Machine Learning Applications in Residential Electricity Data: Forecasting, Clustering, and Synthesis

アヤス マヘル アブドゥルラハマン シャクール

https://hdl.handle.net/2324/7329522

出版情報：Kyushu University, 2024, 博士（工学）, 課程博士
バージョン：

（様式３） Form 3

氏　　名　：Ayas Mahr Abdelrahman Shaqour

論 文 名　：Machine Learning Applications in Residential Electricity Data: Forecasting, Clustering, and Synthesis

(時系列住宅電力データへの機械学習の適用：予測・クラスタリング・データ生成)

区　　分　：甲

論　文　内　容　の　要　旨

Thesis Summary

The residential building sector consumed almost 21% of the global energy demand and contributed to 11% of indirect and 6% of direct global CO2 emissions related to energy in 2022. With such high impacts of both the energy and environment, large investments of 285 billion USD are being made for improving building energy efficiency worldwide. One of the recent timely opportunities for innovations is the wide and fast spread of smart metering infrastructure, providing highly granular data that could be leveraged to improve smart grid functionalities, such as better forecasting of future demands, gaining more insights on user behaviors for pricing policies, enabling demand response mechanisms, and achieving zero-energy housing, through data-driven energy management. While many studies have analyzed various smart meter data from various perspectives, this work utilizes a unique data set of 479 dwellings, with not only the total household demand at the main meter level but also electrical demands of appliances and dwelling spaces (lighting and outlets) as well as water and gas demand. The main question of this thesis is to realize how such a large and unique dataset can be leveraged using various methods mainly focusing on machine learning techniques, to contribute to enabling smarter buildings from the perspective of 3 main paradigms: (1) short-term load forecasting, (2) clustering user behaviors, and (3) data synthesis.

Chapter 2 describes the details of the data set, relating to the building complex, the number of dwellings, as well as the various appliances and dwelling spaces energy demands. It also describes the data aggregation process, the five different aggregation levels, and the produced subsets that are used for the load forecasting analysis.

Chapter 3 comprehensively investigates the accuracy of short-term load forecasting (STLF) of five different demand aggregation levels (3, 10, 30, 100, and 479) based on the 479 residential dwelling data set, with a sample size of (159, 47, 15, 4, and 1) per aggregation level, respectively. It also discusses the underlying challenges in lower aggregation forecasting. Five deep learning (DL) methods are utilized for STLF and fine-tuned with extensive methodological sensitivity analysis and a variation of early stopping, where a detailed comparative analysis is conducted. Furthermore, the model input parameters were selected methodologically based on auto-correlation analysis, weather variables sensitivity analysis, and assessment of the recent literature. The 5 classes of state-of-the-art DL methodologies were assessed comprehensively, where they had highly competitive results due to comprehensive fine-tuning and early stopping, where the deep neural network (DNN) achieved the highest mean absolute percentage errors (MAPEs) for most of the aggregation levels and the Bi-GRU-FCL showed lower root mean squared errors (RMSEs) and had 15% faster-training speed and 40% fewer parameters in comparison to the DNN, which were investigated further for more optimal tunning and increased accuracy. Furthermore, it was assessed that the temperature weather variable had the highest impact on the model accuracy when used without the others. The final results exhibited extremely low MAPEs of 2.44% and 2.47% for the validation and test datasets,

respectively, up to 3.31% on the worst test case for the highest aggregation with 479 dwellings. The results were close to country-level prediction errors and it was found that errors below 9% can be sustained at an aggregation level of 30. In comparison to recent research that suggested 150 to stay below 10%, which indicates that more factors should be considered such as the type of appliances used and levels of electrical demand used by the aggregated houses. Furthermore, the increases in errors with lower levels of aggregation were discussed in detail, which is a challenge for future smart grids with increased penetration of decentralized and renewable energy generation sources. This is important to sizing hybrid renewable energy systems (HRES) or microgrids that need to achieve a supply–load balance at local levels and depend on high STLF accuracies to achieve optimal operational performance.

In Chapter 4 the relationships between the 18 behind-the-meter energy demand attributes (BTMEDAs) and the global demand (GD) data were explored by utilizing the clustering technique. First, a clustering method was applied to the 2-year 24-hour average of the data to unveil the four representative energy consumption behaviors for each of the BTMEDAs and GD. Next, the seasonal change of such behaviors across the selected seasonal categories: hot (summer), cold (winter), and thermally neutral (spring and autumn) were calculated. This analysis revealed how GD behaviors change across seasons, where high-energy consumers can change across seasons. It also highlighted how much energy consumption behaviors are based on many other smaller behaviors related to each appliance and dwelling space. This provides data-driven insights beyond assumptions and standardized data that could be extremely useful for detailed simulations. The second part of this study aimed to quantify the correlation between the BTMEDAs and GD clusters across the different seasons using a multi-method approach to capture their correlation and understand which BTMEDAs dictate GD behavior for every season. This analysis clarified that the energy consumption of the living-dining room outlet and AC are the highest predictors from the BTMEDAs of the GD consumption behavior-levels. Furthermore, the attributes that are indirectly related to the number of dwellers, such as the hallway and WC lighting/outlet and Refrigerator, are also important, although that can change based on the seasons. Finally, the cluster-encoding approach could preserve the correlation information between the GD and BTMEDAs and was a valid approach for decreasing the data dimensionality of the analysis while maintaining crucial information.

Chapter 5 investigates the application of the latest generative adversarial network (GAN) model called the DoppelGANger (DGAN) for synthesizing electrical demand data. The analysis identified a strategy for synthesizing a full 2-year hourly data for a dwelling's total demand, training sub-models on monthly data to capture the short-term time series features, and then aggregating the sub-models results to create the full-time series, to reproduce the seasonal demand features. The model was evaluated for dwellings with different types of demand behaviors based on the clustering results. The data synthesis using the DGAN model was successful in replicating various characteristics of the real demand data such as the overall distribution, auto-correlations, as well as the hourly and monthly demand distributions. However, it was less accurate in learning the weekday features. Furthermore, a thorough sensitivity analysis of the hourly percentile deviation for 5 different percentile ranges (<10%,10-25%,25-75%, and >90%) between the real and synthesized data was conducted. Then the within (5%,10%,20% and 50%) percentile errors were calculated. The results showed that the model makes more errors in the lower 10% percentiles overall while doing better in the middle percentile ranges. And that higher-demand dwellings have more unique distributions that are harder to learn. Although the results were very promising for further analysis.

Finally, Chapter 6, concludes the thesis and discusses the future potential extensions of this work.