MISSING-INSENSITIVE SHORT-TERM LOAD FORECASTING LEVERAGING AUTOENCODER AND LSTM

Kyungnam Park, Jaeik Jeong & Hongseok Kim
Department of Electronic Engineering
Sogang University
35 Baekbeom-ro, Mapo-gu Seoul, Korea
{nami93,jaeik1213,hongseok}@sogang.ac.kr

ABSTRACT

Short-term load forecasting (STLF) is fundamental for power system operation, demand response, and also greenhouse gas emission reduction. So far, most deep learning-based STLF techniques require intact data, but many real-world datasets contain missing values due to various reasons, and thus missing imputation using deep learning is actively studied. However, missing imputation and STLF have been considered independently so far. In this paper, we jointly consider missing imputation and STLF and propose a family of autoencoder/LSTM combined models to realize missing-insensitive STLF. Specifically, autoencoder (AE), denoising autoencoder (DAE), and convolutional autoencoder (CAE) are investigated for extracting features, which is directly fed into the input of LSTM. Our results show that three proposed autoencoder/LSTM combined models significantly improve forecasting accuracy compared to the baseline models of deep neural network and LSTM. Furthermore, the proposed CAE/LSTM combined model outperforms all other models for 5%-25% of random missing data.

1 INTRODUCTION

Greenhouse gas emission causes severe environmental hazards like climate change, and reducing power generation and consumption is an important objective of the smart grid (Khan et al., 2016). In doing this, short-term load forecasting (STLF) plays a pivotal role; STLF is being used by power system operators for preparing the proper amount of electricity supply. Thus, accurate STLF can prevent excessive power generation reserve and lower the use of fossil fuels, which in turn leads to mitigate climate change (Yaslan & Bican, 2017). Furthermore, based on accurate STLF, demand response can actively change the electric usage of users with a high price of electricity in peak hours or by giving an incentive as a reward of lowering the power consumption (Pramono et al., 2019).

Recently, artificial intelligence techniques are widely used for STLF, such as artificial neural network (ANN) (Czernichow et al., 1996), deep neural network (DNN) (Ryu et al., 2017), recurrent neural network (RNN) (Vermaak & Botha, 1998), and long short-term memory (LSTM) (Kong et al., 2017a;b; Choi et al., 2018). However, in practice, data can be lost because of communications error, mechanical failure or loss of power (Li et al., 2018), and missing imputation has become critical. So far missing value of load data is filled with zero or average value using linear regression method, where learning models are usually created by separating missing imputation and other tasks such as forecasting or clustering.

In this paper, we propose a novel method by merging missing imputation into one of the steps in STLF. This model focuses on high forecasting accuracy under random missing and block missing data. In doing this, we leverage the unsupervised learning capability of autoencoder (AE) and the feature extraction of convolutional neural network. The intuition is such that an AE extracts important attributes (Ryu et al., 2019), which are used as an input to LSTM. Thus, even though there are some missing values in time domain, the features extracted by the autoencoder might be *insensitive* to missing values. We consider autoencoder (AE), denoising autoencoder (DAE), and convolutional autoencoder (CAE) for feature extraction.

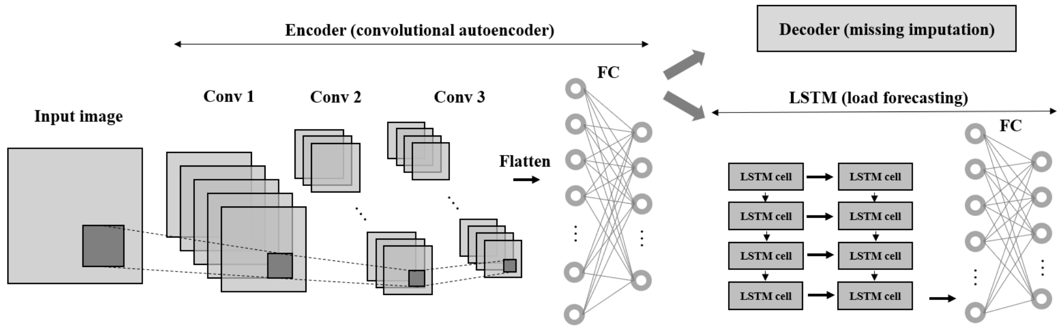


Figure 1: CAE/LSTM combined model.

We summarize our key contributions as follows. First, missing imputation and forecasting need not be performed separately. Instead, we propose a unified forecasting technique that is insensitive to missing values. The proposed method achieves accurate forecasting in the presence of severe missing rate, e.g., up to 25%. Second, using two-dimensional image data and two-dimensional convolution, we extract the features of the data in the presence of severe missing data. As missing rate increases, performance improvement over the conventional method also becomes significant. Third, we perform extensive experiments using not only CAE but also AE and DAE. We confirm that the proposed model outperforms the conventional model separating missing imputation and forecasting.

2 METHODS

2.1 OVERALL PROCESS

The overall process of STLF is divided into three main steps: data preprocessing, training, and test. In the first step, data cleansing is performed, and one-dimensional load time series data undergo min-max normalization to make data in the range [0, 1]. In order, to evaluate the performance when occurs missing, we intentionally make missing data. The data used in our work is demand side load data with 15 minute interval, and the number of data points for one day is 96. We utilize 7 days as an input to reflect one week and forecast the next day. In the case of CAE, we transform one-dimensional time series vector of size 7×96 into two-dimensional load image matrix of size (7, 96). In the cases of AE and DAE, we simply use one-dimensional 7×96 time series load data. The dataset is then partitioned into training set, validation set, and test set. In the second step, training set is used to train the forecasting model, and validation set is used to determine the hyperparameters of each model or each customer. We use the output of various AEs to derive the proper features from the raw time series load data. Then, the features are used as the input of LSTM for STLF. In the final step, we evaluate the performance with test set to demonstrate the feasibility of the proposed model.

2.2 AE/LSTM AND DAE/LSTM COMBINED MODELS

We first consider a model where the feature extraction of AE and the forecasting of LSTM are combined. To utilize AE as feature extraction, unsupervised learning is carried out, and decoder part is discarded after training. The encoder consists of three layers: 7 days one-dimensional load (1×672) data are converted to 500 one-dimensional data in the first layer and 300 one-dimensional data in the second layer and 100 one-dimensional data in the three layer. It is reshaped and applied as an input of (4, 25) to the LSTM. LSTM consists of four cells, which creates a model that forecasts the next one day (1, 96). In the DAE, Gaussian noise is added to the input, and the structure is the same as the AE.

2.3 CAE/LSTM COMBINED MODEL

Next, we consider a model where the feature extraction of CAE and forecasting of LSTM are combined. As shown in Fig. 1, the encoder consists of three layers of convolution (conv1, conv2, conv3) and three layers of pooling (pooling1, pooling2, pooling3). The filters in the convolution layers use gradually increasing structures to 5 filters, 25 filters and 125 filters, and the activation function uses exponential linear unit activation function. In the pooling layer, max pooling is used, and after the last pooling layer, the feature map unfolds, leading to the fully connected layer. Thus, 7 days load image data (7×96) are converted from the encoder output to 100 one-dimensional data. It is reshaped and applied as an input of (4×25) to the LSTM, which consists of four cells to forecast the next day (1×96).

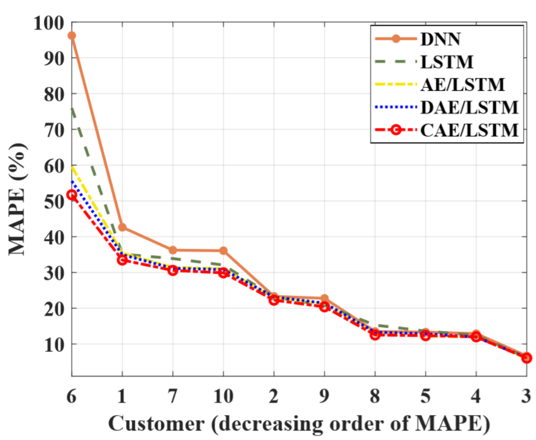
3 Results

The data used in our work is demand-side load data with 15 minutes interval, and is provided by Korea Electric Power Corporation (KEPCO). There are industrial customers in 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. Data set is split into training set for 420 days, validation set for 90 days and test set for 90 days. The peak loads of the customers span from 33kW to 12,342kW. Before using load data set as experimental data, abnormal values and missing values are replaced by the average of highly correlated data to serve as the ground-truth data for our experiment with missing.

In overall, the hyperparameters include the learning rate and the number of iterations. In addition, the hyperparameters of CAE are kernel size, the number of strides, dropout ratio, type of pooling, the number of filters in each layer, encoder output size, etc. In LSTM, we determine the size of the hidden unit vector, sequence length, the number of LSTM cells, etc. Each customer determines their hyperparameters separately. The hyperparameters are determined by comparing the mean absolute percentage error (MAPE) based on the validation set. All frameworks use tensorflow (Abadi et al., 2016), adaptive moment estimation (Adam) (Kingma & Ba, 2014) for optimizer and exponential linear unit (ELU) (Clevert et al., 2015) activation function.

Table 1: MAPE with 10% missing.

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



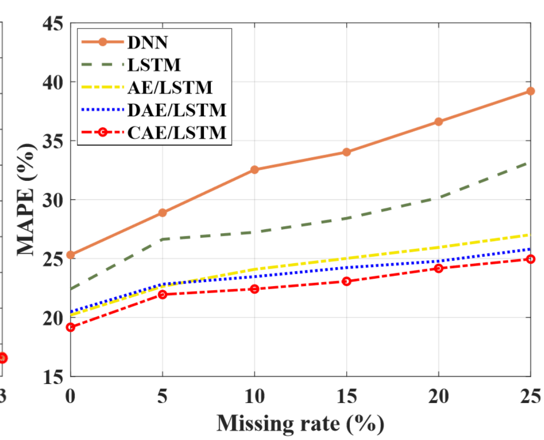


Figure 2: MAPE of each customer (10% missing). Figure 3: Average MAPE in terms of missing rate.

To verify the performance of the proposed models, we analyze the forecasting result by measuring the average, first quartile (0-25%, denoted by Q_1), and the fourth quartile (75-100%, denoted by Q_4) based on MAPE. When 10% missing occurs, the MAPE result for each customer is shown in Table 1 and Fig. 2, which shows that all the combined models of feature extraction and forecasting perform better than the traditional forecasting models of DNN and LSTM. Furthermore, the proposed CAE/LSTM combined model outperforms the other autoencoder/LSTM combined model, which shows the effectiveness of feature extraction using CNN for load image.

Fig. 3 shows the MAPE as the missing rate increases. The MAPEs of DNN and LSTM greatly increase as the missing rate increases. The CAE/LSTM shows the best performance for all the range of missing rate from 0% to 25%, followed by DAE/LSTM and AE/LSTM. Compared to the traditional forecasting models of DNN and LSTM, the combined models of extracting feature and forecasting achieve much smaller error for all missing rates.

Table 2: 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

We also apply the proposed method to block missing. As shown in Table 2, the MAPEs of DNN, AE/LSTM, DAE/LSTM surge when the missing block is relatively important. However, CAE/LSTM shows substantially better forecasting accuracy. In overall, the proposed CAE/LSTM outperforms all other methods.

Table 3: 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

The extracted feature is to prevent overfitting of raw data, so the prediction accuracy is higher than traditional forecasting models. To verify this, we also consider using decoder's output as the input to the forecasting model. The output of the decoder can be used as missing imputation as shown in Fig. 1. The proposed model (feature domain) has high prediction accuracy both in the intact data and the missing data. This result implies that the proposed model does not need to handle missing imputation separately. The corresponding result is shown in Table 3.

4 Conclusion

This paper presents a new forecasting method that is insensitive to missing data. We propose a family of autoencoder/LSTM combined model for missing-insensitive STLF, and the proposed CAE/LSTM generally achieves the best forecasting performance among the proposed models. Also, the higher the discrepancy, the more the proposed models can contribute to the forecasting improvement than the traditional forecasting model. We analyze the forecasting with missing data and show the superiority of the proposed combined models. The results show that, if 10% missing occurs, the baseline DNN model has MAPE 32.54%, whereas the proposed CAE/LSTM model has MAPE 22.41%.

Acknowledgments

Use unnumbered third level headings for the acknowledgments. All acknowledgments, including those to funding agencies, go at the end of the paper.

References

Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, pp. 265–283, 2016.

Hyungeun Choi, Seunghyoung Ryu, and Hongseok Kim. Short-term load forecasting based on resnet and lstm. In *2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, pp. 1–6. IEEE, 2018.

Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*, 2015.

T Czernichow, A Piras, K Imhof, P Caire, Y Jaccard, B Dorizzi, and A Germond. Short term electrical load forecasting with artificial neural networks. *Engineering Intelligent Systems for Electrical Engineering and Communications*, 4(ARTICLE):85–99, 1996.

Ahsan Raza Khan, Anzar Mahmood, Awais Safdar, Zafar A Khan, and Naveed Ahmed Khan. Load forecasting, dynamic pricing and dsm in smart grid: A review. *Renewable and Sustainable Energy Reviews*, 54:1311–1322, 2016.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Weicong Kong, Zhao Yang Dong, David J Hill, Fengji Luo, and Yan Xu. Short-term residential load forecasting based on resident behaviour learning. *IEEE Transactions on Power Systems*, 33(1): 1087–1088, 2017a.

Weicong Kong, Zhao Yang Dong, Youwei Jia, David J Hill, Yan Xu, and Yuan Zhang. Short-term residential load forecasting based on lstm recurrent neural network. *IEEE Transactions on Smart Grid*, 10(1):841–851, 2017b.

Linchao Li, Jian Zhang, Yonggang Wang, and Bin Ran. Missing value imputation for traffic-related time series data based on a multi-view learning method. *IEEE Transactions on Intelligent Transportation Systems*, 20(8):2933–2943, 2018.

Sholeh Hadi Pramono, Mahdin Rohmatillah, Eka Maulana, Rini Nur Hasanah, and Fakhriy Hario. Deep learning-based short-term load forecasting for supporting demand response program in hybrid energy system. *Energies*, 12(17):3359, 2019.

Seunghyoung Ryu, Jaekoo Noh, and Hongseok Kim. Deep neural network based demand side short term load forecasting. *Energies*, 10(1):3, 2017.

Seunghyoung Ryu, Hyungeun Choi, Hyoseop Lee, and Hongseok Kim. Convolutional autoencoder based feature extraction and clustering for customer load analysis. *IEEE Transactions on Power Systems*, 2019.

J Vermaak and EC Botha. Recurrent neural networks for short-term load forecasting. *IEEE Transactions on Power Systems*, 13(1):126–132, 1998.

Yusuf Yaslan and Bahadır Bican. Empirical mode decomposition based denoising method with support vector regression for time series prediction: a case study for electricity load forecasting. *Measurement*, 103:52–61, 2017.