





Detecting Air Conditioning Usage in Households Using Unsupervised Machine Learning on Smart Meter Data

Rodrigo Porteiro¹(✉)  and Sergio Nesmachnow² 

¹ UTE, Montevideo, Uruguay
rporteur@ute.com.uy

² Universidad de la República, Montevideo, Uruguay
sergion@fing.edu.uy

Abstract. This article presents an unsupervised machine learning approach for the problem of detecting use of air conditioning in households, during the summer. This is a relevant problem in the context of the modern smart grid approach under the paradigm of smart cities. The proposed methodology applies data analysis, a thermal inertial model for estimating the temperature inside a household, statistical analysis, clustering, and classification. The proposed model is validated on a real case study, considering households with known use of air conditioning in summer. In the evaluation, the proposed classification methodology reached an accuracy of 0.897, a promising result considering the very small cardinality of the set of households. The proposed method is valuable since it applies an unsupervised approach, which does not require large volumes of labeled data for training, and allows determining characteristics in the electricity consumption patterns that are useful for categorization. In turn, it is a non-intrusive method and does not require investing in the installation of complex devices or conducting consumer surveys.

Keywords: Unsupervised learning · Data analysis · Residential electricity consumption

1 Introduction

Currently, electricity utilities have advanced strongly in the deployment of various smart devices that assist monitoring and decision making. Final consumers have taken a very active role, given that there is greater knowledge of their behavior and their use of the electrical resource [23]. For utilities, it is essential to analyze the large amount of data gathered by using smart devices, to add value to the electricity business. In the residential sector, the existence of smart meters is crucial to improve management and profit. At the same time, it makes it possible to carry out commercial policies in which the consumer is a key player and feels part of the improvements.

Usually, utilities only consider the benefits obtained by being able to operate the smart meter remotely when analyzing the technical requirements of the smart meter to be deployed. For example, the benefits to obtain consume measurements without the presence of an employee in the field, to detect blackouts remotely, or to perform remote power cuts in case of unpaid bills. For this reason, many electric utilities do not properly plan the selection of smart meters considering the benefit they will generate when analyzing the measured data. One of the features of smart meters that is not properly valued is the measurement frequency. Many of the tools that can be developed using consumption data require a high granularity in the values measured from smart meters. Another technical feature of smart meters usually underestimated for the creation of these tools, is the measurement of harmonics.

A very useful approach for building a detailed profile of consumers is to detect which electrical devices are in use. The literature shows that for obtaining a very precise disaggregation of the electrical appliances uses from smart meter data, very high measurement frequency and harmonic measurement are needed [9, 10]. If the smart metering infrastructure is already deployed additional devices would need to be installed in households. However, this type of intrusive intervention is not always financially profitable and is often frowned upon by clients.

This article addresses the problem of detecting the use of air conditioning in summer, using the existing infrastructure of the Uruguayan electricity company (UTE). The company deployed smart meters on residential consumers, so replacing them would be very expensive. The meters are not capable of measuring harmonics, and have a quarterly consumption measurement frequency. However, despite the technological limitations, it is feasible to detect air conditioners using the available information and climate data. Detecting the use of this type of device, which is intensive in the summer, allows designing appropriate energy efficiency policies and commercial products, to optimize the electrical system and reduce the cost for the company, for consumers, and for the country [2, 12].

An unsupervised machine learning algorithm is proposed to detect the use of air conditioning in summer. Given a household, the proposed methodology applies urban data analysis [15] over consumption measurements, and temperature and irradiance measured at the nearest weather station. An approximation of the internal temperature of the household is then obtained a simplified model of thermal inertia for a standard Uruguayan residential building. Thus, a function is obtained that approximates the internal temperature from the external temperature and the irradiance. With the fifteen-minute data on consumption and internal temperature, consumption measurements less than 10% of the average consumption are excluded, as a criterion to consider the moments in which there is activity in the household. The resulting data is classified into two sets using the k -means clustering algorithm, only considering its temperature. A set of consumptions for low temperatures and another one for high temperatures are obtained. The average consumption is calculated for each set and if the difference between these values is greater than a threshold value, the consumer in the studied household is classified as an air conditioning user.

A practical validation of the proposed methodology is presented for a case study with a set of 29 households for whom the use of air conditioning in summer is known. The main results of the research applying the proposed methodology to this set, yields an accuracy of 0.897. This result is promising, especially considering that the study could only be carried out on a very small set of households due to lack of labeled consumers. The proposed method is valuable since it applies an unsupervised approach, which does not require large volumes of labeled data for training and it is non-intrusive. Thus, it is a viable alternative to intrusive methods, which require replacing the currently installed smart meters with more sophisticated ones (a very costly task in economical terms). In turn, having labeled data to use supervised learning or to validate the presented algorithm requires conducting surveys that demand a hard and expensive task. Therefore, the proposed methodology has the advantage that it can be applied immediately using the current installed infrastructure, without incurring in significant investments costs.

The article is organized as follows. Section 2 describes the problem addressed in this article and reviews related works. Section 3 describes the proposed approach, including a description of data sources, the process of data preparation and the definition of the classification algorithm. Details of the developed implementation are provided in Sect. 4. Section 5 present the experimental evaluation of the proposed approach and discusses the obtained results. Finally, Sect. 6 present the conclusions of the research and formulates the main lines for future work.

2 Problem Definition and Literature Review

This section describes the general problem addressed problem and reviews relevant related works.

2.1 General Problem: Energy Disaggregation

Traditionally, electricity companies have mainly worked using static information from consumers. The only source of data that dynamically linked consumers to the company was the measurement that a company official obtained monthly from traditional meter. Nowadays, the massive deployment of smart meters has allowed companies in the energy sector to know in greater detail the behavior of consumers, regarding energy use. It is possible to detect various details regarding the behavior of consumers with greater or lesser accuracy, depending on the technical specifications of the smart meters deployed.

If smart meters with high metering frequency and harmonic measurement are available, it is possible to successfully address the overall problem of energy disaggregation. It consists of identifying the individual consumption of different household appliances, using only aggregate measurements of all the measured variables. Many studies have analyzed the problem of energy disaggregation using smart meters with advanced technological features.

2.2 Problem Description: Detection of Air Conditioning Usage

The addressed problem is detecting the use of air conditioners in summer, from smart meters with a measurement frequency of fifteen minutes and without measuring harmonics. Air conditioner is the only electrically-intensive device to lower the temperature of households in the summer. It is highly correlated with climatic variables and allows addressing the classification problem without the need to have labeled consumers or to have more advanced measurement devices. Then, the solution proposed is considered non-intrusive.

The input data consist of electricity consumption curves with quarterly frequency, the geographical location of the consumer, and weather information to approximate the temperature inside the household. One of the hypotheses assumed is that the only electrical device for thermal comfort in summer with a relevant consumption is the air conditioner. Other thermal conditioning devices for the summer have negligible power consumption respect to air conditioners. For example, the average power of fans is 50 W and coolers 100 W, compared to 1000 W to 2000 W for air conditioner. The other hypotheses is that its use is strongly correlated with the internal temperature of the household.

2.3 Related Work

The application of energy disaggregation tools to the residential sector has developed strongly after the high penetration of smart meters in electrical systems. The addressed problem is to estimate the consumption of each of the electrical devices in the household, considering as input the overall energy consumption. When disaggregation is coarse-grained, the main objectives are related to provide more information to consumers on energy bills, or even offer specific commercial products depending on the type of use. When the disaggregation problem is solved in real time, the main applicability is to identify problems such as electrical losses in the household, detection of overloads, and other relevant issues.

Non-intrusive load monitoring (NILM) and the disaggregation problem in households were introduced by Hart [11], as an alternative to existing intrusive, hardware-based monitoring approaches. The main advantages of NILM is that it does not require installing specific devices, but makes use of existing smart meters, focusing on more sophisticated software for data analysis. Hart also introduced the binary (ON/OFF) variant of the disaggregation problem and proposed the principle of continuity switch, i.e., assuming that in a given small time interval, few appliances change their status (from ON to OFF or vice versa).

Many recent articles have dealt with NILM as a learning problem, applying computational intelligence to solve it, both in supervised and unsupervised fashion. Supervised approaches (e.g., Bayesian learning, neural networks (ANN), patterns similarity) make use of specific datasets of electricity consumption of each device and the aggregate household consumption signal. Unsupervised approaches (e.g., Hidden Markov Models, HMM) seek to learn the ON/OFF state of devices from the aggregate consumption, without explicit knowledge about the consumption of each device [4].

Kelly and Knottenbelt [13] studied ANNs for the NILM problem, using the UK-DALE dataset, which includes the electricity consumption of appliances (fridge, washing machine, dishwasher, kettle, and microwave) in five houses in the UK. A denoising autoencoder ANN computed the best results, outperforming over a long short-term memory (LSTM) and a rectangles ANN. Kolter and Johnson [14] introduced the REDD dataset to study a HMM for the NILM problem. Mixed results were computed over two weeks of data from five households (64.5% accuracy on the training set and just 47.7% in the evaluation test).

Our previous articles [6,7] studied the disaggregation of electricity consumption in residential buildings and proposed a method based on detecting similarities in the electricity consumption patterns from previously recorded labeled datasets. The method was evaluated over four different problem instances that model real household scenarios, reporting accurate results regarding standard prediction metrics.

Computational models are also very valuable for energy demand management and demand response [16,17]. Our previous articles [19,20] applied computational methods for defining a thermal index associated with an active demand management that interrupts domestic electric water heaters. Specific models using Extra Trees Regressor and a linear model were defined for water utilization and water temperature considering continuous power consumption measurements of water heaters, and Monte Carlo simulations to compute the proposed index. The approach was evaluated using real data from the ECD-UY dataset, Uruguay [8]. The thermal discomfort index correctly modeled the impact on temperature, providing accurate inputs for demand response and load shifting. Data analysis and computational intelligence techniques were also applied for the characterization and forecasting of short term electricity consumption on industrial facilities [21,22]. The model was validated for an industrial park in Burgos (Spain), the total electricity demand for Uruguay, and demand from a distribution substation in Montevideo (Uruguay).

3 The Proposed Approach for the Detection of Air Conditioning Usage in Summer

This section describes the data sources, the methodology to approximate indoor temperature, the data preparation and the methodology applied for the detection of air conditioning usage in summer.

3.1 Data Sources

The consumption data used in this article was provided by the Uruguayan National Electricity Company (UTE). It corresponds to “Total household consumption” and “Disaggregated electricity consumption by appliance”, two of the three subsets included in the ECD-UY dataset [8]. The Total household consumption set gathers data of quarterly total consumption from 110953 households and Disaggregated electricity consumption by appliance contains data of 9

households with minuted measures and disaggregated consumption by appliance. These sets have measurements from January 1, 2019 to November 2, 2020. In turn, data obtained from 20 additional households from known consumers was used for validation. The overall dataset includes 29 labeled households (19 use air conditioning in the summer and 10 do not).

To approximate the temperature of the household, which is the variable that has the strongest correlation with the use of air conditioning, climate information on temperature and solar irradiance was used. The data used was obtained manually from Uruguayan Institute of Meteorology (INUMET), disaggregated by weather station. Only data from January 1, 2019 to November 2, 2020 were considered, to match with the consumption data. Therefore, the horizon of data analyzed in this article is determined by these dates.

Likewise, both the consumption and the climate information contains the location of the measurement, which allows households to be associated with the climate data obtained from the nearest station.

3.2 Approximation of the Internal Temperature of the Household from the External Temperature and Solar Irradiance

An approximation of the temperature inside each household is needed for the proposed model. The proposed approach consists in estimating the inside temperature from the curve of the outside temperature and the external solar irradiance.

The thermal inertia that occurs inside the household is considered. The most relevant factors to model this effect are the construction material, the number of windows, and the insulation. Cengel et al. [5] showed that the heat flux is proportional to the magnitude of the temperature gradient, and opposite in sign. This article only requires an approximation of the internal temperature, and for this purpose a simplified model, proposed by Absi et al. [1] is used.

The model by Absi et al. assumes that the effect of the walls of a house produces two transformations in the external temperature curve: a delay (thermal lag) and an attenuation in the amplitude of the curve. Fig. 1 shows that the amplitude of the indoor temperature (A_{ind}) is smaller than the amplitude of the external temperature (A_{ext}). It also shows that indoor temperature is lagged by $\beta \times w_{irr}$, the thermal lag considering the irradiance.

According to the aforementioned model, these parameters depend on the wall material, the wall thickness, and the solar irradiance. Intuitively, the flow of heat from outside to inside is more delayed (thermal lag) and also the indoor thermal amplitude decreases when thicker and more robust walls are used, and the lower the solar irradiance. For instance, if the wall is extremely thin, the indoor temperature and the external temperature are almost equal. This model provides a rough simplification of the real temperature dynamics, which is appropriate for the proposed case study, especially considering that there is not enough information about households to estimate the internal temperature using a more complex model.

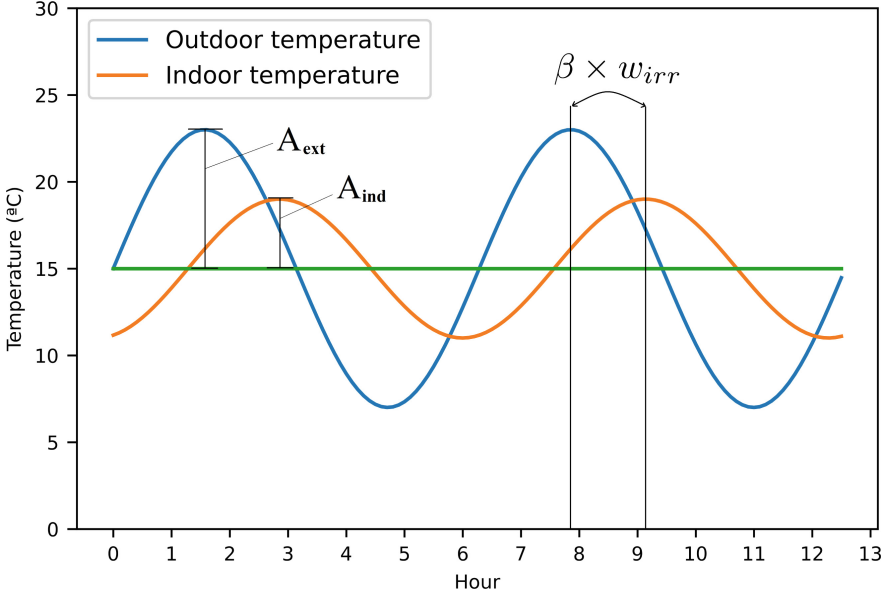


Fig. 1. Amplitude variation and thermal lag of indoor and outdoor temperature.

The internal temperature is represented as a function of the external temperature and the irradiance, according to the formulation in Eq. 1.

$$T_{ind}(t) = (T_{ext}(t - \beta \times w_{irr}) - \bar{T}_{24}(t))\rho \times w_{irr} + \bar{T}_{24}(t) \quad (1)$$

In Eq. 1, the function $T_{ind}(t)$ is the indoor temperature at time t and the function $T_{ext}(t)$ is the external temperature at time t . The function $\bar{T}_{24}(t)$ is the external average temperature of the last 24h, used as a baseline to estimate the amplitude. Then, $\beta \times w_{irr}$ is the thermal lag and the parameter $\rho = A_{ind}/A_{ext} \times w_{irr}$ captures the amplitude reduction. Finally, $w_{irr} = 1$ when irradiance is 0 and a maximum value when irradiance reach its maximum. So β and ρ are the thermal lag and the amplitude reduction factor when solar irradiance is 0.

3.3 Data Preparation

The objective of preparing the data is to obtain a historical bivariate series for the summer, in the considered analysis horizon. The first variable of the series represents the electricity consumption of the considered household and the second variable is an approximation of the internal temperature. First, since the temperature and irradiance data are hourly, they are converted to quarterly simply by using the hourly value. There are no missing values in either the consumption data or the climate data.

To generate the series of indoor temperatures, the parameters β , ρ and w_{irr} must be estimated. β , ρ and w_{irr} are estimated by measuring real temperature curves in three types of buildings (considering the most used construction materials in Uruguay: brick, concrete and wood [3]). First, measures were performed during the night, so $w_{irr} = 1$ (since there is no irradiance).

According to Eq. 1, $\beta \times w_{irr}$ is the thermal lag between indoor and outdoor temperature, considering irradiance. The lag is calculated using the measured curves at night (i.e., the difference between T_{ext} and T_{ind} along the x -axis). Then, setting $w_{irr} = 1$, the value of β is determined. Analogously, $\rho \times w_{irr} = A_{ind}/A_{ext}$; A_{ind} and A_{ext} are measured, and $w_{irr} = 1$ at night, so, the value of ρ is determined.

Then, fixing β and ρ , the value of w_{irr} is estimated for a completely clear day using Eq. 1, so the maximum w_{irr} is determined (w_{irr}^{max}). All estimations are performed using real indoor and external temperature measures. To compute Eq. 1 for an intermediate value of irradiance I_{real} , w_{irr} must be calculated for I_{real} , proportionally. So $w_{irr} = I_{real} \times (w_{irr}^{max} - 1)/(I_{max} - 1)$, where I_{max} is the maximum irradiance measured in a clear day.

Once the three relevant parameters of the temperature model are estimated, the indoor temperature series is obtained from the outdoor temperature series and irradiance series applying Eq. 1. This procedure is performed for each household, using weather data of the closest meteorological station. Figure 2 presents the internal temperature curves for the same external temperature and solar irradiance, depending on the type of construction.

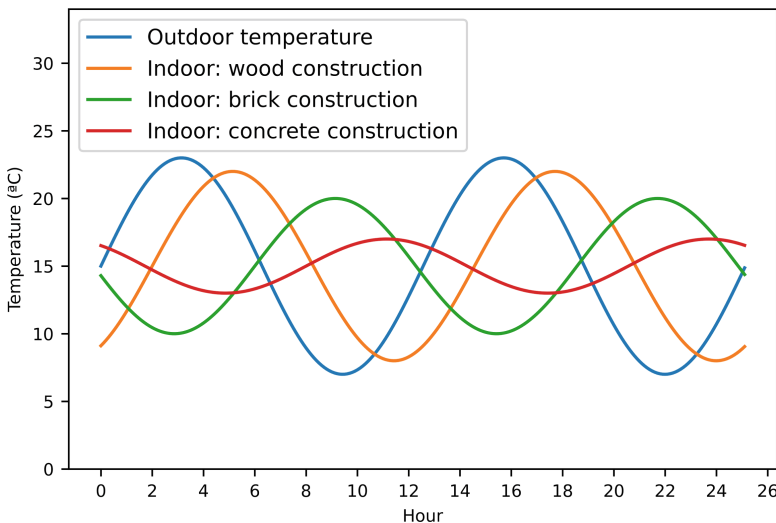


Fig. 2. Indoor temperature curves for wood, brick and concrete constructions.

The graphic in Fig. 2 shows that brick constructions present a thermal inertia between wood and concrete constructions. In the proposed study, since there is no information on the construction material of each household, the set of parameters estimated for a brick construction was used, considering that it represents a non-extreme thermal inertia. Also, brick is widely used in Uruguay, due to solidity and durability [3].

Finally, on the bivariate series with quarterly values, the data that do not correspond to the summer months (December, January, February, and March) are excluded, since they are not useful for the proposed analysis. Then, the maximum consumption for the resulting data is obtained, and all data with consumption less than 10% of the maximum are excluded from the series, considering that in those cases there is no activity in the household.

3.4 Unsupervised Machine Learning Classification Algorithm

The unsupervised classification algorithm must take into account the high correlation between the consumption of the household and its internal temperature when there is activity in it. In the preparation of data, very low consumption was excluded, to focus on the correlation between consumption and temperature in the case of activity in the household. It is important to consider that if there is any device with relevant consumption not associated with thermal conditioning, it will not present a strong correlation with temperature.

The proposed algorithm consists of performing the following seven steps, including data preparation:

1. Construction of the indoor temperature curve of the analyzed household, using the technique described in Sect. 3.2 applied to the meteorological data of the nearest station. As a result, a bivariate quarterly series is obtained with consumption and indoor temperature variables.
2. From the series obtained, entries with consumption less than 10% of the maximum of the series are excluded.
3. A clustering is performed applying k -means, with $k = 2$ in the temperature variable. Thus obtaining a set of consumption values for low temperatures (L) and another set of consumption values for high temperatures (H).
4. A clustering is performed applying k -means, with $k = 2$ in the consumption variable for the set L (obtaining two classes, L_L and L_H).
5. A clustering is performed applying k -means, with $k = 2$ in the consumption variable for the set H (obtaining two classes, H_L and H_H).
6. $center_L = (T_L, E_L)$ is defined as the center of L_H and $center_H = (T_H, E_H)$ as the center of the cluster H_H .
7. if $E_H/E_L \geq \Theta$, then the consumer is classified as user of air conditioning, otherwise it is classified as non-user of air conditioning. Parameter Θ must be calibrated considering the average increase in quarterly consumption when the air conditioning is on. The calibration methodology is described in Sect. 5.3.

Figure 3 presents the main steps of the algorithm. Figure 3a presents the original data for a consumer. Figure 3b shows the data after excluding low consumptions (step 2). Figure 3c presents the data after step 3, the class L in blue and the class H in orange. Finally, in Fig. 3d presents the four classes: L_L , L_H , H_L and H_H after step 6. The blue cross is $center_H = (T_H, E_H)$ and the orange cross is $center_L = (T_L, E_L)$. In this case, the value of E_H is significantly greater than the value of E_L , so with a value of Θ barely greater than 1, the condition $E_H/E_L \geq \Theta$ would be met and the consumer would be classified as an air conditioning user.

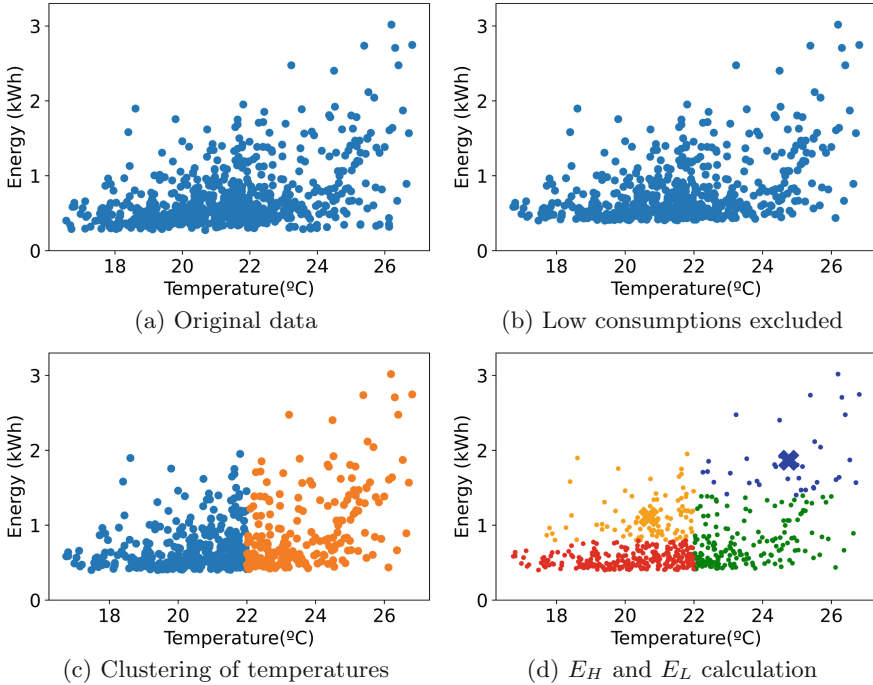


Fig. 3. Main steps of the proposed classification algorithm.

The rationale behind the proposed classification algorithm is that if there is indeed an intensive electrical use associated with thermal comfort, consumption at high temperatures tends to be greater than consumption at low temperatures. By separating the consumption by temperature (low and high) into two sets, and then taking the averages of the highest consumption for each set, if the value associated with high temperatures is significantly higher than that associated with low temperatures, this is due to the use of air conditioning, since no other thermal conditioning device for summer consumes a significant amount of energy. However, if the averages are similar, there is not statistical significance in the reported electricity consumption values with respect to the indoor temperature and the consumer is classified as a non air-conditioning user.

4 Implementation

This section presents the implementation details of the proposed solution.

4.1 Implementation Details

The implementation of the proposed classification algorithm involved four stages, which are described next.

Construction of the Indoor Temperature Curve. The geographical coordinates of each consumer and the weather stations available were processed to find the closest station for each household. The curves of temperature and hourly solar irradiance are obtained from the corresponding station. Next, both curves are converted from hourly to quarterly, assigning each 15-minute time step the value of the corresponding hour. Finally, Eq. 1 is applied using the calibrated parameters to obtain the indoor temperature series. A new quarterly bivariate series S , is constructed with consumption and indoor temperature variables.

Selection of Relevant Consumption Entries. To properly capture the correlation between electricity consumption and internal temperature, the periods of time when there is activity in the household must be considered.

To determine those periods, the method excludes entries from the series S with low consumption. Any entry with a consumption less than 10% of the maximum consumption existing in the considered time horizon is excluded from the series.

Classification by Indoor Temperature. To perform the classification according to the indoor temperature, the *Keras* library is used to apply the the k -means algorithm, using two clusters ($k = 2$). The resulting clusters represent consumptions with low temperatures, and consumptions with high temperatures.

Classification by Consumption. For each cluster found according to the indoor temperature, the k -means algorithm is applied again for classification according to the consumption variable, using two clusters ($k = 2$). The resulting clusters correspond to lower consumptions and higher consumptions values.

4.2 Final Classification

After determining the indoor temperature curve, the selection of relevant consumption entries, the classification by indoor temperature, and the classification by consumption, four sets are obtained. They represent:

1. Samples that have high consumption with high temperatures;
2. Samples that have low consumption with high temperatures;
3. Samples that have high consumption with low temperatures; and
4. Samples that have low consumption with low temperatures.

Sets 2 and 4 are ignored, because the focus is analyzing high consumptions depending on the temperature. The consumption average is calculate for sets 1 and 3. If the quotient between the average obtained for set 1 and the average obtained for set 2 is greater than a specified threshold Θ , the consumer is classified as user of air conditioning in the summer. The value of Θ must be estimated in such a way that it correctly considers the difference in consumption at high and low temperatures. In Sect. 5.3, a value of Θ appropriate for Uruguayan households is estimated.

5 Experimental Analysis and Discussion

This section presents the experimental analysis and discussion of the results.

5.1 Development and Execution Platforms

The proposed solution was implemented on Python. Many scientific libraries and packages were used to handle data, fit the models and visualize results, including Pandas and Matplotlib and Numpy. The experimental evaluation was performed on the high performance computing infrastructure of National Supercomputing Center (Cluster-UY), Uruguay [18].

5.2 General Considerations

The labeled data set, consisting of 29 users, was used for the analysis. In any case, for the calibration of the theta parameter, information from these users was not used, since in this way supervision would be introduced into the algorithm and, given the small amount of available data, this strategy would not have statistical support. For this reason, in this experimental analysis, the set of 29 consumers is used to evaluate the proposed unsupervised methodology but not for designing it. The vast amount of unsupervised data was used to perform an exploratory analysis to design the algorithm.

5.3 Parameter Calibration

To calibrate the parameter Θ , the maximum consumption of the considered household is considered. If it is a consumer that has a relatively low average consumption for the residential sector, the consumption of air conditioning will be relevant. However, if the household has a high base consumption, when turning on an air conditioner, the relative increase in consumption may be small. Therefore, if the parameter Θ is adjusted for households with high average consumption, it will be suitable for households with low consumption. Taking into account that the average energy consumed by an air conditioner in fifteen minutes on extreme situation is 300 Wh (standard power is 1200 W), and assuming an average intense consumption in peak hours of 1000 Wh, an appropriate value

for theta would be $\Theta = 1300/1000$. But the customer does not use the air conditioner in all extreme temperature situations. Therefore, a conservative Θ value could be estimated if it is assumed that on average 1 out of 10 times the customer uses the air conditioner. Then, the average increase in consumption would be $300 \frac{1}{7} 10 = 30$, so $\Theta = 1300/1270 = 1.0237$. This value is the one used for the experimental analysis and subsequent evaluation.

5.4 Analysis of the Proposed Methodology

To validate the proposed methodology, two analysis are presented. On the one hand, to calculate the precision of the algorithm in the set of 29 labeled household data. On the other hand, apply the proposed classification algorithm to the complete set of non-labeled households of the data from ECD-UY and observe the percentage of households that results classified in the category of air conditioning use in summer. The purpose of estimating the number of households that use air conditioning in summer is to compare this value with the continuous survey of households carried out by the National Statistics Institute. To classify the unlabeled data, a sample of 1000 households is randomly taken. In both lines, the procedure described in the Sect. 3.4 is applied to the analyzed set.

5.5 Validation Results on Labeled Data

To validate the results of the application of the algorithm on the labeled data and considering that the sample is balanced, the accuracy metric is used. The result obtained on the set of 29 households is an accuracy of 0.897.

All 16 households were correctly classified as users of air conditioning in summer. However, of the 12 who were classified as not using air conditioning in the summer, 3 were misclassified. This shows that there were no false positives and also allows us to conclude that in order to improve the algorithm it is necessary to avoid the occurrence of false negatives. These considerations are preliminary, due to the small size of the sample used.

5.6 Validation Results on Unlabeled Data

According to the continuous household survey, by 2021, 53% of households have at least one air conditioner. When applying the developed algorithm to the sample of 1,000 non-labeled households, 497 households were classified as users of air conditioning in summer. Bearing in mind that most households that have air conditioning use it in summer, since there are no equivalent alternative thermal conditioning devices in this season (as is the case in winter), the result obtained of 49.7% is reasonable. This result is preliminary since there are no labels and it could happen that, although the total percentage is reasonable, many were misclassified. In any case, it could have happened that this analysis invalidated the algorithm if the percentage obtained was very different from the one shown by the survey.

6 Conclusions and Future Work

This article presented an unsupervised algorithm to detect the use of air conditioning in households, during summer. The proposed approach is valuable because it can be implemented in Uruguay with the existing infrastructure, without incurring large investment costs. The unsupervised algorithm applies the urban data analysis approach. First, it applies a filter by electricity consumption, then a chain of clustering, and finally estimates an indicator related to the variation in consumption with respect to temperature. As a result, the proposed algorithm classifies each consumer as a user (or not) of air conditioning in summer.

The proposed detection methodology was evaluated on a real case study considering data from 29 households in Montevideo, Uruguay. The unsupervised algorithm obtained an accuracy of 0.897 in the considered dataset. This is a promising result, considering the very small cardinality of the set of households.

The main lines of future work are related to improving the accuracy of the air conditioner detection tool, eventually using supervised learning. For this approach, communication with consumers (for example, through a mobile app) would be needed to allow a progressive labeling of households. Another line of future work consists of detecting the use of devices in real time, using the information from various sources and big data analysis.

Acknowledgments. We thank Pedro Moreno, from Universidad Autónoma del Estado de Morelos (México), and Carlos Torres, from Centro Nacional de Investigación y Desarrollo Tecnológico (México), for their valuable comments regarding the thermal inertia model and how to use it in the context of the proposed research.

References

1. Absi, R., Marchandon, S., Bennacer, R.: Thermal-electrical analogy and inertia for thermal performance of building envelopes. In: MATEC Web of Conferences, vol. 330, p. 01037 (2020)
2. Al-Qawasmi, A., Tlili, I.: Energy efficiency and economic impact investigations for air-conditioners using wireless sensing and actuator networks. *Energy Rep.* **4**, 478–485 (2018)
3. Arias, C., Mujica, F., Nicola, C., Menini, A.: Tendencias de diseño, sector materiales de construcción (2019), cámara de Industrias de Uruguay/INEFOP
4. Bonfigli, R., Squartini, S., Fagiani, M., Piazza, F.: Unsupervised algorithms for non-intrusive load monitoring: an up-to-date overview. In: 15th International Conference on Environment and Electrical Engineering (2015)
5. Cengel, Y., Cimbala, J., Turner, R.: Fundamentals of Thermal-Fluid Sciences (SI units). McGraw Hill, New York (2012)
6. Chavat, J., Graneri, J., Nesmachnow, S.: Household energy disaggregation based on pattern consumption similarities. In: Nesmachnow, S., Hernández Callejo, L. (eds.) ICSC-CITIES 2019. CCIS, vol. 1152, pp. 54–69. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-38889-8_5

7. Chavat, J., Nesmachnow, S., Graneri, J.: Non-intrusive energy disaggregation by detecting similarities in consumption patterns. *Revista Facultad de Ingeniería Universidad de Antioquia* (2020)
8. Chavat, J., Nesmachnow, S., Graneri, J., Alvez, G.: ECD-UY, detailed household electricity consumption dataset of Uruguay. *Sci. Data* **9**(1) (2022)
9. Chiang, J., Zhang, T., Chen, B., Hu, Y.: Load disaggregation using harmonic analysis and regularized optimization. In: *IEEE Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, pp. 1–4 (2012)
10. Devarapalli, H., Dhanikonda, S., Gunturi, S.: Non-intrusive identification of load patterns in smart homes using percentage total harmonic distortion. *Energies* **13**(18), 4628 (2020)
11. Hart, G.: Nonintrusive appliance load monitoring. *Proc. IEEE* **80**(12), 1870–1891 (1992)
12. Hu, M., Xiao, F.: Price-responsive model-based optimal demand response control of inverter air conditioners using genetic algorithm. *Appl. Energy* **219**, 151–164 (2018)
13. Kelly, J., Knottenbelt, W.: Neural NILM: Deep Neural Networks Applied to Energy Disaggregation. In: *2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, pp. 55–64 (2015)
14. Kolter, J., Johnson, M.: Redd: A public data set for energy disaggregation research. In: *Workshop on Data Mining Applications in Sustainability*, pp. 59–62 (2011)
15. Massobrio, R., Nesmachnow, S.: Urban mobility data analysis for public transportation systems: a case study in Montevideo. *Uruguay. Appl. Sci.* **10**(16), 1–20 (2020)
16. Muraña, J., et al.: Negotiation approach for the participation of datacenters and supercomputing facilities in smart electricity markets. *Program. Comput. Softw.* **46**(8), 636–651 (2020)
17. Muraña, J., Nesmachnow, S.: Simulation and evaluation of multicriteria planning heuristics for demand response in datacenters. *Simulation*, p. 003754972110200 (2021)
18. Nesmachnow, S., Iturriaga, S.: Cluster-UY: collaborative scientific high performance computing in Uruguay. In: Torres, M., Klapp, J. (eds.) *ISUM 2019. CCIS*, vol. 1151, pp. 188–202. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-38043-4_16
19. Porteiro, R., Chavat, J., Nesmachnow, S.: A thermal discomfort index for demand response control in residential water heaters. *Appl. Sci.* **11**(21), 10048 (2021)
20. Porteiro, R., Chavat, J., Nesmachnow, S., Hernández-Callejo, L.: Demand response control in electric water heaters: evaluation of impact on thermal comfort. In: Nesmachnow, S., Hernández Callejo, L. (eds.) *ICSC-CITIES 2020. CCIS*, vol. 1359, pp. 74–89. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-69136-3_6
21. Porteiro, R., Hernández-Callejo, L., Nesmachnow, S.: Electricity demand forecasting in industrial and residential facilities using ensemble machine learning. *Revista Facultad de Ingeniería Universidad de Antioquia* (2020)
22. Porteiro, R., Nesmachnow, S., Hernández-Callejo, L.: Short term load forecasting of industrial electricity using machine learning. In: Nesmachnow, S., Hernández Callejo, L. (eds.) *ICSC-CITIES 2019. CCIS*, vol. 1152, pp. 146–161. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-38889-8_12
23. Zafar, R., Mahmood, A., Razaq, S., Ali, W., Naeem, U., Shehzad, K.: Prosumer based energy management and sharing in smart grid. *Renew. Sustain. Energy Rev.* **82**, 1675–1684 (2018)