Issam Fares Institute for Public Policy and International Affairs
معهد عصام فارس للسياسات العامة والشؤون الدولية

POLICY BRIEF #2/2020
May 2020

URBAN INSIGHTS THROUGH MACHINE LEARNING IN DATA-SCARCE ENVIRONMENTS

Alaa Krayem, Physics Department, American University of Beirut

Aram Yeretzian, Architecture Department, American University of Beirut

Ghaleb Faour, National Center for Remote Sensing, National Council for Scientific Research, CNRS-L

Ali Ahmad, Issam Fares Institute for Public Policy and International Affairs, American University of Beirut

Sara Najem, Physics Department, American University of Beirut

Summary

Today, the existing building stock constitutes an important component of urban life as it holds much of a city's socio-economic activities. Therefore, comprehensive building survey data is essential to support more effective policymaking related to the sustainable management of cities. However, such a database is not always available and is often limited to a surveyed subset of buildings. This policy brief reports a methodology developed to extend our knowledge to the entire present building stock by adopting data-driven techniques. As an application, the number of floors and the construction period of buildings were predicted at the near-city scale in the Beirut administrative area. The results can help to assess and plan resilience policies at the city level as well as assist demographic and risk management studies.

MAIN RECOMMENDATIONS

- Advances in data science and machine learning provide a real opportunity to fill in gaps in essential data related to the characteristics of the building stock and its inhabitants in Beirut and other Lebanese cities.
- Governmental and public institutions that are concerned with urban planning, social, and economic affairs should invest in building a human capacity that would be capable of leveraging data-driven approaches to support their work.
- For Beirut, a city with significant data gaps, generating reliable data through machine learning can fast track social and developmental interventions and help in risk assessment and disaster management.

Introduction

In developing cities like Beirut, available urban data is often underutilized because of its sporadic nature and/or due to access challenges. In Lebanon, only recently that access to public information and data was facilitated by a Law.[1] Despite that, access to data remains a major challenge because of lengthy bureaucratic processes and lack of digitalization.

To enable the sustainable management of cities, policymakers, urban planners, and academics need to be supported by a detailed statistical outline of the building stock. Information about the buildings' age, use, form, adaptability, and resilience, to list a few, is important for planning and managing the future of the urban context. Existing information and ground surveys constitute the baseline for any database. Unfortunately, this data is sometimes unavailable, insufficient or expensive, and time-demanding. For this reason, benefiting from new data sources, methods, and tools is a central focus area in urban research.

In this respect, crowdsourcing tools that rely on citizens to proactively report their data are gaining interest. One such example is the Coloring Beirut initiative,[2] where residents are encouraged to fill and update buildings' information. Moreover, the automated extraction of building attributes is increasingly facilitated by the development of computational resources, data-driven learning techniques, and remote sensing.

1 Law 28/2017
2 https://coloringbeirut.com/

The work presented in this policy brief demonstrates the utility of data-driven techniques to complement the existing databases. More specifically, this application describes the prediction of missing buildings' features, namely the number of floors and the construction period, as a function of the available attributes in Beirut City.

Machine Learning-based Approach

(1) Collecting existing data

The first step in building a machine learning methodology to inform policymaking on issues related to the existing building stock is to identify the existing database. To create a database for Beirut, information pertaining to 17,742 buildings was collected. The attributes used in this work (footprint's area and perimeter, and building's height) were obtained from the National Center for Scientific Research CNRS Lebanon. Among these buildings, 7,122 of them were surveyed by the Saint-Joseph University, as part of the Libris program. The survey included the buildings' year of construction, type, number of floors, and the number of apartments. In addition, data on the annual electricity consumption for most of the buildings was obtained from the national power utility: Electricité du Liban (EDL), which inherently reflects their vacancy.

(2) Preprocessing the data

The second step that follows the data collection involves data cleaning and reformatting. The corresponding construction years were converted into six construction periods based on Beirut city's architectural history (before 1923, 1924-1940, 1941-1950, 1951-1960, 1961-1990, after 1991). The identification of these 6 periods was based on the definition of particular characteristics related to building form and building envelop materials. The building form relates to the footprint, height, typology and window to wall ratio, whereby the building materials relate to the façade construction materials as well as their texture and color. However, our scheme does not distinguish between buildings from 1941-1950 and 1951-1960 as they are archetypally very similar. The research and information that supported the definition of the characteristic periods are included in the references. Entries with missing fields, incorrect buildings' heights, or atypical floors' height were removed from the dataset. Then, noisy, incorrect or aberrant information was removed from the dataset using an outliers' detection algorithm.

(3) Machine learning

The last step of this pipeline is to feed the cleaned data from the previous steps to the machine learning models. Machine learning algorithms, more specifically Artificial Neural Networks (ANN) and Random Forests (RF), were applied respectively to predict the number of floors and the period of buildings' construction. The prediction was accomplished by establishing the non-linear dependency relationship between the dependent variables, i.e the number of floors and the period of construction, and the independent variables, in this case, the area, perimeter, height, and yearly electricity consumption. This relationship was built using a subset of the data, the buildings which have a complete set of features (number of floors, construction period, area, perimeter, height, and electricity consumption). The predictive model was then tested for accuracy on another subset with complete features, by comparing the predicted outputs to the actual values.

Insights

The model built to predict the number of floors is accurate with a coefficient of determination (R^2) of 87.7 percent, which is a measure of the accuracy of the prediction (ideally 100 percent). While, in the case of the construction period prediction, the accuracy reaches 48.7 percent, suggesting the need for more independent variables and era-specific descriptors. Despite the low accuracy in the construction period's prediction, we still tag the buildings with their corresponding predicted number of floors and construction period, using these models, as can be seen in Figures 1 and 2, respectively. However, this prediction needs to be refined with a closer surveying.

Synthesis and Implications

With the approach adopted in this policy brief, the lack of full datasets is compensated by machine learning interventions that can fill in data gaps and offer policy designers a powerful and verifiable new leverage. Beside the immediate applications reported above related to service provision, efficiency, analytics (e.g. electricity) and buildings' characteristics, the presented methodology can be an effective tool to generate wider policy insights despite data irregularities.

The two main areas where such an approach can particularly be advantageous are: (1) assessing urban resiliency, risk, and emergency planning. For example, having an accurate distribution of the number of floors and building materials would be critical for a rapid assessment of the human loss in the case of a natural disaster such as an earthquake or large fires; (2) generating demographic and socio-economic insights related to population concentration and public services. For example, the number of floors distribution could be used to distinguish between residential, commercial, and industrial units/zones within the city and inform policy experts about electricity rationing strategy (like the case of Beirut where power outages are regular but randomly allocated geographically); or provide information on energy consumption "hot spots" which could help with predicting electricity demand surge and the needed grid reinforcement strategy.

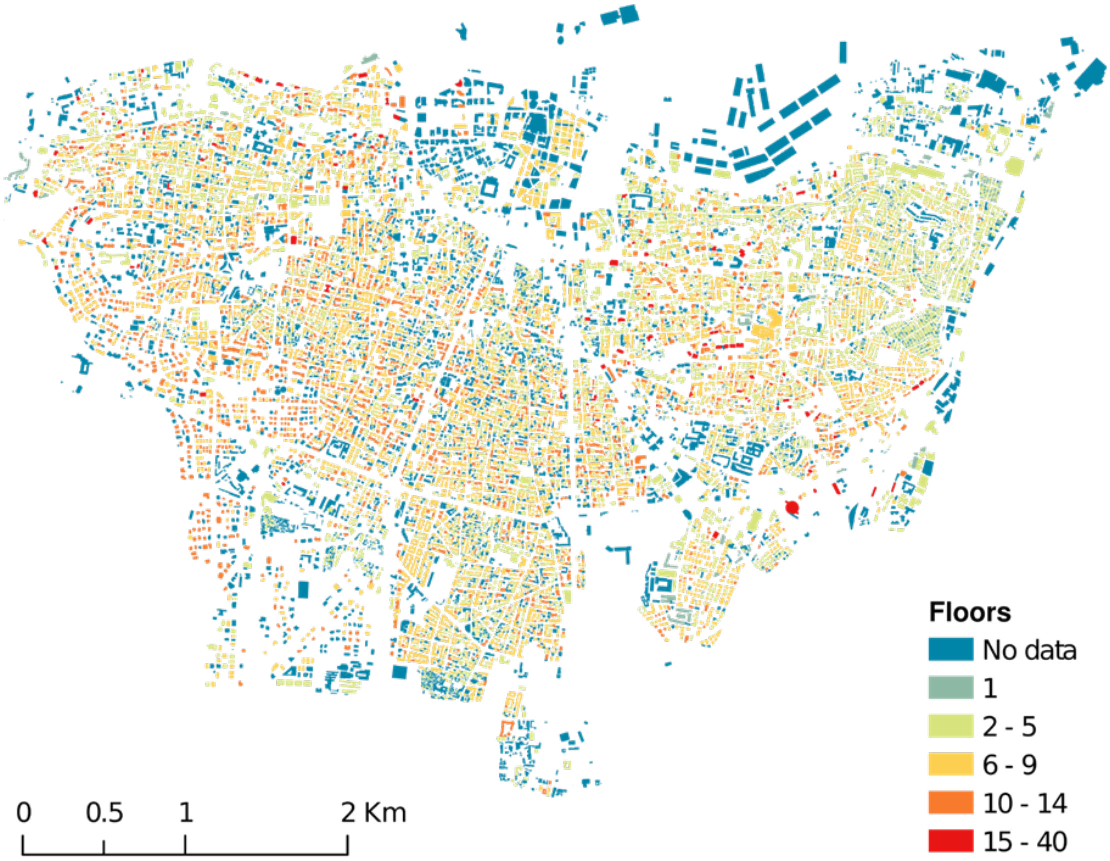

Figure 1: Number of floors in the buildings of the Beirut administrative area.

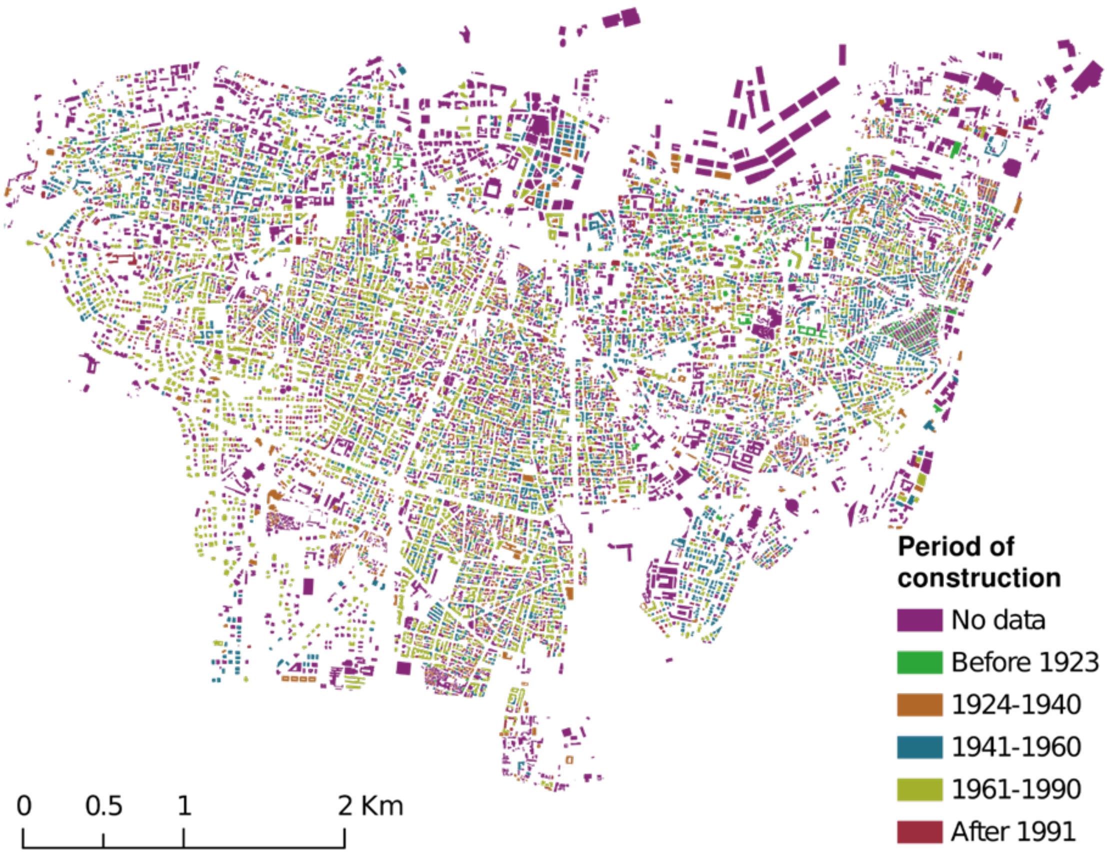

Figure 2: Period of construction of buildings in the Beirut administrative area.

References

[1] T. McPhearson et al., “Advancing Urban Ecology toward a Science of Cities,” Bioscience, vol. 66, no. 3, pp. 198–212, 2016.

[2] M. Batty, “The size, scale, and shape of cities,” Science., vol. 319, no. 5864, pp. 769–771, 2008.

[3] P. Hudson, “Urban Characterisation; Expanding Applications for, and New Approaches to Building Attribute Data Capture,” Hist. Environ. Policy Pract., vol. 9, no. 3–4, pp. 306–327, 2018.

[4] H. Tanikawa and S. Hashimoto, “Urban stock over time: Spatial material stock analysis using 4d-GIS,” Build. Res. Inf., vol. 37, no. 5–6, pp. 483–502, 2009.

[5] N. Kohler, P. Steadman, and U. Hassler, “Research on the building stock and its applications,” Build. Res. Inf., vol. 37, no. 5–6, pp. 449–454, 2009.

[6] G. Meinel, R. Hecht, and H. Herold, “Analyzing building stock using topographic maps and GIS,” Build. Res. Inf., vol. 37, no. 5–6, pp. 468–482, 2009.

[7] G. J. Arbid, “Practicing modernism in Beirut architecture in Lebanon 1946-1970,” Cambridge, Massachusetts : Harvard University, 2002.

[8] R. Saliba, Beirut 1920-1940 Domestic Architecture Between Tradition and Modernity Paperback, The Order of Engineers and Architects - Beirut (1998)

[9] E. El-Achkar. Réglementation et formes urbaines: le cas de Beyrouth. Presses de l’Ifpo, 2013.

[10] J. Urquizo, C. Calderón, and P. James, “Metrics of urban morphology and their impact on energy consumption: A case study in the United Kingdom”, Energy Research & Social Science, vol. 32, pp.193-206, 2017.

[11] L, Frayssinet, et al., “Modeling the heating and cooling energy demand of urban buildings at city scale”, Renewable and Sustainable Energy Reviews, vol. 81, pp.2318-2327, 2018.

[12] Li, Wenliang, et al. “Modeling urban building energy use: A review of modeling approaches and procedures.” Energy, vol. 141, pp. 2445-2457, 2017.

[13] ANR- LIBRIS project (ANR-09-RISK-006) - contribution to seismic hazard assessment in Lebanon. A co-joint project between ISTERRE, IPGP, EDYTEM and RESONNANCE laboratories with the AUB, NDU, USJ universities and the CNRS-L. Collaboration under the task 1.3. “Speleoseismicity and the Lebanese endokarst”.

Issam Fares Institute for Public Policy and International Affairs at the American University of Beirut

The Issam Fares Institute for Public Policy and International Affairs at the American University of Beirut (AUB Policy Institute) is an independent, research-based, policy-oriented institute. Inaugurated in 2006, the Institute aims to harness, develop, and initiate policy-relevant research in the Arab region.

We are committed to expanding and deepening policy-relevant knowledge production in and about the Arab region; and to creating a space for the interdisciplinary exchange of ideas among researchers, civil society and policy-makers.

Issam Fares Institute for Public Policy and International Affairs (AUB Policy Institute)American University of Beirut P.O.Box 11-0236
Riad El-Solh / Beirut 1107 2020
Lebanon

Issam Fares Institute for Public Policy and International Affairs
معهد عصام فارس للسياسات العامة والشؤون الدولية

Issam Fares Institute Building AUB
961-1-350000 ext. 4150
+961-1-737627
ifi.comms@aub.edu.lb
www.aub.edu.lb/ifi
aub.ifi
@ifi_aub