Missing-insensitive Short-term Load Forecasting Leveraging Autoencoder and LSTM

Kyungnam Park, Jaeik Jeong, and Hongseok Kim

2020. 4. 26

Introduction

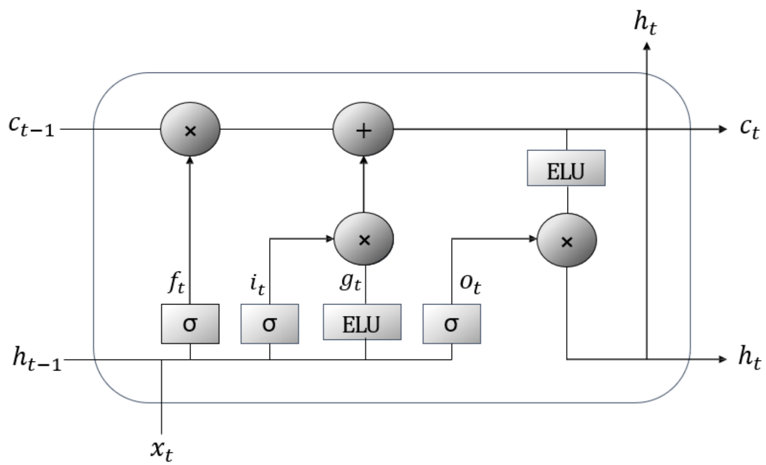
- Short-term load forecasting (STLF) in climate change
  - prevent excessive power generation reserve (power system operation)
  - lower the use of fossil fuels (greenhouse gas emission reduction)
  - mitigate climate change
  - change the electric usage of users with a high price of electricity in peak hours (demand response)

- So far, most deep learning-based STLF techniques require intact data
  - but, many real-world datasets contain missing values
  - learning models are usually created by separating missing imputation and STLF

- In this paper, we jointly consider missing imputation and STLF
  - a family of autoencoder/LSTM combined models
  - autoencoder, convolutional autoencoder, and denoising autoencoder are investigated for extracting features, which is directly fed into the input of LSTM
  - to realize missing-insensitive STLF through various experiments

Deep learning methodologies

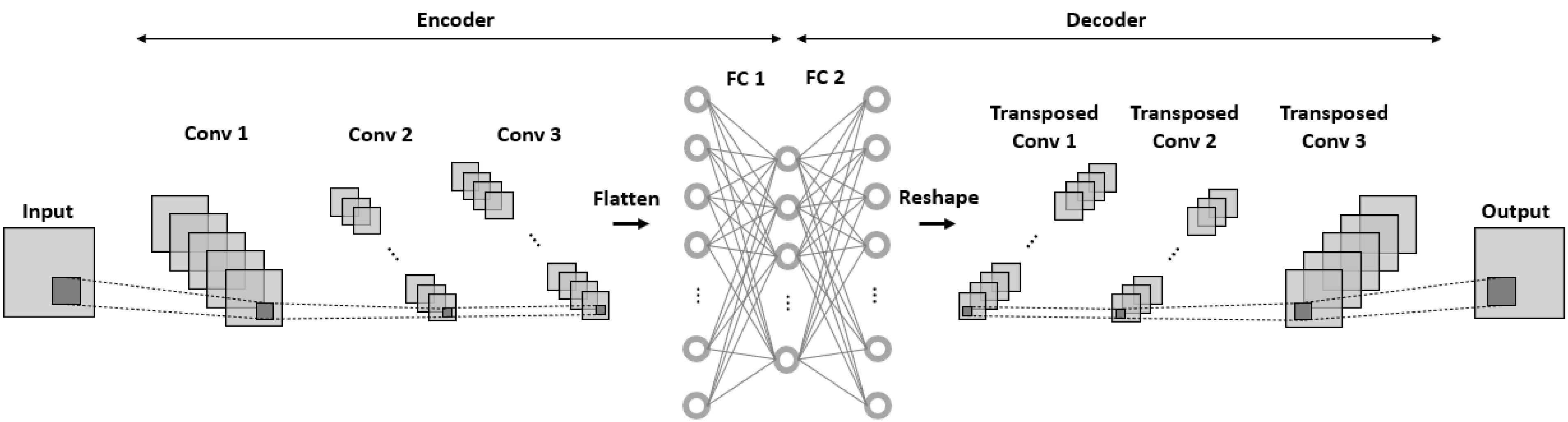
- Long Short-Term Memory (LSTM)
  - load time series data, recurrent neural network (RNN) is often used for forecasting model
  - LSTM solves the problem of gradient vanishing of recurrent neural network (RNN)



- Autoencoder (AE)
  - unsupervised learning version of neural network (only the input value of data is learned)
  - not simply copy the input directly to the output but control to learn how to efficiently represent the data

Deep learning methodologies

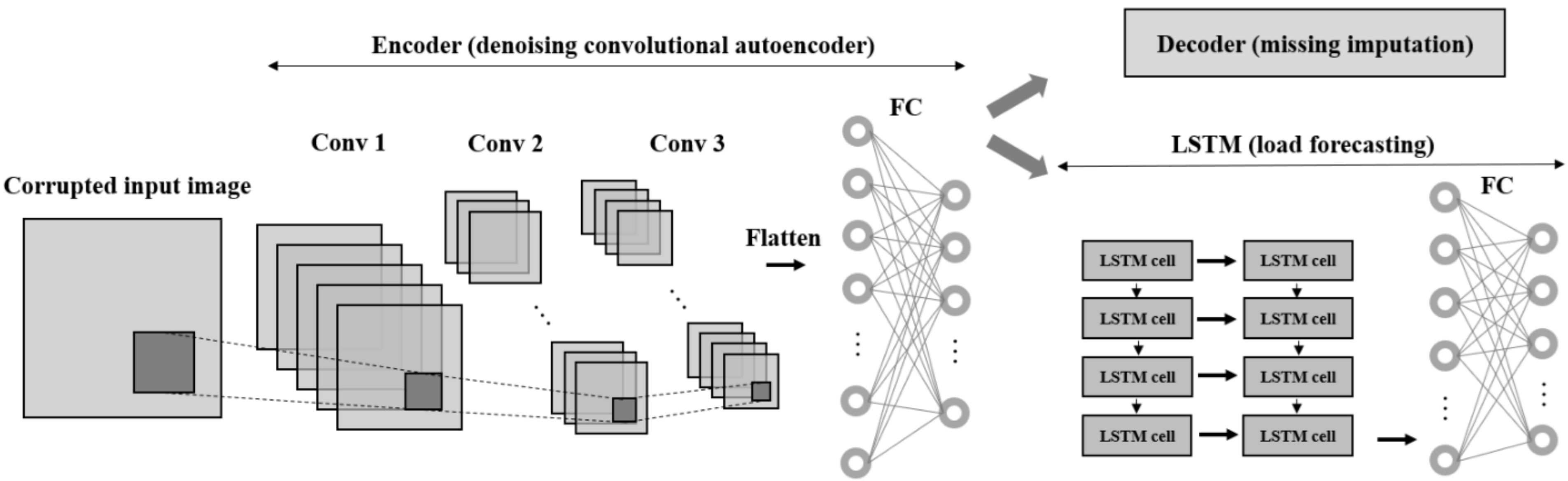
- Convolutional autoencoder (CAE)
  - convolutional neural network (CNN) is used for encoder and decoder
  - CNN is a neural network that uses convolution operations



- Denoising autoencoder (DAE)
  - an autoencoder with denoising capabilities
  - can be used for AE or CAE
  - reconstruct a partially corrupted input instead of the original input
  - achieve robustness to partial destruction of input by learning common latent representations of the original and corrupted data

Proposed models

- Denoising convolutional autoencoder (DCAE)/LSTM
  - the best performing model of the proposed a family of autoencoders



  - the encoder consists of three layers of convolution and pooling
  - the filters in the convolution layers use gradually increasing structures to 5, 25, 125
  - 7 days load image data are converted from the encoder output to 100 one-dimensional data
  - LSTM consists of four cells to forecast the next one day

Simulation Parameters

- Data sets are obtained from Korea Electric Power Corporation (KEPCO)
  - ✓ Data sets consist of seven sectors (mining support service, education service, water supply business, paper products manufacturing, information service, insurance and pensions, and wooden products manufacturing)
  - ✓ each with 600 days of power usage data
  - ✓ peak loads of the customers span from 33kW to 12,342kW
- We leverage data preprocessing such as removing abnormal data points and securing available data as much as possible.
  - – Eventually we use ten sets of customer data for each hyperparameter setting

Industry	User ID	Peak load (kW)	Average load (kW)
Mining support service	1	564.12	246.17
Education service	2	1,792.80	731.88
Water supply business	3	12,342.40	8,558.95
Water supply business	4	2,933.70	1,843.63
Water supply business	5	1,702.56	1,187.80
Paper products manufacturing	6	164.64	70.20
Information service	7	272.64	165.43
Information service	8	151.68	35.71
Insurance and pensions	9	192.24	35.98
Wooden products manufacturing	10	33.00	3.48

Experimental Set-up

- Totally, we set six types of learning model
- To compare results of performance (MAPE)
  - layer configuration, noise, model complexity

ex) DAE, CAE

MAPE : Mean Absolute Percentage Error

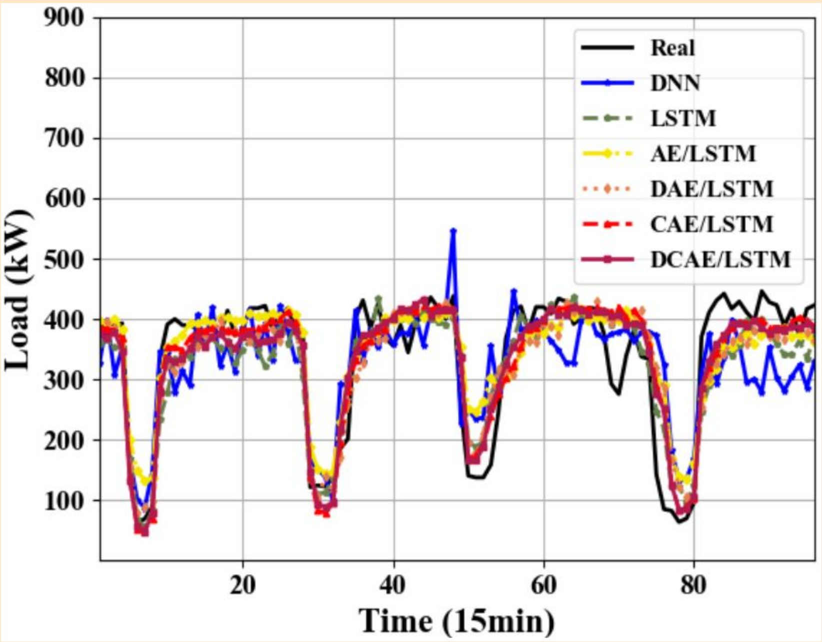
$$MAPE(\%) = \frac{100}{N}\sum_{t=1}^{N}\frac{|y_t - \hat{y}_t|}{y_t}$$

Layer configuration	Noise	Model complexity	Average MAPE (%)
64	0	43,072	37.64
64	0.3	43,072	38.83
64	0.5	43,072	47.13
100	0	67,300	36.76
100	0.3	67,300	38.52
100	0.5	67,300	40.98
200, 100	0	154,700	36.45
200, 100	0.3	154,700	34.47
200, 100	0.5	154,700	34.72
300, 100	0	232,000	35.07
300, 100	0.3	232,000	33.44
300, 100	0.5	232,000	34.60
300, 200, 100	0	282,200	34.77
300, 200, 100	0.3	282,200	33.77
300, 200, 100	0.5	282,200	33.63
500, 300, 100	0	516,900	34.15
500, 300, 100	**0.3**	**516,900**	**32.31**
500, 300, 100	0.5	516,900	34.04

Layout	Layer configuration		Model complexity	Average MAPE (%)
	Convolutional	Fully-connected		
1/1	4@3	100	76,940	37.66
1/1	5@3	100	96,150	34.53
2/1	4@3 16@3	100	77,532	46.76
2/1	5@3 25@3	100	121,300	32.09
2/2	4@3 16@3	300, 100	261,432	38.27
2/2	5@3 25@3	300, 100	391,600	38.06
3/1	4@3 16@3 64@3	100	86,812	32.42
3/1	5@3 25@3 125@3	100	179,550	32.63
3/2	4@3 16@3 64@3	500, 100	444,512	33.57
3/2	**5@3 25@3 125@3**	**500, 100**	**830,050**	**31.48**

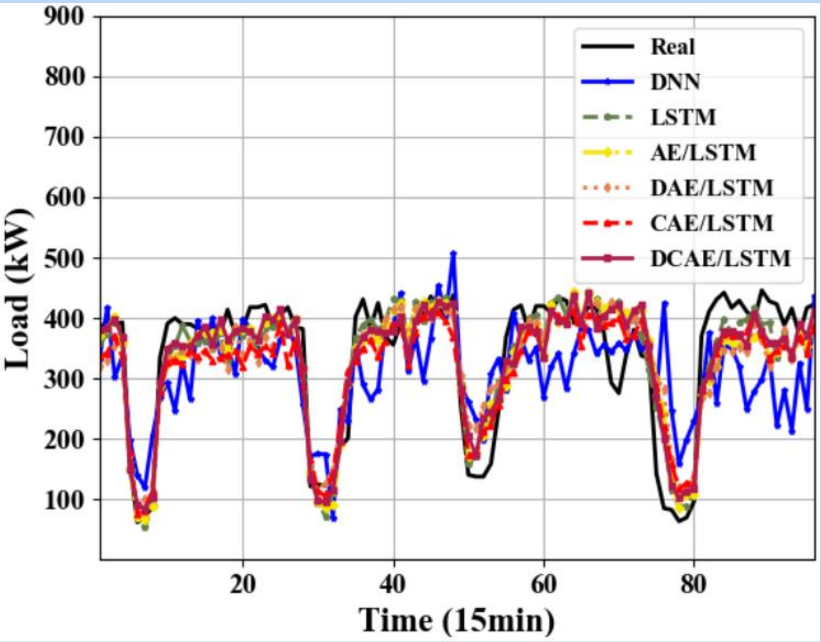
Forecasting results with intact data and missing data

Intact data



Model	MAPE (%)		
	Average	Q_1	Q_4
DNN	25.31	6.40	62.26
LSTM	22.41	6.03	56.22
AE/LSTM	20.20	6.25	46.27
DAE/LSTM	20.49	6.43	45.04
CAE/LSTM	19.17	5.90	42.92
DCAE/LSTM	18.90	6.02	42.41

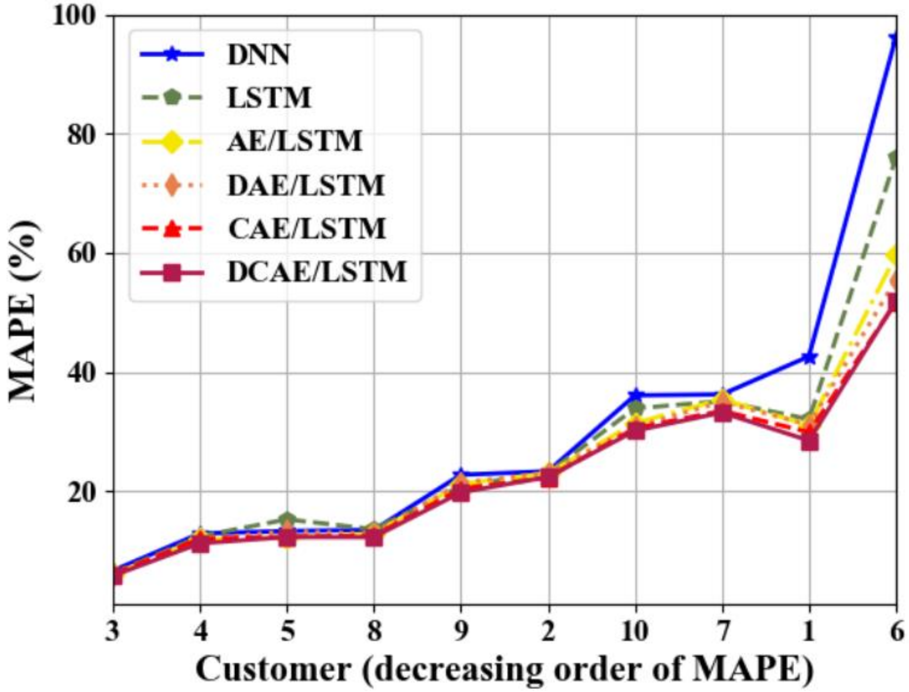
Missing rate 10%



Model	MAPE (%)		
	Average	Q_1	Q_4
DNN	32.54	8.73	78.36
LSTM	27.23	8.29	61.31
AE/LSTM	24.09	8.01	50.09
DAE/LSTM	23.46	8.10	47.41
CAE/LSTM	22.41	8.05	44.44
DCAE/LSTM	22.02	7.62	44.08

Performance evaluation of each customer

- When missing data 10%, the MAPE result for each customer

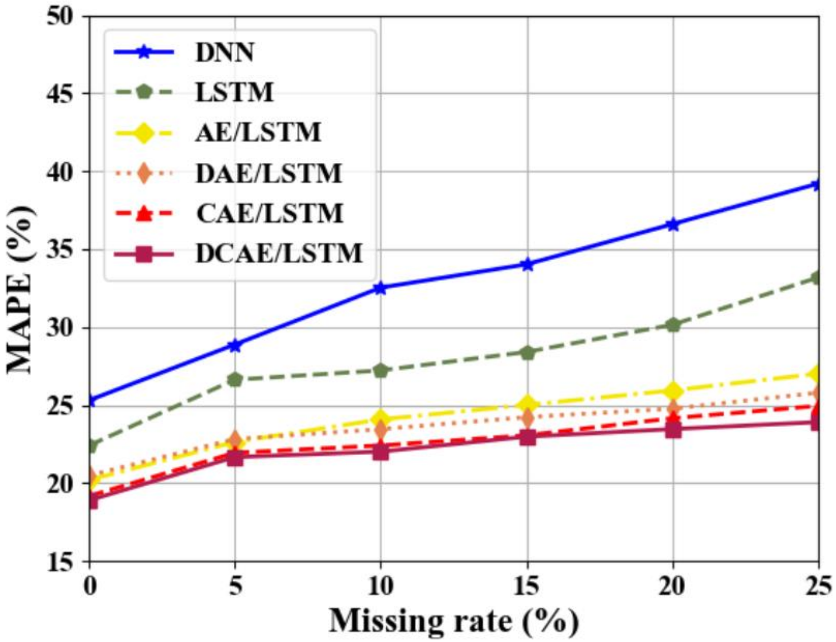


Model	MAPE (%)		
	Average	Q_1	Q_4
DNN	32.54	8.73	78.36
LSTM	27.23	8.29	61.31
AE/LSTM	24.09	8.01	50.09
DAE/LSTM	23.46	8.10	47.41
CAE/LSTM	22.41	8.05	44.44
DCAE/LSTM	22.02	7.62	44.08

↓ 32%, 13%, 44%

Average MAPE with different missing rates

- The convolution combined models show the best performance among other methodologies
- Compared to traditional forecasting models of DNN and LSTM, the combined models of extracting feature and forecasting achieve much smaller error for all missing rates



Model	Missing rate					
	Intact	5%	10%	15%	20%	25%
DNN	25.31	28.89	32.54	34.03	36.61	39.21
LSTM	22.41	26.63	27.23	28.41	30.15	33.19
AE/LSTM	20.20	22.64	24.09	25.01	25.95	27.02
DAE/LSTM	20.49	22.83	23.46	24.23	24.78	25.80
CAE/LSTM	19.17	21.95	22.41	23.07	24.17	24.95
DCAE/LSTM	18.90	21.67	22.02	22.98	23.47	23.91

Additional experiments at a fixed missing rate

- MAPE comparison with 5% block missing

Model	MAPE (%)		
	Intact	random missing	block missing
DNN	25.31	28.89	99.74
LSTM	22.41	26.63	32.17
AE/LSTM	20.20	22.64	60.38
DAE/LSTM	20.49	22.83	57.72
CAE/LSTM	19.17	21.95	25.54
DCAE/LSTM	18.90	21.67	23.47

↓ 25%, 25%, 76%

- Inputs of LSTM and their comparison with 10% missing data

Model	Feature domain		Time domain	
	Intact	Missing	Intact	Missing
AE/LSTM	20.20	24.09	21.48	24.84
DAE/LSTM	20.48	23.46	21.17	24.10
CAE/LSTM	19.17	22.41	21.96	24.07
DCAE/LSTM	18.90	22.02	21.21	23.95

Conclusion

- This paper presents a new forecasting method that is insensitive to missing data
- We propose a family of autoencoder/LSTM combined models for missing-insensitive short-term load forecasting
- The proposed convolution combined models generally achieve the best forecasting performance among the proposed models

Networking
Next

Information
Innovative

Communications
Creative

Energy
Envisioning

NICELAB

Networking for Information Communications and Energy