Data science for building energy management: A review

M. Molina-Solana*, M. Ros†, M.D. Ruiz†, J. Gómez-Romero† and M.J. Martin-Bautista†
*Data Science Institute, Imperial College London, United Kingdom
Email: m.molina-solana@imperial.ac.uk
†Department of Computer Science and Artificial Intelligence, Universidad de Granada, Spain
Email: marosiz@ugr.es, mdruiz@decsai.ugr.es, jgomez@decsai.ugr.es, mbautis@decsai.ugr.es

Summary of published paper: M. Molina-Solana, M. Ros, M. Dolores Ruiz, J. Gómez-Romero, M.J. Martin-Bautista, *Data science for building energy management: A review*, **Renewable and Sustainable Energy Reviews**, Volume 70, 2017, Pages 598-609, ISSN 1364-0321, doi: 10.1016/j.rser.2016.11.132. IF: 9.184. (original source)

***Abstract*—The energy consumption of residential and commercial buildings has risen steadily in recent years, with expected energy loads, transportation, storage and user behaviour influencing the quantity and quality of the energy consumed daily in buildings. Technology is now available that can accurately monitor, collect, and store the huge amount of data involved in this process, as well as analyzing and exploiting such data in meaningful ways. Our paper [1] reviewed how Data Science has been applied to address the most difficult problems faced by practitioners in the field of Building Energy Management. It also discussed the challenges and opportunities that will arise with the advent of fully connected devices and new data processing technologies.**

I. Introduction

Residential and commercial buildings are currently responsible for up to 40% of the energy consumption in the world [2]. Increasing energy efficiency is a two-fold process. Not only does it involve the use of affordable energy sources, but also the improvement of current energy management procedures and infrastructures. The latter includes the optimization of energy generation and transportation based on user demand [3]. In this regard, computer-aided approaches have recently come into the spotlight. In the area of building energy management, Data Science is now used to address problems in different areas, ranging from energy demand prediction to operation optimization and energy fraud detection.

Our paper [1] analyzes the application of Data Science methods to these problems. Although there are other reviews on automatic techniques for building efficiency assessment [4], [5] and classification methods for load and energy consumption prediction [6], this work examined and discussed a broader set of data science techniques and their applications to the different aspects of building energy management. We present here a brief summary of the most relevant techniques identified by our study, some reflections on the possible evolution of the area in the near future and the main conclusions of the work.

II. Applications of Data Science for Building Energy Management

Prediction of building energy load. Energy demand, or energy load, refers to the amount of energy required at a certain time period. Detecting building load patterns can be complex because of the number of interrelated factors. In the literature, we can find different proposals for customer profiling using unsupervised learning, e.g. clustering and association rules, and classification models, such as decision trees. For peak demand anticipation, anomaly detection techniques have been prevalent. Users' behavior and their impact in energy consumption has deserved special attention, and there are some works aiming at activity recognition for automatic energy management (heating/cooling, lighting) and load shifting.

Building operation. Operational data acquired by the building energy management system can be analyzed and exploited to extract patterns describing building operation. In most cases, these patterns are in the form of IF-THEN rules, which can help to generate recommendations for control strategies. Decision trees and association rules have been mostly used in this regard. Similarly, other studies have focused on the improvement of specific components of the building by investigating the relations between control settings and energy usage; e.g. by using regression and neural networks.

Analysis of infrastructures. The latest energy efficiency regulations have fostered the study of architectural aspects related to energy use of new and existing buildings. Clustering algorithms have been applied to discover non-obvious factors of energy loss in building infrastructures. Similarly, classification methods have been applied to estimate relevant indicators of thermal behavior and to support decision-making for retroffiting. Some of these works consider time series analysis to enhance the operation of the infrastructures by extracting trends and episodic relationships.

Economic analysis of electric consumption Numerous companies have resorted to Data Science in an effort to discover and understand how and when their customers use energy. The techniques traditionally used for this task are classification —to identify common consumption categories—, clustering —for segmenting buildings by their use of the energy—, and pattern analysis —mostly by means of association rules.

Streaming data processing has been also considered for real-time analysis and summarization of energy consumption data.

Energy fraud detection Sometimes, energy consumption and services are not appropriately billed because of failures in the measurement equipment. Such failures can either be accidental or the product of fraudulent manipulation. These deviations are commonly referred to as non-technical losses (NTLs), and different techniques have been successfully applied to distinguish legit and fraudulent operation; e.g. tree-based classification, expert systems, association rule extraction and anomaly detection.

III. Discussion

As extensively discussed in our original paper, the techniques that enable an easy visualization of results, such as classification methods or association rules, have been widely used because they are intuitive and easy to understand. This sometimes leads to a sub-optimal selection of techniques for specific problems (e.g. tree-based classification is the most used method by large), or even to a misuse of the techniques (e.g. clustering instead of classification despite the availability of labelled data). In general, the best results have been achieved by using ensambles. Less used techniques often involve processing streaming data, which are more difficult to process with out-of-the-box algorithms.

As a matter of fact, our revision found that the most frequently used data science algorithms in the energy field are those found in commercial toolboxes, such as *Intelligent Miner*, *SPSS*, *Clementine*, *RapidMiner* and *SAS*. Open platforms, such as *WEKA* and *KNIME*, and programming frameworks, such as *Matlab*, *R* or *Python* are not very popular, because they require specific technical skills and knowledge of the underlying algorithms. Still, they would deserve more attention from the community. We also believe that it is necessary to incorporate to building energy management solutions capabilities for managing big data, particularly by relying on the Apache Spark/Hadoop ecosystem.

IV. Current Trends and Future Challenges

Smart metering and Internet of Things. Smart metering is the continuous monitoring of energy consumption with a view to gaining a better understanding of the energy consumption, generation, and transportation stages. As a result, bills are no longer an aggregated value of energy consumption over a long period of time; they are obtained from real-time measurements. This also enables users to better understand their own energy behavior, which can help to reduce energy waste, and encourages users to modify underperforming habits. It is expected that smart meters, as well as other appliances, will soon be connected to the Internet of Things, and equipped with additional intelligent features. At the same time, device hyperconnection brings security and privacy concerns: a hacker could remotely control the devices and get private details of a home (number of occupants, daily routines, etc.)

Big Data and Cloud Computing. The huge amount of data generated by building energy equipment can be exploited to increase energy efficiency, and particularly, smart meters data. This information requires smarter processing, as well as larger and more flexible computational infrastructures. Cloud computing, offering flexible and scalable computational resources, can help to implement Big Data solutions for energy management. It is thus not surprising that large technology corporations are taking an interest in this field, either for the development of energy management solutions for the industry (e.g. Siemens) or to improve their internal energy management procedures (e.g. Google's data centers).

Privacy and security issues. New technological advances in massive data processing naturally lead to security and privacy issues. Therefore, it is necessary to develop mechanisms that ensure that the information remains under control according to the legal framework (i.e. the General Data Protection Regulation) and ethical principles. A first step has already been taken by the European Commission regarding smart grids and smart metering in their "Recommendation on the Data Protection Impact Assessment Template for smart grid and smart metering systems", which provides guidelines on how to support the security of the implementation of smart grids and smart metering by data controllers.

V. Conclusion

Our paper reviewed recent developments in Data Science and their influence on Building Energy Management. We examined the usefulness of data science techniques aimed at improving building energy efficiency. Given the progress and the challenges in this area, it is evident that data science techniques will play a key role in the near future.

Acknowledgment

This work was supported by the University of Granada under the Young Researchers Fellowships Programme (No. P9-2014-ING); the Spanish Ministry for Economy and Competitiveness under the project grant TIN2015-64776-C3-1-R; and the European Union under the project grant Energy IN TIME EeB.NMP.2013-4 (No. 608981).

References

[1] M. Molina-Solana, M. Ros, M. D. Ruiz, J. Gómez-Romero, and M. J. Martín-Bautista, "Data Science for Building Energy Management: A Review," *Renewable and Sustainable Energy Reviews*, vol. 70, pp. 598–609, 2017.

[2] WBCSD, "Transforming the market: Energy efficiency in buildings," The World Business Council for Sustainable Development, Tech. Rep., April 2009. [Online]. Available: http://www.wbcsd.org/transformingthemarketeeb.aspx

[3] H. Allcott and S. Mullainathan, "Behavior and energy policy," *Science*, vol. 327, no. 5970, pp. 1204–1205, 2010.

[4] Z. Li, Y. Han, and P. Xu, "Methods for benchmarking building energy consumption against its past or intended performance: An overview," *Applied Energy*, vol. 124, pp. 325–334, 2014.

[5] T. Nikolaou, D. Kolokotsa, and G. Stavrakakis, "Review on methodologies for energy benchmarking, rating and classification of buildings," *Advances in Building Energy Research*, vol. 5, no. 1, pp. 53–70, 2011.

[6] K. Zhou, S. Yang, and C. Shen, "A review of electric load classification in smart grid environment," *Renewable and Sustainable Energy Reviews*, vol. 24, pp. 103–110, 2013.