Modeling and recommendation system for improving the energy performance of buildings

Loup-Noé Lévy[1,2], Jérémie Bosom[2,3], Guillaume Guerard[4], Soufian Ben Amor[1], and Hai Tran[2]

[1] LI-PARAD Laboratory EA 7432, Versailles University, 55 Avenue de Paris, 78035 Versailles, France
[2] Energisme, 88 Avenue du Général Leclerc, 92100 Boulogne-Billancourt France
[3] EPHE, PSL Research University, 4-14 Rue Ferrus, 75014 Paris, France
[4] De Vinci Research Center, Pole Universitaire Léonard de Vinci, 12 Avenue Léonard de Vinci, 92400 Courbevoie, France

Abstract. The energy performance of building represents a major issue of the 21st century. However, buildings are complex sociotechnical systems because of the massification, heterogeneity and multiplicity of related data. Therefore, evaluating their performance and giving relevant recommendation is a challenging task. In this article, we propose a contributory algorithmic recommendation system applicable to energy performance diagnostic and prediction. Doing so requires to overcome many theoretical and technical issues related to big data infrastructure, machine learning factory, and DevOps. The thesis is done through *Energisme*'s platform. It presents a collection and deployment of algorithmic solutions for optimized energy management in private and public organizations.

Keywords: energy performance, sociotechnical systems, recommender systems

1 Problem statement

In 2019, the French government announced, in the energy renovation plan for buildings, the goal of achieving carbon neutrality by 2050. Reducing the consumption of buildings therefore represents a critical issue.

Improving the energy performance of buildings remains a complex task due to the massification, heterogeneity and multiplicity of related data (Zou et al., 2018). The visualization and analysis of the remotely collected data allows a detailed understanding of the buildings. It is therefore crucial to understand, structure and analyze the data efficiently.

The company *Energisme* presents itself as a trusted third-party for energy measurement and performance. As such, *Energisme* offers a platform for the collection and deployment of algorithmic solutions for optimized energy management in private and public organizations. The problematic proposed by *Energisme* is the following: *How to provide a collective intelligence emerging from a platform gathering algorithms and energy databases?*

This problem raises four technical and scientific challenges:

1. How to organize the storage of massive and heterogeneous data?
2. How to structure the data?
3. How to design a recommendation system based on contributory algorithmic solutions?
4. How to ensure the evolution and reliability of the platform over time?

2 Related work

Data storage, essential before any data analysis, is commonly carried out using a data warehouse. Since the incoming data can vary strongly in its form and its context, the choice of the data lake (Khine and Wang, 2018) is relevant to be more flexible, adaptable and effective than a data warehouse.

In addition to data storage, data analysis as a recommendation system is unintuitive. Specialists have specific business needs, and the data must be preprocessed before it can be utilized. This data stream is called a datamart (Moody and Kortink, 2000). It functionally groups specialized data, aggregated for a particular business.

There are few recommendation systems about energy performance. Moreover, such systems are rapidly worthless since certificates and laws depends on the country and time. The problem of a database of energy building uses, as studied by Mathew et Al. in the U.S.A. (Mathew et al., 2015), is a current challenge over the world since the problem is either a physical problem (Song et al., 2018) or a social problem (Zhang et al., 2018). Some works consider the socio-technical problem using numerical clone to understand the energy waste of a building (Gerrish et al., 2017).

3 Hypothesis & Proposal

The challenge is to define recommendation methods for data analysis. By considering the conceptual map of a mass of data, the objective is to recommend datamarts according to the business needs. This is possible thanks to machine learning coupled with the use of business expertise provided by various experts (data scientists, physicists...).

The recommendation of datamarts and specialized algorithms is only possible thanks to the exploitation of the collective intelligence provided by the business knowledge on the existing datamart. This collective intelligence will have to be reproducible and extensible, the objective being to group together different artificial intelligences (AI), to valorize them and to coordinate them in real time. According to Pierre Lévy (Lévy, 1997), collective intelligence is an "intelligence that is distributed everywhere, constantly valorized, coordinated in real time, and that leads to an effective mobilization of competences." The question to be resolved is: how can we recommend appropriate algorithms to users for the exploitation of datamart ?

Institutions using *Energisme*'s platform will be able to increase the relevance of their AI by soliciting them to more diverse datasets. The latter will be enhanced by all the AIs "browsing" within this contributing platform.

The work of the thesis is focused on three key scientific points in *Energisme*'s platform as shown in Figure 1:

1. To determine some profiles of consumers based on various data (year of construction, type of building, consumption per square meter, etc.) (Abedjan et al., 2015).
2. To extract knowledge from a profile to recommend energy performance improvements and evaluates how a building compares to its peers (Sharma and Gera, 2013).
3. To improve and update the databases and the algorithms in real time to keep in mind the evolution of the business.

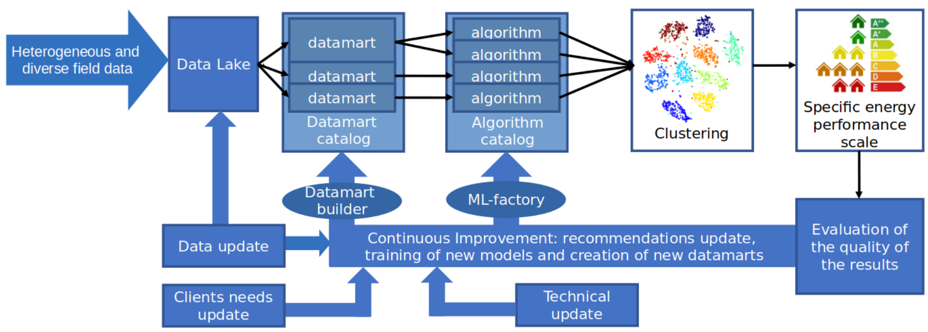

Fig. 1. *Energisme*'s platform generating specific energy performance scale.

4 Preliminary Results

We propose a first tool to study a group of sites to optimize their consumption using to recommendations done on similar sites. To do so the first step is to cluster sites into homogeneous groups. Clustering buildings is a challenging task, as buildings are described by categorical and numerical data as well as consumption time series.

To cluster such systems, we implemented an algorithm based on a theoretical framework presented in 2019 by Julio Laborde (Laborde, 2019). This framework uses the theory of pretopology to perform hierarchical multicriteria clustering.

The clustering was applied on a generated dataset of 2D disks of various sizes, presented in this recent paper (Levy et al., 2021). To evaluate the validity of the clusters determined by the algorithm, our metric was the Adjusted Rand Score also called Adjusted Rand Index (ARI). Since we perfectly identified the clusters on the generated dataset, the ARI of our clustering was 1.

5 Reflections

The three key scientific points (profiling, recommendation system, evolutive platform) must be studied and deepened to propose modern, appropriate and advanced algorithms, methods and techniques to respond to the problem of the recommendation for the energy performance of buildings.

References

[Abedjan et al., 2015] Abedjan, Z., Golab, L., and Naumann, F. (2015). Profiling relational data: a survey. *The VLDB Journal*, 24(4):557–581.

[Gerrish et al., 2017] Gerrish, T., Ruikar, K., Cook, M., Johnson, M., Phillip, M., and Lowry, C. (2017). Bim application to building energy performance visualisation and management: Challenges and potential. *Energy and Buildings*, 144:218–228.

[Khine and Wang, 2018] Khine, P. P. and Wang, Z. S. (2018). Data lake: a new ideology in big data era. In *ITM web of conferences*, volume 17, page 03025. EDP Sciences.

[Laborde, 2019] Laborde, J. (2019). la prétopologie, un outil mathématique pour la structuration des systèmes complexes: méthodes, algorithmes et applications.

[Levy et al., 2021] Levy, L.-N., Bosom, J., Guérard, G., Amor, S. B., Bui, M., and Tran, H. (2021). Application of pretopological hierarchical clustering for buildings portfolio. In *SMARTGREENS*, pages 228–235.

[Lévy, 1997] Lévy, P. (1997). Collective intelligence.

[Mathew et al., 2015] Mathew, P. A., Dunn, L. N., Sohn, M. D., Mercado, A., Custudio, C., and Walter, T. (2015). Big-data for building energy performance: Lessons from assembling a very large national database of building energy use. *Applied Energy*, 140:85–93.

[Moody and Kortink, 2000] Moody, D. L. and Kortink, M. A. (2000). From enterprise models to dimensional models: a methodology for data warehouse and data mart design. In *DMDW*, page 5.

[Sharma and Gera, 2013] Sharma, L. and Gera, A. (2013). A survey of recommendation system: Research challenges. *International Journal of Engineering Trends and Technology (IJETT)*, 4(5):1989–1992.

[Song et al., 2018] Song, M., Niu, F., Mao, N., Hu, Y., and Deng, S. (2018). Review on building energy performance improvement using phase change materials. *Energy and Buildings*, 158:776–793.

[Zhang et al., 2018] Zhang, Y., Bai, X., Mills, F. P., and Pezzey, J. C. (2018). Rethinking the role of occupant behavior in building energy performance: A review. *Energy and Buildings*, 172:279–294.

[Zou et al., 2018] Zou, P. X., Xu, X., Sanjayan, J., and Wang, J. (2018). Review of 10 years research on building energy performance gap: Life-cycle and stakeholder perspectives. *Energy and Buildings*, 178:165–181.