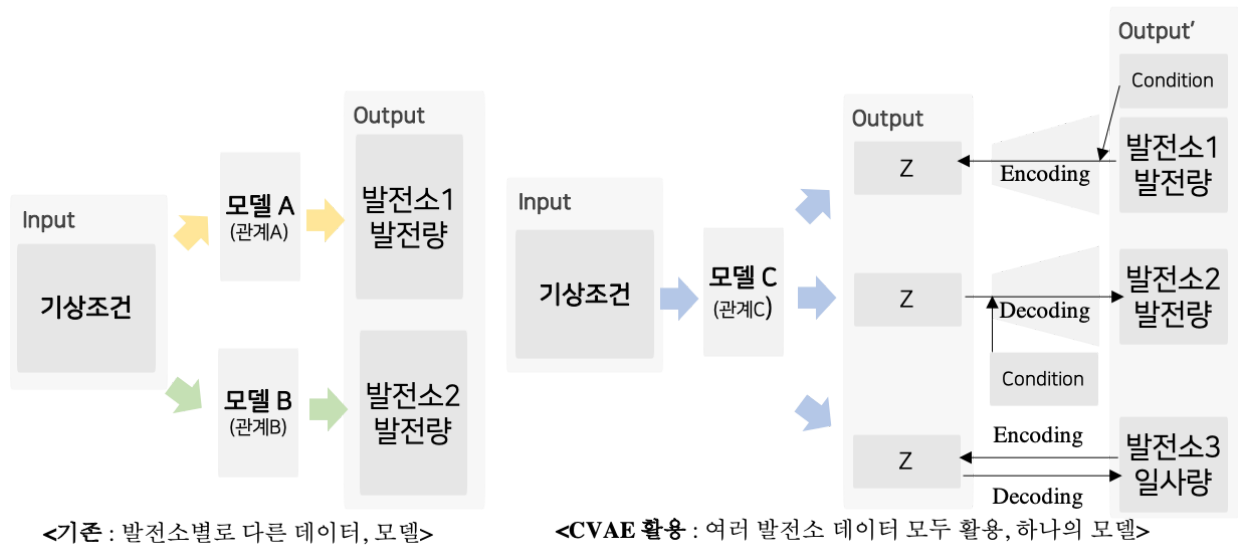
⁶ '태양광발전설비 원격 관제를 위한 빅데이터 분석 및 처리(2018)', 권준아 외 3명

⁷ '수상태양광 발전 시스템의 환경에 따른 모듈의 전기적 특성(2018)', 황수현 외 3명

⁸ AMS 2013-2014 Solar Energy Prediction Contest, Kaggle

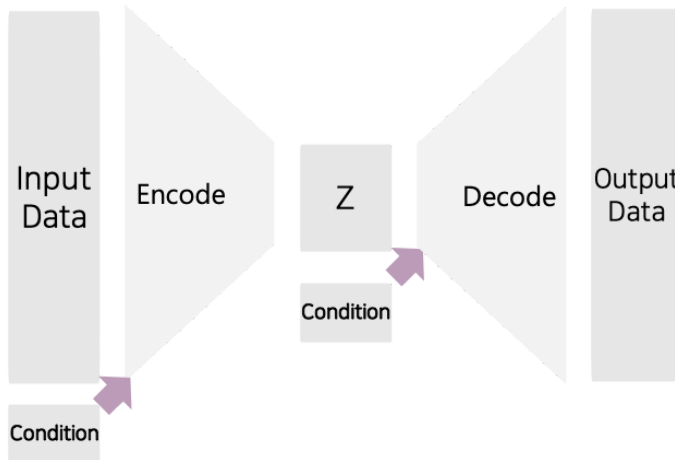
타지역 일사량, 태양광 발전소 데이터⁹ 활용

기상자원기술개발과 Kaggle에 공개된 한국/유럽의 일사량/태양광 발전량 데이터를 활용하여 데이터수 부족 문제를 해결하고자 합니다. 동일한 기상조건에서도 위도차이, 설비차이로 태양광 발전소는 다른 출력의 발전량을 내게 됩니다.



기존의 방식으로는 동일한 기상조건에 대해 각 발전소가 어떤 발전량을 출력하는지에 대한 개별적인 **관계A(모델A)**, **관계B(모델B)**만을 학습할 수 있었습니다.

Conditional Variational Autoencoder (CVAE)



Input Data : 해당 지점의 일사량 또는 해당 지점의 태양광 발전량

Condition: 데이터에서 제거하고 싶은 내재된 요소 (위도 등)

Z : 데이터에서 Condition이 주는 영향이 제거된 Location Invariant한 압축된 데이터

Output Data : Z로부터 Condition을 통해 해당 지점의 태양광 발전량 복원

Input Data와 Output Data는 같다

Conditional Variational Autoencoder(CVAE)¹⁰를 활용하면 각 발전소의 위도/설비 등의 차이를 나타내는 특징을 Condition으로 두어 Encoding함으로써, 각 발전소간의 차이가 제거된 발전량 Z(Output)를 얻을 수 있습니다. 모델C(문제1~문제4를 푸는 모델)는 주어진 기상조건에 대해서 항상 같은 Z(Output)를 예측하도록 트레이닝됩니다. 마지막으로 Z와 각 발전소에 대한 Condition을 이용하여 Decoding함으로써, 각 발전소의 위도/설비가 반영된 최종적인 예측 발전량(Output')을 얻게 됩니다. 이러한 방법을 통해 위치와 설비가 다른 모든 발전소의 데이터(Output')을 이용하여 모델C와 CVAE를 트레이닝할 수 있게 됩니다. 더 많은 기상조건-발전량 관계를 학습했으므로, 단일 발전소의 데이터만을 활용하여 트레이닝시킨 모델A, B보다 성능이 높습니다.

활용 방안

다른 발전소 데이터로부터 **Transfer Learning**으로 신규 발전소도 높은 정확도로 예측

상대적으로 데이터수가 적은 신규발전소도 해당 발전소의 데이터로 CVAE에서 Condition만 학습하여 이미 Training된 모델을 통해 높은 정확도로 발전량을 예측할 수 있다는 장점이 있습니다.

건설 예정인 신규 발전소에 대하여 일사량을 측정함으로써 **Transfer Learning**으로 연간 발전량 예측

건설 예정인 신규 발전소에 대하여 어느정도의 발전량이 나올지 예측하고 싶을때, 해당 위치에서의 일사량을 측정하여 Condition을 구하고, Transfer Learning을 통해 과거 1년간 기상예보로 Inference함으로써 연간 예측 발전량을 구할 수 있습니다.

검증된 예측 모델을 스케줄링 시스템 내 연동

AI 모델의 시스템 내 적용을 위해서는 변화하는 환경에 맞춰 지속적으로 모델을 학습시킬 수 있는 Continual Learning 기술과 다양한 모델을 효율적으로 관리할 수 있는 AI Model Management 기술이 필요. 팀이 보유하고 있는 검증된 AI Operation & Deployment 기술을 활용하여 실 시스템 적용을 타깃으로 함.

⁹ Solar Generation and Demand Italy 2015-2016, Kaggle

¹⁰ 'Learning Structured Output Representation using Deep Conditional Generative Models(2015)', Sohn et al