Design a framework for urban sustainability assessment: methods and challenges

Anonymous Author(s)
Affiliation
Address
email

Abstract

Assessing sustainability in the urban environment is a complex task, mainly due to the amount of heterogeneous data and variables, the presence of context-dependent measures, and the absence, in this domain, of recognized techniques for assessing the quality of the data and of the proposed interventions. Furthermore, expert knowledge is often overlooked in automated analysis. This paper presents the advancements of the doctoral project "INTERPRET: an integrated and adaptive framework to support policy-makers in the urban environment," whose purpose is to define a framework for ascertaining sustainability in cities. Specifically, the paper presents an overview of the preliminary results obtained, focusing on the chosen approaches and defined methodologies rather than the experiments.

1 Introduction

The growing urban data generated by new technologies reshapes our city experiences and aids urban planning. IoT tech, smart devices, IT infrastructures, and data analysis algorithms support the development of comfortable, sustainable cities [1]. This shift requires professionals to interpret data from these tools for effective planning. However, urban studies relied on fewer data before these technologies but still produced performance indicators and structural metrics. Although traditional analysis methods often lack integration with modern techniques, they are the basis of studying the urban environment. They cannot be excluded from automatic analysis, which cannot be separated from a deep understanding of them.

The purpose of the project INTERPRET [2] is precisely to produce a *multidisciplinary methodology* that aims to apply mainly *data modeling* and *data analysis* for optimizing the *sustainability* of urban systems and provide the urbanists with a practical framework for the comprehension and improvement of the existing smart city models. The framework aims to work on various data sources, from the more traditional measurements used by urban planners to sensor data or images from different sources.

The work is based on the research collaboration between the Computer Engineering Area of DEIB[1] and the Integrated Modification Methodology (IMM) designLab[2] of DABC [3], at Politecnico di Milano. In particular, IMM [3] is an innovative design methodology that allows the evaluation and improvement of the environmental performance of cities by modifying their structural characteristics and will be the starting point for the project. Currently, formal relationships between structure and performance have not yet been identified in IMM, and the analysis will focus on this aspect.

This paper presents an overview of the work done so far. The main feature of the proposed framework is the centrality and collaboration of users in both the data design analysis and the validation of the

[1]Dipartimento di Elettronica, Informazione e Bioingegneria: https://www.deib.polimi.it/ita/home
[2]http://www.immdesignlab.com/
[3]Department of Architecture, Built environment and Construction engineering: https://www.dabc.polimi.it/

Submitted to Computational Sustainability Workshop at NeurIPS 2023. Do not distribute.

results. This allows the system to work in a domain characterized by the absence of labeled data or ground truth, uncertainty in evaluating outcomes, and a diversity of aspects to be analyzed. These characteristics are typical of urban sustainability analysis and most problems related to computational sustainability.

This article is structured as follows: Section 2 contains a literature study on urban analysis and a brief overview of the IMM methodology. Section 3 will present the method of building the framework. In the same section, preliminary results will be presented. The results, emerging considerations, and future works will be summarized in the conclusion, Section 4.

2 Background and problem statement

As a starting point for our work, we analyzed different frameworks for assessing sustainability and smart cities. The main limitation revealed by this study is that the concept of urban sustainability, or at least the parameters defining it, are not unique and are ambiguously defined. The city, indeed, is a complex system that includes many subsystems. These involve the most disparate aspects, from morphological and energy systems to complex social and economic ones, and it is essential to consider them all when performing any analysis [4].

The authors of [5] stress that *urban sustainability frameworks* contain many indicators measuring environmental sustainability. In contrast, *smart city frameworks* lack environmental indicators but highlight social and economic aspects. Moreover, the modern smart city frameworks exclude morphological elements from the analysis, using a reductionist paradigm which, through the pervasive use of ICT, tends to control urban systems to optimize the individual subsystems in a logic of efficiency. In contrast, traditional urban diagnostics uses a systematic paradigm that aspires to build cities as robust systems capable of adapting to the emergence of new situations with only minor changes but sometimes produce context-specific results challenging to reproduce [6].

Indeed, from the literature, it seems that still, no work performs the analysis of the city's performance considering all the components describing it. This happens either because of the heterogeneity and quantity of the variables or because of the specificity of the traditional measures used, which are highly context-dependent and often hard to understand completely [7]. Some works, such as the one in [8], attempt to achieve this task but only provide various agendas for research and trace their practical implications without showing a concrete process.

Therefore, this project's main objective is to manage all the above aspects of the city's environmental performance. To do so, we collaborate with the IMM group (DABC Dept), taking their methodology as a starting point for our work. The IMM methodology aims to improve the city's environmental performance by modifying its structural characteristics [4] and considers the built environment as a Complex Adaptive System (CAS), i.e., a system with a significant number of agents that produce a non-linear dynamic behavior which may not be predictable according to the agents' behavior [9]. In IMM, the measures describing the structural properties of a sample area are called *Metrics*. An example of *Metric* is the Building Density (BD), defined as the total number of buildings in an area divided by the total area. *Metrics* belong to different *Key Categories*, structural aspects that guide the choices of the interventions to be made in the built environment. The tools for performance evaluation are called *Indicators* (e.g., Public Transportation Stop Density) organized into *Design Ordering Principle* (DOP) families, i.e., families of actions that designers can perform to improve the current system behavior [4].

3 Planned methodology and contribution made so far

In addition to studying existing methodologies and frameworks, we focused on defining a high-level framework architecture (Fig. 1) that could meet the needs of the IMMdesignLab urban planners and architects with whom we collaborated. The result is summarized below.

3.1 Framework definition

The framework consists of three conceptual levels: (i) project level, (ii) data level, and (iii) data analysis level. The first level refers to systematizing the analysis process, defining user needs and required inputs and outputs, which are currently very vague. The second level is creating a data

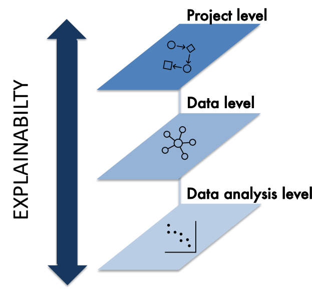

Figure 1: Framework architecture overview

model that can define and unify the data used by experts, which are mainly gained from automatic surveying systems, open geographic databases such as OpenStreetMap or processing through Q-GIS software, and are organized into separate spreadsheets without fixed scheme or methodology; indeed, they are currently full of redundancies inconsistencies. The third level, on the other hand, is the analysis. Across the three levels, we care particularly about explainability that guides the choice of algorithms to use in the framework. Indeed, these must be able to provide results interpretable by domain experts, whose feedback contributes to evaluating the results. The framework's goal is both to accelerate the analysis of the built environment and to enable the simulation of the interventions and the consequent changes and optimizations. Another essential aspect that will be considered is to promote a human-in-the-loop paradigm [10], where domain experts can intervene and refine the analysis process. While designed to support the built environment analysis, the framework can be generalized and used in all scenarios where analysis processes are to be automated and generalized by leveraging human knowledge of the application context.

For what concerns the first two levels, we created a knowledge-based system for the Integrated Modification Methodology (IMM) and we defined the Human-Centred Conceptual Design (HCCD) [11]. Our methodology employs the Entity-Relationship (ER) model as a lingua franca among Human-Centered Design experts, database designers, and application domain professionals, whose implicit and explicit knowledge is identified through a human-centered design process. The output of this work is a conceptual schema and a systematization of the various tasks involved in urban analysis, from the investigation of the built environment through the definition of the drivers that guide the transformations to the formalization of the project and its evaluation. The inputs and outputs of each task have been defined, and the technologies needed to carry them out are identified. The resulting schema includes temporal and spatial dimensions and appropriate connections between timed and static entities, allowing for a more precise understanding and satisfying the requirements. The HCCD approach also helped us identify the minimum subset of attributes needed to describe the raw data, often geometric data obtained from automatic surveying systems or GIS software.

As for the Data analysis level, we want to study the relationships among the different variables, their correlations, and possible causality relationships. This would help us to identify a significant subset of features, the relationships between these and the quality of the proposed interventions, and the prediction of their impact. Regarding this part, we developed the SiMBA system (Systematic clustering-based Methodology to support Built environment Analysis), which is a first attempt to analyze the IMM methodology data and identify the relationships among the different components of the city. The work is described in [12]. We also developed a plug-in exploiting the methodology in the Q-GIS software. The plug-in allows the implementation of the process more efficiently and on a greater variety of data[13]. In [14], we extended SiMBA to the analysis of purely geometric data and spatial patterns identification, and we compared its performances with deep-learning-based approaches. More details on SiMBA are provided in Section 3.2.

3.2 SiMBA

This section presents the SiMBA system and its contribution to the IMM methodology. More details about SiMBA can be found in [12] and [13]. Here, we provide an overview of the system. Fig. 2 represents its methodological workflow. As you can see, the process is composed of three phases:

(i) *Built Environment Decomposition* (BED phase); (ii) *First Level Clustering* (FLC phase); and (iii) *Second Level Clustering* (SLC phase).

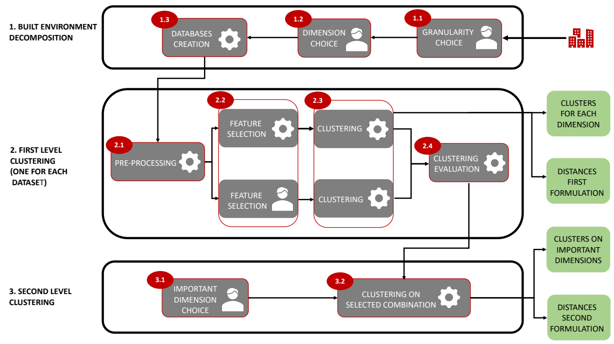

Figure 2: SiMBA system workflow.

The methodology aims to highlight the correspondence between environmental performance and structural patterns. Inputs are data of any type (we tested the system on IMM data and different data sources, tables, images, and spatial data), usually unlabeled. For these reasons, we base the process on clustering. In Step 1.1 (see Fig. 2), the granularity at which the analysis is to be conducted must be chosen. In other words, we must decide which samples to cluster. After the granularity, we must define *Dimensions* (1.2) and retrieve data according to them (1.3). *Dimensions* represents the different aspects we want to analyze. They can be of any type and category, i.e., performances, morphological characters, demographic data, and so on, depending on the kind of analysis we want to carry out. Notice that, according to the *Dimensions* chosen and the amount of the data we must handle, in Step 1.3, we can define more than one dataset. Once the input has been decomposed, we enter the *First Level Clustering* phase where, for each dataset, we perform clustering. Before applying the selected clustering algorithm (2.3), we perform feature selection in Step 2.2. This is done manually, using a set of features chosen by experts for each *Dimension*, and automatically, using the algorithm presented in [15]. Comparing the clusters obtained with the Manual and the Automated procedure, we evaluate our results in Step 2.4. In the third and last phase, *Second Level Clustering* , we combine the evaluation of the obtained results with the IMM expert's knowledge and needs (3.1) to select which are the *Dimensions*, and thus, features, we want to use in the *Second Level Clustering*. In the *First Level Clustering* and *Second Level Clustering* phases, there are algorithms to be chosen for pre-processing, feature selection, and clustering (see [13] and [14] for how choices are made and their justification). These choices may vary depending on the size and characteristics of the specific datasets, but this does not affect the generality of the process.

Regarding the results obtained so far, the experiments have concerned the city of Milan using different granularity and data related to multiple *Dimensions*. Some of the most significant results were the proof of SiMBA's ability to select representative features and create meaningful clusters in any dataset and the correspondence between structural and performance patterns. As mentioned, this last result is critically important for developing the framework and the IMM methodology.

4 Conclusion

The main objective of our research project is to develop a framework that integrates the existing IMM methodology using data science techniques with the final scope of measuring and improving city environmental performances. We designed the framework architecture and the data model and developed the SIMBA clustering-based system. The process and data modeling and analysis rely heavily on collaboration with users, who play a crucial role in the co-design phase of the framework and analysis methodologies. One aspect not currently addressed is the explainability of results, which is critical in unsupervised algorithms. This part, however, is crucial since usable and understandable results are the basis of a fair and sustainable system for users. Nevertheless, thanks to the definition of a precise methodology (HCCD) for the representation of user knowledge and analysis that is spread over several levels and divides a multi-object problem into several individually analyzed ones, these preliminary results are promising for the development of a framework capable of working in the area of - but not limited to - urban environment sustainability analysis.

Acknowledgment

I want to thank Professor Letizia Tanca and Professor Massimo Tadi for the supervision and support they are giving me in this doctoral journey of mine. I also thank Dr. Carlo Andrea Biraghi for his technical and non-technical advice.

References

[1] D. Sikeridis, Iot-enabled knowledge extraction and edge device sustainability in smart cities, 2020, pp. 264–265.

[2] E. Lenzi, Interpret: An integrated and adaptive framework to support policy-makers in the urban environment, in: 2023 IEEE 39th International Conference on Data Engineering (ICDE), 2023, pp. 3918–3922. doi:10.1109/ICDE55515.2023.00381.

[3] M. Tadi, M. H. M. Zadeh, O. Ogut, Measuring the influence of functional proximity on environmental urban performance via integrated modification methodology: Four study cases in milan, International Journal of Urban and Civil Engineering (2020).

[4] Integrated modification methodology (imm): A phasing process for sustainable urban design, World Academy of Science Engineering and Technology (2013).

[5] H. Ahvenniemi, A. Huovila, I. Pinto-Seppä, M. Airaksinen, What are the differences between sustainable and smart cities?, Cities 60 (2017) 234–245.

[6] C. Selada, Smart cities and the quadruple helix innovation systems conceptual framework: The case of portugal, The Quadruple Innovation Helix Nexus: A Smart Growth Model, Quantitative Empirical Validation and Operationalization for OECD Countries (2017) 211–244.

[7] S. Raghavan, S. L. Boung Yew, Y. L. Lee, W. Tan, K. K. Kee, Data Integration for Smart Cities: Opportunities and Challenges, 2019, pp. 393–403. doi:10.1007/978-981-15-0058-9_38.

[8] S. Joshi, S. Saxena, T. Godbole, Shreya, Developing smart cities: An integrated framework, Procedia Computer Science 93 (2016) 902–909. URL: https://www.sciencedirect.com/science/article/pii/S1877050916315022. doi:https://doi.org/10.1016/j.procs.2016.07.258, proceedings of the 6th International Conference on Advances in Computing and Communications.

[9] T. Carmichael, A. Collins, M. Hadžikadić, Complex Adaptive Systems: Views from the Physical, Natural, and Social Sciences, Understanding Complex Systems, Springer International Publishing, 2019. URL: https://books.google.it/books?id=O46dDwAAQBAJ.

[10] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, D. Pedreschi, A survey of methods for explaining black box models, ACM Comput. Surv. 51 (2018). URL: https://doi.org/10.1145/3236009. doi:10.1145/3236009.

[11] E. Lenzi, E. Pucci, F. Cerutti, M. Matera, L. Tanca, Human-centred conceptual modelling for re-designing urban planning, in: Advances in Conceptual Modeling ER 2022 Workshops EmpER Proceedings. (Accepted for pubblication), 2023.

[12] C. A. Biraghi, E. Lenzi, M. Matera, M. Tadi, L. Tanca, Simba: Systematic clustering-based methodology to support built environment analysis, 2022, pp. 453–460.

[13] A. Folini, E. Lenzi, C. A. Biraghi, Cluster analysis: A comprehensive and versatile qgis plugin for pattern recognition in geospatial data, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLVIII-4/W1-2022 (2022) 151–157. URL: https://isprs-archives.copernicus.org/articles/XLVIII-4-W1-2022/151/2022/. doi:10.5194/isprs-archives-XLVIII-4-W1-2022-151-2022.

[14] C. A. Biraghi, E. Lenzi, Spatial pattern analysis through distribution metrics, ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 43B4 (2022) 69–77. doi:10.5194/isprs-archives-XLIII-B4-2022-69-2022.

[15] M. Dash, H. Liu, Feature selection for clustering, in: Knowledge Discovery and Data Mining. Current Issues and New Applications: 4th Pacific-Asia Conference, PAKDD 2000 Kyoto, Japan, April 18–20, 2000 Proceedings 4, Springer, 2000, pp. 110–121.