

시계열 분석

구름

도시공학과 일반대학원

한양대학교

1. 시계열 분석

2. RNN

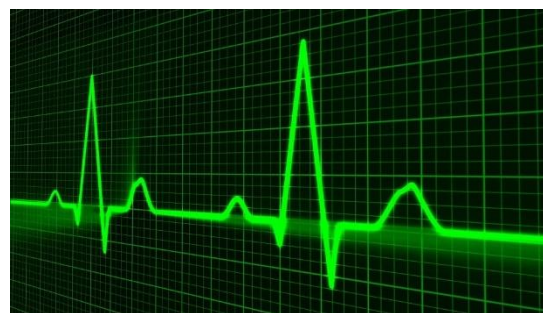
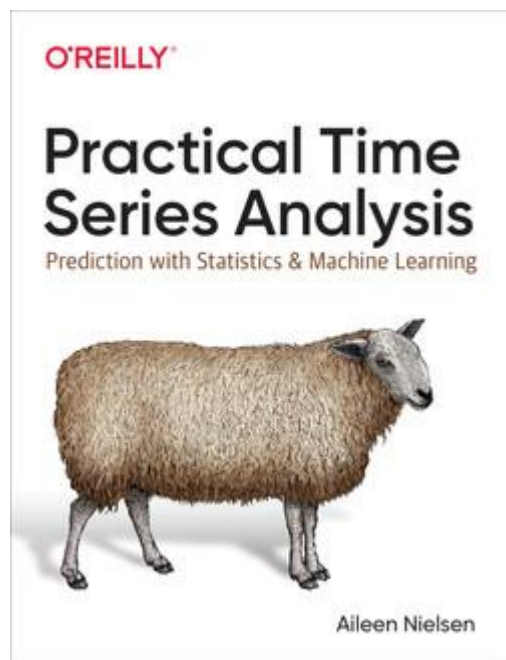
3. LSTM

4. 순환 신경망 응용

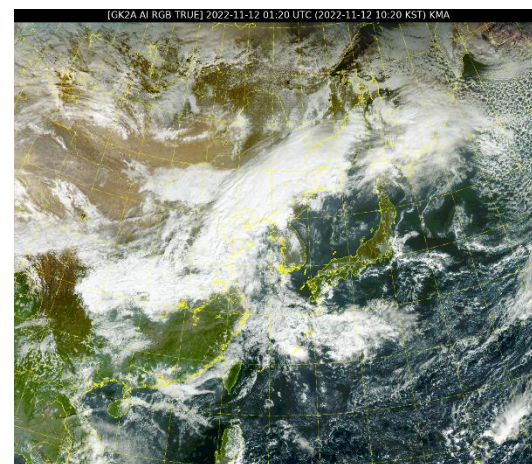
시계열 분석 (Time Series Analysis)

Time series analysis is the endeavor of extracting meaningful summary and statistical information from points arranged in chronological order.

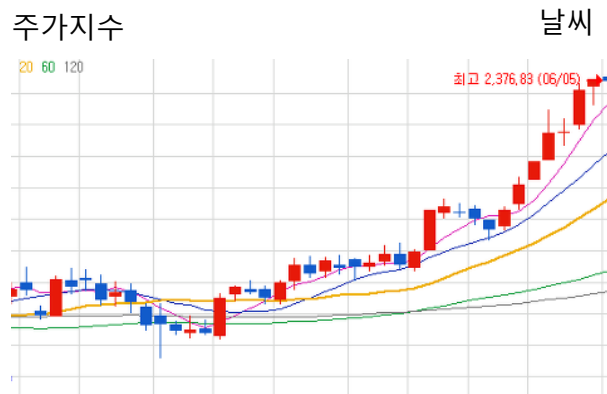
<https://www.oreilly.com/library/view/practical-time-series/9781492041641/>



심장 박동



날씨

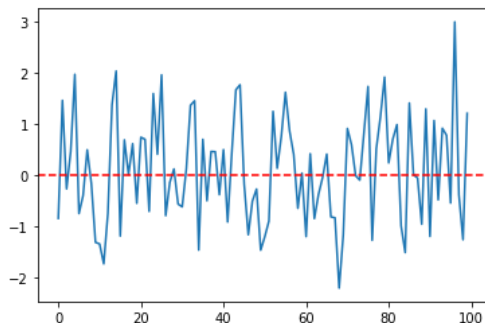


주가지수

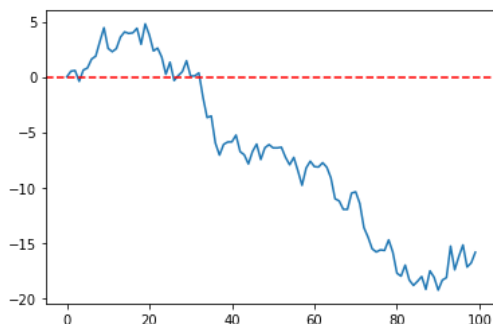


천문학

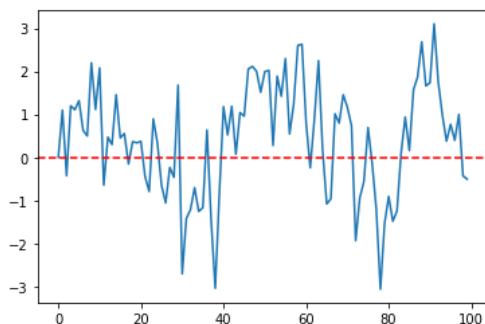
시계열 데이터



```
1 #백색 잡음 (White Noise Process)
2
3 data = np.random.normal(0, 1, 100)
4
5 plt.plot(data)
6 plt.axhline(0, 0, 100, color='red', linestyle='--')
7 plt.show()
```



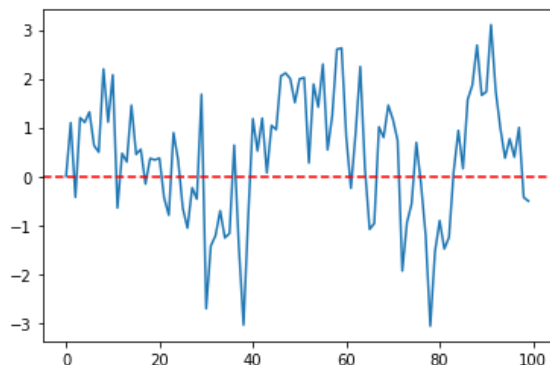
```
1 #확률보행과정 (Random Walr Process)
2
3 data = np.array([0])
4
5 for i in range(99):
6     data = np.append(data, data[-1] + np.random.normal(0, 1, 1))
7
8 plt.plot(data)
9 plt.axhline(0, 0, 100, color='red', linestyle='--')
10 plt.show()
```



```
1 #정상 확률 과정 (Stationary Proecess)
2
3 data = np.array([0])
4 A_corr = 0.5
5
6 for i in range(99):
7     data = np.append(data, data[-1] * A_corr + np.random.normal(0, 1, 1))
8
9 plt.plot(data)
10 plt.axhline(0, 0, 100, color='red', linestyle='--')
11 plt.show()
```

AutoRegressive Integrated Moving Average

ARIMA 모형은 정상시계열 데이터에 적합한 모델로 Box-Jenkins Model을 통해 데이터 정상성 여부를 확인 후 사용



AR 모형 : $x_t = \beta \cdot x_{t-1} + \alpha + \varepsilon_t$

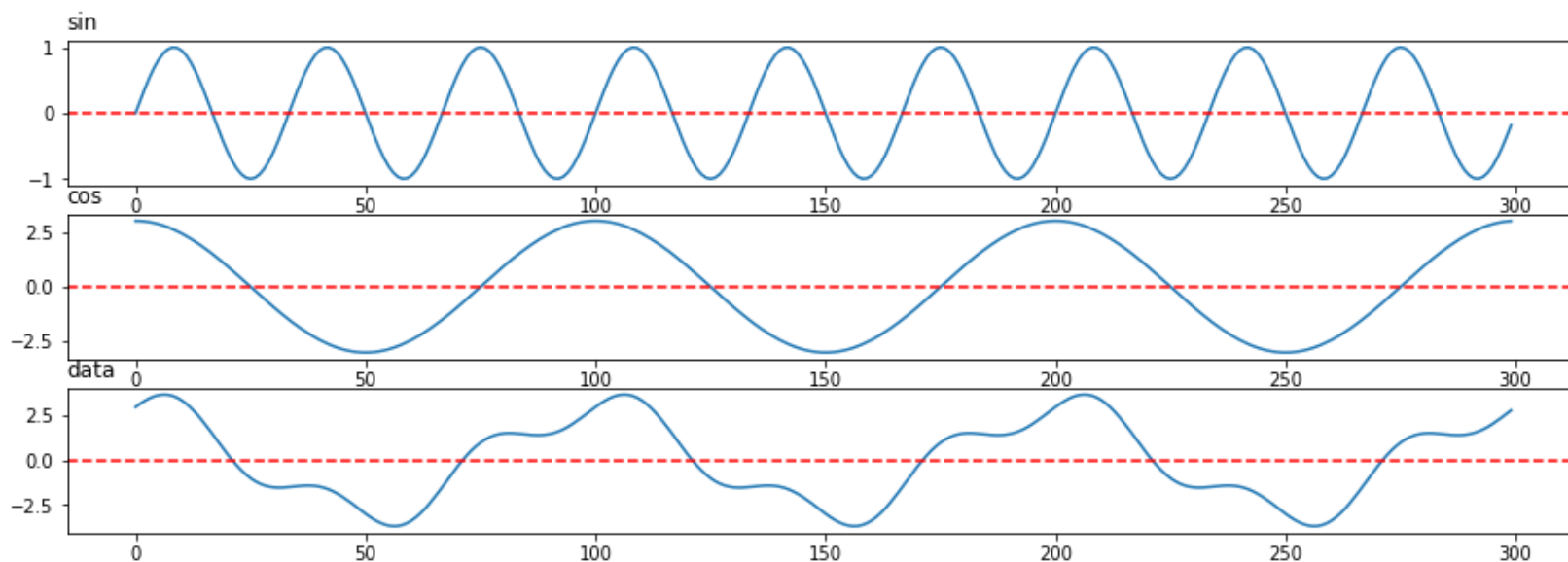
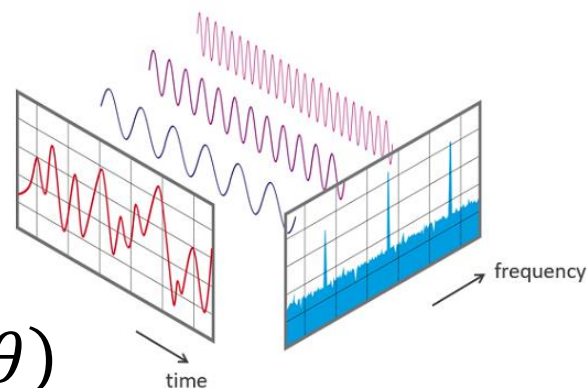
MA 모형 : $x_t = \alpha + \varepsilon_t + \beta \cdot \varepsilon_{t-1}$

Interated : $x_t = x_t - x_{t-1}$

Fourier Transform

Sin, Cos 함수를 이용하여 신호를 분리
주파수, 음성 인식 등에 활용

$$x_t = a \cdot \sin(b \cdot \theta) + c \cdot \cos(d \cdot \theta)$$

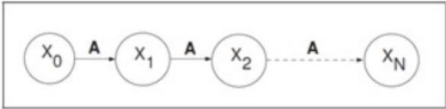


Hidden Markov Model

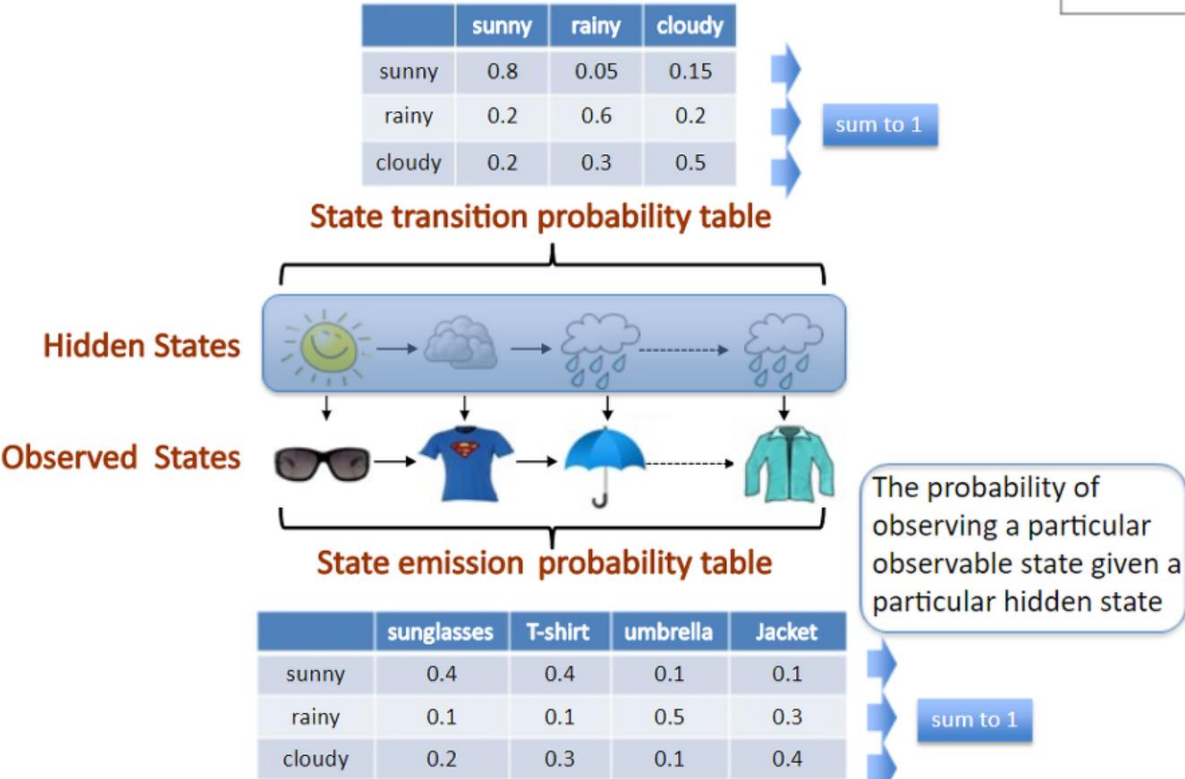
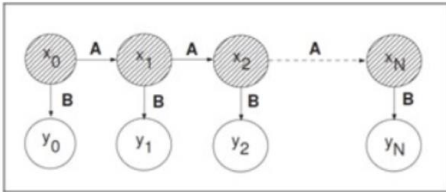
은닉층이 마코프 모델을 따른다는 가정으로 예측
날씨, 음성인식 등에 활용

<https://bioinformaticsandme.tistory.com/53>

A Markov Model



A Hidden Markov Model



브라운 운동 (Brownian Motion)

1827년 식물학자 로버트 브라운이 꽃가루 입자 관찰에서 발견

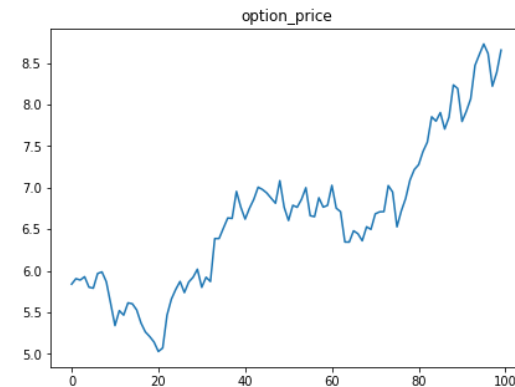
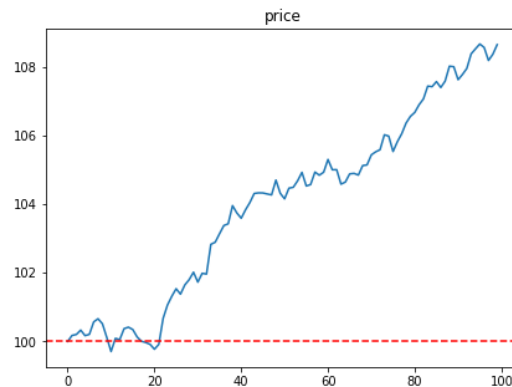
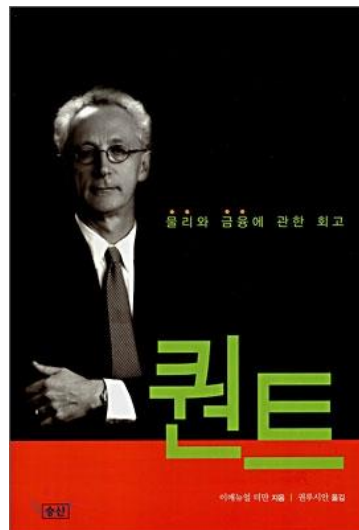
아인슈타인이 브라운 운동 방정식을 제시, 물리학자인 페렝이 증명

루이 바슐리에 1900년 투기이론으로 금융시장 가격 변동을 브라운운동으로 모형화

1950년대 중반 폴 새뮤얼슨이 기하 브라운운동을 발표하며 바슐리에 이론을 수정

1973년 피셔 블랙과 마이런 솔즈는 옵션 가격 결정이론을 발표

<http://www.yes24.com/Product/Goods/2641000>



Prophet - Facebook

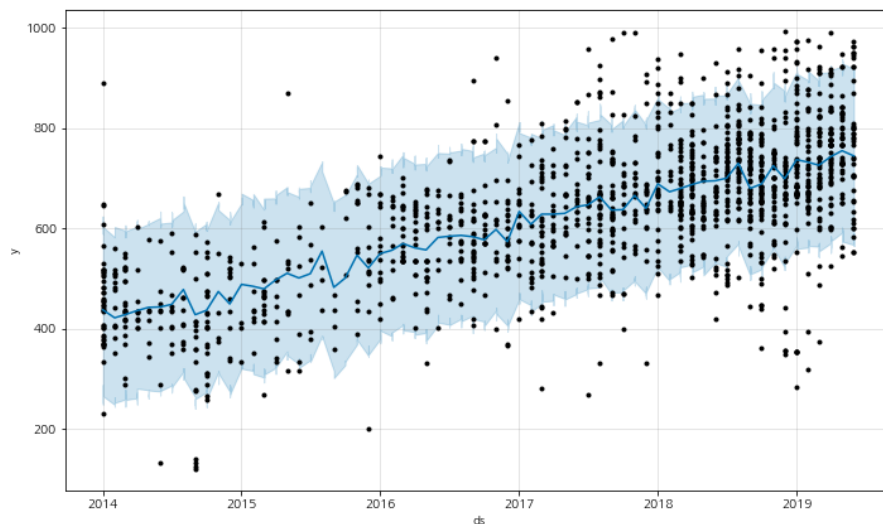


페이스북에서 비즈니스 분야에 적합하게 만든 시계열 예측 알고리즘

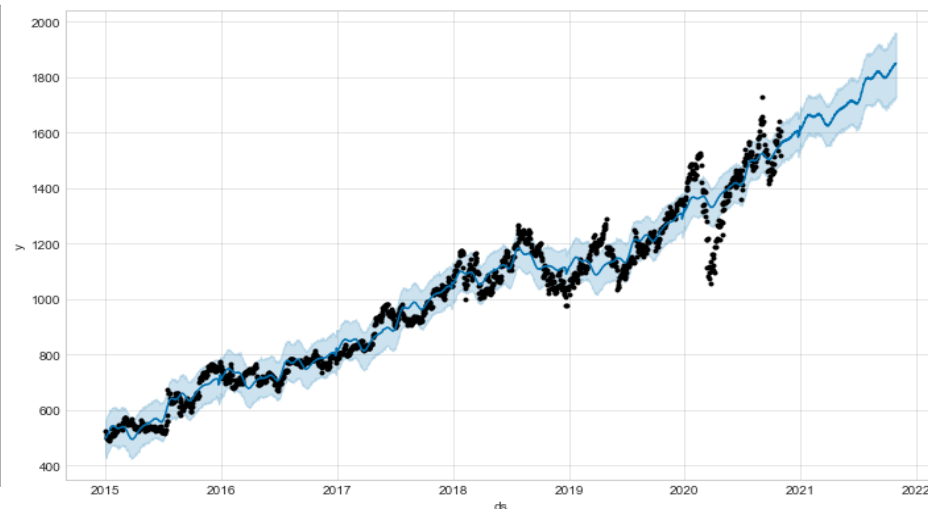
시계열 추세에 계절적 요소와 휴일 요소를 추가하여 비즈니스에 적합하게 구성

데이터에서 변동성 구간, 이상치 검출, 미래 예측, 예측 범위 등 다양하게 제공

<https://facebook.github.io/prophet/>



(빅밸류) 서울 클러스터내 30㎡ 이하 건축년한 0~1년 이하 실거래 단가 Prophet 분석 결과

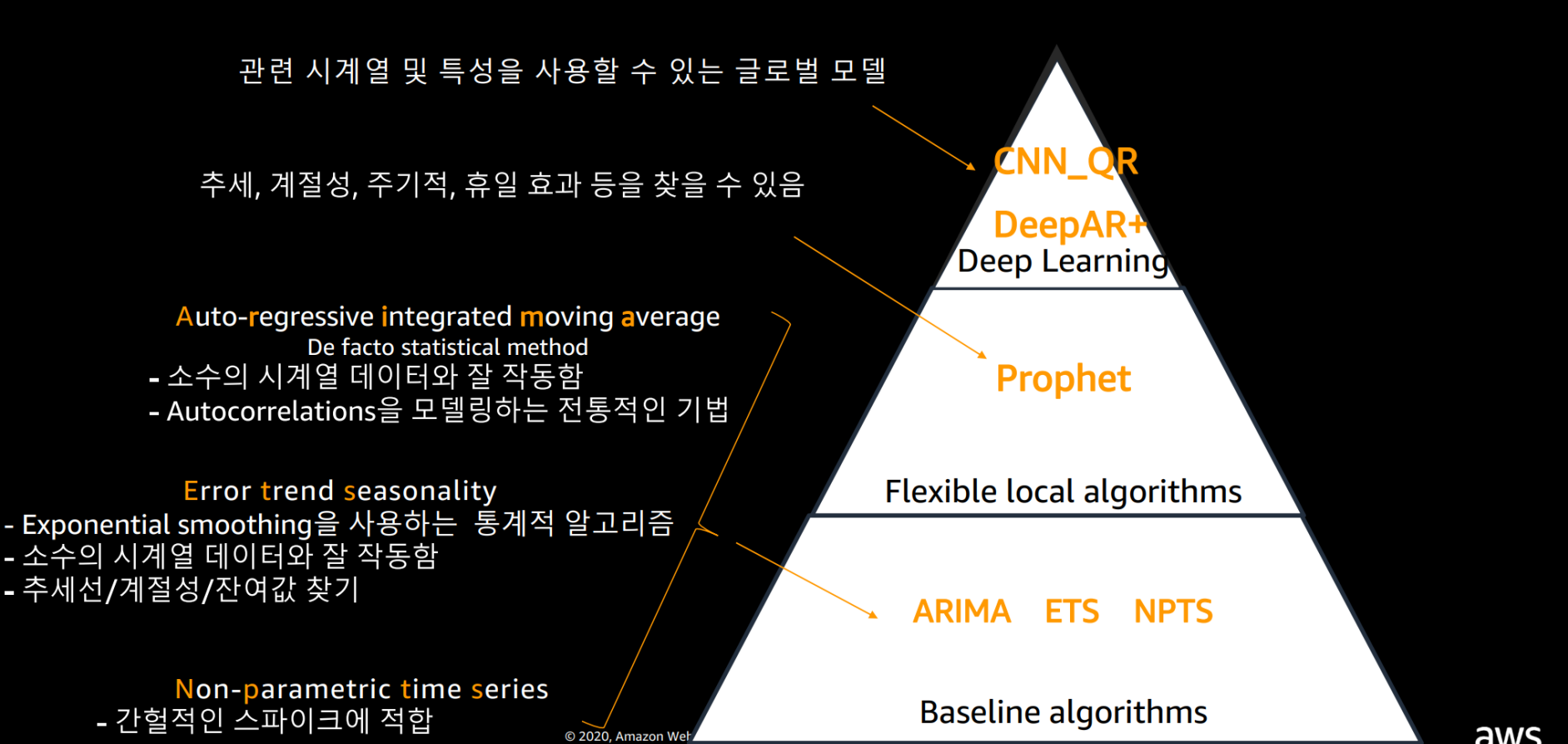


(Medium) Predicting Google's Stock Prices Using Facebook's Prophet
<https://medium.com/mlearning-ai/predicting-googles-stock-prices-using-facebook-s-prophet-4829c83a8590>

Deep AR+ - Amazon Forecast

아마존에서 인공지능경망을 이용해 제시한 시계열 분석 라이브러리.

https://docs.aws.amazon.com/ko_kr/forecast/latest/dg/aws-forecast-recipe-deeparplus.html



1. 시계열 분석

2. RNN

3. LSTM

4. 순환 신경망 응용

부동산원의 주택 매매가격 지수

https://www.reb.or.kr/r-one/statistics/statisticsViewer.do?menuId=HOUSE_21211

부동산 통계 뷰어 - Chrome

reb.or.kr/r-one/statistics/statisticsViewer.do?menuId=HOUSE_21211

R-ONE 부동산통계뷰어

전국주택가격동향조사 > 종합주택유형 > 매매가격지수 > 매매가격지수

지역전체

범위월간

검색기간2021년10월 ~ 2022년09월

확인

정렬오름차순

엑셀전체받기

변동률

작성 통계리스트

전국지가변동률조사

전국주택가격동향조사

월간동향

종합주택유형

매매가격지수

계절조정 매매가격지수

전월세통합지수

전세가격지수

월세가격지수

매매가격

전세가격

월세가격

매매가격 대비 전세가격

전세가격 대비 보증금

수급동향

기타

아파트

연립다세대

단독주택

주간아파트동향

전월세전환율

매매가격지수(종합주택유형)

조회결과

공표보고서

※ 조사·통계 문의: 부동산통계처 주택통계부(053-663-8505, 8516)

(기준월: 2021.06=100)

통계설명

변동률

종합주택 매매가격지수

가격지수

105

104.75

104.5

104.25

104

103.75

103.5

103.7

104.3

104.6

104.7

104.8

104.8

104.8

104.8

104.8

104.8

104.5

103.9

'21.10

'21.11

'21.12

'22.01

'22.02

'22.03

'22.04

'22.05

'22.06

'22.07

'22.08

'22.09

전국

지역	'21.10		'21.11		'21.12		'22.01		'22.02		'22.03		'22.04		'22.05		'22.09
	지수	변동률	지수	변동률	지수	변동률	지수	변동률	지수	변동률	지수	변동률	지수	변동률	지수	변동률	
전국	103.7	0.88	104.3	0.63	104.6	0.29	104.7	0.10	104.8	0.03	104.8	0.02	104.8	0.06	104.8	0.01	104.5
수도권	104.9	1.13	105.7	0.76	106.1	0.33	106.1	0.06	106.1	-0.03	106.1	-0.04	106.1	0.03	106	-0.04	105.5
지방권	102.6	0.67	103.1	0.51	103.3	0.25	103.5	0.14	103.6	0.08	103.6	0.07	103.7	0.09	103.8	0.06	103.5
6대광역시	103.5	0.88	104.2	0.59	104.4	0.26	104.5	0.05	104.5	0.00	104.4	-0.06	104.4	-0.02	104.3	-0.10	104.0
7대광역시	103.0	0.60	103.4	0.40	103.6	0.20	103.6	0.03	103.6	0.03	103.5	0.06	103.5	0.04	103.4	0.08	103.2

월간_매매가격지수....xlsx

전국주택가격동향조사

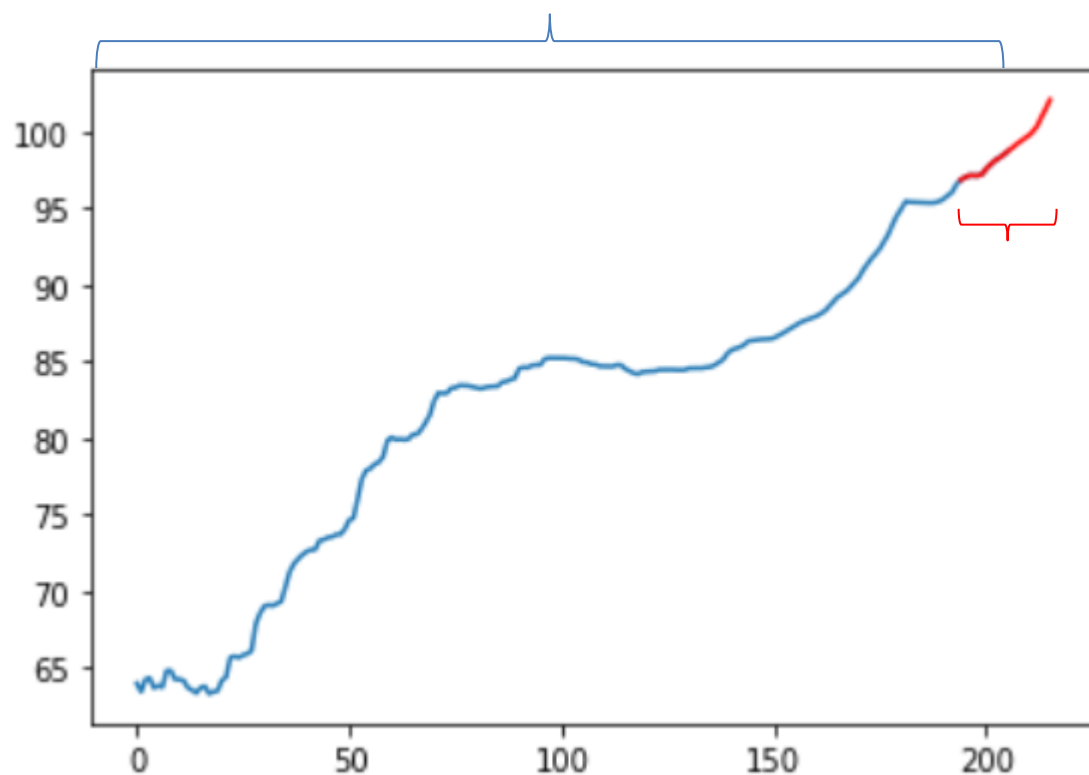
조사개요

> 통계명	전국주택가격동향조사
> 최초작성연도	1986년
> 통계종류	지정·조사통계
> 법적근거	<ul style="list-style-type: none">▪ 국가통계 승인번호: 제304004호▪ 국가통계 승인일자: 1985년 9월 10일▪ 「주택법」 제88조 1항 및 제89조 제2항 제4호 동법 시행령 제91조 제2항 제3호
> 조사목적	전국 주택시장의 매매 및 전세가격을 조사하여 주택시장의 평균적인 가격변화를 측정, 주택시장 판단지표 또는 주택정책 수립의 참고자료 제공
> 유의사항	<ul style="list-style-type: none">▪ 전국주택가격동향조사는 주택가격지수를 산출하여 제공하는 것이 주된 목적이며, 조사항목, 표본추출방법, 지수산출방법 등이 이에 맞게 설계되었음▪ 본 조사는 주택가격 지수산출에 맞게 설계되었기 때문에 평균/중위 주택가격 결과는 참고 사항으로 이용하기 바람
> 조사주기	<div>01 월간 (아파트, 단독, 연립주택)</div> <div>02 주간 (아파트)</div>
> 조사대상	<ul style="list-style-type: none">▪ 대상객체: 아파트, 단독주택, 연립주택(임대 제외)



데이터 분할 : 서울 강북지역 도심권 종로구 주택매매가격지수

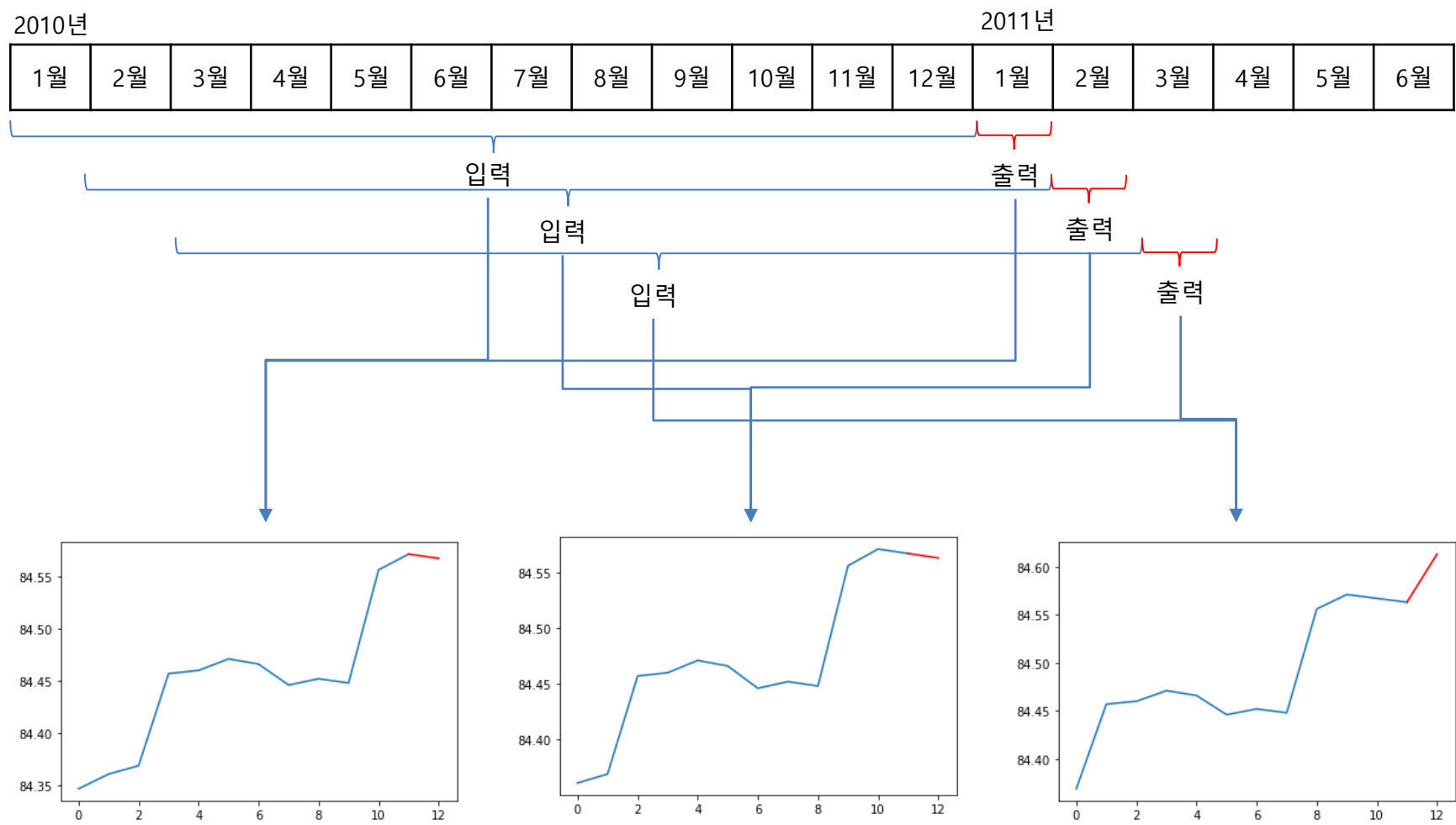
학습데이터 : #2003년 11월~2020년 12월까지



검증데이터 :
#2020년 1월~2021년 5월까지

데이터 구조

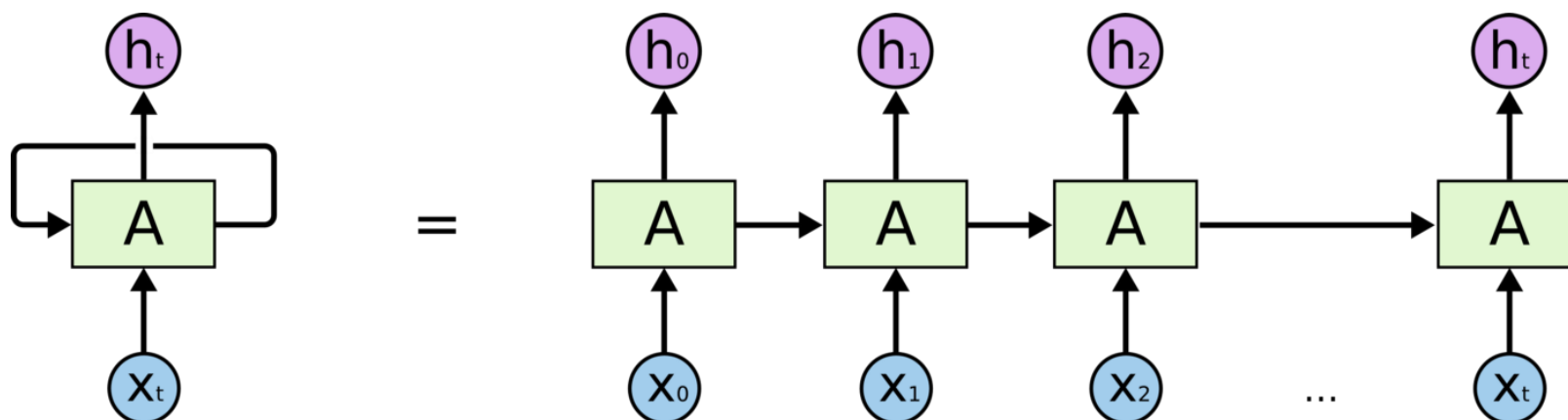
12개월의 지수를 입력하고, 13개월의 지수를 예측하는 모델로 데이터 변경



순환신경망 (Recurrent Neural Network)

시계열 정보를 학습하는 신경망 모델을 의미

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



Simple RNN

```
inputs = layers.Input(shape=(ntraining_x.shape[1], 1))
net = layers.BatchNormalization()(inputs)
net = layers.SimpleRNN(1, activation='relu')(net)
net = layers.Dense(1, activation='linear')(net)
model = Model(inputs, net)

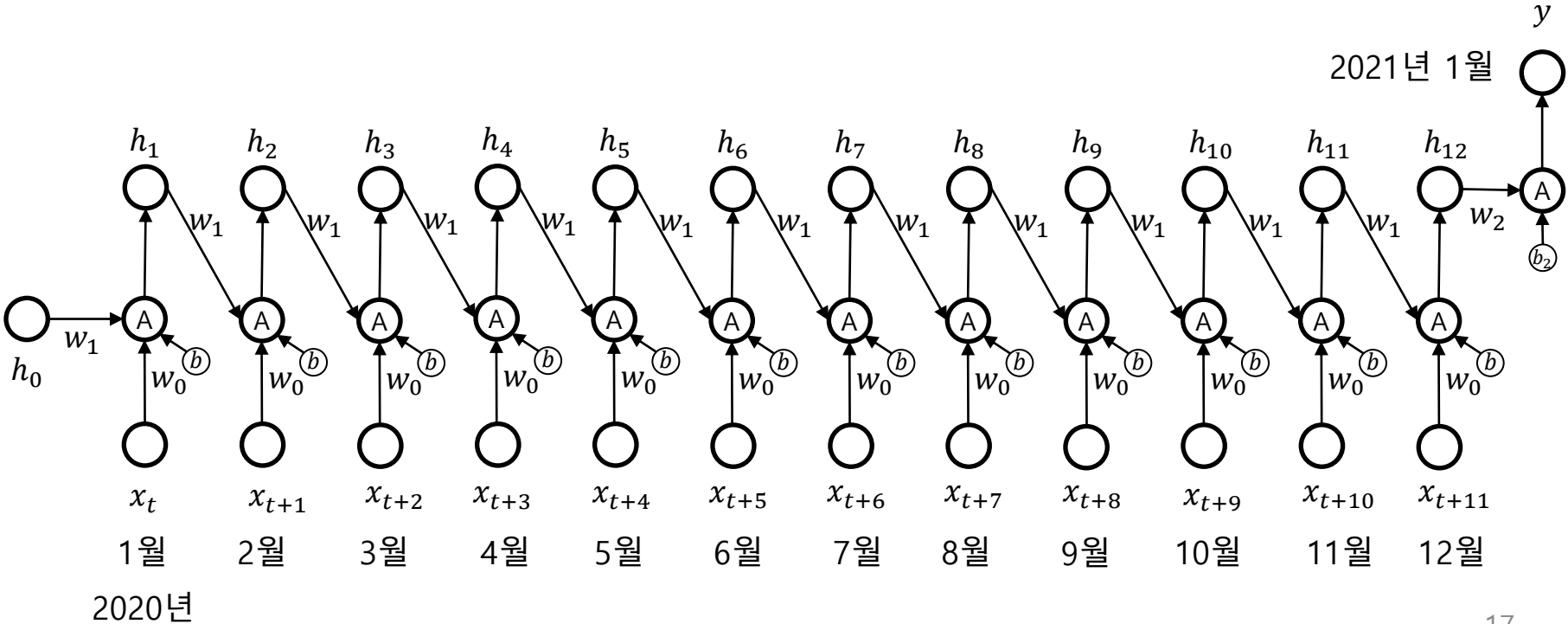
model.compile(optimizer='adam', loss='mse')

model.summary()
```

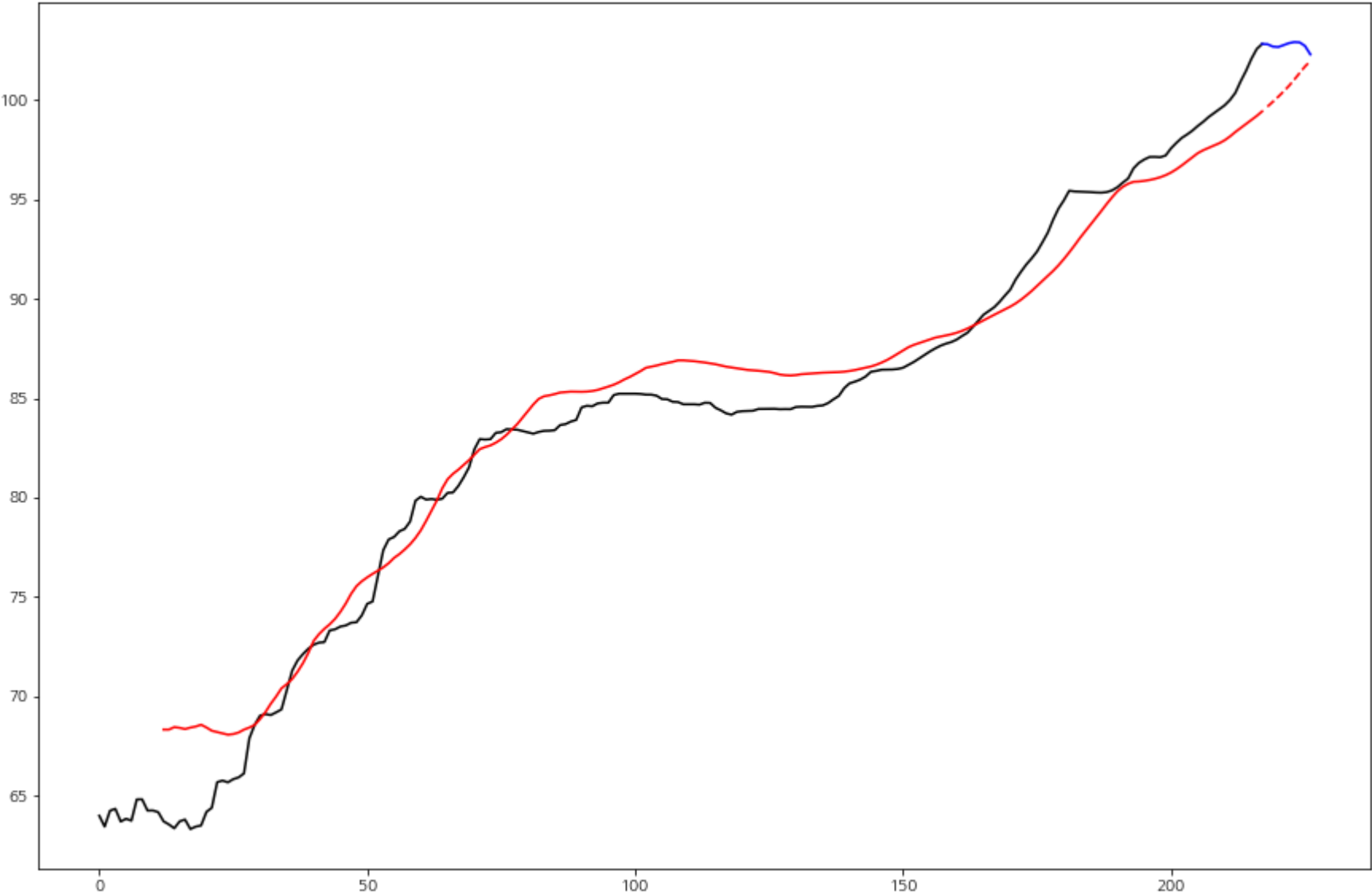
Model: "model_19"

Layer (type)	Output Shape	Param #
input_20 (InputLayer)	[(None, 12, 1)]	0
batch_normalization_9 (Batch Normalization)	(None, 12, 1)	4
simple_rnn_19 (SimpleRNN)	(None, 1)	3
dense_29 (Dense)	(None, 1)	2

=====
Total params: 9
Trainable params: 7
Non-trainable params: 2
=====

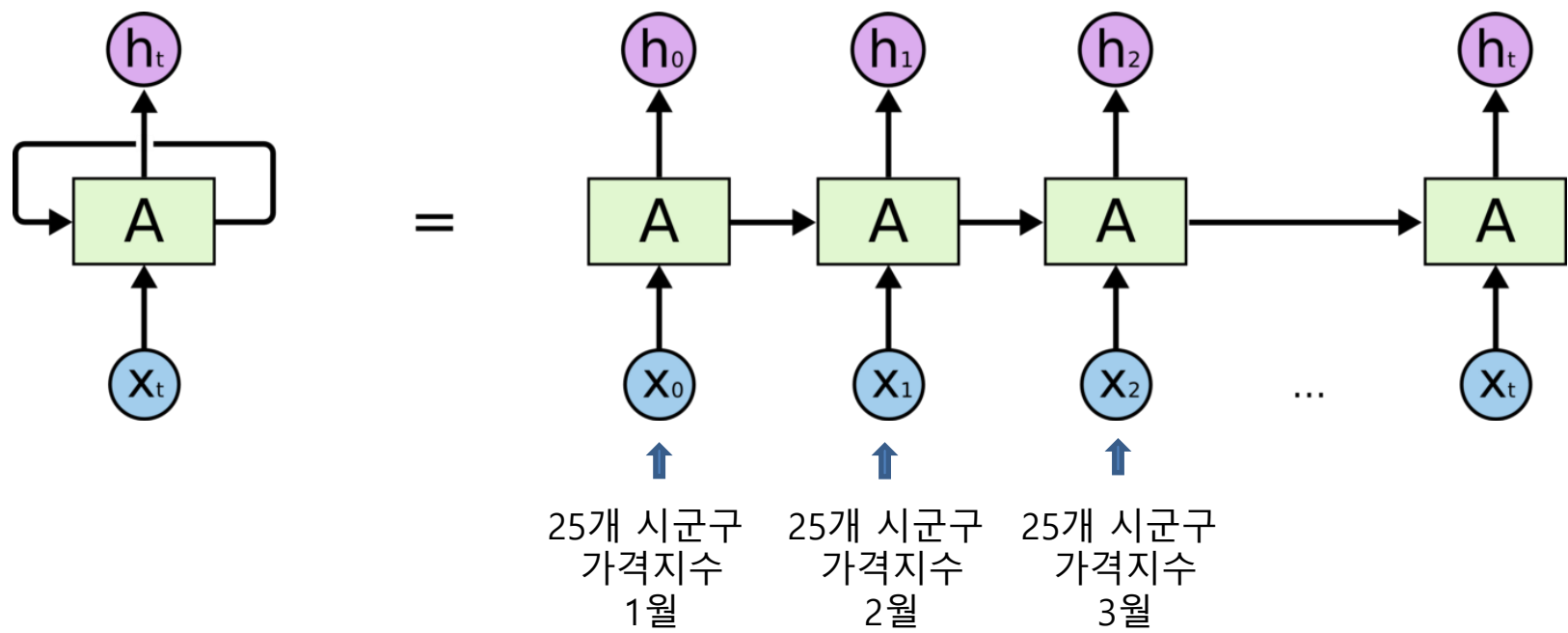


Simple RNN 결과

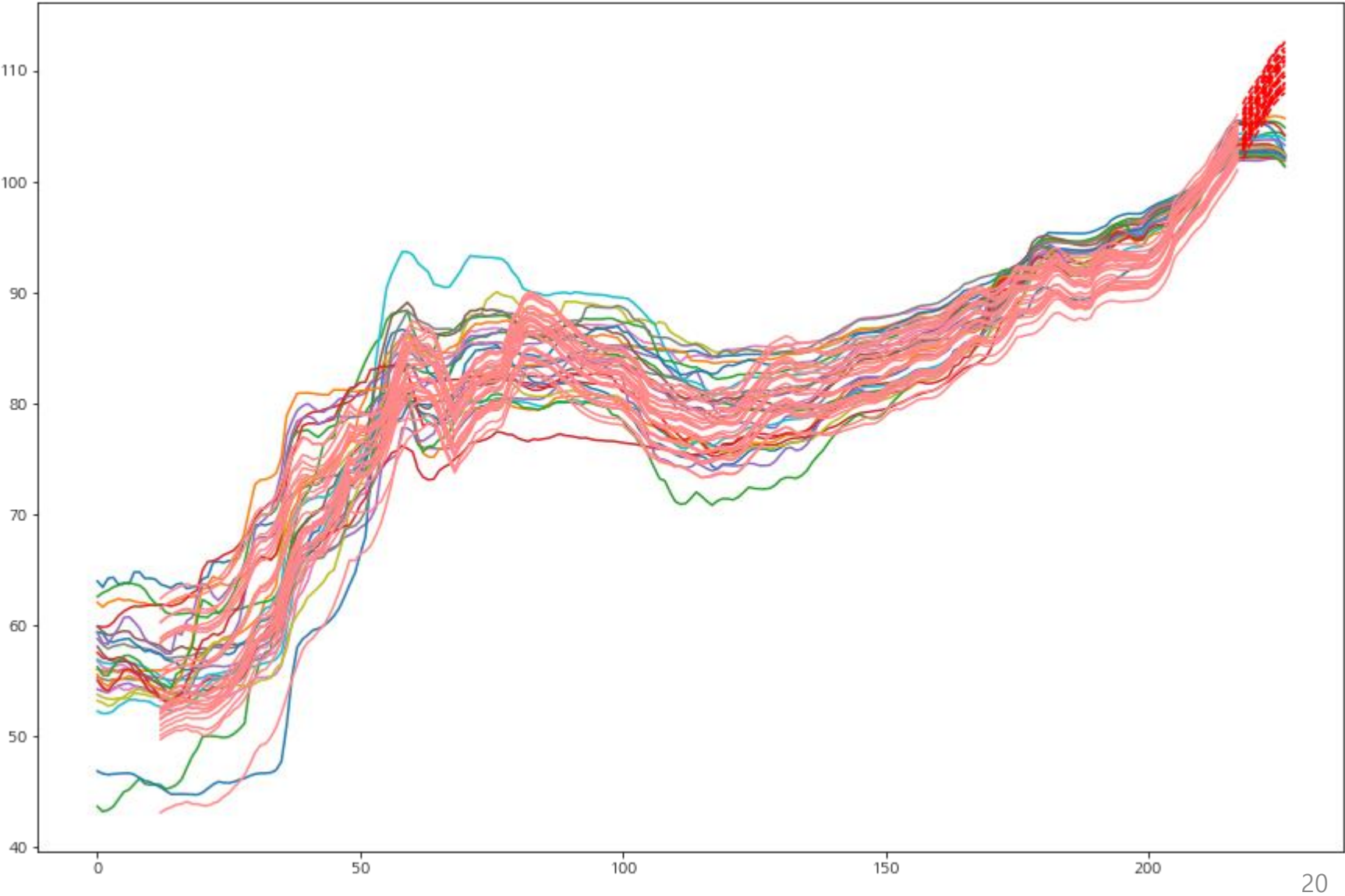


다양한 입력벡터를 사용

입력값은 기존 시계열 분석과 달리 여러 데이터를 전달할 수 있다.



25개 시군구 학습 및 예측



1. 시계열 분석

2. RNN

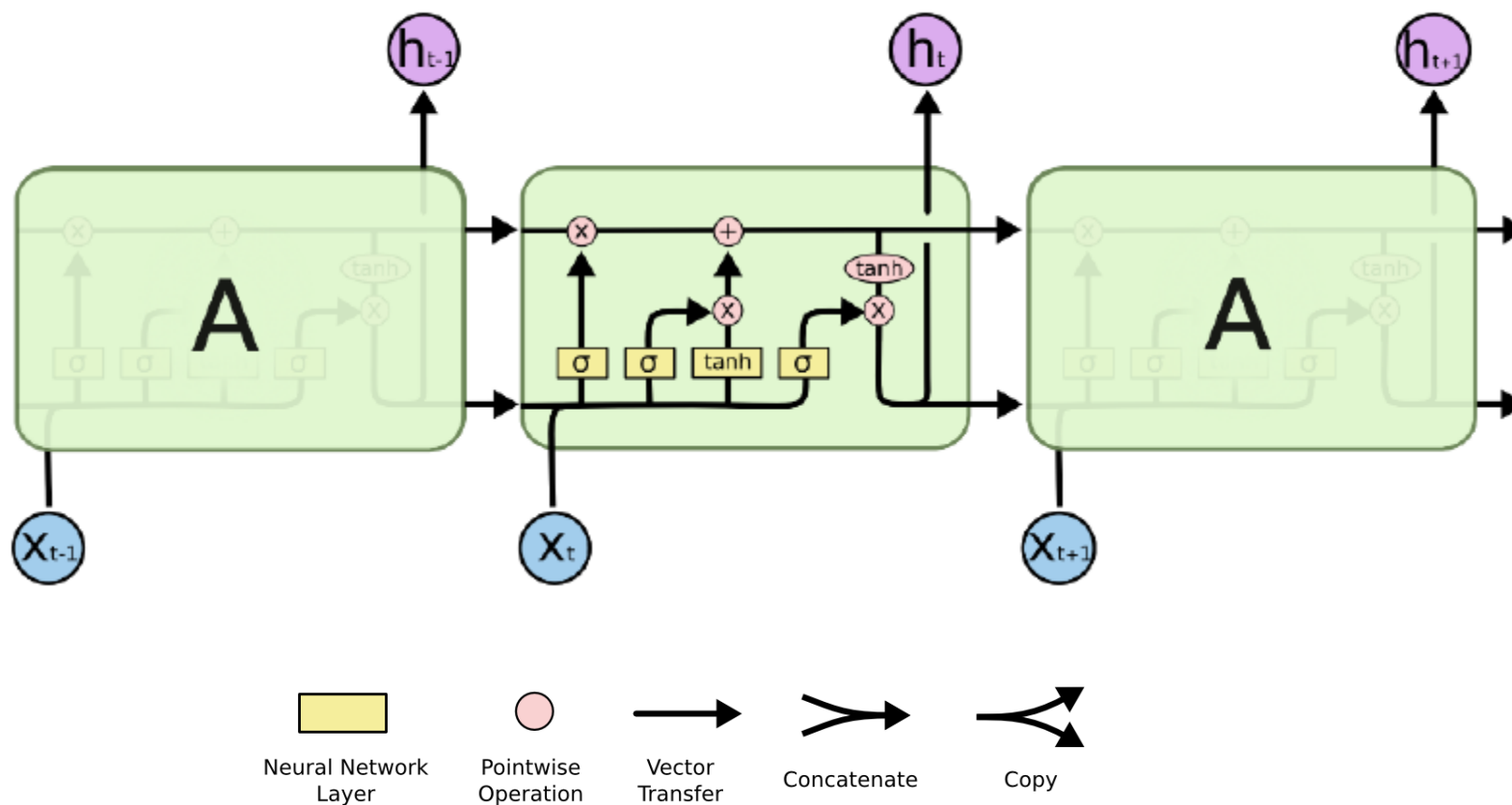
3. LSTM

4. 순환 신경망 응용

Long Short-Term Memory (LSTM)

RNN은 순차 학습 과정에서 초기 데이터에 대한 학습 전파가 어렵고 오래 걸리는 문제 존재
LSTM을 통해 이러한 문제를 해결

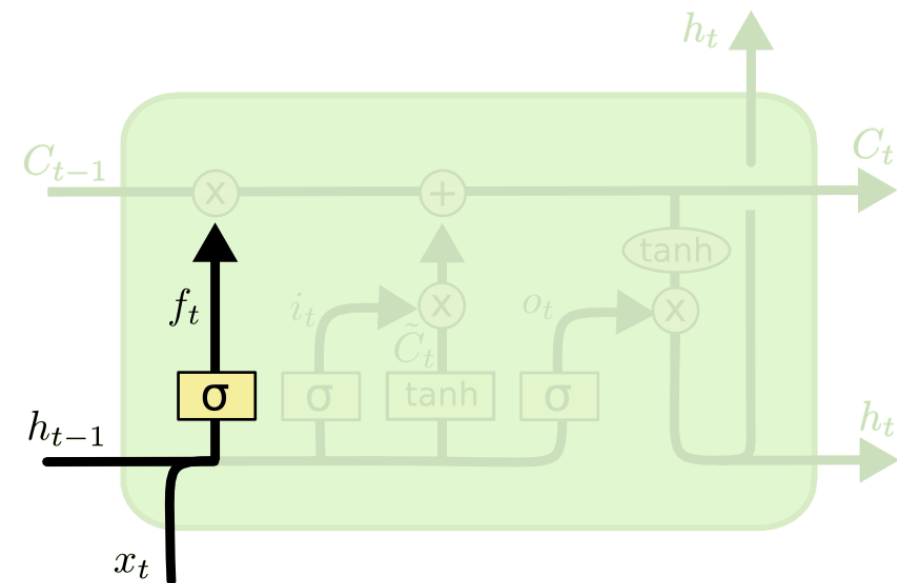
<http://www.bioinf.jku.at/publications/older/2604.pdf>



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

Forget gate

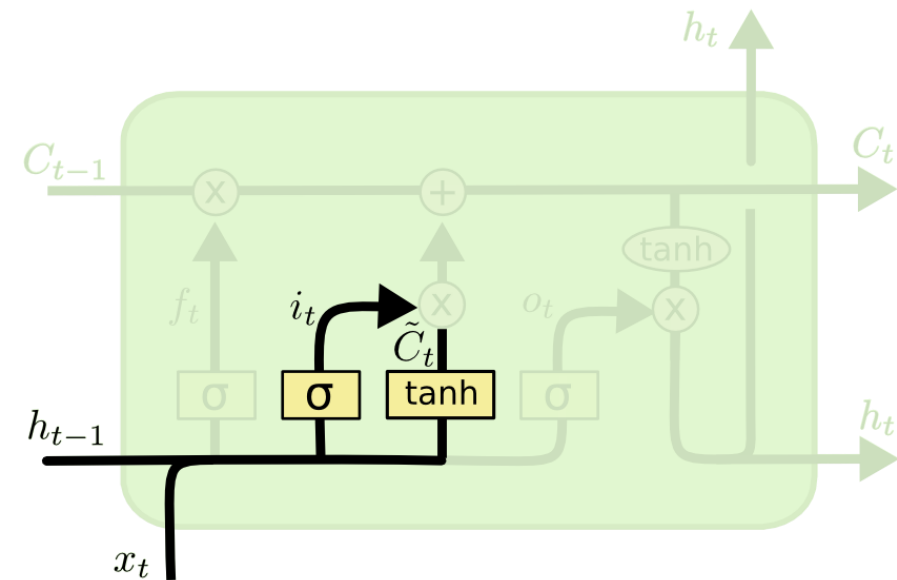
시그모이드 함수를 통해 0 혹은 1의 값을 통해 이전 cell state를 사용할지 안할지를 결정



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

input gate

하이퍼볼릭 탄젠트를 통해 이번 Cell State를 계산하고
시그모이드를 통해 이번 Cell state를 이전 Cell State와 합칠지를 결정

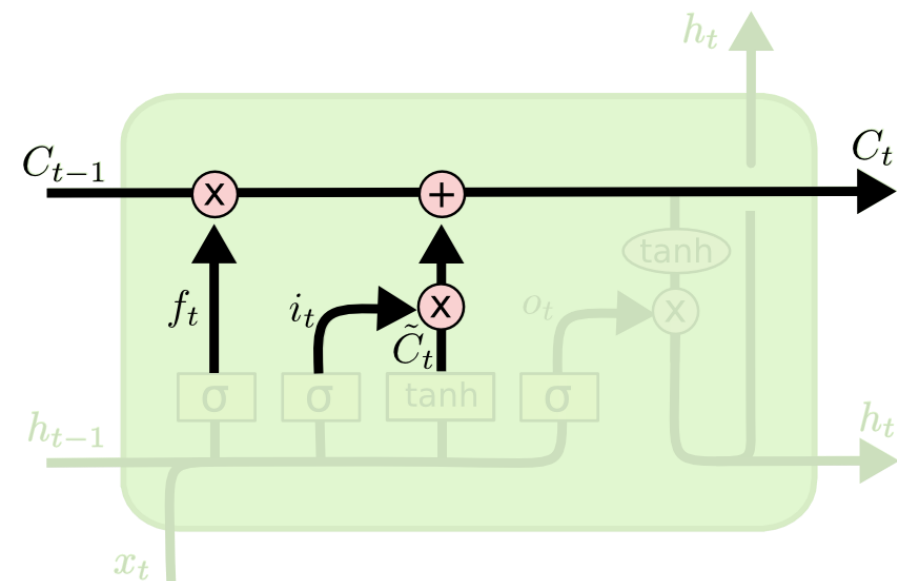


$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

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

Cell State

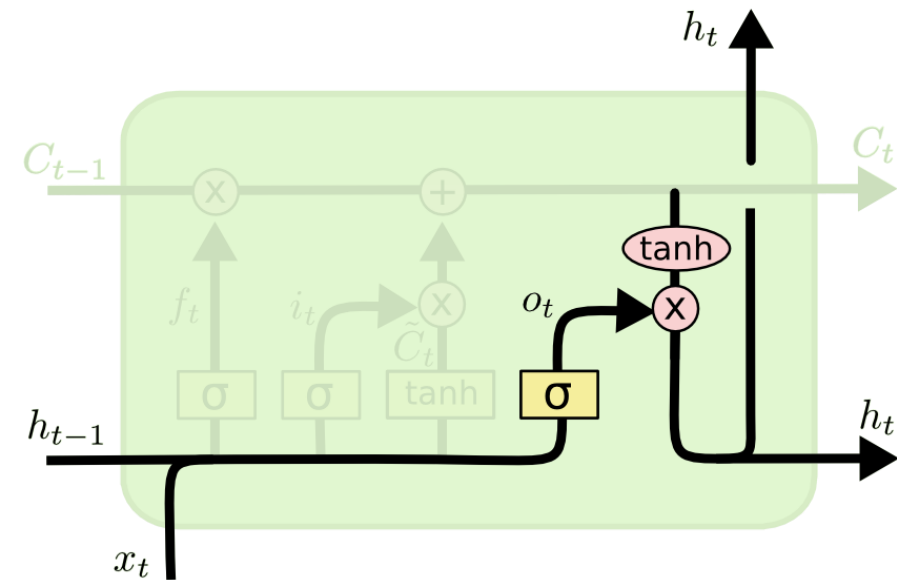
이전 Cell State와 forget gate를 곱하고
 이번 Cell State와 input gate를 곱해서
 두 값을 더하여 이번 Cell State를 확정한다.



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

Hidden State

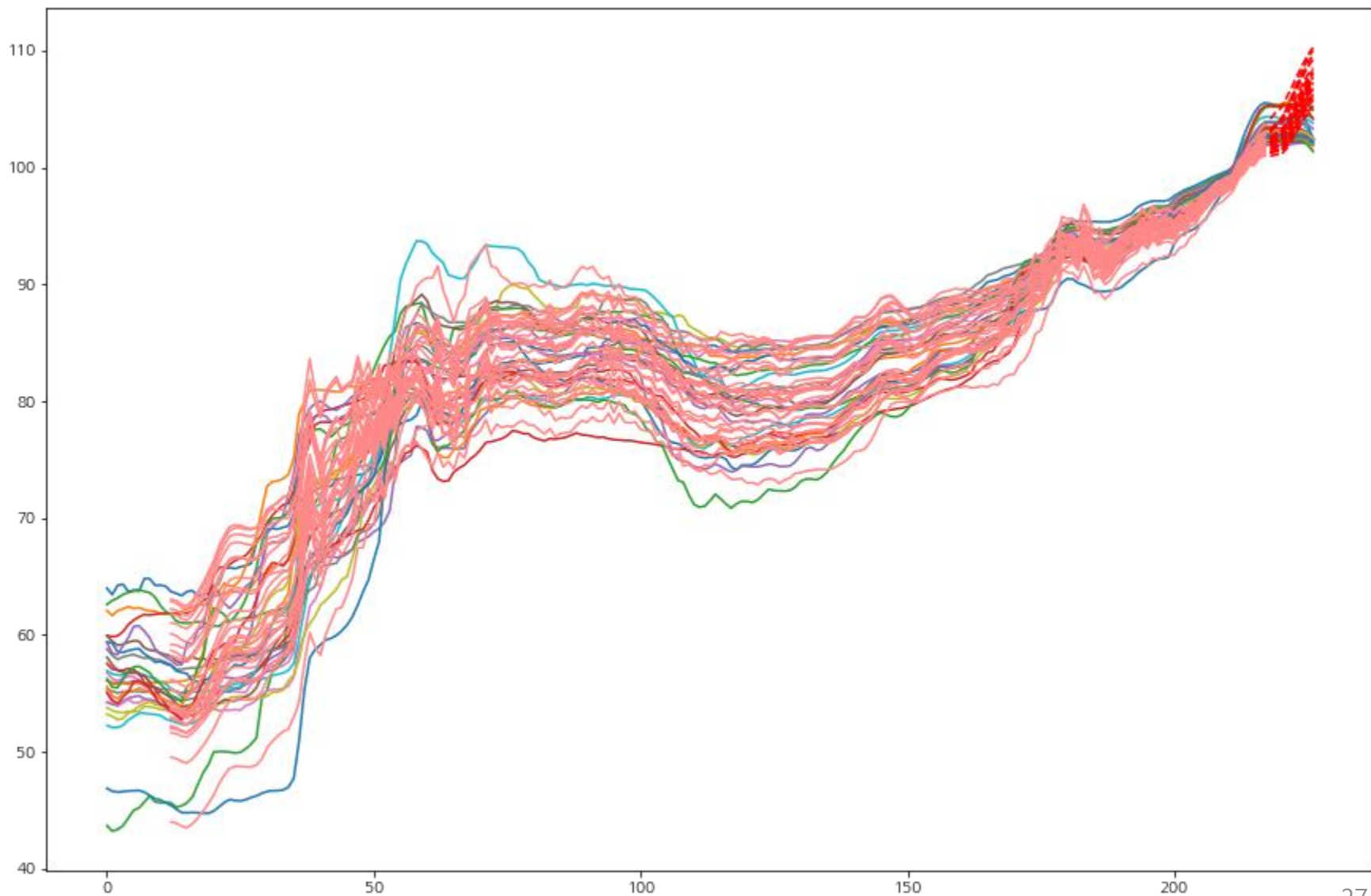
RNN과 동일하게 이전 hidden state와 입력 값을 이용해 히든 스테이트를 계산하고
 앞에서 만든 Cell State를 하이퍼볼릭탄젠트 함수로 계산한 후 곱한다
 새로운 hidden state가 완성



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

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

25개 시군구 학습 및 예측 LSTM



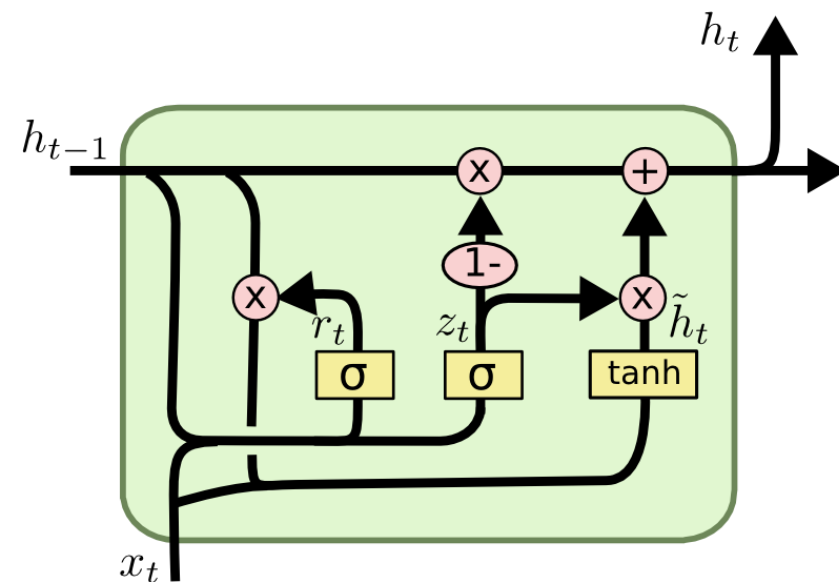
GRU (Gated Recurrent Unit, 2014)

Cell State가 없이 Hidden State 하나로 처리

Reset Gate(r)을 통해 이전 히든스테이트가 참여할지 여부 결정

Update Gate(zt)를 통해 이전 Hiddens State와 이번에 계산한 Hidden State를 몇대 몇의 비율로 합칠지 결정

<https://arxiv.org/pdf/1406.1078v3.pdf>



$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t])$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

1. 시계열 분석

2. RNN

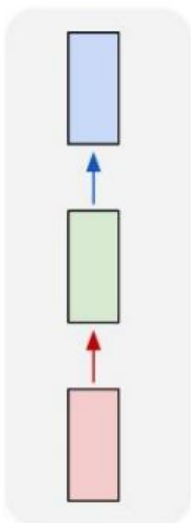
3. LSTM

4. 순환 신경망 응용

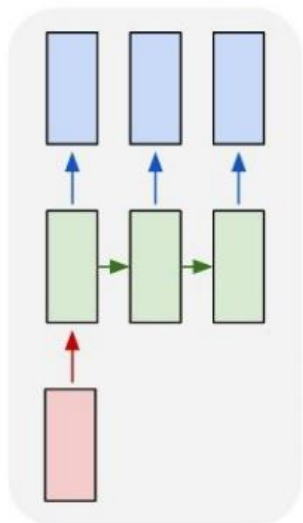
순환 신경망의 응용

순환 신경망을 다양한 형태로 활용이 가능

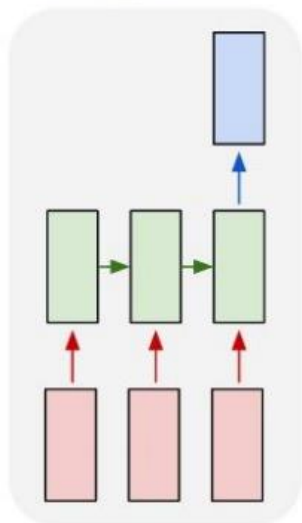
one to one



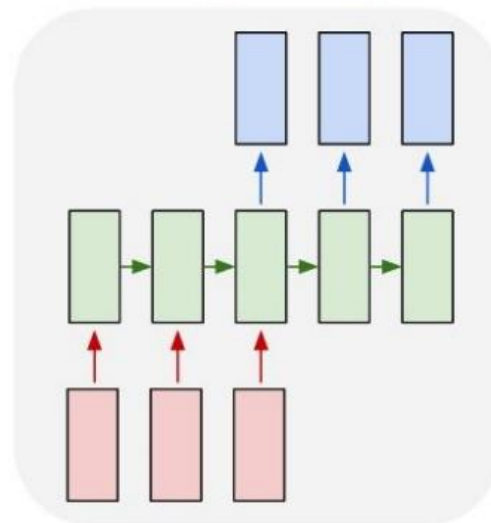
one to many



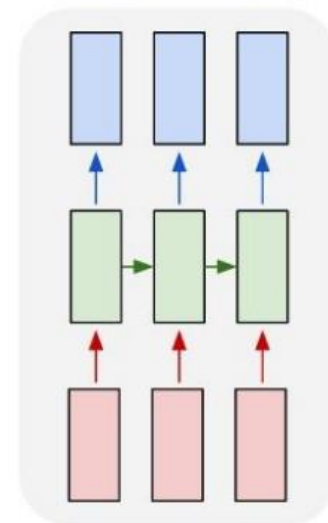
many to one



many to many

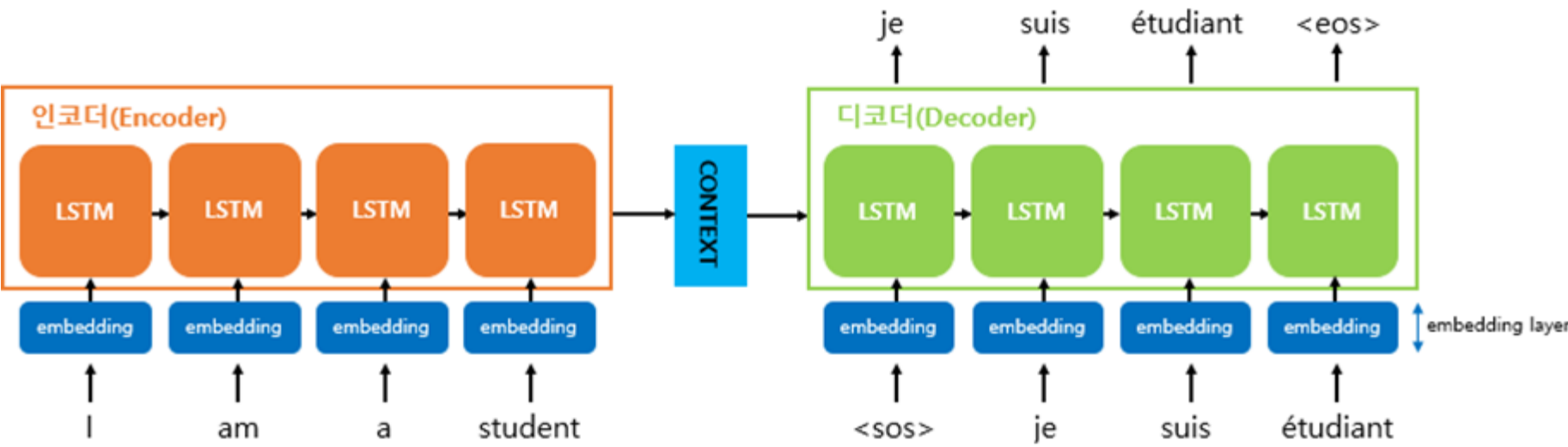


many to many



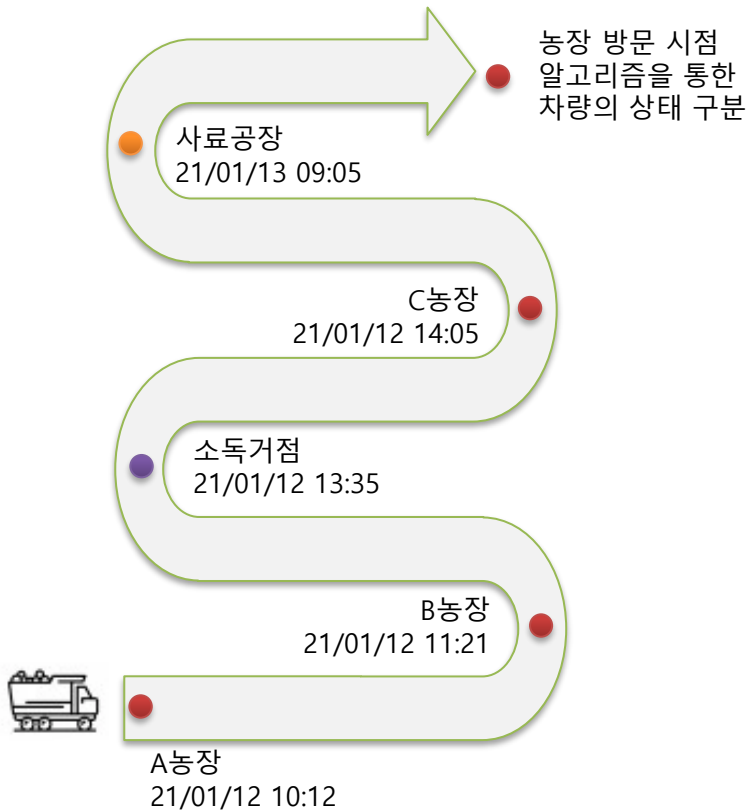
언어 모델에서 활용

인코더 디코더 모델을 이용하여 Context Vector를 추출

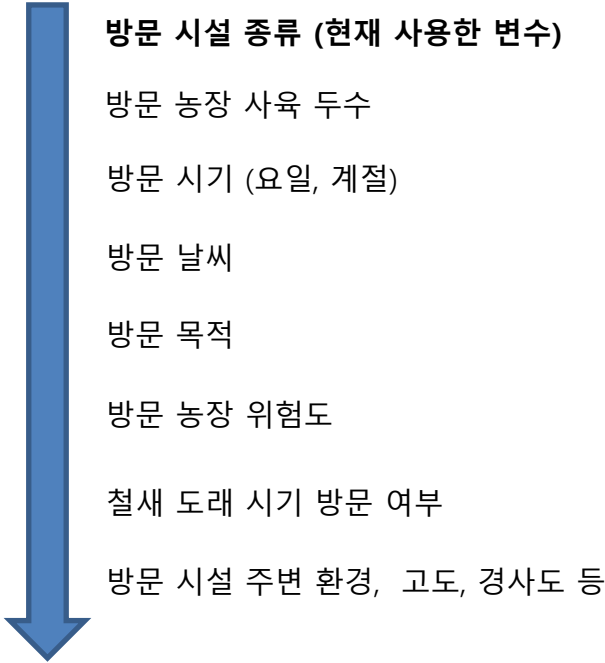


차량 분류 알고리즘

- 차량의 방문 경로 정보를 통해 차량을 분류하는 알고리즘 구축
- 방문 경로에 사용되는 데이터 규모에 따라 분류 정확도가 상승

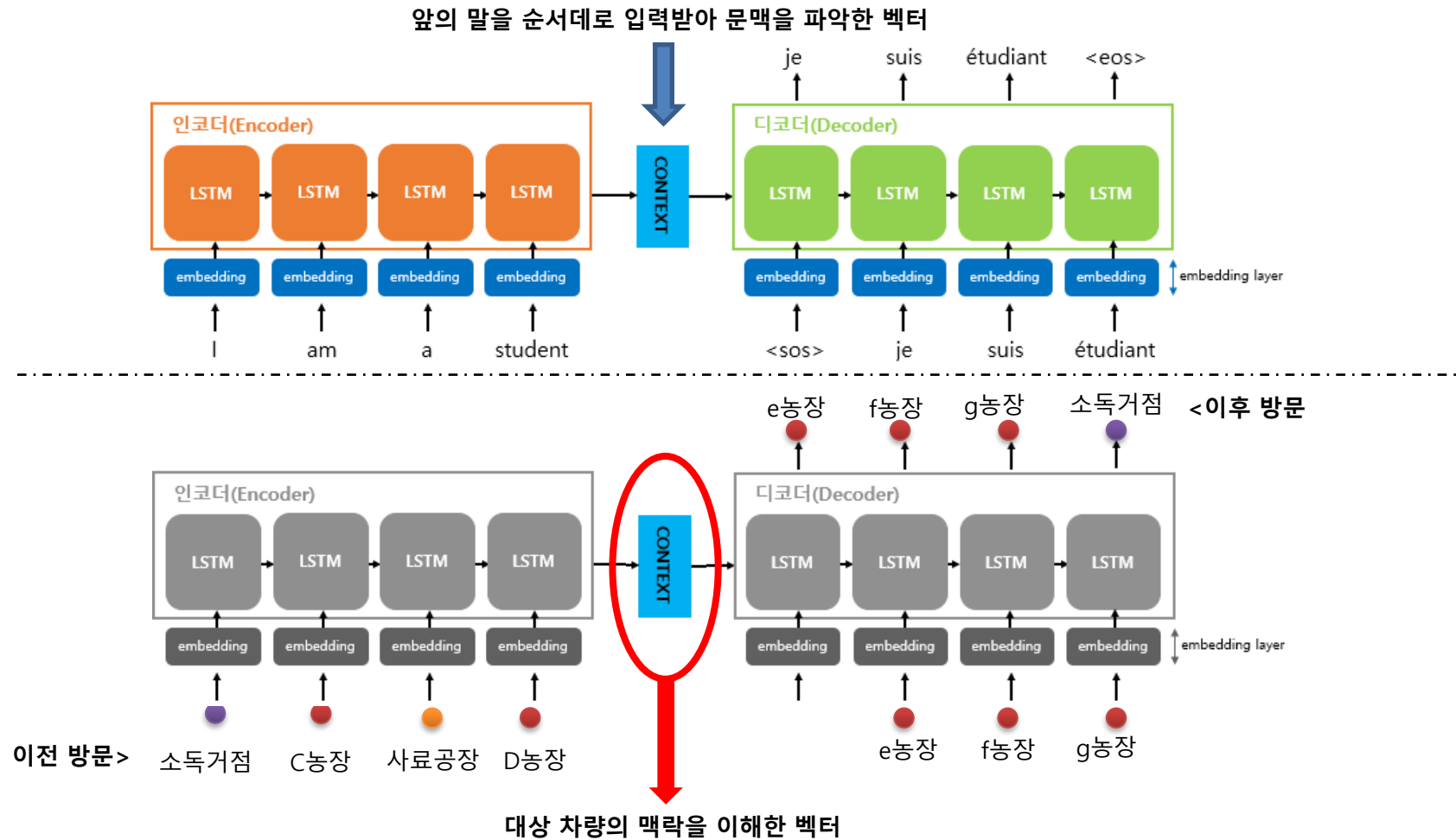


방문 데이터의 고도화에 따라 분류 정교화



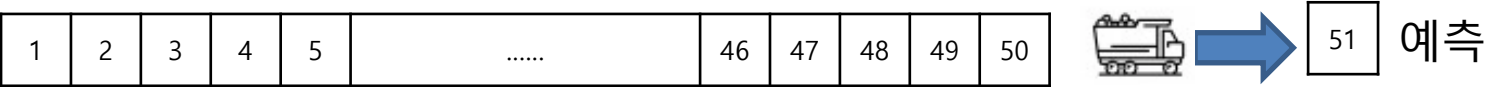
차량 분류 알고리즘 (1)

- 번역, 챗봇 모델 등에서 많이 활용되는 Seq2Seq 모델을 활용하여 방문 시계열을 활용한 분류 모델 학습
- 방문 이력을 통해 다음 방문할 시설을 예측하는 모델을 통해 맥락을 학습



(예시)방문시설 순서만 활용한 분류 결과

- 50회 방문 이력을 통해 차량의 맥락 벡터로 예측한 51번째 방문 시설의 정확도는 76.8%로 맥락 설명력 존재
- 농장, 도계장, 식용란수집판매업의 예측력이 높음
- 각 단계 데이터가 추가될 경우 맥락 설명력이 증가함

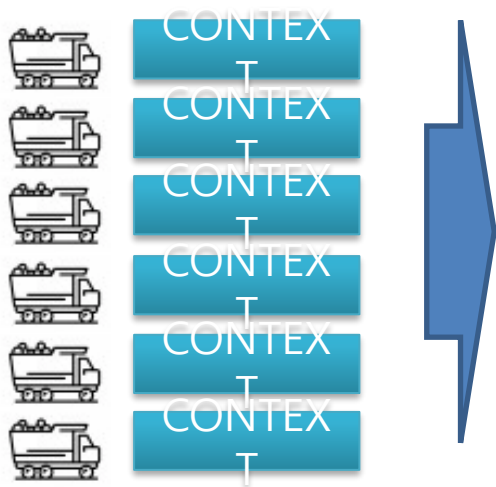


코드	시설종류	전체방문수	예측 성공	예측률	예측실패
101	농장	21,486	19,618	91.3%	1,868
201	종축장	215	45	20.9%	170
202	도축장	770	311	40.4%	459
203	도계장	273	189	69.2%	84
204	도압장	64	35	54.7%	29
205	집유장	323	151	46.7%	172
206	사료공장	1,863	1,047	56.2%	816
207	가축시장	588	54	9.2%	534
208	가축검정기관	89	34	38.2%	55
209	부화장	106	29	27.4%	77
211	비료제조업	305	121	39.7%	184
212	AI센터	14	4	28.6%	10
213	가축분뇨처리장	355	174	49.0%	181
214	철새도래지	9		0.0%	9
215	식용란수집판매업	962	680	70.7%	282
216	거점소독장소	1,972	388	19.7%	1,584
217	전통시장	31	6	19.4%	25
218	가든형식당	10	2	20.0%	8
998	가축인공수정소	488	127	26.0%	361
999	기타시설	77	17	22.1%	60
		30,000	23,032	76.8%	

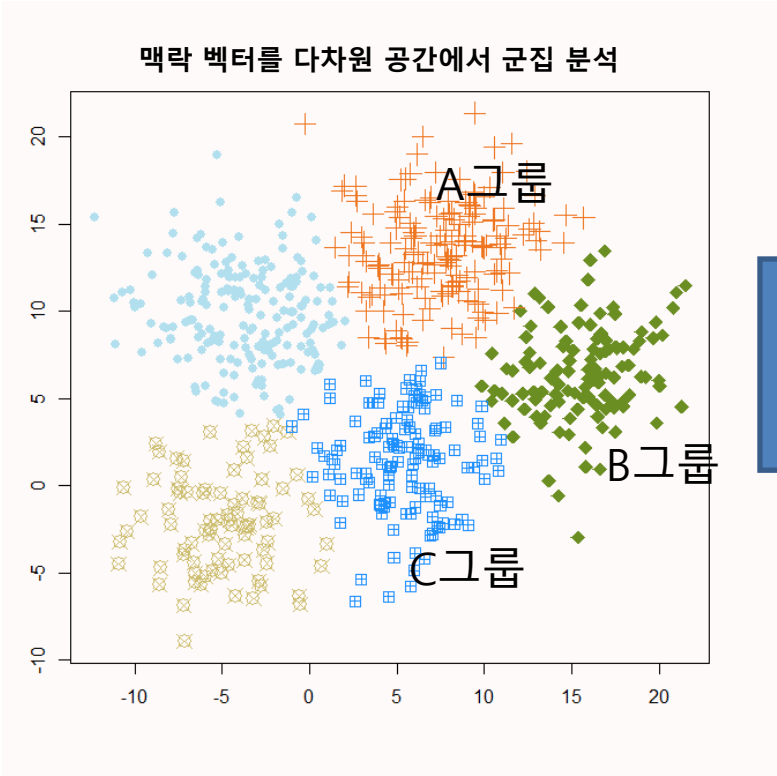
차량 분류 알고리즘 (2)

- 모든 차량의 방문시점에 따른 맥락 분석을 통해 차량을 구분
- 군집분석은 차량의 맥락 벡터가 유사한 그룹끼리 자동으로 묶어줌

모든 차량의 시점별 맥락 벡터 추출



맥락 벡터를 다차원 공간에서 군집 분석



실시간 차량 분류



(예시)방문시설 순서만 활용한 분류 결과







- 3만대의 차량의 과거 50개 방문 시설 정보를 통해 차량을 50개로 분류

1	2	3	4	5	46	47	48	49	50
---	---	---	---	---	-------	----	----	----	----	----



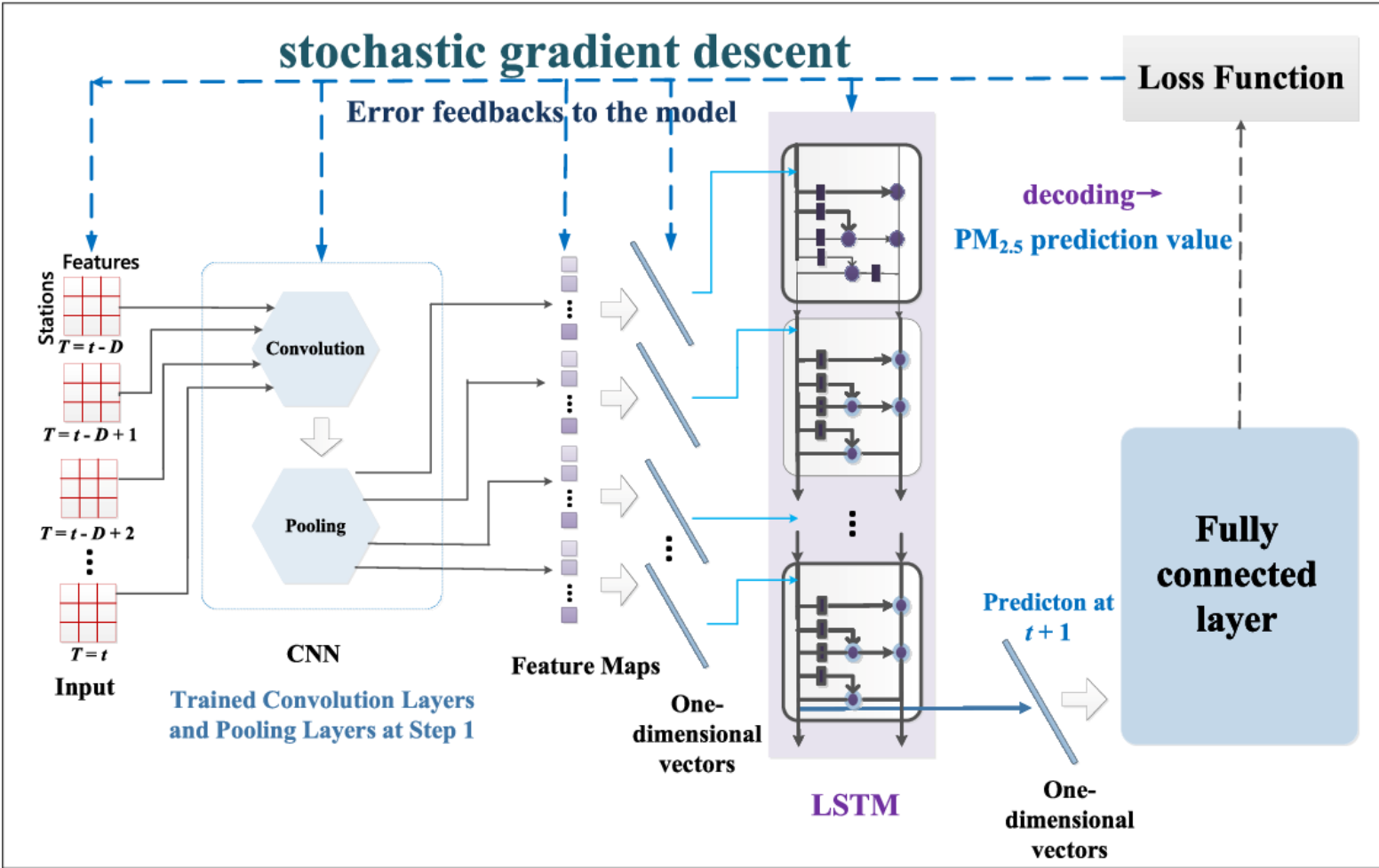
CONTEXT

군집분석

	분류	전체 차량	농장방문	방문 2위
	2번 분류	1,148	796 (69%)	사료공장 (319, 28%)
	0번 분류	797	634 (80%)	가축시장 (71, 9%) 거점소독장소 (60, 8%)
	29번 분류	769	555 (72%)	도축장 (140, 18%)
	10번 분류	655	448 (68%)	집유장 (187, 29%)
	28번 분류	622	443 (71%)	거점소독장소 (138, 22%)
	8번 분류	497	422 (85%)	전부 20건 이하

A Novel Combined Prediction Scheme Based on CNN and LSTM for Urban PM2.5 Concentration

<https://lib.hanyang.ac.kr/#/eds/detail?an=edsee.8632898&dbId=edsee>



DeepMind AlphaStar

<https://paperswithcode.com/method/alphastar>

