Identifying drivers of residential energy consumption by explainable energy demand forecasting

Jiao Jiao
Fraunhofer Institute for Systems and Innovation Research ISI
Breslauer Str. 48
76139 Karlsruhe
Germany
jiao.jiao@isi.fraunhofer.de

Heike Brugger
Fraunhofer Institute for Systems and Innovation Research ISI
Breslauer Str. 48
76139 Karlsruhe
Germany
heike.brugger@isi.fraunhofer.de

Michael Behrisch
Utrecht University
Princetonplein 5
3584 CC Utrecht
Germany
m.behrisch@uu.nl

Wolfgang Eichhammer
Fraunhofer Institute for Systems and Innovation Research ISI
Breslauer Str. 48
76139 Karlsruhe
Germany
wolfgang.eichhammer@isi.fraunhofer.de

Keywords

smart buildings, domestic energy efficiency, energy demand drivers

Abstract

Reducing primary and final energy demand is crucial for substantially decreasing Greenhouse Gas (GHG) emissions and reaching global and European climate targets. To understand what drives energy consumption behaviour is crucial for decision-makers to design policies effectively. Most current studies focus on smart meter data and sensor data. These data should be complemented by contextual, sociological, and behavioural data (acquired for example through surveys), which allow to study more precise user profiles. Integrating user profiles may reveal more valuable information, at the same time too much redundant information may also harm the prediction accuracy. How to select the crucial drivers is still understudied, but has direct impacts on the performance of the prediction.

This paper presents an explainable three-step forecasting method, which identifies long-term as well as seasonal trends and the most important drivers of household energy demand. In the first step, times series analysis (Bayesian Structural Time Series - BSTS) is applied to decompose energy demands into long-term, seasonal and residual components. In the second step, features are selected through a hybrid machine-learning approach (combining Extreme Gradient Boosting - XGBoost and Random Forest - RF), which reveals the key drivers of energy consumption. Finally, the energy consumption for each household is predicted with a deep-learning algorithm (Long Short Term Memory - LSTM). Furthermore, drivers of household energy demand – covering energy usage, building information and user profiles – are extracted and validated by domain experts. We apply this approach to a real-world dataset collected in eight German cities. The results demonstrate significant improvement in the prediction accuracy of electricity demand and interpretability of drivers.

Introduction

Recently, the building sector has become the leading energy-consuming sector: residential and commercial buildings alone take up to 32 % of global energy usage (Molina-Solana 2017). Therefore, policy makers and the public become increasingly aware of the need to address energy sustainability challenges in the building sector. Energy demand prediction is an essential step to achieving energy efficiency, which enables a balance between energy supply and demand in advance with optimized dispatching. Furthermore, identifying drivers of energy demands can provide critical information for decision makers on reducing energy consumption.

The booming of machine learning yields a succession of novel energy demand forecasting approaches, especially short-term forecasting at the regional and national levels (Qiu 2017, Ribeiro 2019, Song 2019, Sideratos 2020). After all, energy demand forecasting for households faces tough challenges due to the inherent high uncertainty of the low-capacity load and high flexibility of demand response systems (Cao 2019).

However, an approach for energy demand prediction with effective driver identification is lacking. On the one hand, the machine learning technique has proven its potential to be highly effective for energy demand prediction but it heavily relies on data quality (Panchalingam 2021). However, data collected from buildings always possesses characteristics such as multisource, polymorphic, heterogeneous and high-dimensional (Cheng 2019). Thus, to simplify the data pre-processing step, most state-of-the-art researches only pick up environment sensor data as the input features by domain knowledge. Very few studies consider survey data from consumers and other contextual, sociological and behavioural data sources (user profiles) for machine learning algorithms. Integrating more external data may reveal more information, but too much redundant information may also reduce model accuracy. How to select essential features for prediction is still an open question (Wang 2018).

On the other hand, the black-box property of current machine learning systems also calls for an approach for residential energy demand prediction with driver identification, which explains the mechanism of the prediction model by visualizing the importance of each input feature. It is difficult for policy makers and domain experts to make a decision with AI assistance, which may lead to a reliability bottleneck and a robustness issue. Besides, the lack of interpretability leads to a failure to effectively embed expertise into training processes, which is a treasure in the energy sector. As Explainable Artificial Intelligence (XAI) springs up in other domains, it should have a huge potential to propose XAI methods for energy demand forecasting and its driver identification, especially for residential buildings.

This paper put forward a novel three-stage framework for short-term forecasting to overcome the above two obstacles. First, embedding intelligence of energy and economic experts to solve the high uncertainty issue and interpretability problem, energy demands represented as smart meter data are decomposed into the long-term trend, seasonal and residual components based on the BSTS model. For each component, a hybrid approach combining XGBoost and RF is proposed for feature selection, where the drivers of each component can be identified in terms of the feature importance. Balancing between approach accuracy and efficiency, an LSTM-based prediction with the visualization of its decision-making process at each timestamp is applied on the residual component. Finally, the prediction of each component is aggregated as the energy demand time series.

The contribution of this work is three-fold. (1) We formulate the energy demand driver identification problem and propose an innovative three-stage framework to link this issue with energy demand prediction. (2) We present a machine learning method to extract crucial features efficiently learned from massive properties of raw user profiles. And we also integrate the LSTM algorithm with the BSTS model, which not only solves the transparency issue but also balances the efficiency and accuracy of our approach. (3) We conduct an extensive visual analysis of user behaviors and their drivers of residential electricity consumption based on the results of our approach on the real-world datasets collected from eight different cities. And the extracted drivers are confirmed by domain experts.

Explainable Energy Demand Forecasting Framework

This paper tackles the driver identification problem in the residential energy demand context by picking up the most important input features for energy demand forecasting. With the superiority of forecasting results, the decisive features of energy demand forecasting can be regarded as the drivers of the actual residential energy demand. In this paper, we work on hourly forecasting, which can be represented that given hourly energy consumption records $E_i=(e_{i1}, e_{i2}, \ldots, e_{it})$ of historical t time steps for each household h_i in H, the forecast model aims to predict the future hour value $e_{i(t+1)}$.

As described in Figure 1, our explainable energy demand forecasting framework consists of three steps: time series decomposition, feature selection and prediction. After pre-processing the smart meter data as the energy consumption per household, these pre-processed time series are decomposed into components with real meanings, which not only provides the depth insights on user behaviour patterns to domain experts but also decreases the uncertainty and increase the transparency of forecasting. The feature selection step aims to pick

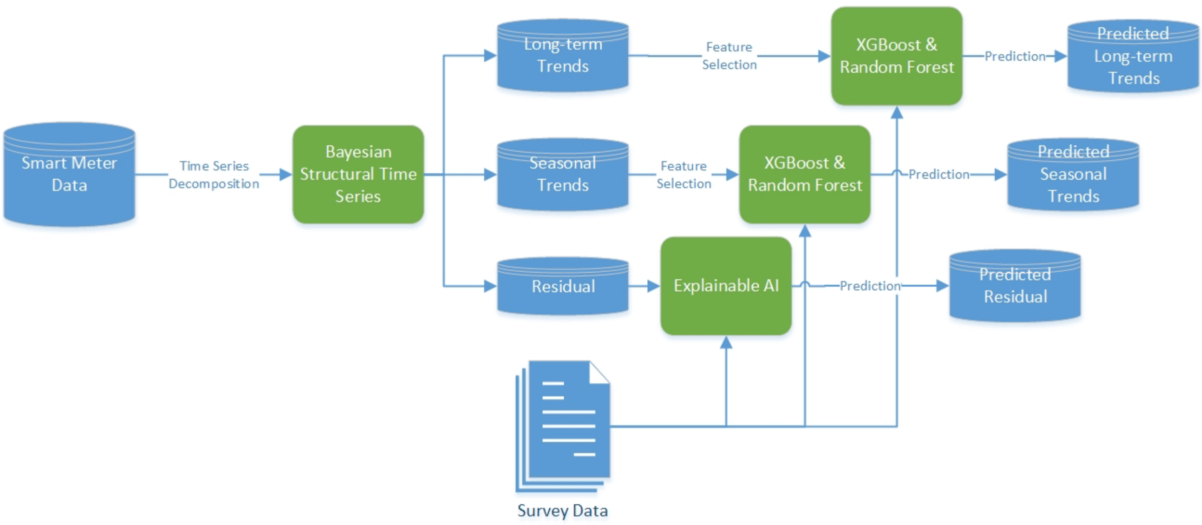

Figure 1. Our explainable three-step forecasting method, consisting of time series decomposition, feature selection, energy demand forecasting with visualization.

up the decisive factors for the energy demand forecasting from the survey data, which is linked with the third step, prediction. We encode the survey data as one-hot input features of the prediction model. As the accuracy of the prediction model is rising, the selected features based on the feature importance are more close to the actual drivers of residential energy demand. Thanks to the nonlinearity of machine learning methods, our framework can figure out indirect drivers, compared with the traditional methods. At the same time, visualization of selected features can reveal how the machine learning algorithms work, which is easy for domain experts to judge whether the outcome makes sense or not.

TIME SERIES DECOMPOSITION

Previous studies have demonstrated that the decomposition method can improve the accuracy of energy demand forecasting based on shallow networks effectively (Qiu 2017, Song 2019, Patidar 2021). Primarily, there are two common algorithms, wavelet decomposition and Empirical Mode Decomposition (EMD), applied to regional or national datasets (Qiu 2017, Song 2019, Xia 2020). However, both decompositions are difficult to explain the actual meanings of each component. In addition, they mainly work on regional and national datasets, which may not solve the high uncertainty and flexibility challenges caused by the low capacity of private households.

BSTS is a highly expressive model with a Bayesian state-space estimation framework, which decomposes time series into long-term, seasonal and residual components based on a clear economic explanation (Shaffery 2020). The long-term trend represents the development of energy demand for each household in a long run. Within each cycle, the user behaviour patterns are recognized as seasonal components. The residual component stands for the rest of energy demands that always come from unregular energy use. Shaffery et al. applied BSTS to decompose the summed photovoltaic generation and gross energy consumption for the Pecan Street AMI dataset, which is household level and 1-min resolution (Shaffery 2020). Since it has been proven on the high-resolution datasets for households, this paper adopts the BSTS model to decompose energy demand with more interpretability.

FEATURE SELECTION

In general, there are three categories of feature selection approaches: filter, wrapper and embedded algorithms (Pirbazari 2019). Filter methods rank features by statistical measures, such as mutual information (Pirbazari 2019), Pearson correlation (Kim 2020) and so on, independent of any machine learning method. However, it needs experiences to find matched feature selection methods and models for the best results (Sun 2015). In contrast, wrappers utilize learning algorithms to evaluate a combination of features forming during a search process and score a subset of features based on their predictive power (Pirbazari 2019). Recursive feature elimination (Ayub 2019, Pirbazari 2019) and RF (Gonzalez-Vidal 2019, Kim 2020) are typical wrappers for time series prediction. As for embedded methods, they have a built-in mechanism to perform variable selection, which means that feature selection is performed as a part of the training process for prediction (Pirbazari 2019). However, Pirbazari et al. has proven by experiments that Elastic Net, the up-to-date embedded network, generally lags behind the other two approaches (Pirbazari 2019). Thus, this paper works on wrapper algorithms.

Compared to previous studies, Ayub et al. proposed a hybrid feature selector with XGBoost and decision tree, which is more robust (Ayub 2019). XGBoost is an ensemble model of decision trees. Similar to RF, XGBoost generates a weak decision tree learner at each step and accumulates it into the total model (Pan 2018). Trees in RF are built independently, while XGBoost adds a new tree to complement already built ones based on the gradient direction of the loss function (Pan 2018). Therefore, parallel and distributed computing can speed up the learning process. When the forecasting model learns, XGBoost and RF look for the best tree splitting. Each feature represents a node that can be split. The feature importance then is calculated based on the number of times a feature is used to split the data across all trees, the average gain across all splits the feature is used in, and the average coverage across all splits the feature is used in. Since tree-based models are naturally designed for feature selection, this paper takes advantage of the hybrid combination of these two ensemble tree models.

ENERGY DEMAND FORECASTING

Over the past decades, a wide range of deep learning algorithms has been developed for short-term energy demand forecasting, especially at regional and national levels (Pirbazari 2019, Kim 2020, Sideratos 2020).

But for household-level prediction, LSTM is dominant (Bouktif 2018, Gangopadhyay 2020, Kim 2020). However, in our case, long-term and seasonal components are relatively regular, which can be well predicted by the feature selectors, XGBoost and RF. Thus, only the residual component with sudden changes is applied LSTM for hourly forecasting.

As for opening the black-box of the LSTM algorithm, state-of-the-art XAI studies for energy demand forecasting only dig into parts of household energy demand, such as Heating, Ventilation and Air-Conditioning (HVAC) system (Gangopadhyay 2020), transportation (Amiri 2021) and so on. Our work attempts to fill this gap by applying XAI on electrical smart meter data. Amiri et al. applied the Local Interpretable Model-Agnostic Explanation (LIME) to visualize the feature importance of transportation energy demand forecasting by LSTM (Amiri 2021). Considering its effectiveness on household travel survey data, we also apply LIME together with LSTM in the context of residential energy demand.

Visual Analysis on Drivers of Residential Electricity Consumption

Although our three-step framework is designed for energy demands, due to the data availability, we can also conduct visual analysis on the real-world electricity consumption dataset collected from a field study between 2009 and 2010 on smart metering within the German research project Intelliekon (Schleich 2011). 1720 households from eight German municipalities located in five federal German states are picked up to collect hourly electricity consumption at least 12 months. These households were interviewed at the beginning and the end of the field experiment about detailed user profiles with 70 features including building information (e.g. living space, house situation, location), household appliance stock, the use of household appliances, attitudes towards energy efficiency and socio-demographic characteristics (e.g. the number of residents, gender, job, education, age, living status, income) (Schleich 2011).

The decomposition of hourly electricity demands of 1720 households reveals the yearly and daily user behaviour patterns by long-term trend and seasonal component respectively. As for residual components, they represent the sudden or individual uses of each household, where uncertainty mainly comes from.

LONG-TERM TRENDS

As depicted in Figure 3, both XGBoost and RF predict the long-term trends of electricity demands with high precision. XGBoost even can forecast the sudden drops accurately so that the top drivers identified by these two selectors are close to the real user behaviours.

Size of living spaces is selected as the first driver of electricity demand in the long run, which set up the baseline of the long-term trend for each household. The larger the living spaces are, the higher baselines of electricity demands are. Apart from that, there are two categories of top 10 drivers for long-term trends: the number of house appliances and the number of residents. Thus, except for holiday periods, long-term trends are stable within the specific range. There is no distinct increase or decrease in electricity demand per household between 2009 and 2010, since the top 10 drivers almost remain constant. Energy efficiency measures also cannot change so much with the long-term trends.

SEASONAL COMPONENTS

As for seasonal components, they are quite similar among households. Within one day, electricity demand for each household has a clear peak in the evening (around 8 pm) and a weak peak in lunchtime (around noon). Hour is the most decisive factor for seasonal components, since daily user routines have a direct impact on their electricity consumption. Energy efficiency measures can have a deeper influence on the seasonal components than the long-term trends, since more electricity usage behaviours become top drivers, such as how many hours computers are used, washing machine usage per week and so on, which can be optimized by demand-side management, dynamic tariffs and other energy efficiency measures. Thus, identifying the most important drivers of seasonal components can be the basis for personalizing the nudging measures of energy efficiency.

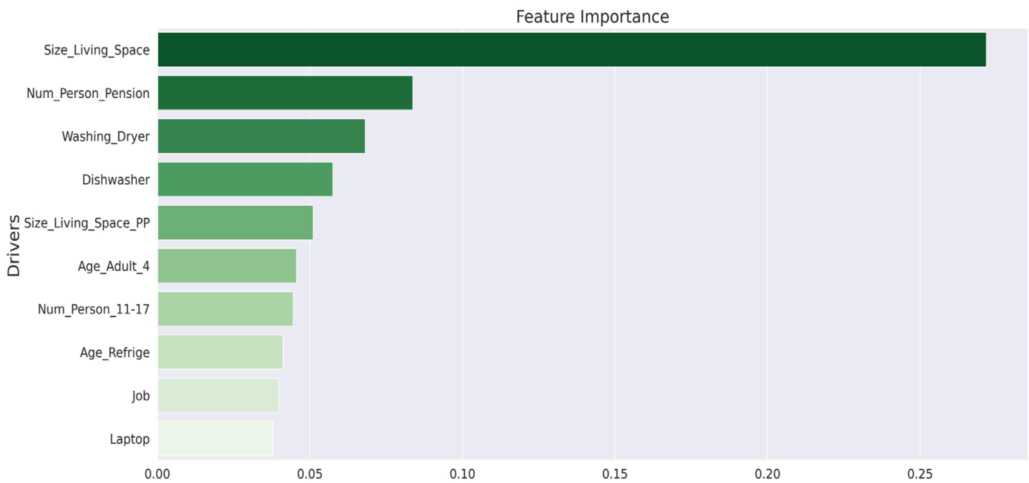

Figure 2. Top 10 important drivers for long-term trends of residential electricity demand.

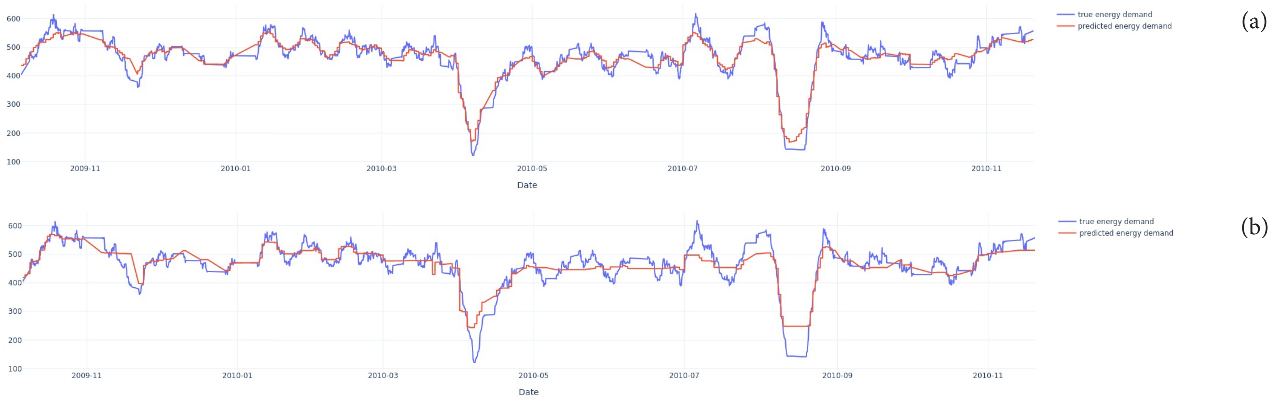

Figure 3. Hourly forecasting results for long-term trends of residential energy demand. (a) XGBoost (b) RF.

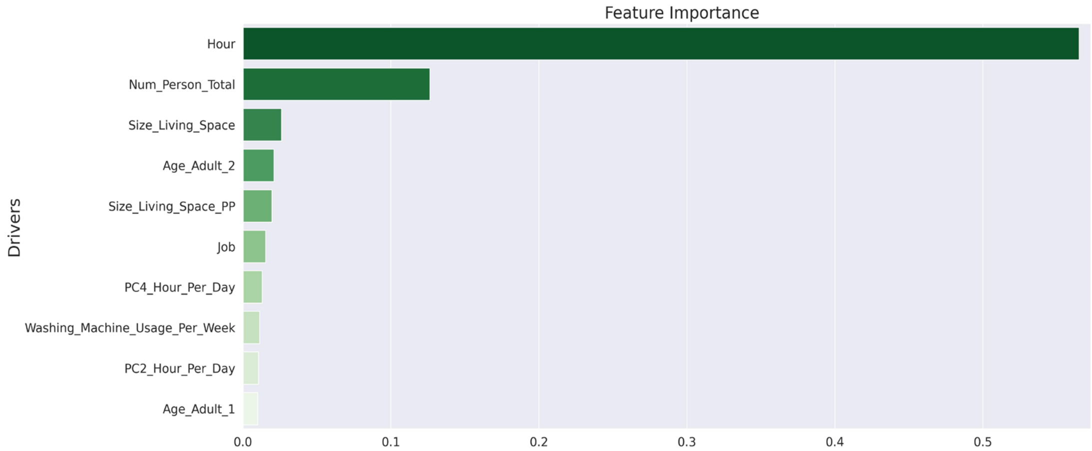

Figure 4. Top 10 important drivers for seasonal components of residential electricity demand.

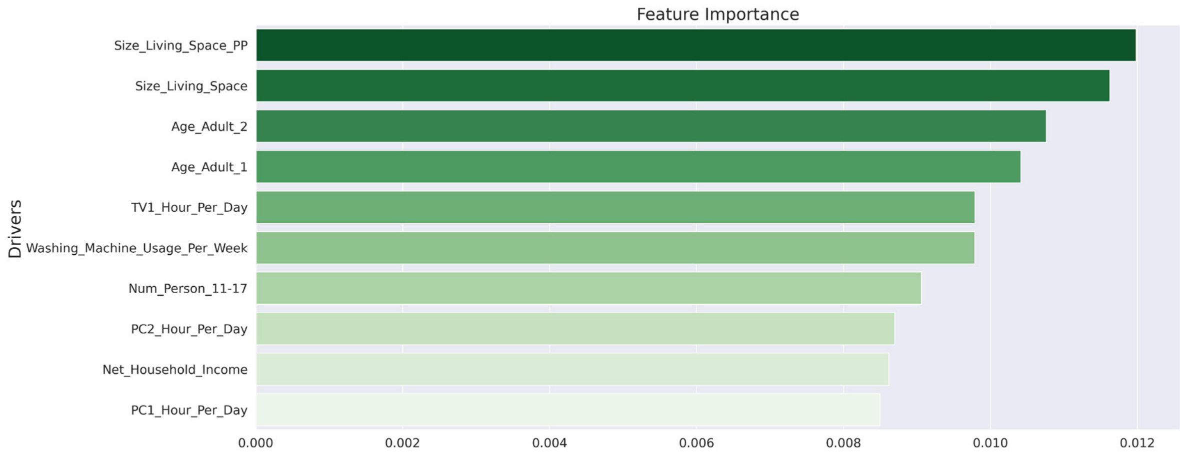

Figure 5. Top 10 important drivers for residual components of residential electricity demand.

RESIDUAL COMPONENTS

After forecasting with LSTM, LIME only provides interpretations for the individual predictions at each timestamp. Thus, we calculate the averages of feature importance for all the households at all the timestamps. Similar to seasonal components, there are multiple drivers related to electricity usage behaviours, such as watching TV, washing clothes, using computers and so on. Their energy efficiency can be easily lifted by demand response systems. Luckily, compared to long-term trends, this component takes a large share of the whole electricity demands, which offers a large space for optimization. However, the impacts of those drivers are quite weak (Figure 5), which may be caused by the hourly resolution. The higher resolution is processed, the more residual components fluctuate. Thus, the accuracy of the model decreases and then the impact on the sudden electricity usage of a single known driver turns to be weak.

In conclusion, the size of living spaces, the number of residents, the number of appliances (in particular washing dryers, refrigerators and dishwashers) and the use of appliances (particularly for washing machines and IT equipment) have a deep influence on the residential electricity demand, which has been widely reported in the literature (Jones 2015, Fan 2017, Cassarino 2018). However, the contribution of this paper is to quantify the impacts of those drivers directly. Surprisingly, income does not play an important role in the residential energy demand as expected, which supports the minority of opinions from the USA (Kavousian 2013, Jones 2015). Besides, attitudes towards energy efficiency, such as interests in electricity saving offers, taken measures of saving energy and so on, also do not significantly influence the residential electricity demand within a year, which makes up the lack of state-of-the-art research.

Conclusion

To advance the state-of-the-art studies on driver identification of residential energy demand by integrating data-driven approaches, this paper proposes an explainable three-stage energy demand forecasting framework with time series decomposition, feature selection and forecasting. The challenge of identifying the drivers for residential energy demand is solved by picking up the most important features for energy demand forecasting.

From a **methodological perspective**, this paper introduces XAI into the energy domain, in particular for the linkage between energy demand forecasting and driver identification. First, our time series decomposition approach based on BSTS can directly reveal yearly and daily user behaviour patterns per household by long-term trend and seasonal component respectively. Second, we import a hybrid feature selector with XGBoost and RF to identify the drivers of energy demand, especially for yearly and daily user behaviours. Thanks to the nonlinearity of machine learning algorithms, apart from direct drivers, indirect drivers can also be identified. Third, for visualizing how the energy demand forecasting system works, we set up a LIME-based XAI model with LSTM layers, which unveils the black-box of deep learning methods for domain experts and reveals the potential personalized nudging measures of energy efficiency.

From a **content perspective**, this paper draws the following two conclusions regarding the drivers of residential electricity demand: First, in general, there are two categories of top 10 drivers for long-term trends: the number of household appliances and the number of residents. Thus, the base of yearly user behaviour almost remains constant. Furthermore, energy efficiency measures on user behaviours also cannot change the electricity demand in the long run in a significant manner. Second, there are multiple drivers of seasonal and residual components related to electricity usage behaviours, which offers a large space to be easily optimized by energy efficiency measures, in particular demand response systems.

Lastly, although our proposed approach has already brought various technical and thematic contributions, there are two areas in particular could be explored further: First, current visualization still cannot express the work principles of deep learning methods clearly to domain experts who are not trained in AI methods. In the next step, we focus on designing a more user-friendly visualization format for domain experts. Second, the individual drivers identified by LIME at each timestamp can be a good base for personalizing the nudging measures of energy efficiency. In the future, we plan to link these results to further nudging measure personalization.

References

Amiri S S, Mottahedi S, Lee E R, et al. Peeking inside the black-box: Explainable machine learning applied to household transportation energy consumption[J]. Computers, Environment and Urban Systems, 2021, 88:101647.

Ayub N, Javaid N, Mujeeb S, et al. Electricity load forecasting in smart grids using support vector machine[C]//International Conference on Advanced Information Networking and Applications. Springer, Cham, 2019: 1–13.

Bouktif S, Fiaz A, Ouni A, et al. Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches[J]. Energies, 2018, 11(7): 1636.

Cao Z, Wan C, Zhang Z, et al. Hybrid ensemble deep learning for deterministic and probabilistic low-voltage load forecasting[J]. IEEE Transactions on Power Systems, 2019, 35(3): 1881–1897.

Cassarino T G, Sharp E, Barrett M. The impact of social and weather drivers on the historical electricity demand in Europe[J]. Applied energy, 2018, 229: 176–185.

Cheng L, Yu T. A new generation of AI: A review and perspective on machine learning technologies applied to smart energy and electric power systems[J]. International Journal of Energy Research, 2019, 43(6): 1928–1973.

Fan H, MacGill I F, Sproul A B. Statistical analysis of drivers of residential peak electricity demand[J]. Energy and Buildings, 2017, 141: 205–217.

Gangopadhyay T, Tan S Y, Jiang Z, et al. Interpretable deep attention model for multivariate time series prediction in building energy systems[C]//International Conference on Dynamic Data Driven Application Systems. Springer, Cham, 2020: 93–101.

Gonzalez-Vidal A, Jimenez F, Gomez-Skarmeta A F. A methodology for energy multivariate time series forecasting in smart buildings based on feature selection[J]. Energy and Buildings, 2019, 196: 71–82.

Jones R V, Fuertes A, Lomas K J. The socio-economic, dwelling and appliance related factors affecting electricity consumption in domestic buildings[J]. Renewable and Sustainable Energy Reviews, 2015, 43: 901–917.

Kavousian A, Rajagopal R, Fischer M. Determinants of residential electricity consumption: using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior. Energy 2013;55:184–94.

Kim J Y, Cho S B. Electric energy consumption prediction by deep learning with state explainable autoencoder[J]. Energies, 2019, 12(4): 739.

Kim W, Han Y, Kim K J, et al. Electricity load forecasting using advanced feature selection and optimal deep learning model for the variable refrigerant flow systems[J]. Energy Reports, 2020, 6: 2604–2618.

Molina-Solana M, Ros M, Ruiz M D, et al. Data science for building energy management: A review[J]. Renewable and Sustainable Energy Reviews, 2017, 70: 598–609.

Pan B. Application of XGBoost algorithm in hourly PM2. 5 concentration prediction[C]//IOP conference series: earth and environmental science. IOP publishing, 2018, 113(1): 012127.

Panchalingam R, Chan K C. A state-of-the-art review on artificial intelligence for Smart Buildings[J]. Intelligent Buildings International, 2021, 13(4): 203–226.

Patidar S, Jenkins D P, Peacock A, et al. A hybrid system of data-driven approaches for simulating residential energy demand profiles[J]. Journal of Building Performance Simulation, 2021, 14(3): 277–302.

Pirbazari A M, Chakravorty A, Rong C. Evaluating feature selection methods for short-term load forecasting[C]//2019 IEEE International Conference on Big Data and Smart Computing (BigComp). IEEE, 2019: 1–8.

Qiu X, Ren Y, Suganthan P N, et al. Empirical mode decomposition based ensemble deep learning for load demand time series forecasting[J]. Applied Soft Computing, 2017, 54: 246–255.

Ribeiro G T, Mariani V C, dos Santos Coelho L. Enhanced ensemble structures using wavelet neural networks applied to short-term load forecasting[J]. Engineering Applications of Artificial Intelligence, 2019, 82: 272–281.

Schleich J, Klobasa M, Brunner M, et al. Smart metering in Germany–results of providing feedback information in a field trial[J]. Proceedings of the ECEEE 2011 Summer Study, 2011:1667–1674.

Shaffery P, Yang R, Zhang Y. Bayesian Structural Time Series for Behind-the-Meter Photovoltaic Disaggregation[C]//2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT). IEEE, 2020: 1–5.

Sideratos G, Ikonomopoulos A, Hatziargyriou N D. A novel fuzzy-based ensemble model for load forecasting using hybrid deep neural networks[J]. Electric Power Systems Research, 2020, 178: 106025.

Song Z, Cao Z, Wan C, et al. An ensemble wavelet deep learning approach for short-term load forecasting[C]//2019 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia). IEEE, 2019: 1205–1210.

Sun Y, Li J, Liu J, et al. Using causal discovery for feature selection in multivariate numerical time series[J]. Machine Learning, 2015, 101(1): 377–395.

Wang Y, Chen Q, Hong T, et al. Review of smart meter data analytics: Applications, methodologies, and challenges[J]. IEEE Transactions on Smart Grid, 2018, 10(3): 3125–3148.

Xia C, Wang Z. Drivers analysis and empirical mode decomposition based forecasting of energy consumption structure[J]. Journal of Cleaner Production, 2020, 254: 120107.