

PAPER • OPEN ACCESS

Source Load Forecasting of Customer-side Multi-energy System Based Feature Engineering

To cite this article: Qian Liu *et al* 2021 *IOP Conf. Ser.: Earth Environ. Sci.* **632** 042078

View the [article online](#) for updates and enhancements.

You may also like

- [Enhancing the generalizability of predictive models with synergy of data and physics](#)
Yingjun Shen, Zhe Song and Andrew Kusiak
- [Energy storage capacity configuration in multi-energy complementary systems](#)
Yafei Ding, Xuhui Shen, Yunting Song et al.
- [Iterative dynamic dual-energy CT algorithm in reducing statistical noise in multi-energy CT imaging](#)
Yidi Yao, Liang Li and Zhiqiang Chen



ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

Source Load Forecasting of Customer-side Multi-energy System Based Feature Engineering

Qian Liu*, Jingtao Wang and Yuanbo Zhang

NARI Technology Company Limited Beijing Energy Technology Branch, Beijing, China

*Corresponding author e-mail: liuqian3@sgepri.sgcc.com.cn

Abstract. Existing research mainly focuses on the single forecasting target and algorithm optimization, rarely mentions how to extract features for forecasting. For the source and load data of the customer-side multi-energy system, the influencing factors are complex and coupled. So, the quality of features gradually becomes the bottleneck that limits its forecasting accuracy. In this regard, based on EDA technology, this paper proposes a source and load forecasting method of customer-side multi-energy system based on feature engineering. First, combined with domain knowledge, perform a more systematic and complete feature analysis for the original observation data. Through this process, we can extract time features, statistical features and combined features. Subsequently, according to the characteristics of the data, the corresponding algorithm is selected to build a forecasting model. Finally, conduct two experiments based on source-side data and load-side data respectively, which proved that this approach can significantly improve the forecasting accuracy.

1. Introduction

In recent years, with the increasing shortage of energy supply and the promotion of policies related to energy conservation and emission reduction, the proportion of renewable energy, mainly wind power and photovoltaic, in the customer-side power supply structure has continued to increase. Through accurate source and load forecasting, it can provide support for the economic dispatch of multi-energy systems on the customer-side, coordinate and control resources, and realize the mutual complementarity of multiple energy sources. Because the load and source of the multi-energy system on the customer-side are affected by weather, environment, and energy-using behavior, it has strong randomness and volatility. Therefore, accurate and reliable customer-side multi-energy system source and load forecasting is of great significance to ensure the intelligent use of energy [1].

Research on source and load forecasting is currently focused on the prediction of a single target. Such as photovoltaic [2], wind power [3] and other equipment, energy storage charge state forecasting [4], enterprise electrical load forecasting [5], etc. For the forecasting of the source and load of the multi-energy system on the customer-side, we need to feature extraction and model construction separately, and there is no universal method. In fact, the data characteristics of different energy types and load types are mostly similar. Based on a universal method, it can reduce unnecessary repetitive operations and ensure the prediction accuracy to a certain extent.



In addition, with regard to the extraction of source and load data features, most of the existing studies have only performed simple operations such as correction, clustering and standardization. With the access of distributed energy, energy storage equipment and various forms of loads such as photovoltaic and wind power, the source and load data characteristics of multiple energy systems on the customer-side are becoming more and more complicated. Only simple processing of the data will obviously bring greater predictions error.

In summary, there are two main problems in the study of customer-side multi-energy system source and load forecasting: on the one hand, there is a lack of a general modeling method, and each independent forecasting target needs to be analyzed and model constructed one by one; on the other hand, less attention is paid to feature extraction, which is not conducive to improving the accuracy of source and load forecasting in the comprehensive energy field.

For this purpose, the paper proposes a customer-side multi-energy system source and load forecasting method based on feature engineering. Through systematic feature engineering, construct time features that reflect time series and change trends, statistical features that reflect data distribution characteristics, and combined features that reflect multivariate coupling relationships, and so on. Then, considering the characteristics of prediction targets and feature items, select suitable algorithm to predict. Experiments show that the method can significantly improve the accuracy.

2. Related Work

For the source and load forecasting of customer-side, multi-energy system, the optimization of the forecasting algorithm and the extraction of source and load features are the two main research contents.

For the source and load forecasting algorithms, existing researches have both predictions based on a single algorithm and combined method based on multiple forecasting algorithms. Paper [6] proposed a load forecasting method based on similar days and SVM. Considering the chaotic characteristics of load time series, paper [7] proposed a short-term load forecasting method based on chaos theory and support vector machines; paper [8] applied data mining, clustering and other algorithms to the forecasting method, and proposed a short-term load forecasting model based on data mining and fuzzy neural network. However, these forecasting models are built for a specific business scene and may not be suitable for other requirements. This paper analyzes the experimental data, summarize the extracted features, using different forecasting methods to build models based on the characteristics of different features, to ensure the accuracy and efficiency of the source and load forecasting.

For the extraction of source and load data features, most of the existing studies only performed simple correction, clustering, standardization and other operations. Paper [9] considers the continuity of photovoltaic power generation data and introduces the daily maximum photovoltaic output power and the daily average photovoltaic output power as the new characteristics of the previous day. Considering that the statistical values of meteorological data can reflect the characteristics of data distribution to a certain extent. The paper [10] introduced intra-day temperature difference and daily maximum temperature as new features. On this basis, paper [11] also considered the continuity of meteorological data and introduced daily average temperature and average solar irradiance on the current day and the previous 5 days as new features. Paper [12] uses the Pearson Coefficient method to obtain solar irradiance, atmospheric turbidity, and relative humidity as features, and then uses clustering to find similar days to be measured, using historical data of similar days to forecast the daily PV output power. Despite the simple processing of the data, considering the challenges and complexity of feature engineering, it is clear that these operations cannot guarantee the quality of the features. In view of the decisive role of feature quality in the forecasting model, the feature extraction process must be optimized.

In order to solve the above problem, paper proposes a source and load forecasting method of customer-side, multi-energy system based on feature engineering. Through systematic feature engineering, accurate feature modeling for the forecast target, and adopted the most suitable forecasting algorithm according to its energy using characteristics, which is more efficient and accurate, and has practical significance.

3. Approach

3.1. Overall Workflow

The overall workflow of this approach is shown in Figure 1. **First**, based on the exploratory data analysis (EDA) methods, perform feature engineering on the original data. Specifically: (1) Pre-process the original data. (2) Perform feature analysis, feature design, construction, and extraction based on EDA technology to obtain the data feature set. (3) Filter features with high correlation to form feature subsets as the final feature. **Second**, according to the data characteristics of final features, use a suitable algorithm to forecast. For example, for the targets with strong regularity of feature items, a simple algorithm can be used. For the target with high volatility of feature items, a more complex algorithm needs to be used to ensure accuracy.

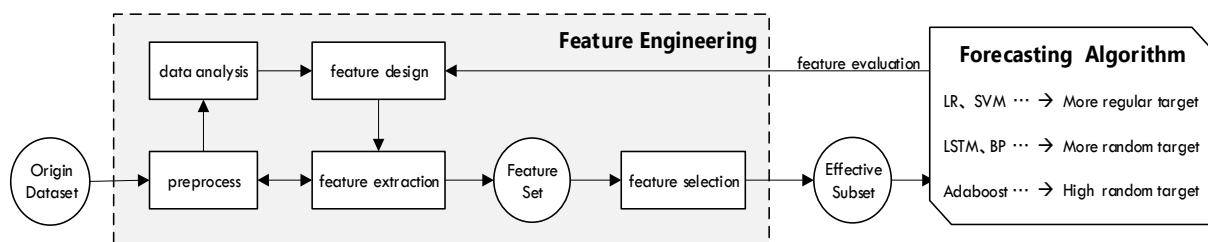


Figure 1. The overall workflow of approach.

3.2. Feature Expression

Combined with the domain knowledge, we designed and constructed three types of features: statistical feature, time feature, and combination feature, which respectively reflect the overall distribution characteristics of data, timing and change trend of data, and coupling relations among multiple variables.

3.2.1. Statistical Feature. Make statistics and description of quantitative data by using the statistical indicators in traditional statistics. As shown in Figure 2, the scatter plot can be used to find the distribution characteristics of the data, including the central trend, the off-center trend, the distribution pattern, and so on.

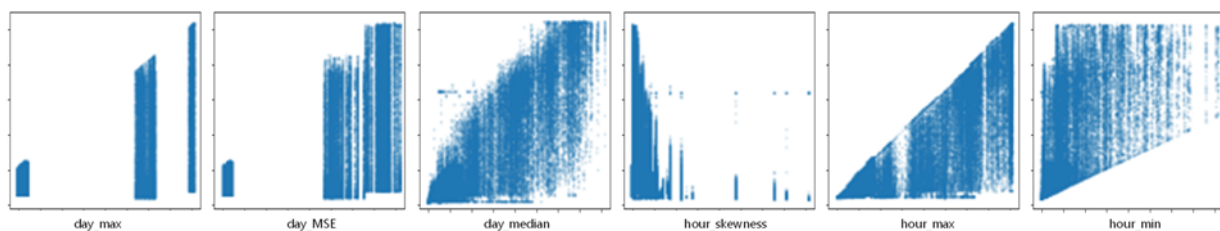


Figure 2. The data scatter diagram.

The statistical indicators are mainly divided into three categories. (1) The average index is used to reflect the general level or distribution trends. Commonly used are arithmetic mean, mode, median, etc. (2) the mutation index is used to reflect the mutations of overall distribution or the degree of dispersion. Commonly used are quartile, standard deviation, discrete coefficient, etc. (3) Moments, skewness and kurtosis, to reflect the overall distribution.

3.2.2. Time Feature. Although, it is known that there has a probability of mutation in the source and load data of customer-side, multi-energy system. It still has strong regularity. Therefore, the idea of time series model can be used for reference to carry out feature construction. On the one hand, parse the time item. We can get year, month, day, time, season and other information. On the other hand, calculate the rate of change. Although the source and load data of the customer-side and multi-energy system is prone

to random fluctuations, there is no mutations between adjacent moments. Therefore, the rate of change can be considered as a new feature.

3.2.3. Combination Feature. There is often a certain correlation between multiple features. By combining a single feature to form a combined feature, the features can be connected and interacted with each other, thereby expressing the non-linear characteristics that the single feature does not have, and enhancing the ability to express features. Using historical data to draw contour maps, the horizontal axis and vertical axis are the feature items to be combined. If there is a monotonous trend in the contour plot, it is basically a good combination. As shown in Figure 3, temperature and humidity can form a better second-order combination feature.

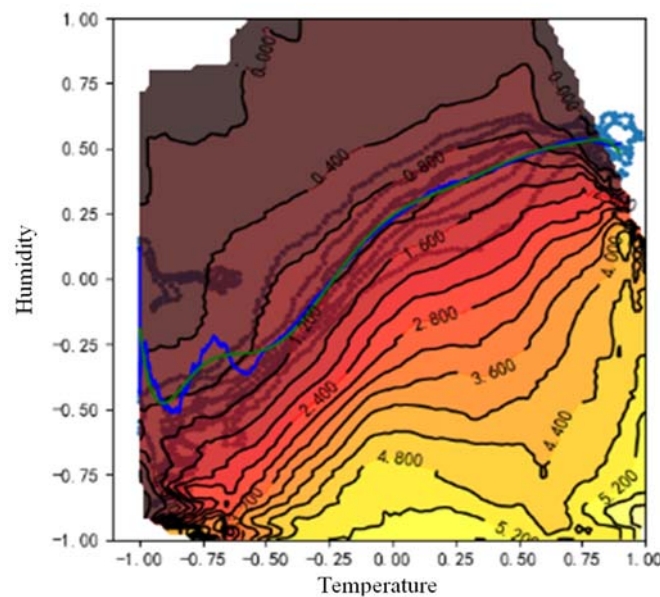


Figure 3. The temperature-humidity contour map.

4. Experiment

In order to verify the effectiveness for source and load forecasting in customer-side and multi-energy system, we perform two experiments, respectively: (1) Do the photovoltaic output forecasting experiment. (2) Do the electric load forecasting experiment.

4.1. Data Sources

The experimental data of the paper comes from Open Power System Data Platform [13]. This dataset contains measured time series data for several small businesses and residential households relevant for household or low-voltage-level power system modeling. The data includes solar power generation as well as electricity consumption (load) in a resolution up to single device consumption. The experiment chooses the dataset of “*industrial_building_institute*” to predict the electric load and photovoltaic output of the institute. The sum of data items “*grid_import*” and “*p_v*” is regarded as the electrical load for this institute. The data item “*p_v*” is regarded as the photovoltaic output.

The time range of the dataset is from February 2016 to February 2017. This experiment uses 2016 historical data for training, and uses January and February 2017 data to verify the accuracy of model.

4.2. Evaluation Indicators

In order to evaluate the performance of the prediction, paper sets root mean square error (RMSE) as the indicators. RMSE is more sensitive to the maximum and minimum values in the data, and can better

reflect the degree of error dispersion of the prediction model and reflect the accuracy of the fitting. The closer the value is to 0, the better the model fit. The specific formula is as follows:

$$\text{RMSE} = \sqrt{\sum_{i=1}^n (X_{\text{act}}(i) - X_{\text{pred}}(i))^2 / n} \quad (1)$$

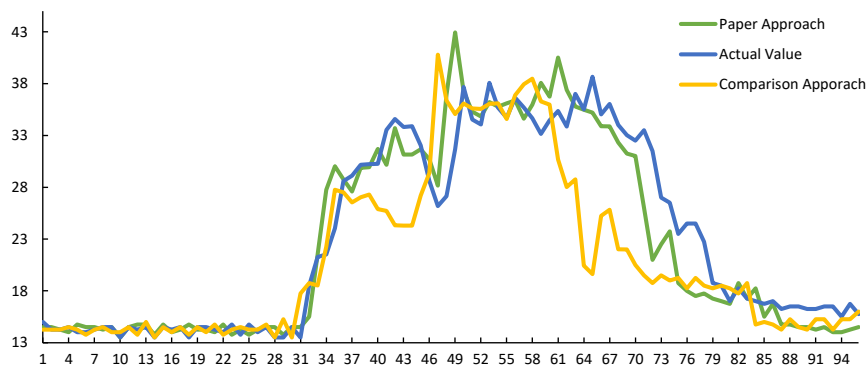
Where n is total count, $X_{\text{act}}(i)$ and $X_{\text{pred}}(i)$ are real and predictive value at time i , respectively.

4.3. Experimental results

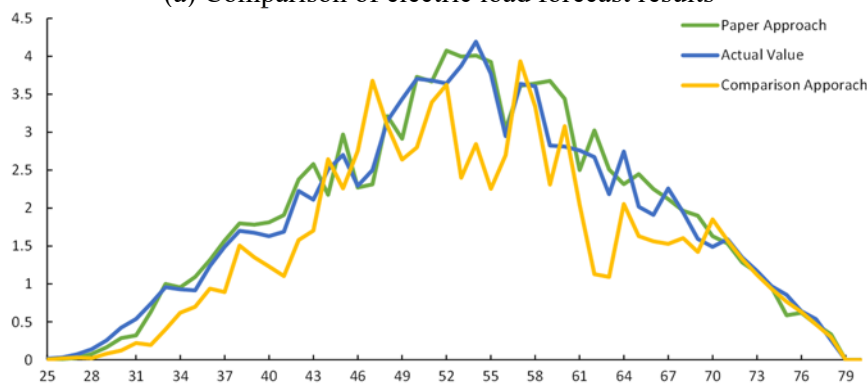
Through analysis, the feature items extracted from two experimental datasets have a high mutation rate, so both experiments use LSTM algorithm for feature learning and model building. Experiments use the method without feature engineering as a comparison. Finally, experiments show that the approach we proposed can significantly improve the forecasting accuracy.

4.3.1. Electrical load forecasting experiment. The RMSE of approach paper proposed is 0.21, while the comparative approach without feature engineering is 0.73. The results of a certain day are shown in Figure 4(a), where curve Paper Approach is the method based on feature engineering proposed in this paper, curve Comparison Approach is the method without using feature engineering, and curve Actual Value is the actual measured electric load.

4.3.2. PV output forecasting experiment. The RMSE of approach paper proposed is 0.13, while the comparative approach without feature engineering is 0.54. The results of a certain day are shown in Figure 4(b), where curve Paper Approach is the method based on feature engineering proposed in this paper, curve Comparison Approach is the method without using feature engineering, and curve Actual Value is the actual measured PV output.



(a) Comparison of electric load forecast results



(b) Comparison of PV output forecast results

Figure 4. The experimental results.

5. Conclusions

The existing research on source and load forecasting methods mainly focus on the single forecasting target, and focus on the improvement of algorithms. Most of the data used are processed data, rarely involving the feature engineering of the original data. However, feature engineering is the basis of building forecasting models, which determines the upper limit of prediction accuracy.

Based on the above, this paper proposes the source and load forecasting of customer-side multi-energy system based on feature engineering. With the classical DEA technology, combined with the domain knowledge to understand and express the data, paper summarizes three kinds of features: the time characteristics reflecting the timing sequence and change trend, the statistical characteristics reflecting the data distribution characteristics, and the combined characteristics reflecting the multivariable coupling relationship. Finally, combined with its energy characteristics, choose appropriate algorithm to build the prediction model.

Experimental results show that the approach proposed in this paper can significantly improve the accuracy of other methods.

Acknowledgments

This work was financially supported by foundation of NARI Technology Company Limited. (No.524608200058).

References

- [1] GONG Yingfei, LU Zongxiang, QIAO Ying, et al. An overview of photovoltaic energy system output forecasting technology [J]. Automation of Electric Power Systems, 2016, 40 (4): 140 - 151. DOI: 10.7500/AEPS20150711003.
- [2] Chen Changsong, Duan Shanxu, Cai Tao, et al. Short-term photovoltaic generation forecasting system based on fuzzy recognition [J]. Transactions of China Electrotechnical Society, 2011, 26 (7): 83 - 87.
- [3] Yang Qi, Zhang Jianhua, Wang Xiangfeng, et al. Wind speed and generated wind power forecast based on wavelet-neural network [J]. Power System Technology, 2009, 33 (17): 44 - 48.
- [4] Shi Qingjun. Research on optimal sizing and optimal energy management for microgrid [D]. Hangzhou: Zhejiang University, 2012.
- [5] Fahad Javed, Naveed Arshad, Fredrik Wallin, Iana Vassileva, Erik Dahlquist, et al. Forecasting for demand response in smart grids: An analysis on use of anthropologic and structural data and short term multiple loads forecasting [J]. Applied Energy, 2012, 96: 150 - 160.
- [6] Application of Support Vector Machine and Similar Day Method for Load Forecasting [C]. //Advances in Natural Computation pt.2: First International Conference on Advances in Natural Computation (ICNC 2005) August 27-29, 2005 Changsha, China. 2005: 602 - 609.
- [7] Zhang Zhisheng, Ma Long, Sun Yaming. Load forecasting model using chaos theory and support vector machine [J]. Proceedings of the Chinese Society of Universities for Electric Power System and Automation, 2008, 20 (06): 31 - 35.
- [8] Cui Herui, Song Xiuli, Ge Manqian. Research on FNN short-term electric load forecasting based on data mining technology [j]. Power System Protection and Control, 2009, 37 (22): 54 - 57.
- [9] CHEN Jinming, GUO Yajuan, WU Wangsong, et al. Multi-kernel SVM Short-term PV Power Prediction Based on Data Pre-processing and Characteristic Representation [J]. Water Resources and Power, 2018, 36 (09): 153+211 - 214.
- [10] ZHAO Shuqiang, ZHANG Tingting, LI Zhiwei, et al. Distribution Model of Day-ahead Photovoltaic Power Forecasting Error Based on Numerical Characteristic Clustering [J]. Automation of Electric Power Systems, 2019, 43 (13): 36 - 45.
- [11] SUN Yonghui, FAN Lei, WEI Zhinong, et al. Short-term Forecasting of the PV Output Power Based on Wavelet Analysis and Ensemble Learning [J]. Proceedings of the CSU-EPSA, 2016, 28 (4): 6 - 11, 30.
- [12] YU Feihong, CHEN Yongqiang, LEI Xia, et al. Short-term Prediction of Photovoltaic Output

Based on Meteorological Data Fuzzy Clustering and GAPS-BP Algorithm [J]. Computer Simulation, 2019, 36 (2): 447 - 451, 457.

- [13] Open Power System Data. 2020. Data Package Household Data. Version 2020-04-15. https://data.open-power-system-data.org/household_data/2020-04-15/. [EB/OL].