# TOWARDS A UNIFIED STANDARDS FOR SMART INFRASTRUCTURE DATASETS

**Abdulrahman A. Ahmed**
Faculty of Computers and Artificial Intelligence
Cairo University
5 Dr. Ahmed Zewail street, Orman, Giza, Egypt. Postal code: 12613
{a.elsayed}@fci-cu.edu.eg

## ABSTRACT

Development of smart devices and smart home appliances allowed us to harness more data about energy patterns inside households, overtime this amount will increase. There are contributions published to address building datasets, working for objective of energy consumption optimization. Yet there are still factors if included could help in understanding problem better. This proposal tries to annotate missing features that if applied could help in a better understanding energy consumption in smart buildings impact on environment. Second, to have a unified standards that help different solutions to be compared properly.

## 1 INTRODUCTION

Buildings energy consumption is accounted for a considerable percentage from total energy consumption on electric grid. For example, in US, commercial and residential energy consumption accounted for around 40% of total energy consumption. Meanwhile in UK, Germany residential sector accounted for 28%, 26%, of total energy consumption in each country respectively Iskandarova & Genus (2019),Grealis et al. (2019). The increase of electronic devices allowed more data to be gathered about power consumption especially within residential buildings. Though different data sources have been gathered, still there are missing points to have a machine learning models work better for optimizing energy consumption. and lacking for agreed standards in optimizing energy consumption problem makes the problem more harder to attack and solution methods will not appear on the same basis. Pecan street Inc. dataset collects data for residential area it includes: type of rooms, number of windows, availability of some devices, etc. and such diversity enabled different contributions in this area. However, it lacked conclusive indicators for weather conditions which play important rule in energy consumption and other environmental impact to electric devices. Social game dataset Konstantakopoulos et al. (2019) addressed the problem form a little different perspective as it recorded unit features with indicative features for weather conditions. Still we can't have full picture of residential energy consumption behavior as it's a residential building inside university campus. In addition to points mentioned, neither of both datasets addressed energy sources (i.e. how much solar power, wind power, etc. contribute to the electric grid). This proposal tries to address gaps (between datasets) that could help in better understanding of energy consumption in smart infrastructure and more unified standards in comparison.
In the following subsections I'll discuss briefly main previous datasets that are strongly related to proposal.

### 1.1 PECAN STREET INC. DATASET

A dataset collects different appliances data minute by minute from residential buildings (mainly Austin, TX). Dataset shows how much each component consumes over the time domain. Dataset also records how is the size of house, house construction, number of windows, etc. Dataset also can be combined with records from time of use energy tariffs for energy consumption like used in Mocanu et al. (2018). Where they implemented two Deep Reinforcement Learning (DRL) methods to solve two problems: Reduce load peaks, and minimize cost of energy.

### 1.2 SOCIAL GAME DATASET

Dataset in Konstantakopoulos et al. (2019) is collected from residential housing of Nanyang Technological University. Dataset is comprised from a record of minute by minute for different possible energy resources including desk light, room light and ceiling fan/air conditioning (a.k.a Heating, ventilation, and air conditioning (HVAC)) while labeling in data important times for students like break, midterm , or final exam.
The authors extended the work and built web portal so user can interact with his own info about energy consumption and do actions. While those who can reduce their energy consumption get rewarded. This gamification process allowed to record both the electronic devices behavior and the human behavior within a non-cooperative game theory framework.

## 2 PROPOSAL

Park et al. (2019) in an effort to define a load shape (e.g. either residential, commercial, etc.) gathered different data sources in their work. I'll mention in the following points what I think aimed dataset must contain both for optimizing energy consumption and dealing with climate change:

- Structure of main energy sources within the smart-grid unit(where unit is: city, neighborhood, etc.): how much percentage it's generated from fossil fuels, solar energy, wind farms, etc. To assess energy consumption harm to environment. As a big consumption sourced mainly from renewable energy isn't a threat as compared relatively to a small consumption but from fossil fuels.
- Unified energy time step: either it's second, or minute, etc. Surely it's better to have data by seconds but it may be hard due to different technical issues but investigating migration to minutes will cause feature loss for dataset is necessary (especially for Artificial Neural Networks (ANN) models)
- Water consumption as an auxiliary data and relation with appliances like heater, dishwasher, clothes washing machine, etc.
- Considering other environmental impact for electrical devices like consumed water and the effect on clean water resources, Chlorofluorocarbon (CFC) emissions from refrigerators, HVAC, etc.
- Augmenting data gathered from personal mobile phones or wearable to predict better individual needs and actions. It'll help indicate better timing of consumption for each unit or predicting actions like turning on HVAC ( Liu et al. (2019) showed potential for that)

Many of mentioned points could be gathered from different devices or sources yet not included (to my knowledge) in dataset for optimizing energy consumption. Still there are many limitations to structure those points. Like risk of compromising personal privacy, or quantifying impacts other than energy consumption for electrical devices (like polluted water).

## 3 CONCLUSION

While there is an advancement in building datasets for smart infrastructure. There are still missing points to have a better understanding of energy consumption and impact on environment. In this paper I tried to mention points that once being components in a dataset will help building better machine learning models to either understand behavior, consumption consequences or optimize not only energy consumption but also other environmental impacts for electrical devices and home appliances. Lastly, providing a unified standards (or at least a group of consistent points) will lead to better understanding of the problem and clear comparison for different machine learning methods.

### ACKNOWLEDGMENTS

I would like to thank Hari Prasnna Das for his great help in writing this proposal. And CCAI ICLR mentorship commitee to enable me this such useful opportunity.

## REFERENCES

E. Grealis, A. Musch, and H. Rau. *The Impact of German Energy Policy on Household Energy Use*, pp. 21–33. Springer International Publishing, Cham, 2019. ISBN 978-3-030-20339-9.

M. Iskandarova and A. Genus. *UK Responses to the Energy Challenge: Dominant Framings and New Imaginaries*, pp. 59–69. Springer International Publishing, Cham, 2019. ISBN 978-3-030-20339-9.

I. Konstantakopoulos, H. Das, A Barkan, S. He, T. Veeravalli, H. Liu, A. Manasawala, Y. Lin, and C. Spanos. Design, benchmarking and explainability analysis of a game-theoretic framework towards energy efficiency in smart infrastructure. *ArXiv e-prints*, 2019.

S. Liu, S Schivano, H. P. Das, M. Jin, and C. J. Spanos. Personal thermal comfort models with wearable sensors. *Building and Environment*, 2019.

E. Mocanu, D. C. Mocanu, P. H. Nguyen, A. Liotta, M. E. Webber, M. Gibescu, and J. G. Slootweg. On-line building energy optimization using deep reinforcement learning. *IEEE Transactions on Smart Grid*, pp. 211–229, 2018.

J. Y. Park, X. Yang, C. Miller, P Arjunan, and Z. Nagy. Apples or oranges? identification of fundamental load shape profiles for benchmarking buildings using a large and diverse dataset. *Applied Energy*, pp. 1280–1295, 2019.