



Short-term multiple power type prediction based on deep learning

Ran Wei¹ · Qirui Gan¹ · Huiquan Wang¹ · Yue You² · Xin Dang³

Received: 28 March 2019 / Revised: 7 July 2019 / Published online: 6 September 2019
© The Author(s) 2019

Abstract This paper proposes a method based on a 4-layer deep neural network model by stacked denoising auto-encoders to analyze four types of power data: current (I), voltage (U), active power (P) and reactive power (Q). We collect 7 days of household power data. In the beginning, the prediction accuracy rate can reach 82.45% when 1-h historical data are used to predict the data for the following 5 min. In order to optimize the parameters of this model, data over a 3-month period are collected. The prediction accuracy rate is 95.52% when three-day historical data are used to predict the data for the next hour. Finally, supplemental experiments are added to verify that the current change has a greater impact on the model. The 3-month data set is used as the training set. Extract 2 weeks of data

from 3 months of data, and the 2-week data is divided into two test sets. The effect of the model on the prediction accuracy from 7:00 in the morning to 24:00 in the evening, and from 0:00 in the evening to 7:00 in the morning is studied. The accuracy rates are 95.05% and 99.02%, respectively. It shows that the prediction accuracy of the model is higher for the period with a lower frequency of power consumption than the period with a higher frequency of power consumption, and that the change of the current has a greater impact on the prediction of the model. Finally, we prove that the effect of the 4-layer network is better than that of the 3-layer, 5-layer and 7-layer network models.

Keywords Deep learning · Neural network · Stacked denoising auto-encoder · Power prediction

This work is supported by the Tianjin Key Laboratory of Optoelectronic Detection Technology and Systems, the National Natural Science Foundation of China (61403276), Ministry of Education Returned Overseas Students to Start Research Fund (the 49th).

✉ Qirui Gan
qiruigan@stu.tjpu.edu.cn

Ran Wei
ranwei_tjpu@126.com

Huiquan Wang
huiquan85@126.com

Xin Dang
xindang_tjpu@126.com

- ¹ School of Electronics and Information Engineering, Tianjin Polytechnic University, Tianjin, China
- ² Integrated Energy Division, State Grid Electronic Commerce Company, LTD, Beijing, China
- ³ School of Computer Science and Software Engineering, Tianjin Polytechnic University, Tianjin, China

1 Introduction

The methods of power forecasting can be divided into two categories: short-term load forecasting methods, and mid- and long-term load forecasting methods (Yalcinoz and Eminoglu 2005; Santos et al. 2003). Common short-term load forecasting methods include: multiple regression, time series, artificial neural network (ANN), and expert system analysis. Multivariate regression methods are divided into linear and non-linear; multivariate linear regression is the need to use two or more influencing factors as independent variables to explain the dependent variable changes, and a linear relationship develops between multiple independent variables and dependent variables (Moghram and Rahman 2002; Han et al. 2015). The non-linearity between the independent variable and the dependent variable is a multiple nonlinear regression. This method can only

calculate the development level of comprehensive electricity load, and cannot predict the development level of load in each area (Wu et al. 2013; Adamowski et al. 2012). The time series method is a method of predicting the future using chronological data. This method only focuses on fitting mathematical models through data, does not combine deep neural networks, and does not deal with regularity (Safari et al. 2017; Emmert-Streib and Dehmer 2007). Artificial neural network (ANN) is a mathematical model that simulates neuronal activity and is an information processing system based on the structure and function of the brain neural network (Mohammadi et al. 2015; Giorgi et al. 2011). However, the analyzed power types are relatively simple. The method of Zhang et al. (2017) uses a 7-layer SAE based deep neural network for cerebral microbleed detection. But for different research objects and situations, not the more layers, the better the effect. The classification of Jia W's method has a good effect, but it may take a long time, and the loss of data may affect the generalization of the model (Jia et al. 2017). Based on the above deficiencies, this study uses a stacked denoising auto-encoder to construct a 4-layer deep neural network and combines it with time series prediction methods to analyze and predict four types of power. The cross-validation method was used to construct the training set and verification set (Turney 1994; Allix et al. 2014). Using sigmoid as the loss function (Shamir 2016; Rudin et al. 2011), this study is divided into three groups of experiments. The data sets of the first two sets of experiments use 1 week of data and 3 months of data to study the prediction accuracy of four kinds of power data: current (I), voltage (U), active power (P), reactive power (Q). The third set of experiments is based on 3 months of data to study the impact of the frequency of electricity consumption on power prediction.

2 Data collection and preprocessing

2.1 Collection of data

The data is collected electricity by using the power equipment of Tianjin TransEnergy Technology Co., Ltd. It is also can be expanded. And it is the electricity consumption data for the average household from February 26th to March 5th, 2018, and for the 3 months from December 2017 to February 2018. These two dataset experiments are recorded as Experiment 1 and Experiment 2, respectively. The supplementary experiment was recorded as Experiment 3, and the training data sets were based on the 3-month data used in Experiment 2. In addition, the data from January 5, 2018 to January 20, 2018 are extracted and divided into two parts, which are from 7:00

am to 12:00 p.m. and 12 p.m. to 7:00 p.m. every day. The two parts of the data are used as the test set of Experiment 3-1 and Experiment 3-2, respectively. The data types are current (I), voltage (U), active power (P), and reactive power (Q), and are recorded once every minute. Figure 1 shows a random sample of data for a given day. The abscissa is time in minutes and the ordinate is the corresponding value of the four type of power consumption data measured. The figure shows that the amplitude of current and active power change in different periods is relatively large. The figure also shows that power frequency and power consumption are varied among different periods.

2.2 Data preprocessing

The data is sorted by time series. After preparing the data, experiment 1 has 10,037 samples, experiment 2 has 106,887, experiment 3-1 has 122,095, and experiment 3-2 has 114,859. The time represented by each sample can be adjusted by altering the dimensions of the samples in the original data.

The data of experiment 1 is calculated as a new single sample every 5 min, and 2005 samples are obtained. The data is then normalized and replaced with data between 0 and 1. The dimension of each sample is transformed from the previous 4-dimensional data to the 48-dimensional data of each sample. The 4*12-dimensional representation of time is 5*12 for 1 h, the second set of samples starts from the following 5 min and ends at the 1-h mark. The number of samples is 2005, using the cross-overlay method, take the first to 48-dimensional data as a new single sample dimension, that is, the first sample to the 12th sample are superimposed together to form a new single sample. Then from the second sample to the 13th sample together as a new second sample, that is, cross-oversampling. As an example, suppose the number of new samples is 1993, and the each sample is 48 dimension. Finally, we used the first 1500 of the 1993 samples as the training set and the last 493 as the test set. Cross-sampling is shown in Fig. 2.

The rest of the experiments are averaged in 60 min. Experiment 2 uses 48 h of historical data to predict the

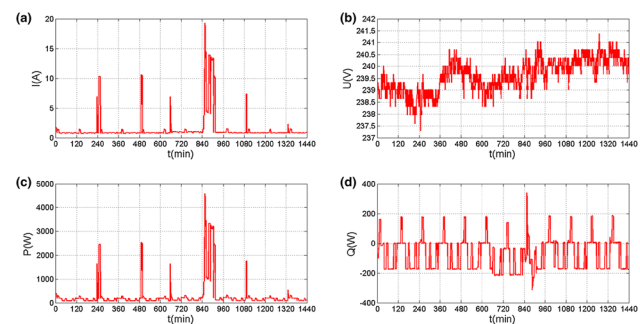


Fig. 1 Single day data graph

49th hour. After processing, the new sample size was 1733, and each sample had 192 dimension. Take the first 1500 samples as the training set, and the remaining 233 samples as the test set. Experiments 3-1 and 3-2 predict the 13th hour with 12 h of historical data. The new sample sizes are 2022 and 1818 respectively, the training set is 1800 and 1700 samples respectively, and the test set is 222 and 118 samples respectively.

3 Deep neural network model of power prediction

This chapter will introduce the construction of the power forecasting network model, including the relationship between the input and output relations of the power prediction model, the introduction of the training phase, and the setting of the model parameters after optimization. The overall structure of the DNN power prediction model experiment is shown in Fig. 3.

The overall process of the model experiment is divided into data collection, data preprocessing including changes in dimension, establishment of DNN model and input of sample data, and output layer output prediction data.

3.1 The relationship between the input and output of power prediction model

In the training phase, the number of neural nodes in the input layer is consistent with the dimension of the single sample after the pretreatment phase, and the middle two hidden layers are given optimal settings obtained through continuous training optimization. The output layer corresponds to the input layer, and the number of neuron nodes is 4. This establishes a link in the time series of data. During the test phase, the error between the predicted value and the real value is reduced by iterative and fine-tuning.

3.2 Stacked denoising auto-encoder (SDAE)

3.2.1 Denoising autoencoder

The traditional auto-encoder trains the network by the output value of the output layer and the error of the input layer data. This way of training while it is possible to obtain good training effect, but there is still a small error,

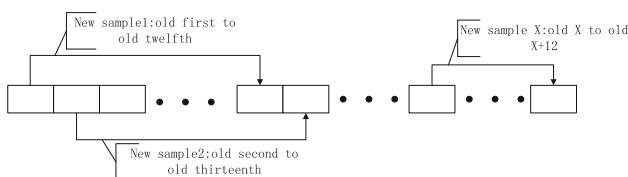


Fig. 2 Experimental sampling principle

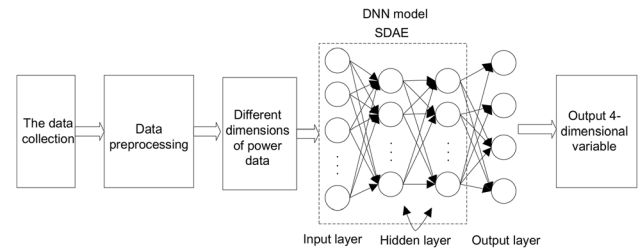


Fig. 3 Experimental sampling principle

the existence of small errors in the system of data of the study is not perfect, sometimes even contains the error of the data itself, tiny fluctuation is more sensitive to data, namely the poor robustness.

In this paper, on the basis of the previous research, optimization, the stacked sparse since the encoder into stack denoise auto-encoder (SDAE), the stack noise reduction since the encoder is a kind of training to get the encoder by way of special training. The specific method is to add some noise to the input data to learn and train the auto-encoder so as to generate anti-noise capability, thus obtaining more robust data reconstruction effect. The model is shown in Fig. 4.

Unlike the traditional auto-encoder, the input layer of the de-noising self-encoder is not the original data I , but the data that has been added to the noise, while the data of the output layer is still I . This approach can force network restore the original data from the defect data, from the error data found in the stable characteristics of mode greatly improves the robustness of the network, reduces the network to the sensitivity of the tiny differences in the input data.

3.2.2 *InputZeroMaskedFraction* is used to adjust the ratio of hidden layer nodes

InputZeroMaskedFraction is used in the input layer to adjust the noise ratio. In the hidden layer, it is similar to the

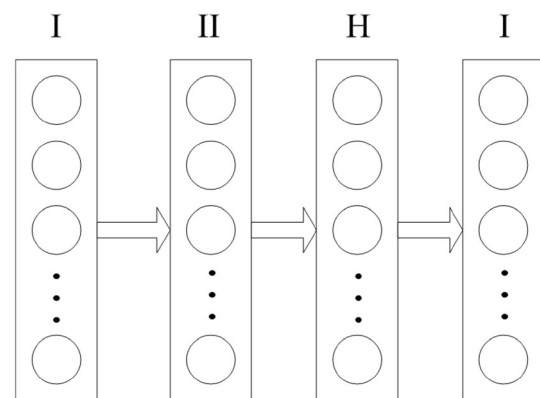


Fig. 4 A schematic diagram of noise reduction autoencoder structure

role of dropout, but the difference between the dropout and the dropout is that the dropout is to set the partial weight bias parameter to zero, while inputZeroMaskedFraction is to set a part of the data to zero. Both are anti-interference, dropout is external interference, and inputZeroMaskedFraction is internal anti-interference, which can increase the robustness of the system.

3.3 Network parameter optimization

The number of input and output nodes, the structure of the hidden layer, the number of layers and nodes of the hidden layer, the learning rate, the number of trainings and the batch quantity, and the number of iterations are all set. The effect of the model has a great influence on the prediction accuracy. In this study, a stacked denoising auto-encoder is used in the network layer. After optimization, the parameter settings of the DNN power prediction model in the two groups of experiments are summarized in Table 1.

In the experimental pre-training stage, we use the SDAE to iteratively update to get a set of initial parameters. Then the parameters are put into the neural network for training, the parameters are updated, and finally a better gradient fit is achieved. the activation function and noise coverage were Sigmoid and 100%, respectively. Numepochs, Learningrate1, and Batchsize1 are set to 50, 2, and 100, respectively. In the fine-tuning phase the iterations, Learningrate2, and Batchsize2 are 10, 1, and 100, respectively.

4 Power prediction neural network model results and analysis

4.1 Experiment 1 result

The overall test error of experiment 1 is 0.1755, i.e., the accuracy rate is 82.45%. The error of each feature vector is shown in Table 2.

Experiment 1 compares the predicted values of the eigenvectors on the test set with the graphs of the true values. The abscissa unit is time in minutes and the ordinate is the unit value of each eigenvector as shown in Fig. 5.

Table 1 Model parameters

Experiment	Input	Hidden 1	Hidden 2	Output
One	48	13	11	4
Two	192	13	10	4
Three-1	48	13	10	4
Three-2	48	13	10	4

4.2 Experiment 2 results

The overall test error of experiment 1 is 0.0448, i.e., the accuracy rate is 95.52%. Each error of each feature vector is shown in Table 3.

Experiment 2 compares the predicted value of each eigenvector on the test set with the graph of the true value. The abscissa unit is time in minutes, and the ordinate is the unit value of each eigenvector as shown in Fig. 6.

4.3 Experiment 3 result

In order to verify the impact of the electrical consumption at different times of day on the model, the 3-month data set is used as the training set. A total of 15 days of data from January 5 to January 20, 2018 are collected. The test set is taken from 7:00 am to 12:00 pm every day as test 3-2. Separating the test set for experiment 3-2 that uses less electricity from 0:00 to 7:00 in the morning, the results and analyses of the two supplementary experiments will be discussed separately.

4.3.1 Experiment 3-1 results

The overall error in Experiment 3-1 is 0.0495, which means the accuracy rate is 95.05%. Each error type of the feature vector is shown in Table 4.

Experiment 3-1 plots the predicted and true values, as shown in Fig. 7.

4.3.2 Experiment 3-2 results

The overall error in Experiment 3-2 is 0.0098, the accuracy rate is 99.02%. Each error type of the feature vector is shown in Table 5.

Experiment 3-1 plots the predicted and true values, as shown in Fig. 8.

4.4 Deep stacked denoising auto-encoder (SDAE) experiment

In order to better explore the model, take the data of experiment two as an example. we decrease and increase the number of layers in the network, from one auto-encoders to three auto-encoders and five auto-encoders. The

Table 2 The error of the E1 feature vector

Experiment 1	I	U	P	Q
meanerror	0.0610	0.0692	0.0531	0.0704
varerror	0.0084	0.0026	0.0095	0.0041
maxerror	0.6532	0.3362	0.6239	0.3398

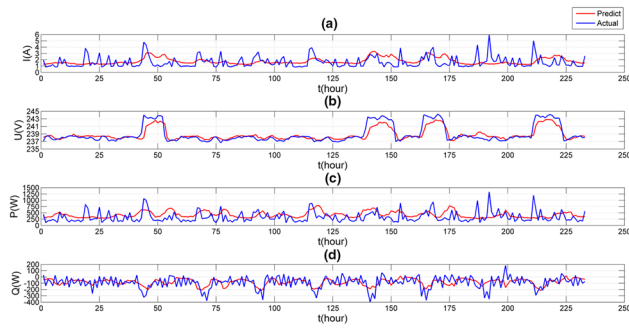


Fig. 5 Experiment 1 trends between predicted values and real values

Table 3 The error of the E2 feature vector

Experiment 2	I	U	P	Q
meanerror	0.0530	0.0920	0.0595	0.1156
varerror	0.0020	0.0118	0.0019	0.0069
maxerror	0.3455	0.7800	0.3120	0.5659

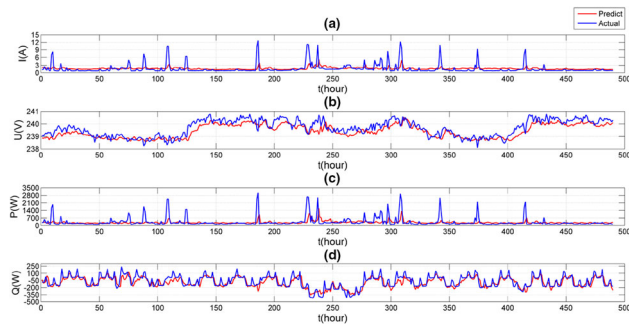


Fig. 6 Experiment 2 trends between predicted values and real values

Table 4 The error of the E3-1 feature vector

Experiment 3-1	I	U	P	Q
meanerror	0.0538	0.0711	0.0520	0.1132
varerror	0.0015	0.0107	0.0015	0.0072
maxerror	0.2965	0.7110	0.2842	0.4992

number of layers in the network is divided into three layers, five layers and seven of neural networks.

4.4.1 3-layer deep stacked denoising auto-encoder experiment 4-1

The overall error in experiment 4-1 is 0.4020, which means the accuracy rate is 59.80%. Each error type of the feature vector is shown in Table 6.

Experiment 4-1 plots the predicted and true values, as shown in Fig. 9.

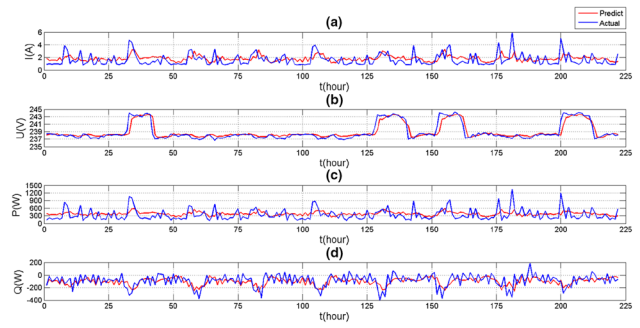


Fig. 7 Experiment 3-1 trends between predicted values and real values

Table 5 The error of the E3-2 feature vector

Experiment 3-2	I	U	P	Q
meanerror	0.0495	0.0118	0.0477	0.1158
varerror	0.0018	0.0001	0.0023	0.0101
maxerror	0.2606	0.0403	0.3096	0.5232

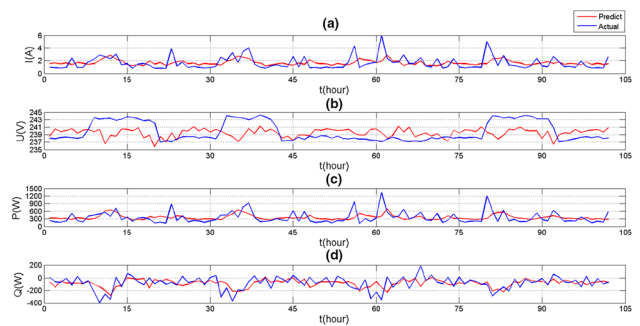


Fig. 8 Experiment 3-2 trends between predicted values and real values

4.4.2 5-layer deep stacked denoising auto-encoder experiment 4-2

The overall error in experiment 4-2 is 0.4041, the accuracy rate is 59.59%. Each error type of the feature vector is shown in Table 7.

Experiment 4-2 plots the predicted and true values, as shown in Fig. 10.

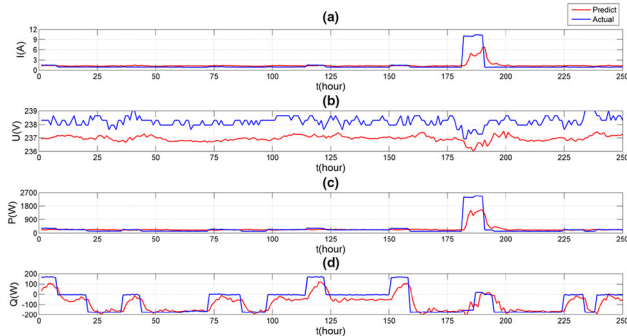
4.4.3 7-layer deep stacked denoising auto-encoder experiment 4-3

The overall error in Experiment 4-3 is 0.4041, the accuracy rate is 59.59%. Each error type of the feature vector is shown in Table 8.

Experiment 4-3 plots the predicted and true values, as shown in Fig. 11.

Table 6 The error of the E4-1 feature vector

Experiment 4-1	I	U	P	Q
meanerror	0.0211	0.1042	0.0165	0.0286
varerror	0.0018	0.0009	0.0015	0.0009
maxerror	0.3124	0.1937	0.3264	0.2607

**Fig. 9** Experiment 4-1 trends between predicted values and real values

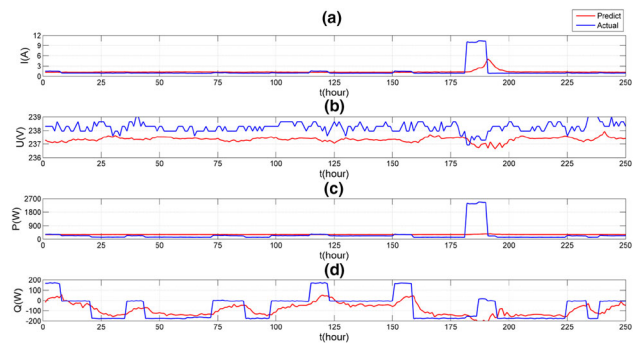
Comparing the results of Experiments 2 and 1, we can see that with the increase of data sets, all types of errors have been reduced as a whole, and the accuracy rate has greatly improved. The Experiment 3 shows that the prediction effect of the model on peak hours of electricity consumption is lower than that of the stable period of electricity consumption, i.e., the change of current has a great influence on the model. The fourth set of experiments shows that for the different number of network layers. Increasing or decreasing the number of layers in the network does not necessarily achieve better results. We should choose the most suitable number of structural layers and parameters. Too few layers can result in insufficient feature extraction. Too many layers can lead to over-abstraction of high-order feature extraction, so finding the right number of layers can achieve better results.

5 Conclusion

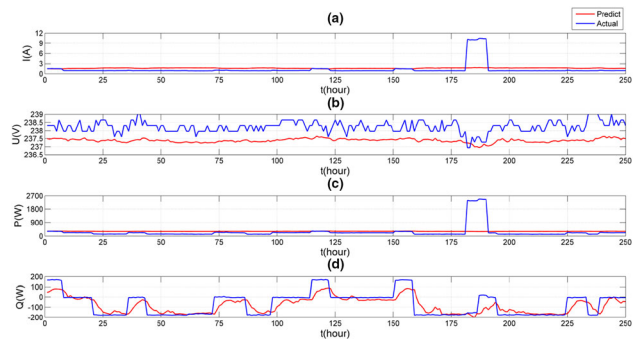
Based on previous research, we improve the optimization model and add noise based on the stacked auto-encoder to form a stacked denoising auto-encoder (SDAE). Using

Table 7 The error of the E4-2 feature vector

Experiment 4-2	I	U	P	Q
meanerror	0.0222	0.0739	0.0277	0.0382
varerror	0.0027	0.0010	0.0030	0.0010
maxerror	0.3345	0.1676	0.3359	0.2653

**Fig. 10** Experiment 4-2 trends between predicted values and real values**Table 8** The error of the E4-3 feature vector

Experiment 4-2	I	U	P	Q
meanerror	0.0325	0.0695	0.0287	0.0273
varerror	0.0027	0.0008	0.0030	0.0009
maxerror	0.3279	0.1560	0.3387	0.2493

**Fig. 11** Experiment 4-3 trends between predicted values and real values

auto-encoder can extract a large amount of information into nonlinear features, and this structural information loss is less. Combining the auto-encoder with the neural network architecture can establish a higher degree of fitness, and the information retention is higher after extracting features. Using SDAE can improve the robustness of the model. For the actual power collection process, special cases such as incomplete data collection and partial missing data may be encountered. SDAE can better predict and achieve better results than SAE. One-week data sets use 1-h historical data to predict the next 5 min. The prediction accuracy can reach 82.45% using this setup. The 3-month dataset uses 2 days of historical data to predict the next 1 h. The accuracy of prediction can reach 95.28% using this setup. Compared with the first group of experiments, the overall progression has been greatly improved. Supplementary experiments, using the 3-month data as a training set,

divided the 2-month data into two periods of frequent use and relatively stable data, respectively, as a test set. It verifies that the prediction effect of the model on peak hours is lower than that of the stable period. It has been proven that the change of the current has the greatest impact on the training prediction effect in the entire power prediction model. Then, we prove that the choice of the network model should choose the most appropriate number of network layers, not the larger or smaller the number of layers, the better the effect. The experimental results show that the DNN model based on stacked denoising auto-encoder (SDAE) is more scientific and accurate than the traditional power prediction method and general deep learning DNN model. In the output stage, we produce the average error, error value and maximum error of each feature vector based on the output of the overall prediction accuracy. This provides effective information for home users and the power supply sector. We still have no discussion about the generalization of the model. In the next work, we will try to add other types of data, such as climate, temperature and humidity, in addition to power data, in order to explore the generalization of rich models. With the continuous improvement of this method, DNN based on stacked denoising auto-encoder (SDAE) has a very high practical application in today's energy prediction demands.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Adamowski J, Fung Chan H, Prasher SO et al (2012) Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resour Res* 48(1):273–279
- Allix K, Klein J, State R, et al. Large-scale machine learning-based malware detection: confronting the “10-fold cross validation” scheme with reality. In: *ACM conference on data and application security and privacy*. ACM, 2014, pp 163–166
- Emmert-Streib F, Dehmer M (2007) Nonlinear time series prediction based on a power-law noise model. *Int J Mod Phys C* 18(12):1839–1852
- Giorgi MGD, Ficarella A, Tarantino M (2011) Assessment of the benefits of numerical weather predictions in wind power forecasting based on statistical methods. *Energy* 36(7):3968–3978
- Han Y, Li Z, Zheng H et al (2015) A decomposition method for the total leakage current of MOA based on multiple linear regression. *IEEE Trans Power Deliv* 31(4):1422–1428
- Jia W, Muhammad K, Wang SH et al (2017) Five-category classification of pathological brain images based on deep stacked sparse autoencoder. *Multimed Tools Appl* 1:1–20
- Moghran I, Rahman S (2002) Analysis and evaluation of five short-term load forecasting techniques. *IEEE Trans Power Syst* 4(4):1484–1491
- Mohammadi K, Shamshirband S, Yee PL et al (2015) Predicting the wind power density based upon extreme learning machine. *Energy* 86:232–239
- Rudin C, Letham B, Kogan E et al (2012) A learning theory framework for sequential event prediction and association rules. *Massachusetts Inst Technol Oper Res Cent* 43(5):A294
- Safari N, Chung CY, Price GCD (2017) A novel multi-step short-term wind power prediction framework based on chaotic time series analysis and singular spectrum analysis. *IEEE Trans Power Syst* 33:590–601
- Santos PJ, Martins AG, Pires AJ (2003) On the use of reactive power as an endogenous variable in short-term load forecasting. *Int J Energy Res* 27(5):513–529
- Shamir O (2016) Without-replacement sampling for stochastic gradient methods: convergence results and application to distributed optimization
- Turney P (1994) Theoretical analyses of cross-validation error and voting in instance-based learning. *J Exp Theoret Artif Intell* 6(4):331–360
- Wu Y, Shen Y, Lin C et al (2013) Maximum power supply capability evaluation based on multivariate nonlinear regression model for medium-voltage loop distribution network. *Dianli Zidonghua Shebei* 33(12):7–13
- Yalcinoz T, Eminoglu U (2005) Short term and medium term power distribution load forecasting by neural networks. *Energy Convers Manag* 46(9):1393–1405
- Zhang YD, Hou XX, Lv YD, et al (2017) Sparse autoencoder based deep neural network for voxelwise detection of cerebral microbleed. In: *IEEE international conference on parallel & distributed systems*. IEEE

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.