Context Analysis in Energy Resource Management of Residential Buildings

Bruno Madureira[1], Tiago Pinto[2], *Member, IEEE*, Filipe Fernandes[1], Zita Vale[1], *Senior Member, IEEE*

[1]GECAD research group, Polytechnic of Porto (ISEP/IPP), Porto, Portugal

{brmad, fijgf, zav}@isep.ipp.pt

[2]BISITE research group, University of Salamanca, Spain

tpinto@usal.es

***Abstract*—This paper presents a context analysis methodology to improve the management of residential energy resources by making the decision making process adaptive to different contexts. A context analysis model is proposed and described, using a clustering process to group similar situations. Several clustering quality assessment indices, which support the decisions on how many clusters should be created in each run, are also considered, namely: the Calinski Harabasz, Davies Bouldin, Gap Value and Silhouette. Results show that the application of the proposed model allows to identify different contexts by finding patterns of devices' use and also to compare different optimal k criteria. The data used in this case study represents the energy consumption of a generic home during one year (2014) and features the measurements of several devices' consumption as well as of several contextual variables. The proposed method enhances the energy resource management through adaptation to different contexts.**

Keywords — Artificial Intelligence, Context Analysis, Data-Mining, House Managament, Residential Energy Management

I. Introduction

The current electric power grid only allows a reductive single direction interaction between the consumer and distribution service providers. With the recent technological development, the growing consumption of electricity and the use of renewable energy sources, it is necessary to adapt the grid to the energy needs of the new century. In order to respond to this challenge, the concept of smart grid emerged, an intelligent network that enables a two-way interaction between the consumer and distribution service providers, including a variety of new operational and measurement technologies [1]. Smart Grid can then be defined as a network that generates the search for sustainable energy, reliable and efficient manner, constructed to facilitate the integration of all components [2].

The development of smart grids requires, at the same time, the development of new other concepts such as the smart meter or the smart home. The concept of smart home can be seen then as the lowest level of management of energy resources in a smart grid. In the near future, it will be possible for residential consumers to manage their consumption and production of electricity autonomously and intelligently, considering all equipment of consumption, production and electric energy storage inserted in their homes. But a smart home system to be considered an intelligent system, should have at least the following three elements: an internal communication network, intelligent control systems, and automation [3]. It should, however, also interact with external factors such as changes in energy prices, giving the residential resource management system better decision making according to these interactions.

To this end, the implementation of residential management system is essential, House Management System (HMS), able to effectively manage the full and micro generation consumption, maximizing energy efficiency, without ignoring the comfort levels required by the user, are urgently required [4]. In order to improve the efficiency of HMS, it is essential to acquire information about consumer behavior to adjust the management process according to your preferences and needs. Hence, decision support systems for residential energy management should be made taking into account contexts of use of the different devices in the house. In this sense, the intelligent system resource management of the house should be context aware, being able to feel and react based on the surrounding environment.

Taking into account the needs identified in this field, the aim of this paper is to study and develop a method for analyzing, identifying and defining different contexts of resource usage in homes. For this, a data-mining approach is used, namely a clustering method [5], which groups different historical events related to consumption measures in a home, the associated devices usage at each moment, and related contextual variables (e.g. temperature, humidity, external luminosity, etc). This clustering process allows identifying situations with similar characteristics, and thus enable the identification of different contexts of consumption and devices' usage. The proposed method is used to improve home resource management methods of the Supervisory control and data acquisition (SCADA) House Intelligent Management (SHIM) system [3], giving them ability to adapt to different contexts.

After this introductory section, section II presents an overview of related work in the field, section III describes the proposed method, section IV presents the achieved results, and section V discusses the main conclusions of this work.

The present work was done and funded in the scope of the following projects: European Union's Horizon 2020 research and innovation programme, under the Marie Sklodowska-Curie grant agreement No 703689 (project ADAPT); EUREKA - ITEA2 Project M2MGrids (ITEA-13011), Project SIMOCE (ANI|P2020 17690), and has received funding from FEDER Funds through COMPETE program and from National Funds through FCT under the project UID/EEA/00760/2013

II. RELATED WORK

Studies of house management systems began to be developed in the eighties, it consisted of primitive remote control equipment and data relating to energy management and security applications. During this period, one work stood out, the Residential Energy Usage Comparison project by using innovative techniques for analyzing and monitoring data with collection done in five minute intervals. Were also used different models in this project, as the hourly load model and model-peak hours in order to assess how to load the wave and the resulting economic impacts [6].

In 1989, it was developed a system by [7] that allowed the user to manage the use of electricity for heating or cooling of the housing through an interface, taking into account the comfort of the user and the price of energy. During the nineties the study on energy management in residential environment continued and in 1991 the concept of a system called Smart Home [8] emerged, describing the communication and new control methods, system architecture, protocols and structures needed to implement it. Another big development was done in 1999, introducing the idea of connecting services and an equipment through a home network [9]. Currently, there are intelligent systems such as the system developed by [10] that specifies access control for personal wireless networks with low transmission rates.

The smart home concept evolved and according to [11] today for a home to be considered intelligent should include three main elements: the internal communication network, intelligent control systems and automation housing. However, this work will be developed in order to improve housing resources management methods included in SHIM. Which is a simulation platform, developed and implemented in GECAD. This platform is designed to operate as a housing management system, including advanced features that allow the control of various types of loads in a home. It also allows the control of micro units and storage units such as electric vehicles. Another interesting feature of this platform is the ability to manage real and virtual loads in the same simulation [3].

A smart home should provide the level of comfort required in each context considering the minimum energy consumption and minimum price operation. But the uncertainty of the location of the residents in the house makes it difficult to identify their activities and therefore hinders the intelligent resource management of the home. There are also some activities that have common events that complicate the determination of which of the contexts are inserted, for example, turning on the stove might imply the resident is going to cook food but to know if he's making breakfast, lunch or dinner at least another variable is needed.

From here emerges the challenge of finding variables that can help to overcome these problems in order to successfully identify and differentiate contexts, to later define the priorities to be given to each device to perform a good resource management without ignoring, as was previously stated, the comfort of the user [12]. These variables could be categorized into the following indicators: user indicators, time indicators, load indicators and comfort indicators.

In the United States, Americans spend about 90% to 92% [13] of their time indoors. The same applies to most of the populations in developed countries. This has led to a greater awareness of the importance of comfort and quality of life indoors. In recent years there have been various studies on the characteristics which influence comfort indoors. The study conducted by [13] specifically identified through questionnaires to the Danish population that the comfort in residential environment is fundamentally organized in the following areas: lighting, air quality, acoustic and thermal comfort.

Lighting comfort is defined as people's satisfaction with the luminous environment they occupy. According to [14], access to sunlight prevents diseases caused by lack of vitamin D which affects the visual perception and the mood of residents. The exposure to sunlight improves the mental health of people and their productivity and that the use of artificial light for long hours is normally indicative of a decrease in the visual comfort level.

Air quality is a term that describes the quality of the air inside and surrounding the residence, related to the health and comfort of the occupants. To ensure this, there are heat, ventilation and air conditioning (HVAC) systems which purpose is to provide clean air to residents and remove odors, particles and pollutants through the use of hoods or attenuating them to acceptable levels. The EPA has these pollutants classified as the six principal pollutants (or "criteria pollutants" - as they are also known): particle pollution (often referred to as particulate matter), photochemical oxidants and ground-level ozone, carbon monoxide, sulfur oxides, nitrogen oxides, and lead. [15] The standards to assure and maintain the National Ambient Air Quality Standards (NAAQS) are listed in the Clean Air Act, which was last amended in 1990.

Acoustic comfort is defined as the absence of audible noise or as the lowest noise sound as possible to ensure the comfort of the users. How humans perceive sounds and loudness is a subjective measure. However, it's possible to create a comfortable environment by controlling objective measures like decibel level (sound pressure), reverberation time, and the sound reflection and damping properties of materials [16] In the perspective of energy management in residential buildings, acoustic quality does not seem to be very relevant but if there is microgeneration unit in the smart home. Having an acoustic sensor, allowing the monitoring of the noise is recommended. This is particularly important during night time where the presence of loud noises can prevent residents from having good sleep quality.

Thermal comfort can be defined as the perception of thermal well-being in an environment. In determining thermal comfort, usually it's considered six factors divided into two categories: personal factors (metabolic rate and insulating clothing) and environmental factors (mean radiant temperature, air velocity and humidity) [17] which are used in the most recognized thermal comfort models like Predicted Mean Vote (PMV). But due to comfort being a subjective measure that doesn't just depend on factors like the metabolic rate and the isolation of clothing each habitant, but also their behavior, it originated adaptive models that take into account the human actions such as taking off clothing or opening a window.

III. Proposed Methodology

The activity detection problem at the residential environment is characterized by a great diversity and complexity of possible types of activities. This implies that there are no pre-defined outputs and a large amount of data obtained by multiple sensors in the smart home. This section describes the methodology used to develop a method of analysis and definition of contexts. The proposed method is divided into four steps, as presented in Figure 1.

Fig. 1. Proposed methodology overview

A. Problem Definition

The activity detection problem at the residential environment is characterized by a great diversity and complexity of possible types of activities. This implies that there are no pre-defined outputs and a large amount of data obtained by multiple sensors in the smart home.

B. Data Gathering

The collected data, provided by GECAD, represent the energy consumption of a generic home during the year 2014 (64 samples) and feature the measurements of the following variables: month, day of the week, hour, season, temperature, number of residents inside the home, total power usage, indication ON/OFF of lighting by division (bedroom, living room, hall, kitchen) and devices (microwave, oven, coffee machine, fridge 1, fridge 2, washing machine, HVAC, water heater, room TV, living room TV).

C. Normalization

The process of normalization of the collected data is in most cases required prior to the application of a data mining algorithm, improving its performance and results. Magnitude differences or ranges of different variables can result in some attributes having more "weight" than others. This process aims to prevent this problem by reducing the scale of the collect data to pre-determined values, usually between 0 and 1 or between -1 and 1.

Assuming each variable is disposed in a database by column, the standard normalization method is applied by dividing the value of each measurement by the maximum value of the corresponding column, which will reduce the scale of all the data to values between 0 and 1.

D. Data Mining

Clustering methods seek to organize a set of objects into clusters so that the objects inserted in a given cluster have high similarity level between them, and a low level of similarity with objects belonging to different groups [18].

A method of the unsupervised learning branch of machine learning has been chosen for the processing of the data; more precisely, a grouping method by iterative optimization (partitional clustering), k-means clustering [19]. K-means has been chosen because of the characteristics of the identified problem and because this model can be applied to large databases without suffering major counterparts, which is an essential requirement for this problem.

1) *Clustering process*

Given a set of observations $(x_1, x_2, \ldots, x_n)$ where n is the number of observations considered and in which each observation is a d -dimensional vector. The clustering process aims to perform the division of n observations into k clusters $(C_1, C_2, \ldots, C_k)$ to minimize the sum of square errors, WCSS (Within Cluster Sum of Squares) [5].

$$min \sum_{i=1}^{k} \sum_{x \in C_i} \|x - \mu_i\|^2 \tag{1}$$

Where μ_i is the average points in C_i and C_i being the centroid of the cluster. With the aim of minimizing the equation (1), the iterative k-means clustering process consists of two steps:

- Association, each observation x is linked to the cluster C whose mean value can WCSS minimum;
- Update the centroids, considering new observations associations, new means of each cluster values are calculated, given a new centroid μ_i.

The algorithm execution continues until the convergence process has finishes, in other words, when the combination of observations to different clusters no longer varies, finding a fixed point or local minima.

2) *Optimal k*

K-means requires the specification of the number of clusters (k) beforehand. The success or failure of the method will depend on this indication. Unfortunately, there is still no general procedure to find the correct number of clusters. However, there are some criteria to suggest an optimal value of k, i.e., the number of groups that allows the best grouping for the collected data. These criteria are mainly based on two aspects:

- Compaction: a group of measures that assess the compactness of the cluster based on variance. Measures how closely related objects in a cluster are;
- Separation: a group of measures that assess the distinctive as well or separate a cluster is from other clusters.

The following criteria are considered: Calinski Harabasz [20], Davies Bouldin [21], Gap Value [22] and Silhouette [23].

The Calinski Harabasz criteria (CH) also known as variance ratio criterion (VRC) evaluates the quality of the grouping based on the mean variance between cluster sum of squares (BCSS) and within cluster sum of squares (WCSS). The equation (2) represents the CH index:

$$i_{CH} = \frac{trace\ (SS_B)}{trace\ (SS_W)} \times \frac{n-k}{k-1} \tag{2}$$

Where SS_B is the overall between-cluster variance, SSW is the overall within-cluster variance, k is the number of clusters, and n is the number of observations. The overall between-cluster variance SS_B is defined as (3):

$$SS_B = \sum_{i=1}^{k} n_i \|m_i - m\|^2 \quad (3)$$

Where k is the number of clusters, m_i is the centroid of cluster i, m is the overall mean of the sample data, and $\|m_i - m\|$ is the L2 norm (Euclidean distance) between the two vectors. The overall within-cluster variance SSW is defined as (4):

$$SS_W = \sum_{i=1}^{k} \sum_{x \epsilon c_i} \|x - m_i\|^2 \quad (4)$$

Where k is the number of clusters, x is a data point, c_i is the i_{th} cluster, m_i is the centroid of cluster i, and $\|x - m_i\|$ is the euclidean distance between the two vectors. Well-defined clusters have a large between-cluster variance (SSB) and a small within-cluster variance (SSW). The larger the VRCk ratio, the better the data partition. To determine the optimal number of clusters, maximize VRCk with respect to k [20].

The Davies-Bouldin criterion (DB) is based on a ratio of within-cluster and between-cluster distances. The Davies-Bouldin index is defined as (5):

$$DB = \frac{1}{k} \sum_{i=1}^{k} max_{j \neq i}\{D_{i,j}\} \quad (5)$$

Where $D_{i,j}$ is the within-to-between cluster distance ratio for the ith and jth clusters. In mathematical terms (6):

$$D_{i,j} = \frac{(\bar{d}_i + \bar{d}_j)}{d_{i,j}} \quad (6)$$

Where $\bar{d}_i$ is the average distance between each point in the ith cluster and the centroid of the ith cluster. $\bar{d}_j$ is the average distance between each point in the jth cluster and the centroid of the jth cluster. $d_{i,j}$ is the Euclidean distance between the centroids of the ith and jth clusters.

The maximum value of $D_{i,j}$ represents the worst-case within-to-between cluster ratio for cluster i. The optimal clustering solution has the smallest Davies Bouldin index value [21].

A common graphical approach to cluster evaluation involves plotting an error measurement versus several proposed numbers of clusters, and locating the "elbow" of this plot. The "elbow" occurs at the most dramatic decrease in error measurement. The gap criterion (GAP) formalizes this approach by estimating the "elbow" location as the number of clusters with the largest gap value. Therefore, under the gap criterion, the optimal number of clusters occurs at the solution with the largest local or global gap value within a tolerance range. The gap value is defined as (7):

$$Gap_n(k) = E_n^*\{\log(W_k)\} - \log(W_k) \quad (7)$$

Where n is the sample size, k is the number of clusters being evaluated, and W_k is the pooled within-cluster dispersion measurement which can be represented by (8):

$$W_k = \sum_{r=1}^{k} \frac{1}{2n_r} D_r \quad (8)$$

Where n_r is the number of data points in cluster r, and D_r is the sum of the pairwise distances for all points in cluster r. The expected value $E_n^*\{\log(W_k)\}$ is determined by Monte Carlo sampling from a reference distribution, and log(W_k) is computed from the sample data. The gap value is defined even for clustering solutions that contain only one cluster, and can be used with any distance metric. However, the gap criterion is more computationally expensive than other cluster evaluation criteria, because the clustering algorithm must be applied to the reference data for each proposed clustering solution [22].

The silhouette value for each point is a measure of how similar that point is to points in its own cluster, when compared to points in other clusters. The silhouette value for the ith point, is defined as (9):

$$s_i = (a_i - b_i)/\max(a_i, b_i) \quad (9)$$

Where a_i is the average distance from the ith point to the other points in the same cluster as i, and b_i is the minimum average distance from the ith point to points in a different cluster, minimized over clusters [23].

The silhouette value ranges from -1 to +1. A high silhouette value indicates that i is well-matched to its own cluster, and poorly-matched to neighboring clusters. If most points have a high silhouette value, then the clustering solution is appropriate. If many points have a low or negative silhouette value, then the clustering solution may have either too many or too few clusters.

IV. Results

The data used in this case study represents the energy consumption of a generic home during the year 2014 (64 samples) and features the measurements of the following variables: month, day of the week, hour, season, temperature, number of residents inside the home, total power usage, indication ON/OFF of lighting by division (bedroom, living room, hall, kitchen) and consumption of devices (microwave, oven, coffee machine, fridge 1, fridge 2, washing machine, HVAC, water heater, room TV, living room TV).

The set of 64 samples of 21 variables is decomposed into groups that represent different situations (or contexts). The considered contexts are of different natures, trying to gather information about different situations. Out of several scenarios that have been analyzed, this focuses on the experimental findings considering three scenarios that relate specific variables, namely: Temperature vs. HVAC, Hour vs. Power and Number of Residents vs. Power.

A. Temperature vs. HVAC

The purpose of this grouping is to find patterns in the usage of the HVAC system by residents depending on the temperature, the following variables were selected:

- Outside temperature
- Indication ON / OFF of the HVAC system

Although there is not a consensual optimal *k* among all four criteria, two of them (DB and S) indicate the same number of clusters that permits the best clustering results, as shown in Table I.

TABLE I
CRITERIA COMPARISON – TEMPERATURE VS. HVAC

Number of clusters	Criteria
10	*CH*
3	*DB*
10	*GAP*
3	*S*

Analyzing the results obtained and the context of this scenario, it can be said that the criteria Calinski Harabasz and Gap Statistic overestimated the optimal *k* value.

Figure 2 presents the clustering results using an optimal *k* equal to 3 (resulting from DB criterion), considering the xx-axis as the outside temperature (normalized) and the yy-axis as the indication ON / OFF of the HVAC system. Each color represents the observations that have been grouped in each cluster.

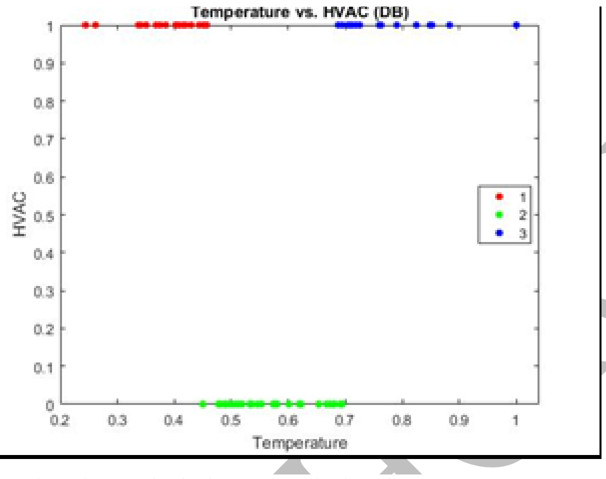

Fig. 2. Clustering results for k = 3 (DB criterion) for Temperature vs. HVAC

From Figure 2 it is visible that the three achieved groups represent: 1 - the use of the HVAC system when the temperature is low; 2 - the HVAC system turned OFF; and 3 – the use of the HVAC system when the temperature is high. Three significant contexts of use have, therefore, been identified.

B. Hour vs. Power

This clustering process aims to find power usage patterns by relating the total power consumption with the hour of day. To this end, the following variables were selected:

- Hour of the day
- Total power consumption

Excluding the Calinski Harabasz criterion, the remaining three criteria point towards an optimal k of four or five clusters, as shown in Table II.

TABLE II
CRITERIA COMPARISON – HOUR VS. POWER

Number of clusters	Criteria
10	*CH*
5	*DB*
4	*GAP*
4	*S*

Figure 3 presents the clustering results using an optimal k equal to 4 (resulting from GAP criterion), considering the xx-axis as the hour (normalized) and the yy-axis as the total power consumption (normalized). Each color represents the observations that have been grouped in each cluster.

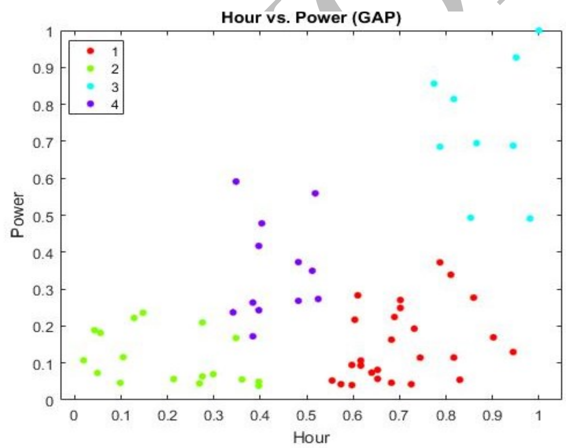

Fig. 3. Clustering results with k = 4 (GAP criterion) for Hour vs. Power

These four groups verified from Figure 3 represent the consumption in: morning, noon, evening and night. It is possible to see that the hours of higher power consumption are mainly in the noon and night groups. Being the maximum total power consumption value during nighttime and the minimum during the morning and afternoon.

C. Number of Residents vs. Power

This grouping aims to find usage patterns by relating the total power consumption with the number of people in the residence at a given time. In order to achieve this goal, the following variables were selected:

- Number of residents
- Total power consumption

Similarly to the first scenario, despite not existing a consistent optimal *k* among all four criteria, DB and S both indicate 3 groups as the best grouping, as shown in Table III.

TABLE III
CRITERIA COMPARISON – HOUR VS. POWER

Number of clusters	Criteria
8	*CH*
3	*DB*
10	*GAP*
3	*S*

Analyzing the results from Table III, it can be said that the criteria Calinski Harabasz and Gap Statistic overestimated again the value of optimal k.

Figure 4 presents the clustering results using an optimal *k* equal to 3 (resulting from S criterion), considering the xx-axis as the number of residents (normalized) and the yy-axis as the total power consumption (normalized). Each color represents the observations that have been grouped in each cluster.

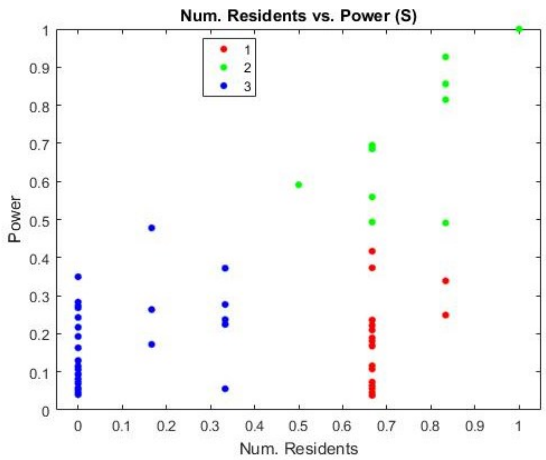

Fig. 4. Results with k = 3 (S criterion) for Number of Residents vs. Power

From Figure 4 it is visible that three groups are achieved. The minimum total power consumption value is verified when the number of residents in the house of the household is below 40% and maximum when all members of the household are at home at the same time.

V. CONCLUSIONS

This paper focuses on finding factors that influence decision making in the residential environment. The main objective of the study is therefore, the development of context analysis methodology that can be able to improve the management of residential energy resources by making the decision making process adaptive to different contexts. A model has been proposed and described, using a clustering process and a set of clustering quality assessment indices, which support the decisions on how many clusters should be created in each run.

Results show that the application of the proposed model allows to find patterns of usage and compare different optimal k criteria. The considered combinations of variables have allowed to identify different contexts and take conclusions on the relationship between Temperature and usage of HVAC; Hour of the day and Consumption Power; and Number of Residents and Total Consumption Power. Results also show that the DB and S indexes are, in general, those that are able to achieve the most promising results. These indexes enable reaching a smaller amount of different clusters, thus allowing energy management systems to deal with a smaller number of well-defined distinct contexts, which eases the decision making process, as well as the interpretation of results, in opposed to having a large number of different contexts, with not so much difference between them.

REFERENCES

[1] Lund, H., “Renewable Energy Systems, Renewable Energy Systems – A Smart Energy Systems Approach to the Choice and Modeling of 100% Renewable Solutions”, Academic Press, 2nd Edition, May 2014.

[2] U.S. Department of Energy, “What is the Smart Grid?”. [Online]. Available: https://www.smartgrid.gov/the_smart_grid/smart_grid.html

[3] F. Fernandes, H. Morais, P. Faria, Z. Vale, C. Ramos, SCADA house intelligent man-agement for energy efficiency analysis in domestic consumers, in: 2013 IEEEPES Conference on Innovative Smart Grid Technologies (ISGT Latin America). 2013.

[4] Y. Fei, B. Jiang, Dynamic residential demand response and distributed generation management in smart microgrid with hierarchical agents. 2011.

[5] Anil K. Jain, Data clustering: 50 years beyond K-means. Pattern Recognition Letters. 2010.

[6] J. T. Powers, M. S. Shirilau, R. T. Uhlaner, and B. A. Smith, "Design of a utility competitive assessment experiment: the Residential Energy Usage Comparison Project", IEEE Transactions on Power Systems, 1988.

[7] T. C. Matty, "Advanced energy management for home use", IEEE Transactions on Consumer Electronics, 1989.

[8] H. B. Stauffer, "Smart enabling system for home automation", IEEE Transactions on Consumer Electronics, 1991.

[9] D. Gann, J. Barlow, and T. Venables, "Digital Futures: Making Homes Smarter", Chartered Institute of Housing, 1999.

[10] H. Dae-Man and L. Jae-Hyun, "Design and implementation of smart home energy management systems based on zigbee", IEEE Transactions on Consumer Electronics, 2010.

[11] B. Jiang and Y. Fei, "Dynamic Residential Demand Response and Distributed Generation Management in Smart Microgrid with Hierarchical Agents", Energy Procedia, 2011.

[12] Tiago Pinto, Zita Vale, Tiago M. Sousa, Isabel Praça, Negotiation context analysis in electricity markets, Energy, Volume 85, 1 June 2015, Pages 78-93, ISSN 0360-5442

[13] US Environmental Protection Agency, “Buildings and their impact on the Environment: A Statistical Summary, 2009” report. [Online]. Available: https://archive.epa.gov/greenbuilding/web/pdf/gbstats.pdf

[14] Bernstein, T. Municipal green building policies: Strategies for transforming building practices in the private sector. Environmental Law Institute. 2008.

[15] Monika Frontczak, Human comfort and self-estimated performance in relation to indoor environmental parameters and building features. DTU Civil Engineering report. 2011.

[16] Peng Xue, C.M. Mak, H.D. Cheung, The effects of daylighting and human behavior on luminous comfort in residential buildings: A questionnaire survey, Building and Environment, Volume 81, 2014.

[17] US Environmental Protection Agency, “Criteria Air Pollutants”. [Online]. Available: https://www.epa.gov/criteria-air-pollutants.

[18] B Teixeira, F Silva, T Pinto, I Praça, G Santos, Z Vale, Data mining approach to support the generation of Realistic Scenarios for multi-agent simulation of electricity markets, Intelligent Agents (IA), 2014 IEEE Symposium on, 2014

[19] Faia, Ricardo; Pinto, Tiago; Vale, Zita; ,Dynamic fuzzy estimation of contracts historic information using an automatic clustering methodology,International Conference on Practical Applications of Agents and Multi-Agent Systems,270-282,2015,Springer International Publishing

[20] Mathworks, “Calinski-Harabasz criterion clustering evaluation object”. [Online].Available:http://www.mathworks.com/help/stats/clustering.evaluation.calinskiharabaszevaluation-class.html

[21] Mathworks, “Davies-Bouldin criterion clustering evaluation object”. [Online].Available:http://www.mathworks.com/help/stats/clustering.evaluation.daviesbouldinevaluation-class.html

[22] Mathworks, “Gap criterion clustering evaluation object”. [Online].http://www.mathworks.com/help/stats/clustering.evaluation.gapevaluation-class.html

[23] Mathworks, “Silhouette criterion clustering evaluation object”. [Online].http://www.mathworks.com/help/stats/clustering.evaluation.silhouetteevaluation-class.htm