Energy Proceedings
Vol 52, 2025
ISSN 2004-2965

Day-ahead prediction of household multi-appliance usage trajectory considering individual heterogeneity[#]

WU Jialin[1], GAO Jiajing[1], ZHANG Yi[1*]

1 Institute of Future Human Habitats, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, 518055, PR China

(Corresponding Author: zy1214@sz.tsinghua.edu.cn)

ABSTRACT

Demand response is currently an effective strategy to address the mismatch between electricity supply and demand. In residential buildings, there is a wide variety of household appliances with different usage characteristics. It is necessary to be aware of occupants' daily appliance usage behavior in order to better manage home energy system. In this study, a home energy consumption tracking application is developed for a longitudinal survey based on self-monitoring, during which household appliance usage data of 166 users in Shenzhen are collected. Then day-ahead prediction model for appliance usage trajectories is established. CNN+GRU is used to identify appliance usage characteristics as well as the logical relationships among different appliances. In addition, due to the limited amount of data, this paper utilizes random forest regression for day-ahead prediction for multi-appliance usage states based on occupants' social attributes, weather parameters and home awake state. It is shown that the predicted multi-appliance usage trajectories are more accurate and logical to user behavior. The results provide a reference for the incentive recommendation mechanisms of different appliances under demand response.

Keywords: household appliances, panel study, feature extraction, trajectory prediction

1. INTRODUCTION

With the development of smart grid technology, demand response has become a key component of power system management, which is used to balance the load on the grid by encouraging users to use less electricity when the supply is tight or more electricity when the supply is sufficient[1]. For residential buildings, household appliances are diverse with different usage characteristics because of their flexible adjustment based on time shifting or power shifting[2]. Thus they have a very large potential for flexible adjustment when confronted with dynamic actual demand. At present, the photovoltaic storage direct flexible buildings are developing vigorously. It becomes significant for day-ahead prediction of appliance usage based on occupants information, weather parameters and other data. Then energy management system can make accurate regulation based on the prediction of household appliances and PV power generation curve. Therefore, photovoltaic energy self-consumption reaches maximum so as to reduce grid withdrawal.

Large amount of literature are studied on building energy consumption prediction[3,4] while the granularity is accurate up to sub-consumption prediction, and it is difficult to predict certain appliance usage. Compared with aggregation-level prediction, appliance-level prediction has not received much research attention. However, in recent years, a number of works have been published exploring this area[5,6], especially due to the need to consider the stochastic nature of user behavior and external factors that affect the consumption patterns of individual appliances. Some scholars[7] have applied LSTM deep neural networks for day-ahead prediction of different household appliances such as TV, refrigerator, laptop, electric heater, stove and microwave oven. Some research[8] has also proposed an LSTM-based sequence-to-sequence model that uses historical data from the past day to predict energy consumption for one hour at a 10-minute resolution. The proposed model was evaluated for several appliances in four households, including dishwashers, glass washers, televisions, and refrigerators. The results show that the proposed LSTM-based sequence-to-sequence model outperforms other techniques, including convolutional neural networks and 2-layer LSTM networks. However, these studies aimed to predict the future state based on the historical data of appliances, which does not take into account the variability of users' habits. In addition, they have ignored that there may be interactions between appliances, for example, there is a high likelihood that a

[#] This is a paper for the 16th International Conference on Applied Energy (ICAE2024), Sep. 1-5, 2024, Niigata, Japan.

hair dryer will be used after the water heater has been used.

Therefore, it is necessary to study the appliance usage behaviors of different users under various living habits, and finally make day-ahead predictions of multiple appliances so as to evaluate the potential for flexible regulation at the household scale.

2. METHOD

2.1 Panel study

Existing methods for collecting residential energy consumption *behavior* data are primarily based on questionnaire surveys, which can lead to discrepancies between user recall and actual conditions, and also fail to explore the interactions between appliances. With the advancement of technology, researchers have started to explore more refined monitoring using smart devices, such as smart plugs or cameras. However, due to privacy concerns, these methods are not easily promoted for large-scale deployment. Therefore, to ensure the accuracy of users' daily appliance usage behavior while facilitating large-scale surveys, self-monitoring tracking was chosen for data collection in this study. A household energy consumption tracking application was developed for this research, allowing users to record the on and off status of each appliance in real-time on their mobile phones.(The interface can be seen in Fig.1) This survey was conducted from April to July 2024 and recruited 166 participants totally after rigorous screening, each of whom choose 2-4 days to fully record their daily routines and appliance usage status.

2.2 Data cleaning

For appliances which are frequently and briefly operated within a certain period, such as toilet lights and induction cookers, they are processed as being in continuous use during that time. Then, the state of each appliance is converted into a binary format with a 15-minute precision, where on state is marked as 1 and off state as 0. Therefore, the usage trajectory of each appliance during the whole day is represented as a binary list with a length of 96. In addition, user’s daily wake-up time, departure time, return home time, and bedtime are also processed into binary lists, serving as indicators of home awake state.It is worth noting that home awake rate denotes the condition where occupants are present at home and not engaged in sleep (designated as '1' when the condition is fulfilled, '0' otherwise).

Fig. 1 Energy-Tracker APP user interface

2.3 DL autoencoder-based feature extraction

In this study, it is necessary to predict the on/off status of 8 types of appliances simultaneously (including living-room light, bedroom light, TV, rice cooker, fan, washing machine, water heater and bedroom air-conditioner). Therefore, each data is presented in the form of a 96 $\times$ 8 two-dimensional binary array, where 96 denotes the number of time steps, and 8 represents the types of appliances at each step. The challenge is to extract key features from high-dimensional time series data to support subsequent prediction tasks. To address this issue, we employ two-dimensional convolution, max pooling, and Gated Recurrent Units (GRU) for feature extraction. The two-dimensional convolutional layers effectively capture local spatial features within the time series data, while the pooling layers reduce the spatial dimensions of the features, distilling the essential characteristics. The GRU layers further extract

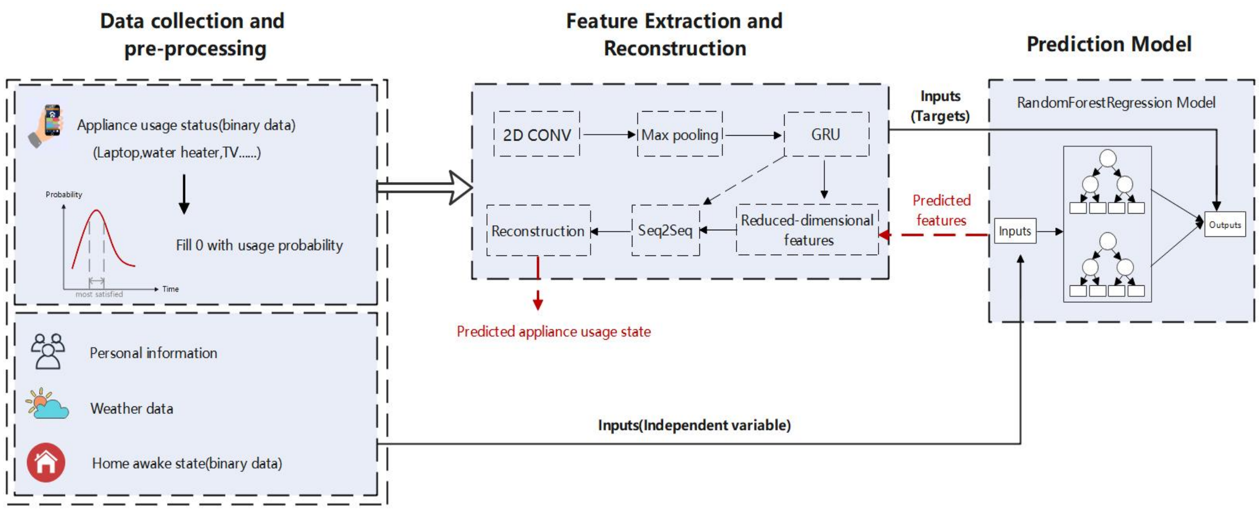

Fig. 2 Research framework

key dynamic features from the time series data. Compared to traditional Recurrent Neural Networks (RNN), GRU is more adept at handling long-term dependencies and exhibit less temporal lag.

Given the requirement for dimensional reduction and feature extraction in this study, upsampling followed by deconvolution for data reconstruction[9] is not suitable. Therefore, based on the features extracted before and the key temporal features captured by the GRU, the Seq2seq approach is employed to reconstruct the original data on a per-timestep basis.

Moreover, due to the sparsity of elements in the two-dimensional data, the design of the loss function lays more emphasis on the appliances which used less frequently used in order to enhance the accuracy of predictions. The formula for the loss function is as follows.

$$loss = \sum_a w_a \sum_t (y_{t,a} - \hat{y}_{t,a})^2 + r \cdot \frac{1}{at} \sum_a \sum_t (|y_{a,t}|) \quad (1)$$

Where a means types of appliance; t means time step with 15-minute precision; y denotes real data and $\hat{y}$ is the construction data; w is the penalty coefficient, used to focus on the reconstruction results of key appliances; r is the sparsity coefficient, which helps to avoid prolonged predictions due to the short usage time of some appliances.

In the reconstruction model, 80% of the data is used for training, and 20% for testing. The sparsity coefficient is set to 0.2, the learning rate is 0.005, and the number of training epochs is 20000.

2.4 Random forest regression and reconstruction

Due to the constrained sample size (373 totally), the Random Forest algorithm was employed for prediction model. The characteristic parameters include workday/rest day, gender, age, educational background, job, annual income, own/rental house, layout, number of household occupants, residential population composition, building area, outdoor temperature, and home awake state. The label value are the aggregated data after feature extraction of multiple appliances. In this study, 80% of the data set was allocated for training, with the remaining 20% reserved for testing. Furthermore, the hyper-parameter for the number of trees, n_estimators, was assigned a value of 100. The predicted feature data were then reconstructed with *DL autoencoder-based* model to restore the original data configuration. Thus, multi-appliance usage trajectory can be obtained considering individual heterogeneity. Figure 2 illustrates the framework of this study.

3. RESULTS AND ANALYSIS

3.1 Distribution of main electrical appliances usage

Due to space limitations, Figure 3 displays the probability distribution of some typical appliances usage throughout the day. It can be observed that the usage probability of some appliances exhibits a unimodal pattern, with the peak occurring mainly in the evening; kitchen appliances such as induction cookers and rice cookers show a bimodal pattern, with peaks at noon and in the evening; for desktop computers, a trimodal pattern is presented, with peaks appearing in the morning, afternoon, and evening, and the peak gradually increasing.

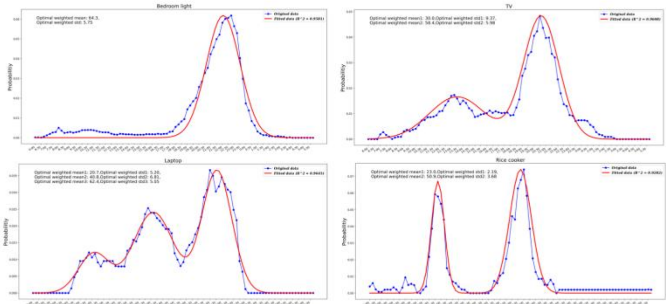

Fig. 3 Probability distribution of some appliances

3.2 Analysis of feature extraction and reconstruction

Dimensionality reduction for feature extraction can lead to loss of information during reconstruction. However, too many feature dimensions can cause "curse of dimensionality" in later predictions, easily resulting in overfitting. To resolve this dilemma, this study first attempted to extract features with different dimensions, considering reducing the number of feature to 16, 32, 48 and 64. Figure 4 shows the training loss curve with 16 feature dimensions. A comparison between the original data and reconstructed data is made, as shown in Figure 5. R-squared and MSE are calculated, with the results presented in Table 1.

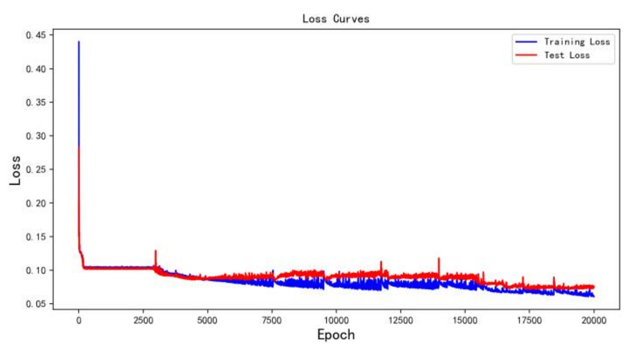

Fig. 4 Training and testing loss curves

Table 1 Reconstruction performance of various dimensions

Dimension	R-squared	MSE
16	0.60	0.036
32	0.61	0.02
48	0.57	0.031
64	0.65	0.016

It can be observed that as the feature dimension increases, reconstruction performance does not significantly improve. It is quite challenging to reduce a high-dimensional array of 96*8 to a dimension of 20-50. Through subsequent adjustments, it was found that when the dimension is reduced to 128, reconstruction performance significantly improves, while it is not conducive to making prediction. Therefore, this study extracts 16-dimensional features. Figure 5 shows the real appliance usage trajectory of one occupant in the test set and the reconstructed trajectory based on 16-dimensional features.

Overall, the reconstruction performance is acceptable. We calculated the R-squared values between original data and construction data for different appliances: living-room light is 0.47, the bedroom light is 0.51, television is 0.17, rice cooker is 0.04, electric fan is 0.32, washing machine is 0.84, water heater is 0.9 and bedroom air conditioner is 0.73. It can be seen that reconstruction information of television and rice cooker is completely lost, mainly because not many users have used these two appliances among 373 samples, thus the reconstruction model is not well trained. In contrast, the majority of users are using these appliances such as washing machine and water heater and the reconstruction effect is quite good although the usage time is not long.

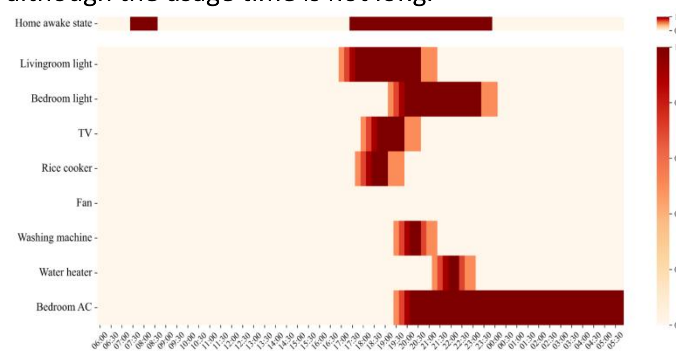

Fig. 5(a) Original data

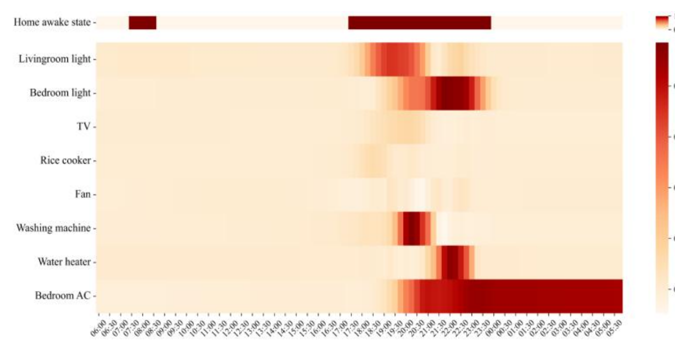

Fig. 5(b) Reconstruction data

3.3 Analysis of prediction

The predictive feature parameters including outdoor temperature and home awake state are also multi-dimensional arrays. Therefore, typical parameters need to be extracted before prediction. In this study, average temperature and standard deviation of temperature throughout the day are used as features for outdoor temperature. The number of consecutive occupancy instances, the first occupancy time and the duration of occupancy are used as features for home awake state.

The R-squared of the prediction model is 0.55. For each appliance, the R-squared results are as follows: living-room light is 0.27, the bedroom light is 0.34, television is -0.08, rice cooker is -0.13, electric fan is 0.17, washing machine is 0.42, water heater is 0.28 and bedroom air conditioner is 0.30. Figure 6 shows the predicted data from the same user in Figure 5. It can be seen that prediction effects of living-room light and bedroom light are acceptable mainly because the usage behavior of these two types of appliances is likely to depend on user's returning home time and bedtime. Appliances such as televisions and rice cookers may not be suitable for unified prediction. To solve this problem, we use probability distribution (Figure 3) to replace the predicted results for televisions and rice cookers combined with home awake state.

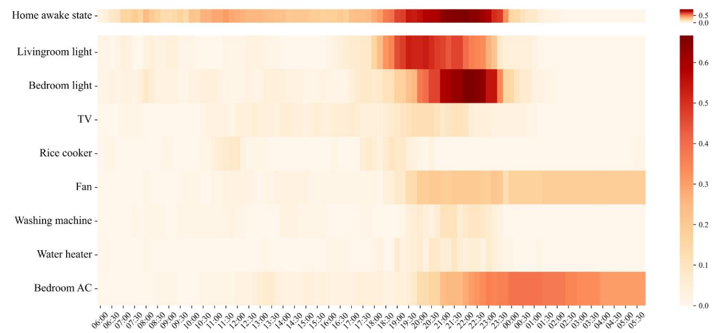

Fig. 6(a) Distribution of real data

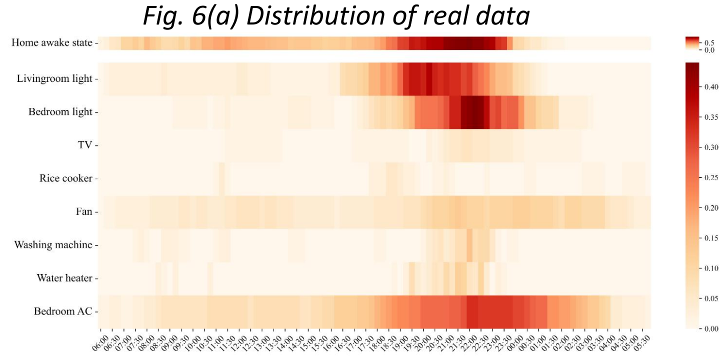

Fig. 6(b) Distribution of predicted data

In fact, it is not quite appropriate to measure the accuracy of appliance usage trajectories by using R-squared, since it is not necessary to make accurate prediction for the usage state of each 15-minute interval. The appliance's usage period and duration should be as close to the actual situation as possible. Therefore, this study employs Dynamic Time Warping (DTW) technology to evaluate the accuracy of time series data prediction, which calculates the similarity between two time series by Euclidean distance[10], with a smaller distance indicating greater similarity. To validate the effectiveness of the method proposed in this paper, conventional sampling methods are also used to obtain the usage time of each appliance. Figure 7 presents the DTW distances for each appliance using both methods, demonstrating that the CNN+GRU prediction approach significantly outperforms the sampling method.

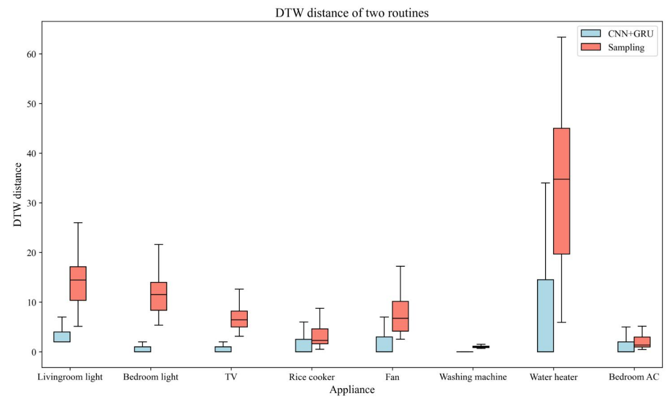

Fig. 7 Dynamic time warping distance

3.4 Analysis of feature importance

Figure 8 illustrates the importance of various factors on the prediction results, where home awake state, outdoor temperature and building area are the three most significant factors. Regarding the users' social attributes, education background, annual income, and number of occupants may slightly influence users' appliance usage habits.

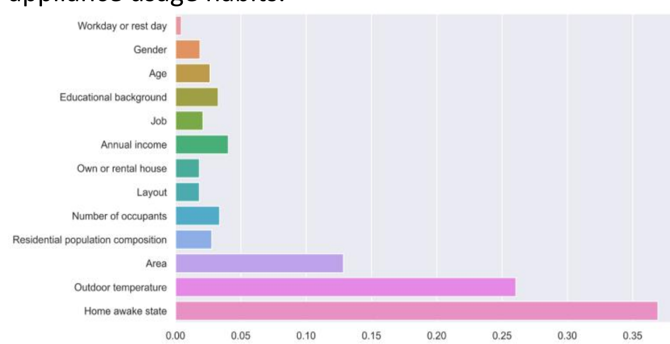

Fig. 8 Feature importance

4. DISCUSSION

Large amount of literature on energy consumption behavior prediction are based on extensive historical data with limited objects while few studies have explored the impact of user heterogeneity on appliance usage patterns[11]. This study endeavors to identify disparities in appliance usage habits across diverse populations, thereby providing a basis for precise and flexible appliance regulation. The results in this study indicate that the prediction accuracy is not high. Those appliances with lower usage frequencies or with short usage duration cannot be accurately predicted, primarily due to the difficulty in data acquisition and the inherent randomness of appliance usage.

However, this study introduces a probability-driven multi-appliance prediction method that addresses user heterogeneity. Future surveys can refine this survey, integrating devices such as smart plugs to monitor the status of appliances and expanding the number of subjects surveyed. In terms of predictive evaluation metrics, the R-squared might not be entirely suitable because appliances trajectories have randomness. Therefore, it is necessary to explore fuzzy evaluation indicators. Moreover, other forecasting methods could be considered, such as time-rolling forecasting.

5. CONCLUSIONS

In this study, full-day multi-appliance usage trajectories through panel studies were obtained. Eight types of appliances were selected to establish a prediction model considering individual heterogeneity. The conclusions are as follows:

1. The dimensionality reduction and feature extraction of multi-appliance trajectories by CNN+GRU is acceptable, with R-squared of 0.61 for the reconstructed data. However, the reconstruction performance is not acceptable for those appliances with low usage frequency or short usage time.
2. The prediction model has an R-squared of 0.55, where home awake state, outdoor temperature, and building area have the greatest impact on the prediction results. In terms of user social attributes, education background, annual income, and the number of occupants have a certain influence on appliance usage habits.

This model can provide theoretical support for day-ahead forecasting of flexible potential in residential buildings, offering data support for flexible regulation of electricity loads.

ACKNOWLEDGEMENT

This research was supported by Shenzhen Science and Technology Program KCXST20221021111608020.

REFERENCE

[1] Guo, L. , Tan, Z. , & Li, X.. (2006). Demand response based model and method for optimal design of time-of-use electricity price. Power System Technology, 30(5), 24-28.

[2] Ghasem Azar, A. , & Jacobsen, R. H.. (2016). Appliance Scheduling Optimization for Demand Response.

[3] Natarajan, Y. , Sri, P. K. R. , Wadhwa, G. , Choi, Y. , Chen, Z. , & Lee, D. E. , et al. (2024). Enhancing Building Energy Efficiency with IoT-Driven Hybrid Deep Learning Models for Accurate Energy Consumption Prediction.

[4] Berriel, R. F. , Lopes, A. T. , Rodrigues, A. , Varejao, F. M. , & Oliveira-Santos, T.. (2017). Monthly energy consumption forecast: A deep learning approach. 2017 International Joint Conference on Neural Networks (IJCNN). IEEE.

[5] Andrade, V. , Morais, H. , & Pereira, L. . (2024). Industrial kitchen appliance consumption forecasting: hour-ahead and day-ahead perspectives with post-processing improvements. Computers and Electrical Engineering, 116.

[6] Werthen-Brabants, L. , Dhaene, T. , & Deschrijver, D.. (2022). Uncertainty quantification for appliance recognition in non-intrusive load monitoring using bayesian deep learning. Energy and Buildings.

[7] Basu, K. , Hawarah, L. , Arghira, N. , Joumaa, H. , & Ploix, S.. (2013). A prediction system for home appliance usage. Energy&Buildings, 67(dec.), 668-679.

[8] Razghandi, M. , Zhou, H. , Erol-Kantarci, M. , & Turgut, D.. (2021). Short-Term Load Forecasting for Smart Home Appliances with Sequence to Sequence Learning. ICC 2021 - IEEE International Conference on Communications. IEEE.

[9] Kim, J. , Song, K. , Lee, G. , & Lee, S. H.. (2024). Time-series data clustering with load-shape preservation for identifying residential energy consumption behaviors. Energy & Buildings, 311.

[10]Ardan Hüseyin Elik, Akarslan, E. , & Doan, R. . (2024). A novel approach for residential load identification based on dynamic time warping. Sustainable Energy, Grids and Networks, 38.

[11] Lei, M. , Cai, W. , Liu, W. , & Wang, C.. (2022). The heterogeneity in energy consumption patterns and home appliance purchasing preferences across urban households in china. Energy, 253.