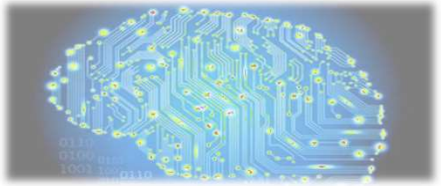
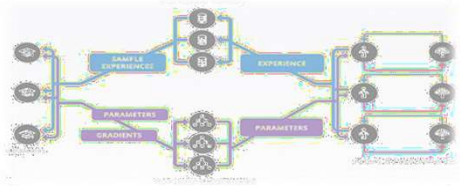
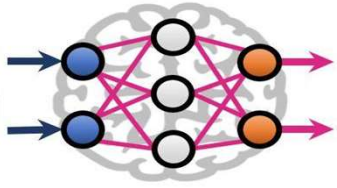
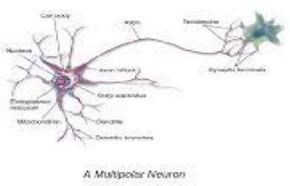
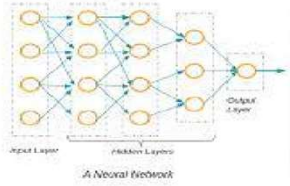







Comparison Long Short-Term Memory Network and Convolution Neural Nework, with additional data

LSTM 네트워크, CNN 네트워크 구현 및 데이터 추가

Geonil Yun,
rjsdlf45@gmail.com
2021. 06. 22.

Introduction

- **Missing Value Imputation**
  - ✓ **Last Observation Carried Forward (LOCF)**

dataset

index	x	y
0	10	7
1	1	7
2	6	missing
3	missing	10
4	3	7

fill 6

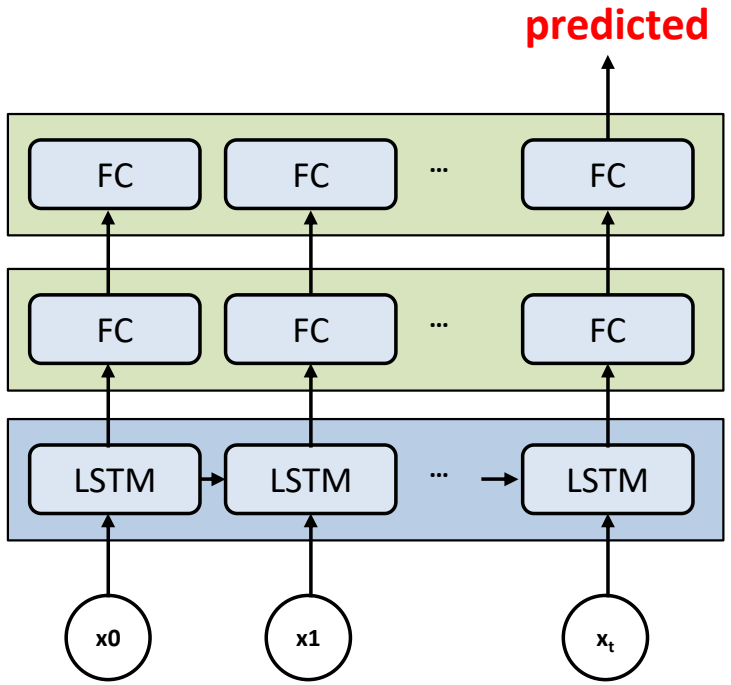
fill 7

- filling missing data with last observation value

Experimental setting

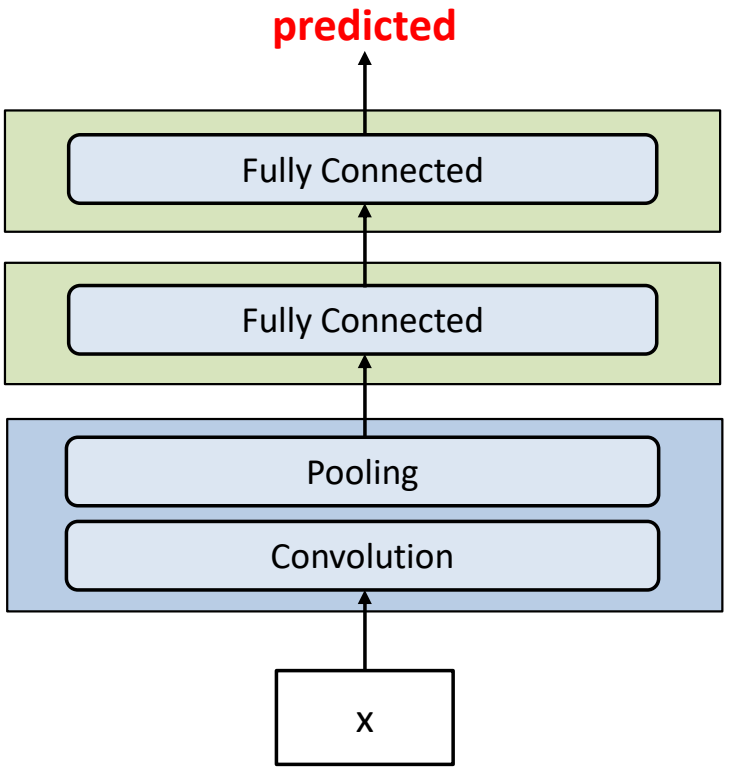
Structure (LSTM)

MSE for Loss = $\frac{1}{2}\sum(y-t)^2$



- input data shape : (batch, timestep, features)

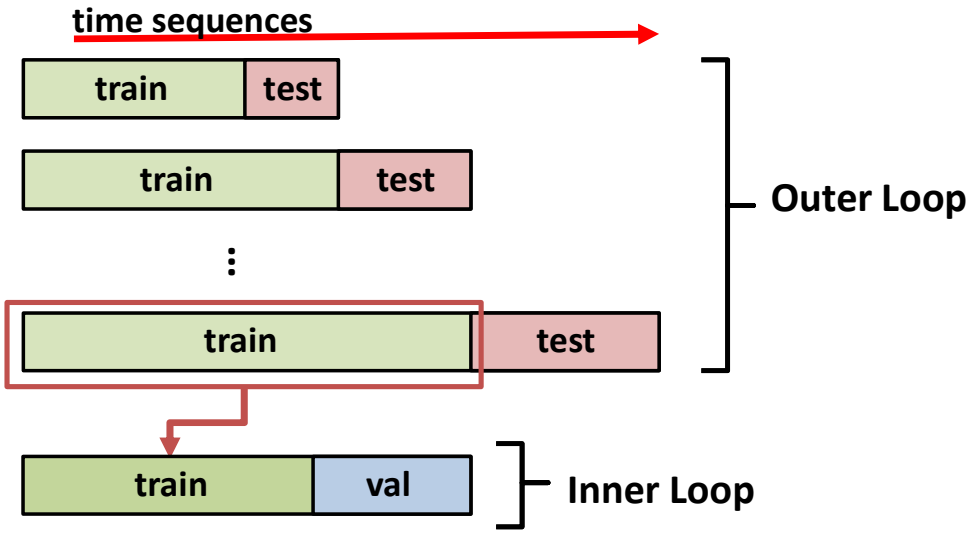
Structure (CNN)



- input data shape : (batch, 1, timestep, features)

Experimental setting

- **Nested Cross Validation for time series**



- ➢ **train past data & predict future data**
- ➢ **handling time series data in Nested CV**

- hyperparameter

Model	CNN	LSTM
fold num	3	
filter size	3	
filter num	30	
hidden size	64	
input length	24	
output length	1	
epoch	30	
batch size	64	
learning rate	**0.01, 0.001**	

Experimental setting

- **Statistics**

Dataset		Mean	Standard Deviation	Skewness	Kurtosis	Jarque Bera	ADF
Air Quality		1099.833	217.080	0.755	0.334	897.811	-9.823
Appliances Energy		97.694	102.524	3.386	13.664	191240.781	-21.616
Beijing Air Pollution		98.613	92.050	1.802	4.768	62162.743	-20.606
Hungarian Chickenpox		101.245	76.354	0.948	1.400	121.010	-6.933
Electric Consumption		1.483	1.194	1.141	1.732	3421.560	-5.360
Eye State		0.448	0.497	0.205	-1.957	2497.788	-4.803
Istanbul Stock Exchange	ISE	0.0016	0.016	-0.094	1.391	44.022	-22.746
	SP	0.0006	0.014	-0.090	3.068	211.013	-11.031
	DAX	0.0007	0.014	-0.106	2.104	99.894	-23.056
Metro Volume		3259.818	1986.860	-0.089	-1.309	3506.110	-28.016
Occupancy Detection		0.212	0.408	1.406	-0.020	2686.285	-3.430
SML 2010		20.493	3.312	-0.169	-0.105	21.792	-4.314

*ISE : Istanbul Stock Exchange, *SP : Standard & Poor's, *DAX : Germany

Experimental setting

- **Data split (8:2)**

Dataset	N	Train	Val	Test
Air Quality	8,582	5,140	1,721	1,721
Appliances Energy	19,687	11,803	3,942	3,942
Beijing Air Pollution	43,752	26,242	8,755	8,755
Hungarian Chickenpox	474	276	99	99
Electric Consumption	9,952	5,962	1,995	1,995
Eye State	14,932	8,950	2,991	2,991
Istanbul Stock Exchange	488	284	102	102
Metro Volume	48,156	28,884	9,636	9,636
Occupancy Detection	8,095	4,849	1,623	1,623
SML 2010	4,089	2,445	822	822

Comparison results

- RMSE (Root Mean Squared Error)

Data		LSTM	Standard Deviation	CNN	Standard Deviation
Air Quality		109.080	17.514	105.242	9.892
Appliances Energy		65.520	5.514	80.046	3.822
Beijing Air Pollution		22.545	1.628	28.076	2.941
Hungary Chickenpox		76.973	31.905	72.272	35.092
Electric Consumption		0.525	0.344	4.488	5.797
Eye State		1.257	0.633	5.431	3.540
Istanbul Stock Exchange	ISE	6.402	7.395	19.167	21.672
	SP	6.711	7.778	19.091	21.328
	DAX	6.149	6.977	19.615	21.960
Metro Interstate Traffic Volume		500.132	127.115	1279.006	855.703
Occupancy Detection		5.006	5.854	2.924	1.942
SML 2010		0.258	0.102	7.297	9.684

Comparison results

- MAE (Mean Absolute Error)

Data		LSTM	Standard Deviation	CNN	Standard Deviation
Air Quality		81.485	14.005	79.173	8.538
Appliances Energy		35.454	10.560	45.601	4.144
Beijing Air Pollution		12.167	0.899	16.333	0.653
Hungary Chickenpox		59.167	28.547	55.613	30.438
Electric Consumption		0.186	0.100	2.321	3.036
Eye State		1.172	0.669	2.891	2.807
Istanbul Stock Exchange	ISE	6.400	7.396	19.159	21.669
	SP	6.709	7.779	19.084	21.323
	DAX	6.147	6.978	19.607	21.956
Metro Interstate Traffic Volume		334.276	83.755	598.471	137.168
Occupancy Detection		4.920	5.897	2.427	1.436
SML 2010		0.183	0.063	7.191	9.675

Thank you

Appendix : Missing Value

Deletion

- Likewise deletion

dataset

index	x	y
0	10	7
1	3	5
2	6	missing
3	missing	10
4	3	2

remove

➢ **Remove index 2, 3 from the dataset**

- Dropping variable

dataset

index	x	y	z
0	10	7	missing
1	3	5	missing
2	6	missing	missing
3	missing	10	missing
4	3	2	7

remove

➢ **Remove 'z' column from the dataset**

Appendix : Missing Value Imputation

Mean-Mode imputation

- fill missing data with mean or mode from data set (or median)

dataset

index	x	y
0	10	7
1	1	7
2	6	missing
3	missing	10
4	3	7

fill 5 (mean)

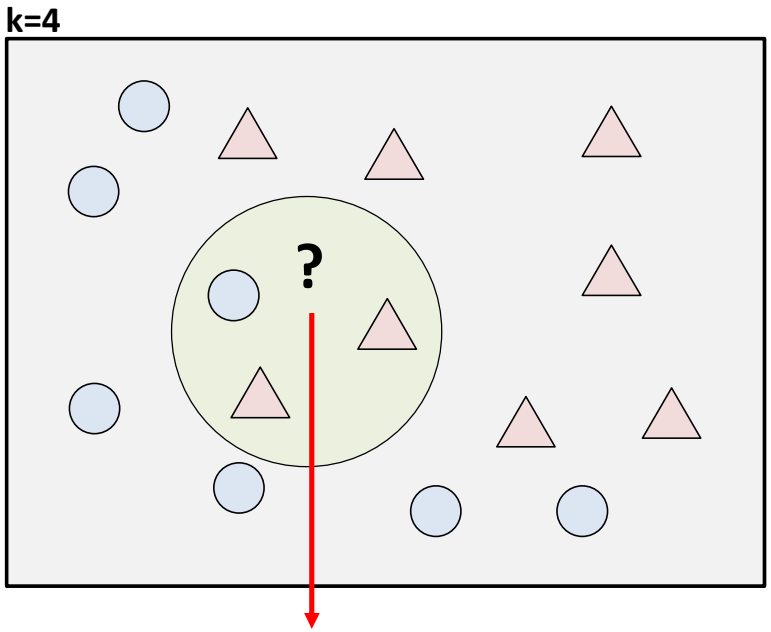
fill 7 (mode)

- using representative value : fast
- not consider the correlation with other features
- low accuracy
- mean reduce variance
- mode create bias

Appendix : Missing Value Imputation

KNN imputation

- fill missing value by using KNN(K-Nearest Neighbor)



- More accurate than M-M impute
- Sensitive to feature scale
- Too many calculations

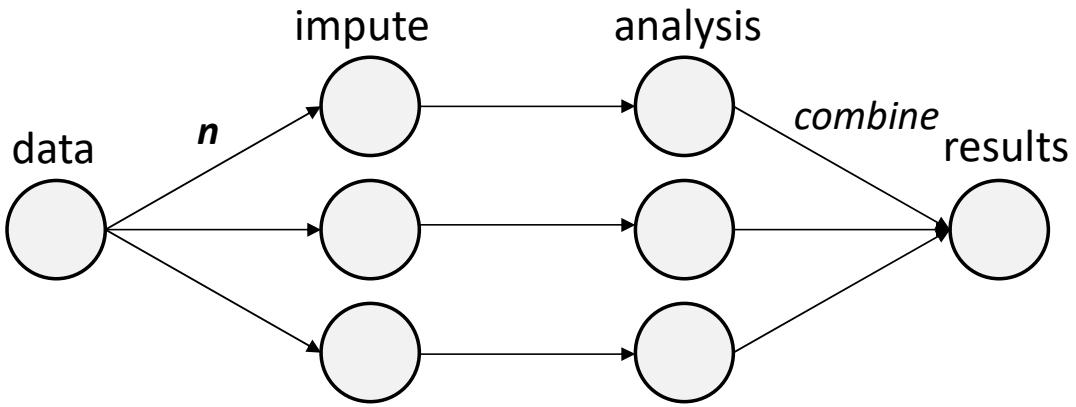
use mean or median of near data (mode : categorical)

Appendix : Missing Value Imputation

Multiple Imputation

- fill missing value by using Multiple Imputation

① **Imputation** : impute *n* datasets.
② **Analysis** : analysis result of *n* datasets.
③ **Combine** : combine result of *n* datasets.



Appendix : Code

Model	view with code
LSTM	https://colab.research.google.com/drive/1oNW4HzBzmrjyWcWZbjFMtksP1vTV9IO2?usp=sharing
CNN	https://colab.research.google.com/drive/1Ccd7l65bJwiD2HW8XmOKTHO1PEUMuQI3?usp=sharing
LSTM-Beijing air pollution	https://colab.research.google.com/drive/1De0Tn_iD5NQBKAu_eKr3PWZwvcnNTu97?usp=sharing
CNN-Beijing air pollution	https://colab.research.google.com/drive/1HmjdyXbwUASZxzS40v4opc9myAuuhcrF?usp=sharing
LSTM-Hungarian chickenpox	https://colab.research.google.com/drive/1eeLLtCJGzQ2u_V09pXF0dHg2Oz4YHfxn?usp=sharing
CNN-Hungarian chickenpox	https://colab.research.google.com/drive/1E4PywHa-xACYhHTiMiHRwEkCHT9eBz4A?usp=sharing
LSTM-Power consumption	https://colab.research.google.com/drive/1HVLLJMPGJeixJ4H9O3IvzfUhUWp8DAXN?usp=sharing
CNN-Power consumption	https://colab.research.google.com/drive/1d7wu9u7zo_URAXYd2E01uROGcsdoSc87?usp=sharing

Appendix : Dataset

Dataset	references
Air Quality	https://archive.ics.uci.edu/ml/machine-learning-databases/00360/
Appliances Energy	https://archive.ics.uci.edu/ml/machine-learning-databases/00374/
Beijing Air Pollution	https://archive.ics.uci.edu/ml/machine-learning-databases/00381/
Hungarian Chickenpox	https://archive.ics.uci.edu/ml/machine-learning-databases/00580/
Electric Consumption	https://archive.ics.uci.edu/ml/machine-learning-databases/00235/
Eye State	https://archive.ics.uci.edu/ml/machine-learning-databases/00264/
Istanbul Stock Exchange	https://archive.ics.uci.edu/ml/machine-learning-databases/00247/
Metro Volume	https://archive.ics.uci.edu/ml/machine-learning-databases/00492/
Occupancy Detection	https://archive.ics.uci.edu/ml/machine-learning-databases/00357/
SML 2010	https://archive.ics.uci.edu/ml/machine-learning-databases/00274/