



지상기기 상시감시 머신러닝서비스

김윤승 연구원

INDEX

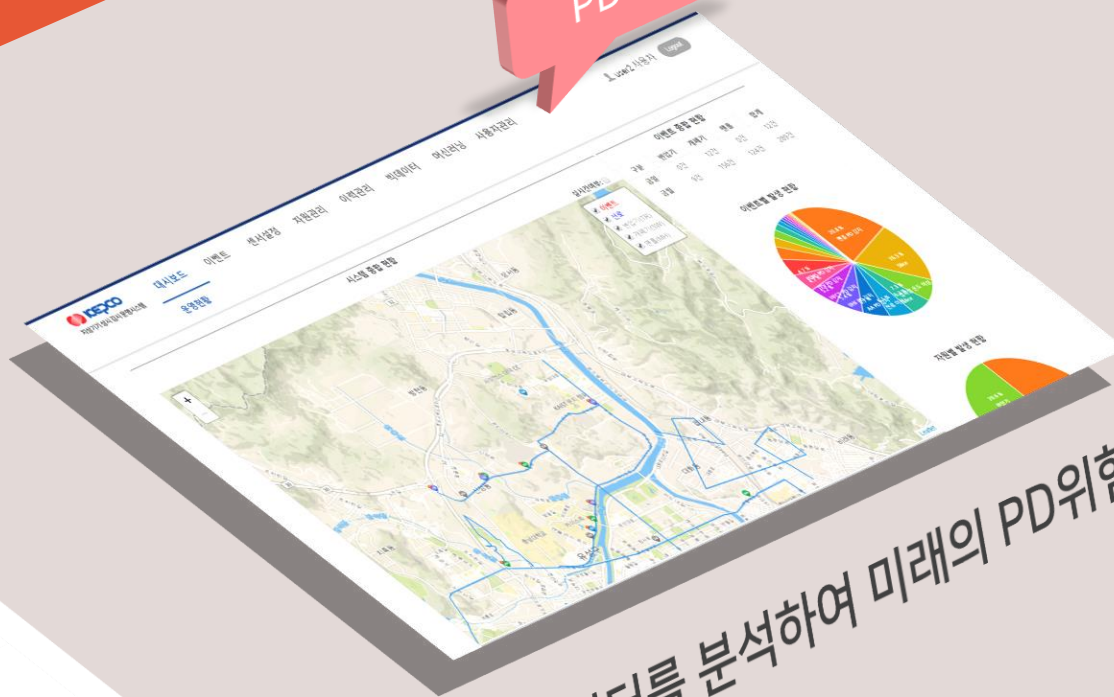
서비스별 분류

PD분류



PD 발생시 어떤종류의 PD인지 자동분류

PD예측



데이터를 분석하여 미래의 PD위험 예측

PD분류

정형 데이터 학습예시

날짜	오늘날씨	오늘 중국 미세먼지 수치	오늘 온도	미세먼지 위험 여부
2019-3-1	흐림	299	10	0
2019-3-2	흐림	400	19	0
2019-3-3	맑음	99	10	X

오늘 날씨 : 날씨가 맑을 경우, 미세먼지 위험이 없었다

오늘 중국 미세먼지 수치 : 100 이하인 경우, 미세먼지 위험이 없었다

⋮

2019-3-9	맑음	80	10	?
----------	----	----	----	---

데이터가 행렬에 맞춰 규칙적으로 구성됨

PD분류

PD 데이터 예시

	0	1	2	3	4	5	6	...	4130	4131	4132	4133	4134	4135	4136	
X	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	360.00	360.00	360.00	360.00	360.00	360.00	360.00	
Y	2.82	4.24	5.65	7.06	8.47	9.88	11.29	...	2.82	4.24	5.65	7.06	8.47	9.88	11.29	

	0	1	2	3	4	5	6	7	8	9	...	5067	5068	5069	5070	5071	5072	5073	5074	5075	5076
X	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	360.00	360.00	360.00	360.00	360.00	360.00	360.00	360.00	360.00	360.00
Y	2.82	4.24	5.65	7.06	8.47	9.88	11.29	12.71	14.12	15.53	...	2.82	4.24	5.65	7.06	8.47	9.88	11.29	12.71	14.12	15.53

	0	1	2	...	1736	1737	1738
X	1.55	1.55	1.55	...	358.73	358.73	358.73
Y	4.38	5.79	7.20	...	4.38	5.79	7.20

모양이 제각각이다

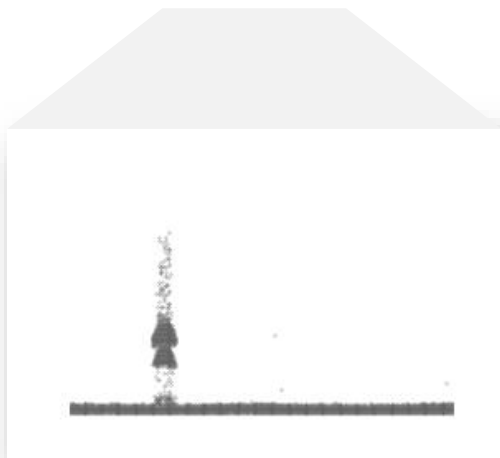
PD발생시에 만들어 지기 때문에 균일하지 못하다.

정형데이터 학습에 적합하지 않다!

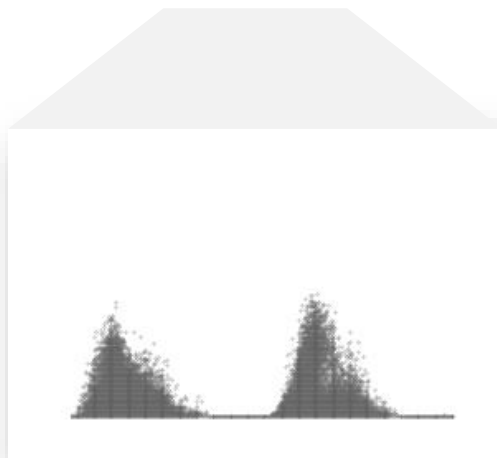
PD분류

시각화이미지로 변환하면 어떨까?

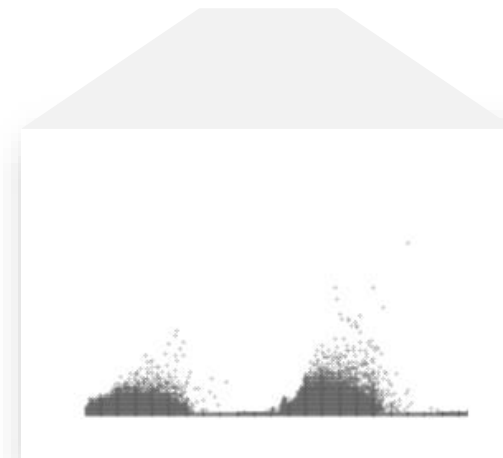
	0	1	2	3	4	5	6	...	4130	4131	4132	4133	4134	4135	4136
X	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	360.00	360.00	360.00	360.00	360.00	360.00	360.00
Y	2.82	4.24	5.65	7.06	8.47	9.88	11.29	...	2.82	4.24	5.65	7.06	8.47	9.88	11.29



	0	1	2	3	4	5	6	7	8	9	...	5067	5068	5069	5070	5071	5072	5073	5074	5075	5076
X	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	360.00	360.00	360.00	360.00	360.00	360.00	360.00	360.00	360.00	360.00
Y	2.82	4.24	5.65	7.06	8.47	9.88	11.29	12.71	14.12	15.53	...	2.82	4.24	5.65	7.06	8.47	9.88	11.29	12.71	14.12	15.53



	0	1	2	...	1736	1737	1738
X	1.55	1.55	1.55	...	358.73	358.73	358.73
Y	4.38	5.79	7.20	...	4.38	5.79	7.20



이미지를 활용한 학습이 가능하다!

데이터 수는 각각 다르지만
산점도로 표시했을때 전체적인 모양으로 학습가능

PD분류

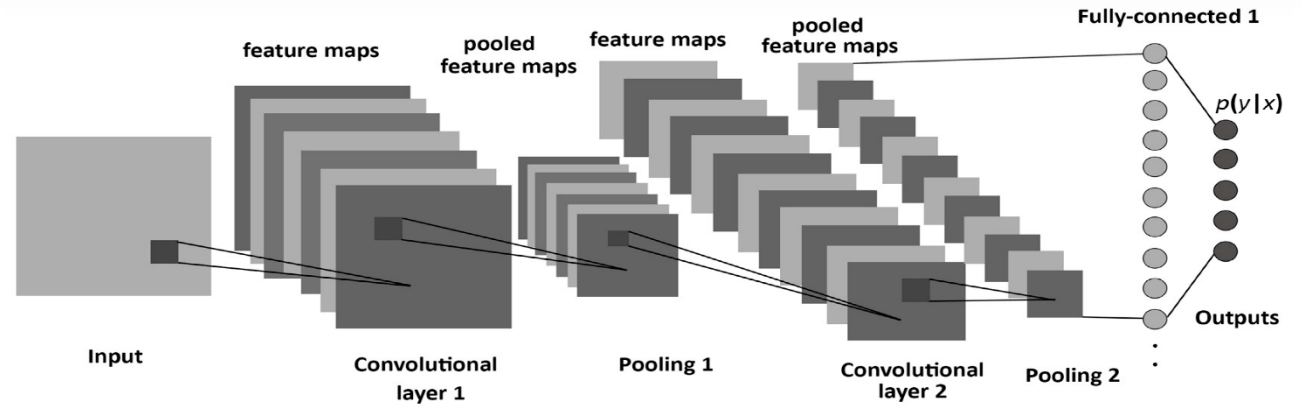
알고리즘소개

CNN

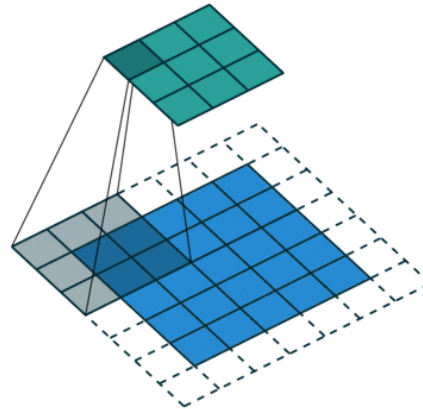
이미지인식에 유용한 딥러닝 알고리즘

합성곱 연산을 통해
이미지의 특징을 추출
차원을 축소시키며 연산을 줄인다
축소, 회전, 크롭 등
다양한 변화에도 안정적
학습해야할 가중치가 줄어들므로
빠른 학습가능

Architecture



What is "Filter"



PD분류

알고리즘소개

RESNET

단점을 개선한 CNN 모델

ILSVRC 2015년 1등

Gradient vanishing 해결

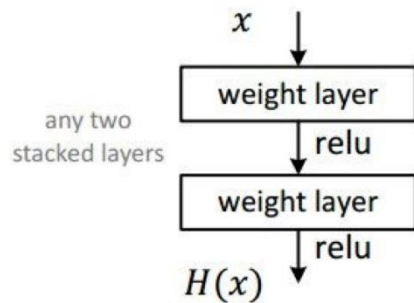
잔차를 학습하는 방식

복잡하지 않은 구현방식

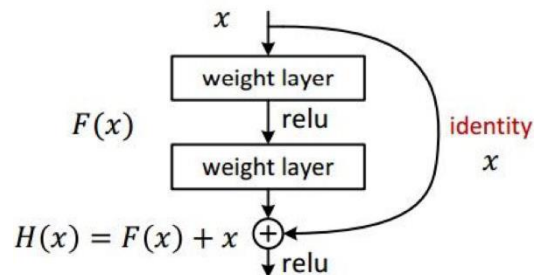
Bottleneck 방식

CNN vs RESNET

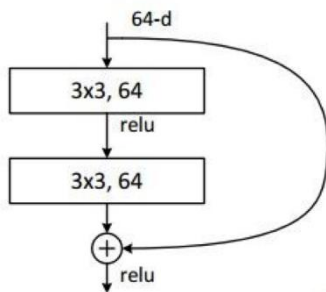
• Plain net



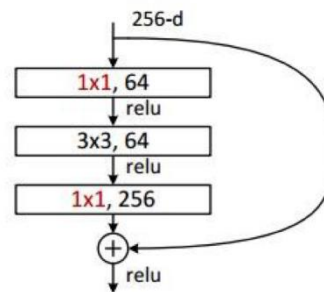
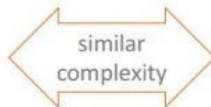
• Residual net



What is "bottleneck"



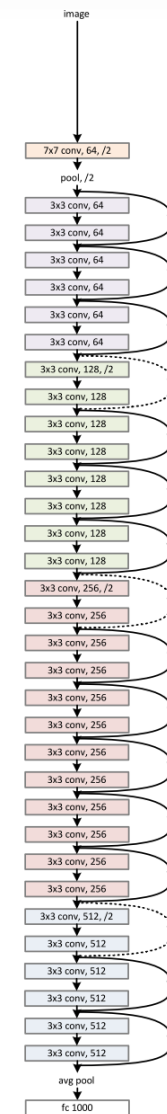
all-3x3



bottleneck
(for ResNet-50/101/152)

34 layers

34-layer residual



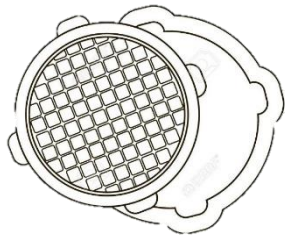
PD분류

데이터플로우

PD 발생시

Bigdata Server

Local Sensor



Manhole

PD database



데이터 추출

Raw data

	0	1	2	3	4	5	6	...	4130	4131	4132	4133	4134	4135	4136
X	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	360.00	360.00	360.00	360.00	360.00	360.00	360.00
Y	2.82	4.24	5.65	7.06	8.47	9.88	11.29	...	2.82	4.24	5.65	7.06	8.47	9.88	11.29

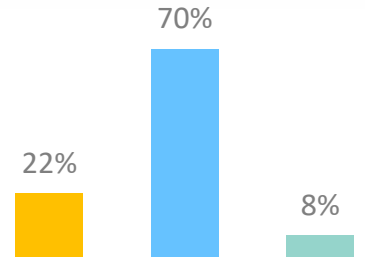


결측치처리
이상치처리
데이터표준화
시각화

Refined data



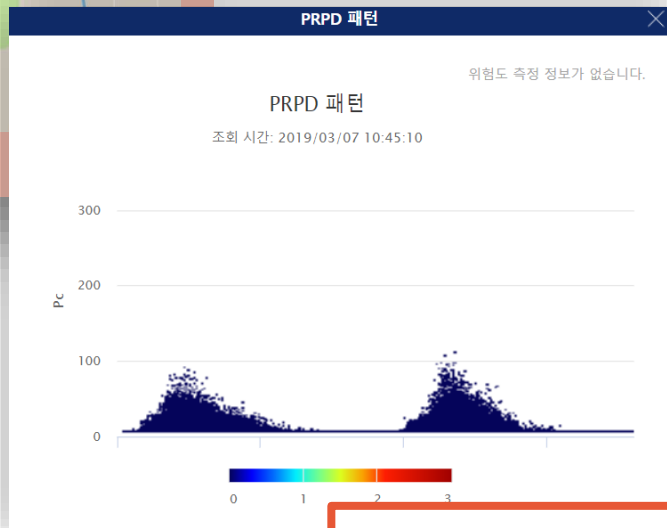
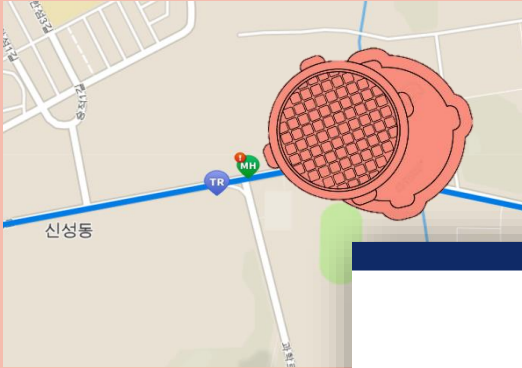
Classification



Surface!

PD분류

구현예시



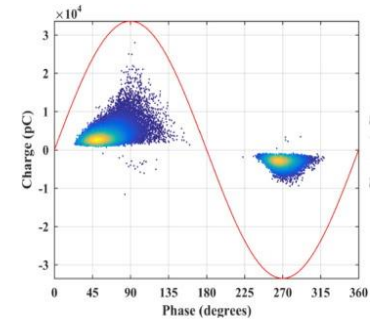
머신 러닝 예측 결과

PD 패턴	예측 율
Corona	3.74 %
Void	4.06 %
Surface	92.19 %

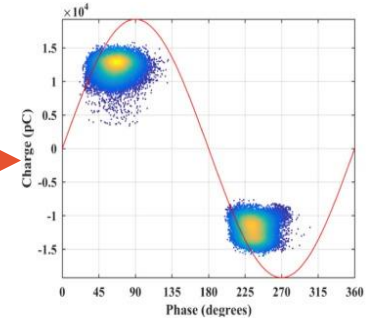
모델 정확도 99%

학습데이터 적용시

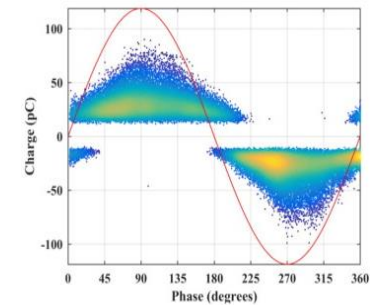
CORONA



SURFACE



VOID



PD예측

데이터플로우

Bigdata Server



실시간 데이터DB



일일 최대 기록
일일 평균 기록



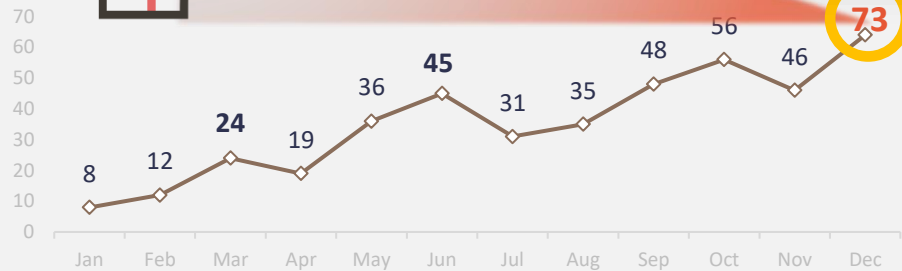
일일 데이터DB

Machine Learning Server



단위기간 데이터
e.g. 이전 7일치

Predict



임계점 초과

Warning



PD분류

알고리즘소개

LSTM

이전 정보를 기억하는 딥러닝 알고리즘

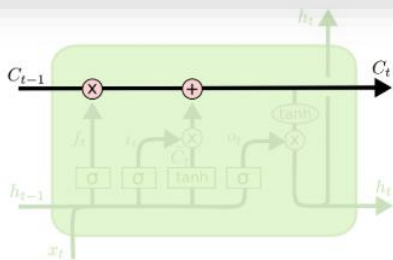
현재 데이터를 연산할때 이전 데이터를 반영

비선형연산을 통해 기존 시계열 모델보다 정확함

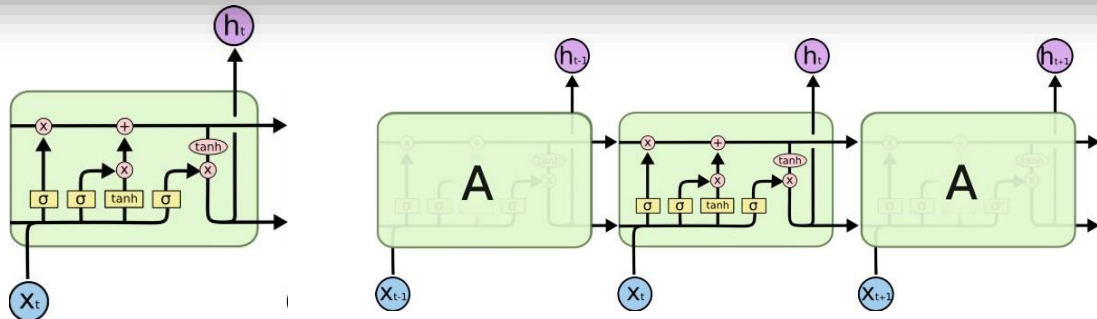
Cell state를 통해 Long-term dependency 해결

LSTM 층을 겹겹히 쌓으며 더욱 견고한 예측 가능

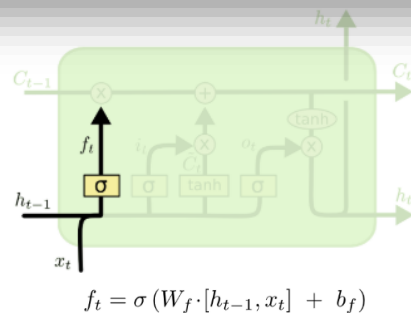
Cell state



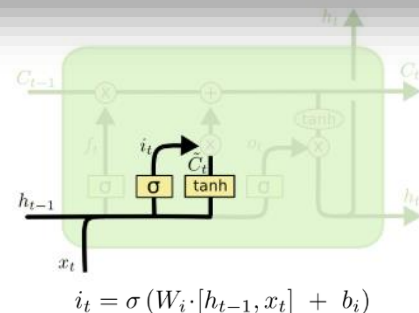
Architecture



Detail part

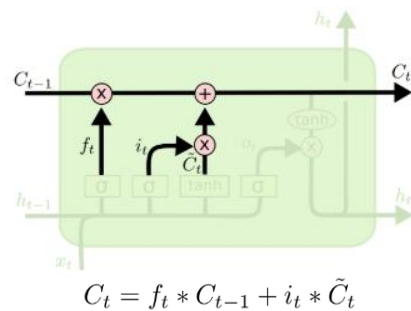


$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

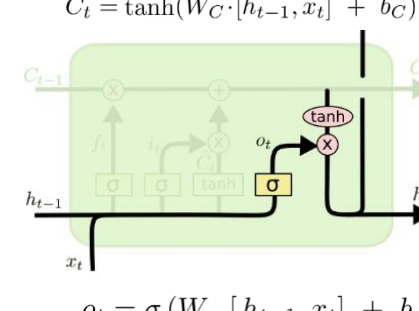


$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

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



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



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

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

PD분류

알고리즘소개

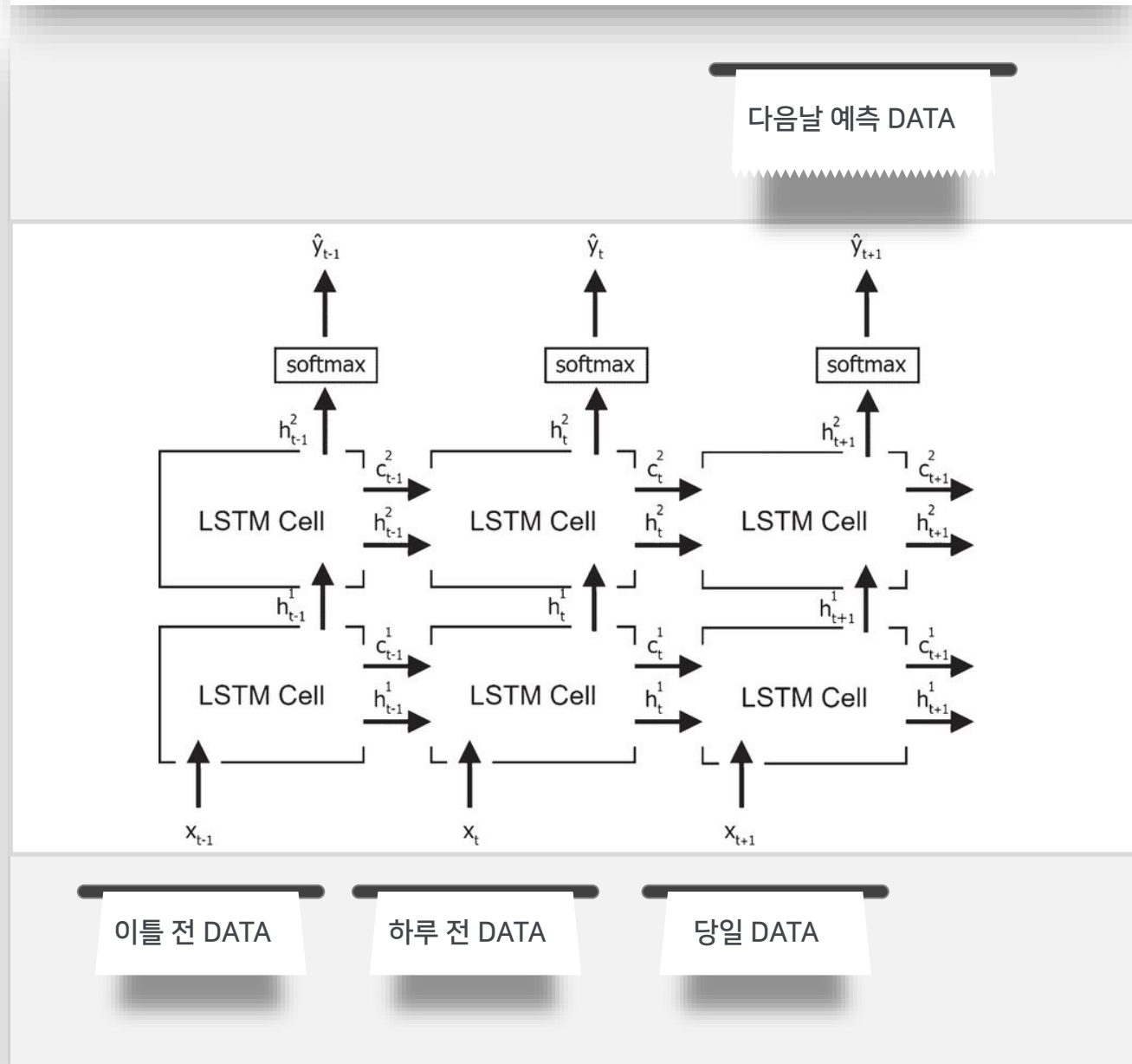
Stateful Stacked LSTM

상태유지 스택 순환신경망

LSTM 층을 쌓아올려 예측을 유연하게 만듦

상태유지를 통해 기존 학습정보를 유지함

적용 예시





감사합니다

