Energy & Buildings 178 (2018) 61–70

Contents lists available at ScienceDirect

Energy & Buildings

journal homepage: www.elsevier.com/locate/enbuild

Revealing the relationships between the energy parameters of single-family buildings with the use of Self-Organizing Maps

Krzysztof Cebrat[a,*], Łukasz Nowak[b]

[a] Department of Environmental Management, Faculty of Architecture, Wrocław University of Science and Technology, Wrocław, Poland
[b] Department of Building Physics and Computer Design Methods, Faculty of Civil Engineering, Wrocław University of Science and Technology, Wrocław, Poland

ARTICLE INFO

Article history:
Received 22 May 2018
Revised 4 July 2018
Accepted 15 August 2018
Available online 1 September 2018

Keywords:
Useful energy demand
Self-Organizing Maps
Building energy efficiency
Energy parameters
Energy performance
Procedural modelling
Random parameters
Unsupervised data mining

ABSTRACT

With a large number of factors affecting the energy efficiency of buildings, the importance of analyzing the growing amount of data becomes important. The aim of this research is to check whether the use of Self-Organizing Maps will allow the indication of the relationships between building features which are considered important from the point of view of their energy performance.

The research was carried out on a sample of 5040 randomly generated variants of single-family buildings with a fixed volume and location. These models were next subject to clustering, based on selected features, with the use of Self-Organizing Maps.

The results prove the suitability of the method used. Grouping analysis shows the dependencies between particular parameters, confirming the importance of the U-value of partitions, the thermal mass of the building and its air-tightness for energy efficiency, while discovering unexpected relationships such as the irrelevancy of a building's orientation.

In addition to expanding knowledge about the relationships between building features affecting their energy performance, it may allow optimal parameters for the given initial conditions to be found.

1. Introduction

The location of a building, its orientation, form, internal structure, construction of partitions; these are just some of the features that are believed to have an impact on energy consumption for heating, cooling, and ventilation. An analysis of the solar gains, ventilation, and other physical parameters of a building allow many more factors affecting the energy performance of a building to be listed.

It is known that most of these factors are in close, but not necessarily linear, relationships. One can determine the impact of a separated factor on the building's energy efficiency (comparing different building models with one variable), but it is very difficult to define the most desirable set of many features (comparison of building models with many variables) under specific external conditions.

The problem is to achieve the assumed level of energy consumption in a building, while modifying more than one parameter affecting its energy balance.

It is not possible to increase the surface of windows without changing the surface of the partitions, the average U value of the partitions, their tightness, etc. The form of the building cannot be changed, without changing the length of the thermal bridges or the building's mass.

With the proposed change of one of the building's parameters, the others will also change, but the remaining ones can be adjusted to maintain the assumed level of energy consumption in the building. This adaptation should be a conscious and knowledge-based action, which, according to our research, can be supported by SOM.

The problem exists as designers usually face the dilemma of choosing not one, but many variables, so that the design meets not only the requirements of the Vitruvian triad – structurally sound, beautiful, and functional – but also the need of having a low impact on the environment.

Since this is a non-linear issue, it is legitimate to look for procedures that best deal with solving such problems. Neural maps are considered such methods.

They are mentioned in many reviews generally related to energy efficiency [1], or specifically to energy efficiency in buildings [2], among other data mining techniques so urgently needed to understand the growing amount of data that can be obtained.

---

[*] Corresponding author.
E-mail address: krzysztof.cebrat@pwr.edu.pl (K. Cebrat).

https://doi.org/10.1016/j.enbuild.2018.08.028
0378-7788/

Table 1
Parameter values of three randomly generated buildings with the lowest energy demand.

	Share of walls facing south [%]	*Roof surface* [m^2]	*Surface of windows facing south* [m^2]	*Share of windows facing north [%]*	*Window surface to wall surface ratio*	*Share of heat loss through thermal bridges in transmission heat loss [%]*	*Total heat capacity* [J/K]	*Useful energy demand for heating index* [kWh/m^2/year]
1	*25.71*	*87.78*	*13.73*	*12.80*	*0.20*	*7.42*	63.82×10^6	*31.44*
2	*21.94*	*94.77*	*22.73*	*5.99*	*0.32*	*9.18*	29.98×10^6	*31.90*
3	*23.78*	*73.48*	*23.97*	*21.09*	*0.40*	*16.75*	139.13×10^6	*32.28*

They are used to predict building energy demand [3,4]; however, the main field of their current application is the analysis of occupant behavior and understanding the patterns of energy use [5-7].

The Self-Organizing Map (SOM) as described by Kohonen [8] belongs among the techniques of learning neural maps described in [9,10]. It is a recognized method of data analysis, but for reasons that are incomprehensible, SOM is rather used in the field of biological sciences [11] or even urban planning [12] but has not yet been applied to the science of buildings.

2. SOM and its alternatives

To effectively analyze the large set of data that one has to deal with for buildings, it should be simplified so that its internal structure becomes intelligible. One therefore looks for a method that can present multidimensional data in a one- or two-dimensional form, and which at the same time clusters data so that the intragroup variance will be minimal and the between-group variance will be at a maximum [2]. This is exactly what the SOM does, although it projects data from a multidimensional space into a smaller number of dimensions by local smoothing, interpolation, and extrapolation, with a smaller or larger external group variance being the side-effect of the goal sought in the mapping process.

Because the clustering performed by SOM is unattended, it is not known exactly how it is made. Attempts are made to omit this problem by using the *Decision Tree Method* (DTM) [13]. However, the disadvantage of this method is that it mainly operates on categorical variables, while buildings are described mainly by numerical data. In addition, the DTM algorithm works better with discrete rather than continuous variables. Meanwhile, most parameters affecting the energy performance of the building are continuous. Of course, continuous variables can also be grouped – this is what the SOM does. But if the discretization is done arbitrarily, as it is before starting the DTM, the clustering could be considered as biased. Thus, the DTM algorithm is applicable rather after grouping with the SOM and not instead of it. Some researchers [14] performed a similar analysis – except that they use the *K-means* algorithm for grouping (where *K* local averages from a finite data set are calculated).

In fact, the SOM could be understood as a *K-means* algorithm expanded with smoothing. What's more, with certain settings of the SOM algorithm, it works exactly like *K-means* [15], which was used in studies such as the impact of user behavior on energy consumption in a building [16].

Another possibility was to use one of the *ranking methods* [17]. However, even if one chooses (on a subjective basis) to compare only a few buildings with the lowest energy demand from among a randomly generated large number of variants (Table 1), it can be seen that the features that are responsible for this demand take on very different values and it is very difficult to indicate the relationships between them. With more variants considered, differences are likely to rise. The data included in Table 1 are the values of the parameters of building models generated during the described research. All data is freely available at https://osf.io/qsjzy/ and from Mendeley Data.

The implementation of *Genetic Algorithms* (GA) was also considered. GA are usually used to search for an optimal solution (e.g. energy demand) for a non-linear equation of several variables [18,19]. In this particular problem, one could use GA to look for a relationship between a larger number of parameters, on a "population" of e.g. 5000 individuals (buildings), agreeing to obtain results within a certain range of values, not only the best ones. Then one could search for general dependencies between the parameters within the received models. However, with 15 uncorrelated variables, one should expect that as a result of the use of GA, many solutions will be local optimums, characterized by very different variable configurations. This means that they will constitute separate "species" whose recombination will not provide "individuals" that would have the chance to survive the "evolution" (achieve a better result for energy demand). Analyzing the chosen individual solutions will result in conclusions similar to the analysis of the *ranking method*.

3. Aim and objectives of the research

This research proposes a slightly different approach than the one most frequently encountered.

The aim of the research is, first and foremost, to indicate the relationships between the features of buildings which are very likely to have a significant impact on their energy performance. This is not the optimization of energy consumption, or even its prediction. It is the step that comes before.

Once this is known, one can, using only the selected parameters as variables, conduct an optimization, for example using *Genetic Algorithms* or the *Decision Tree Method*.

Since the goal is to acquire new knowledge and discover the general laws governing the relationships between building parameters, it is necessary to use a large data set and therefore to use unsupervised methods of analysis [2]. This is because any data selection, or parameter determination, limiting the number of input variables and expected results (characteristic of supervised techniques) limits the possibilities of discovering the dependencies, correlations, and the internal structure of the data itself.

Although the SOM is not an optimization algorithm, its use allows interdependencies between the values of particular parameters of buildings to be identified and describes their importance to their energy performance, which is difficult to achieve with other methods.

4. Methodology – acquiring data

The research was carried out on a sample of 5040 procedurally-generated models of single-family buildings with a constant heated volume of 600 m^3, naturally ventilated, and located in the climatic zone of Wroclaw, Poland.

The building variants were described by means of randomly generated parameters necessary for the subsequent calculation of energy indicators and performance. The drawn parameters of building models, together with the search domain for their values and type of data are shown in Table 2.