

Behavioral similarity of residential customers using a neural network based on adaptive resonance theory



Daniela S. Justo*, Carlos R. Minussi, Anna Diva P. Lotufo

Electrical Engineering Department, Faculty of Engineering of Ilha Solteira (FEIS), UNESP, Univ Estadual Paulista "Júlio de Mesquita Filho", Av. Brasil 56, PO Box 31, 15385-000, Ilha Solteira, SP, Brazil

ARTICLE INFO

Keywords:

Household
Electricity consumption
Consumer similarity
Fuzzy ART neural network
Adaptive resonance theory

ABSTRACT

This article presents a methodology aiming at the comprehension and analysis of the residential electricity consumption habitual behavior by means of a similarity analysis, based on the use of an ART (Adaptive Resonance Theory) neural network, which is a neural network composed of two fuzzy ART modules whose training is performed in an unsupervised mode. ART neural networks are stable and plastic and these properties, combined with the processing of essentially binary data, give the neural system a wide capacity for producing objectives that may be easily modified to satisfy requirements predetermined by consumer. The expected result is to obtain information regarding the similarity of consumers. Thus, some benefits may be derived by consumers, such as improved habits of electricity consumption and better strategies for negotiating more favorable rates, especially in the case of smart grid systems. In this new electricity sector paradigm, there is a strong consumer trend for free choice among electricity suppliers. This methodology also benefits load forecasting studies at grid points, where there is greater uncertainty, e.g., the busbars that are closest to consumers, i.e., the uncertainties in the context of the total load forecasting system are increased from the global load to the final consumer.

1. Introduction

Electricity consumption is constantly increasing, especially in the residential sector. This sector presents significant importance in the demand for electricity caused by the growth of consumer units and the increase in electronic equipment use. As stated in (The World Bank (IBRD-IDA), 2014), electricity consumption increased by approximately 24% *per capita* in Brazil from 2006 to 2013, whereas the increase in the global demand for electricity was approximately 14% over the same period. The expectation is that this demand will continue to grow in the coming years. In the Brazilian residential sector, this is caused by the growing number of consumers (2% per year) and the growth of consumption per consumer (2.2% per year), with a demand forecast of 4.3% per year from 2014 to 2024 (Energy Research Company (EPE), 2015a).

To meet this growing demand, obviously, the country must invest in new sources of electricity (e.g., hydroelectric, wind power, photovoltaic sources). In partnership with consumers, it also must heavily invest in developing new electricity consumption habits. The efficient control of consumption by consumers thus comes as a solution to save energy. A more effective participation of consumers in the energy market is a

subject of studies because it plays a crucial role in the equation demand for electricity and the expansion of the energy sector. The role of the final consumer in this context includes energy efficiency and distributed generation actions, which can help reduce the need for expanding the Brazilian electricity sector over the long term (Energy Research Company (EPE), 2015b).

Moreover, one of the most significant challenges for electricity distribution companies is reducing electricity consumption peaks. To alleviate this problem, these institutions have invested in modern technologies to create a smart grid (Fan & Borlase, 2009).

Given the aforementioned scenario, it is necessary to better understand the residential consumption habits to identify them as standard and to predict future consumption aiming at maximum efficiency in the clustering of similar consumers. Foreknowledge of the load in an electric power system with highest accuracy is essential. That is, the load dictates the operating modes of the system and the strategies that must be implemented to meet it continuously, economically (both for companies and consumers) and with quality (e.g., voltage, frequency, waveform, selected as appropriate according to the electrical energy).

As a consequence, this article aims to present a neural system that makes it possible to know residential consumers' profiles regarding

* Corresponding author.

E-mail addresses: sbizera@yahoo.com (D.S. Justo), minussi@dee.feis.unesp.br (C.R. Minussi), annadiva@dee.feis.unesp.br (A.D.P. Lotufo).

their consumption habits and ownership of electronic devices, by means of an artificial neural network (ANN) ART (Carpenter, Grossberg, & Rosen, 1991a) used to group consumers together. The preference for this class of ANN is due to its excellent capacity for stability, plasticity and low computational cost for implementation. Stability refers to the guarantee of always producing solutions; i.e., there is no problem regarding poor convergence. Plasticity is the characteristic that enables continuous training without destroying the previously acquired knowledge and it is a property rarely observed in most ANN available in scientific and technical literature. Data grouping addresses the identification of groups in dataset (objects) according to similarity in objects belonging to this set. These groups are named clusters. This is a technique from an unsupervised paradigm that consists of identifying relevant information into data without the presence of an external agent (teacher) to guide learning; i.e., it observes, finds out and performs an exploratory data analysis. The dataset used was obtained by means of a survey questionnaire handled to a cluster of Brazilian residential consumers, with questions about socioeconomics, demography, consumption and ownership, and habits of using electronic devices.

2. Use of clustering techniques for characterization of residential electricity consumers

Clustering techniques are applicable when there are no classes to predict but there is the need to separate objects into groups (Witten & Frank, 2011). The data clustering process involves several steps from data preparation (objects) to interpretation of the results obtained by the technique (clusters) (Jain, Murty, & Flynn, 1999). One of the steps is defining the similarity and dissimilarity measures among objects. These measures are used to decide the allocation of an object to a cluster or to unite or divide clusters. A review of several types of similarity measures may be found in (Jain & Dubes, 1988). In this paper, Manhattan Distance and Euclidian Distance have been used because they are appropriate to the quantitative nature of the features that describe objects. Another step of the clustering process is the algorithm selection. The present work is based on using the unsupervised fuzzy ART ANN algorithm Carpenter et al., (1991a) and Carpenter, Grossberg, and Roden, (1991b).

In the electricity sector, a large amount of research has been carried out using the unsupervised data clustering approach for residential consumers, aiming to understand how they use electricity and what it can cause in low-voltage networks (Beckel, Sadamori, Staake, & Santini, 2014; Stephen & Galloway, 2012; Chicco, 2012). Several studies have been conducted regarding available and emergent technologies from the electricity sector (such as smart meters), and they propose solutions that will contemplate the expected tools not only for consumers but also for electricity distribution companies in the near future. For these companies, it is crucial to know the residential consumption profiles to obtain parameters in order to plan and manage energy supply with greater efficiency, safety and stability. For residential consumers, having a better knowledge about their consumption will stimulate decision-making to save energy, especially at peak times when the cost is higher.

Beckel et al. (2014) and McLoughlin, Duffy, and Conlon (2012) proposed tools to identify which household characteristics may be related to electricity consumption habits. Flath, Nicolay, Conte, and Dinther (2012) presented the implementation and evaluation of a clustering analysis approach to help companies define rates more accurately according to consumption profiles. The definition of consumer profiles was the aim of Räsänen and Kolehmainen (2009) and Stephen, Mutanen, Galloway, Burt, and Järventausta (2014).

To identify consumers who may be candidates for energy saving programs such as Demand Side Management and Demand Response programs, several authors have used clustering techniques and proposed different solutions (Albert & Rajagopal, 2013; Cao,

Beckel, & Staake, 2013; Dent, Craig, & Aickelin, 2012; Kwac, Flora, & Rajagopal, 2014). The identification of nontechnical losses caused by the violation of restrictions of the electricity network was also a subject of studies using data clustering methods (Angelos, Saavedra, Cortés, & Souza, 2011; Babu, Murthy, & Sivaiah, 2017; Jokar, Arianpoo, & Leung, 2015).

To implement the aforementioned solutions, different unsupervised clustering techniques, described below, were applied. Studies Benitez, Quijano, Diez, and Delgado, (2014), Dent et al. (2012), Flath et al., (2012), Räsänen and Kolehmainen (2009) and McLoughlin, Duffy, and Conlon, (2015) are about the use of data clustering of residential consumers by means of *k-means* and *k-medoid* methods. In Stephen and Galloway (2012), Stephen et al. (2014) and Haben, Singleton, and Grindrod (2016), the authors used the clustering technique called *Finite Mixture Models* (FMM). The SOM (*Self-Organizing Maps*) technique was applied to studies (Beckel et al., 2014) and (McLoughlin et al., 2015).

Equally important to choosing the method to be used in data clustering is the definition of characteristics that will be used in the clustering. In (McLoughlin et al., 2012) and (McLoughlin et al., 2015), the specifics of households and the way consumers use electricity may be characterized by means of individual consumer attributes.

ÓDoherty, Lyons, and Tol (2008) applied a generalized linear model to infer the relation between the ownership of electronic equipment and the electricity consumption. The analysis presented results that had great significance for electricity consumption such as characteristics of households (location, price and type) and residents' characteristics (income, age, residence time, social class and type of ownership).

In)Yohanis, Mondol, Wright, and Norton (2008), a significant relation of attributes related to characteristics of households and electricity consumption was used. Electricity load profiles were characterized based on the type of household, usable area, number of residents, number of bedrooms, ownership of devices, age of residents and family income.

Aydinalp, Ugursal, and Fung (2004), Aydinalp, Ugursal, and Fung (2002) and Aydinalp-Koksal and Ugursal (2008) developed an ANN to model the electricity consumption of electronic appliances, lighting and cooling of residences. The variables used in the ANN model that were developed and influenced electricity consumption were the ownership and use of electronic equipment, income, type of household and family composition.

In terms of the aforementioned research, the present study applies, in an original way, the clustering technique implemented by means of an artificial neural system composed of fuzzy ART ANN, which is commonly used for data clustering. However, literature has not shown its use to group residential electricity consumers according to their socioeconomic and demographic characteristics, habits and ownership of electronic devices, as this study presents.

In contrast to other methods, fuzzy ART neural network does not need a determined number of clusters. The number of clusters obtained by the creation of clusters in ART networks happens dynamically by means of its algorithm.

3. Methodology

The objective of this study is to apply a fuzzy ART neural network as a data clustering algorithm of residential electricity consumers and identify the consumer profiles in the studied group. To reach this goal, the steps described below were taken to perform the whole clustering process presented by Jain et al. (1999).

3.1. Acquisition and restructuring data

Data preparation is a foundational step in a clustering process that involves data collection, preprocessing and the proper representation to use in a clustering algorithm.

Data acquisition was performed by means of field research. A survey

on residential consumers' electricity consumption habits was carried out in a group of 61 Brazilian residential consumers, being 90% located in the state of São Paulo, Brazil. After collecting data, they were tabulated and preprocessed to be adjusted and used in the neuro-fuzzy system. Pre-processing consists of manipulating data present in answers given by the responding consumers in order to perform the database construction including normalization, data wrangling and reduction of attributes (Jain et al., 1999). The normalization process consists of adjusting values on different scales, while data wrangling and reduction are processes of cleaning and unifying data for analysis. The data is represented by attributes that can be of different types: binary, discrete or continuous. Binary attributes can assume exactly 2 values: 0 or 1, indicating presence or absence of a certain characteristic. Discrete attributes usually have a finite set of possible values. Continuous attributes, on the other hand, can assume any real value within a predefined range. Thus, the transformation performed in data has made the attributes relevant, according to the focus of this research, which is of a quantitative nature, with reduced domain of values and discrete values.

Firstly, data from one consumer was eliminated, since that was the only one to reside in a rural area and the electricity bill encompassed not only the consumption of his/her residence present in that place, but also the consumption for other purposes in the rural property.

Then, a reduction of domain values was performed for some attributes. This reduction was necessary to avoid a large number of combinations among those values. Such combinations would lead to an increase in the amount of clusters and, therefore, could make it difficult to perform the analysis.

The performed data manipulation also included the definition of categories for the attributes. The categorization process consists in assigning a numeric value for each possible questionnaire response and for each range of values that has been established in some of the questions.

Following these steps, the questionnaire data that were transformed in order to obtain categories were:

- ConsumptionAvg: average electricity consumption (Table 3);
- RoomsNumber: number of rooms in the residence (Table 1);
- ApplianceConsumpRange: appliance consumption (Table 3);
- SalaryRangeID: family monthly wage range (Table 2).

Electricity consumption average was calculated based on the consumption of the last 12 months by means of the arithmetic means of reported values. Then, eight consumption ranges were enumerated from 0 to 7 as shown in Table 3.

The attribute that represents the size of a residence is derived from a question (in the applied questionnaire) which asks the respondent about the number of rooms in the residence. This attribute was defined in order to obtain only three possible responses as shown in Table 1.

In a similar way to the categorization of monthly consumption average of electricity, the electronic device consumption was categorized to define consumption ranges. For the definition of categories, appliance consumption was calculated based on the information given

Table 2
Categories for residence size representation.

Category	Number of rooms	Residence size
1	up to 6	Small
2	from 7 to 12	Medium
3	above 12	Large

Table 3
Categories for family monthly wage range representation.

Category	Amount of monthly minimum wages (MW)
1	Up to 2 MW
2	from 2 to 4 MW
3	from 4 to 10 MW
4	from 10 to 20 MW
5	above 20 MW

by the respondents. The number of days of use, the number of hours of use per day and the consumption per hour based on Brazilian electronics consumption data were considered for each device. Then, eight consumption ranges were determined, measuring, in this case, 100 kWh each, and numbered from 0 to 7 (Table 4).

Later, data were structured and stored in a database created with the MySQL Database Management System (DBMS). The diagram in Fig. 1 represents the data structure that was defined for the system. This data structure consists of a set of 9 entities and their relationships. Each entity has its set of attributes and corresponding data types, and it is represented by a rectangle divided horizontally by a line. The attributes located above the line represent the primary key. The attributes identified with a "(FK)" indicate the foreign keys which represent the relationship among the corresponding entities.

3.2. Definition of proximity measures

Selecting a similarity measure involves, among other features, knowing the nature of data. When data are quantitative, as is the case of this system, distance measures are the most convenient. Thus, two of the most popular proper distance measures were applied: Manhattan Distance, also known as City-Block, and Euclidian Distance. These simple distance measures can be often used to reflect dissimilarity between two patterns (Jain et al., 1999).

3.3. Definition of ARTab neural network for clustering residential electricity consumers

Hereinafter, the fuzzy ART neural system modeling and implementation used for the identification of clusters of residential electricity consumers are presented. This neural system is referred to as an ARTab neural network; i.e., it is composed of two fuzzy ART modules whose training is unsupervised, and it is intended to find patterns in data that enable knowledge discovery. These modules are called ARTa

Table 1
Composition of the ARTa module input vectors.

Attributes of input subvectors (α)				
	Position	Name	Description	Size (Number of bits)
Input variables (α) of ARTa module	α_1	ResidentsNumber	Number of residents in residence	3
	α_2	SalaryRange	Family salary range in number of minimum wages (MW) (see Table 3).	3
	α_3	RoomsRange	Range amount of rooms in the residence (see Table 2).	2
	α_4	ConsumptionRange	Range of average power consumption (in kWh) (see Table 4).	3
	α_5	ResidentsDuringDay	Whether residents remain home more than 16 h during a day	1
	α_6	ResidentsBelow15	Whether there are inhabitants below 15 years old	1
Total size of vector				13 bits

Table 4
Categories for monthly energy consumption range representation.

Category	Monthly consumption (kWh)
0	0–128
1	129–256
2	257–384
3	385–512
4	513–640
5	641–768
6	769–896
7	897–1024

and ARTb and are described below. The computer implementation may be constructed as described in [Appendix A](#):

- ARTa Module: in this module, the neural network processes data to group consumers into different clusters according to socioeconomic, demographic and consumption data.
- ARTb Module: in this module, data processing is relating to the ownership and use habits of electronic devices by residents of each of residential consumer unit. In addition to these data of ownership and habit, the result obtained by the ARTa module is incorporated, i.e., the cluster identification to which the consumer is associated in the ARTa module.

The structure of the neural system was designed to have an essentially binary architecture of data processing. Therefore, in each of the fuzzy ART modules, after reading the data from the database, a routine that performs a conversion of numeric values to binary values is followed. Although the fuzzy ART neural network permits the working of both analog and binary data, creating an essentially binary system is more advantageous because it is faster in processing and achieves more accurate results than do mixed neural networks (binary and analog data) or purely analog networks ([Martinho, Nunes, & Minussi, 2013](#)).

Moreover, it is a conception with great flexibility in the use of the database, allowing changes in its structure, e.g., including new important information to improve solutions.

The two described modules are interrelated ([Fig. 2](#)); i.e., the result obtained in the ARTa module, which consists of defining the cluster to which each consumer belongs, is used as part of the pattern vector of the ARTb neural network embedded in the ARTb module.

In both modules, in addition to the parameters for configuring the neural network behavior (vigilance parameter, choose parameter, training rate), data that will be processed by the neural network must be informed.

3.3.1. ARTa module

The pattern input vector of the ARTa module expressed by Eq. (1) consists of six attributes that are represented by the subvectors a_1, a_2, \dots, a_6 and has a total size of 13 bits. It is important to highlight that we adopted row vectors and not column vectors because they are more frequent in scientific literature:

$$a = [a_1 a_2 a_3 a_4 a_5 a_6] \quad (1)$$

These subvectors contain the binary representation of the socioeconomic characteristics of residents, part of the demographic data and the average consumption calculated based on monthly consumption reported by the respondents. The composition of the subvectors of the ARTa module of the ARTb neural network can be seen in [Table 1](#).

3.3.2. Output of the ARTa module

The output of the ARTa neural network (y_a) is converted via active code ([Carpenter et al., 1991a](#)), forming the 5-bit subvector (representing up to 31 clusters), which will be used as part of the input vector of the ARTb module.

3.3.3. ARTb module

The pattern input vector b in the ARTb module, shown in Eq. (2)

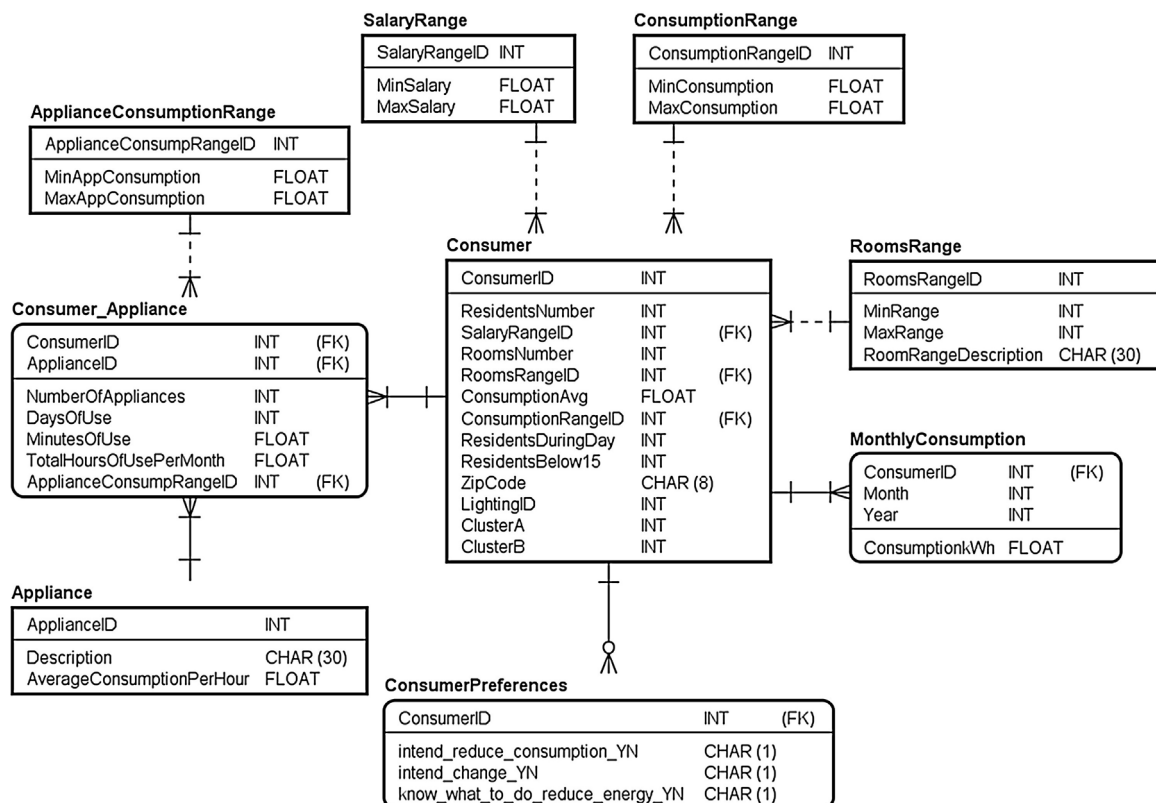


Fig. 1. ER diagram of design database.

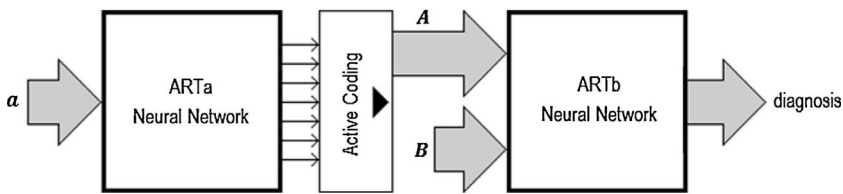


Fig. 2. ARTb Neural Network Architecture.

consists of 16 attributes,

$$\mathbf{b} = [\mathbf{A} \ \mathbf{B}] \quad (2)$$

which are

\mathbf{b} : input vector of the ARTb neural network comprising the attributes used in this module;

$$\mathbf{A} = \mathbf{b}_1 \quad (3)$$

$$\mathbf{B} = [\mathbf{b}_2 \mathbf{b}_3 \dots \mathbf{b}_{16}]. \quad (4)$$

The subvectors of this module also consist of the binary representation of the features used. Such features are from the following:

- 1) The cluster identified in the ARTa module (subvector \mathbf{A});
- 2) Data about ownership and electronic device habits as well as the type of the lighting in the residence (subvector \mathbf{B}).

However, to allow a full evaluation of the results of the implementation of this module, three different approaches were applied (Table 5) regarding the nature of the features that refer to electronic devices (subvectors $\mathbf{b}_2 \dots \mathbf{b}_{15}$). In the three approaches presented, the subvectors \mathbf{b}_1 and \mathbf{b}_{16} receive the cluster identifier number assigned to each sample (obtained as the output of the ARTa module) and the

Table 5
Approaches to the constitution of the input pattern vector for the ARTb module.

Approach	Subvectors $\mathbf{b}_2 \dots \mathbf{b}_{15}$	Vector \mathbf{b} size
I	The subvectors represent the number of each of the electronics that was reported by the respondents	30 bits
II	Consider only if the electronics are present in residence	21 bits
III	Consider calculating the estimated usage of consumer electronics based on the pattern of each data type and the hours and days provided in the response by the respondent.	49 bits

Table 6
Composition of other input vectors of the ARTb module in three different approaches.

Attributes of input subvectors (\mathbf{b})		Subvector	Name	Description	Subvector size (number of bits)		
					Approach I	Approach II	Approach III
Input variables (\mathbf{b}) of ARTb module	\mathbf{b}_1	ConsumerCluster	Cluster to which the consumer unit fits		5	5	5
	\mathbf{b}_2	Appliance1	Refrigerator		2	1	3
	\mathbf{b}_3	Appliance2	Freezer		2	1	3
	\mathbf{b}_4	Appliance3	Washing machine		2	1	3
	\mathbf{b}_5	Appliance4	Clothes dryer		1	1	3
	\mathbf{b}_6	Appliance5	Dishwasher		1	1	3
	\mathbf{b}_7	Appliance6	Microwave		1	1	3
	\mathbf{b}_8	Appliance7	Electric shower		3	1	3
	\mathbf{b}_9	Appliance8	Air conditioner		3	1	3
	\mathbf{b}_{10}	Appliance9	Solar heater		1	1	3
	\mathbf{b}_{11}	Appliance10	Gas heater		1	1	3
	\mathbf{b}_{12}	Appliance11	TV smaller than 21 inches		2	1	3
	\mathbf{b}_{13}	Appliance12	21-inch TV or bigger		2	1	3
	\mathbf{b}_{14}	Appliance13	Iron		1	1	3
	\mathbf{b}_{15}	Appliance14	Electric oven		1	1	3
	\mathbf{b}_{16}	Lighting	Predominance of the type of lamp used		2	2	2
Total size of vector					30 bits	21 bits	49 bits

predominant lighting in the residence, respectively.

The representative value for subvectors is expressed in binary numbers according to what is shown in Table 6.

4. Results and discussion

The application of the proposed neural system resulted in the placement of electricity consumers in different clusters, as expected. Several simulations were performed by applying changes to the network parameters to find the best performance and then define the parameters for system operation.

The idea is to extract the initial knowledge (considering a little quantity of consumers) and incrementally add new consumers, which is called the continuous-learning. This training conception is possible by using the Fuzzy-ARTMAP neural network, according to what is proposed in reference (Marchiori, Silveira, Lotufo, Minussi, & Lopes, 2011). This resource improves the analysis system, i.e., the knowledge extraction is done permanently, without destructing the previously acquired knowledge (most of the neural networks proposed in the literature require the initialization of the process destructing the previously acquired knowledge). This way, the neural network is constantly auto improving, and consequently providing solutions that are more precise.

In order to increase (or decrease) the quantity of clusters it is enough to increase (or decrease) the vigilance parameter. A better calibration can be done, for example, by “trying and error”.

In the ARTa module, the created clusters aim to relate the socio-economic characteristics of residents, their permanence in the residence throughout the day and consumption data. The implementation of the ARTb neural network resulted in the 19 consumer clusters shown in Table 7.

Clusters 2 and 9 were randomly selected for the descriptive analysis of the results to assist the semantic comprehension of these clusters.

Table 7
Results of consumer clustering obtained by the ARTa module.

ARTa Clusters	Number of consumers in cluster	Consumer identification
1	3	1, 4, 14
2	7	2, 3, 11, 29, 41, 57, 59
3	4	5, 6, 49, 60
4	5	7, 9, 13, 18, 30
5	4	8, 10, 27, 39
6	4	12, 19, 21, 58
7	4	15, 16, 22, 55
8	2	17, 40
9	5	20, 25, 34, 35, 54
10	3	23, 26, 38
11	1	31
12	2	28, 48
13	4	24, 32, 42, 56
14	2	33, 37
15	2	36, 50
16	1	43
17	4	44, 45, 46, 52
18	2	47, 61
19	2	51, 53

Table 8
Description of consumers in cluster 2.

Cluster	Residents Number	Salary Range	Rooms Range	Consumption Range	Residents During Day	Residents below 15
C2	2	2	1	1	0	0
C3	2	3	2	1	0	0
C11	2	2	1	1	1	0
C29	2	3	2	0	0	0
C41	2	3	1	1	0	0
C57	2	3	2	1	0	0
C59	2	3	1	1	1	0

The cluster identified by number 2 consists of seven consumer units (residences) (Table 8). These households are identified by the numbers 2, 3, 11, 29, 41, 57 and 59 and consist of only two residents older than 15 years of age who rarely stay at home throughout the day (in 5 of the 7 residences in this cluster, no one is at home throughout the day). The family income range is between 2 and 10 minimum wages (ranges 2 and 3). The residence is small or medium-sized, and the monthly consumption range is less than 256 kWh.

On the other hand, the cluster identified by the number 9 is composed of five consumer units (Table 9). These households are identified by the numbers 20, 25, 34, 35 and 54 and consist of four residents of

Table 9
Description of consumers in cluster 9.

Cluster	Residents Number	Salary Range	Rooms Range	Consumption Range	Residents During Day	Residents below 15
C20	4	3	1	2	0	1
C25	4	2	1	2	3	1
C34	4	3	1	2	1	1
C35	4	3	3	2	0	1
C54	4	3	1	2	3	0

whom at least one stays at home throughout the day (60%), and 80% of the residences have one resident who is younger than 15 years of age. The family income range is from 4 to 10 minimum wages (range 3). The residence is small, and the monthly consumption range is between 257 and 384 kWh.

The monthly electricity consumption profile (in kWh) of consumers belonging to clusters 2 and 9 is illustrated in Fig. 3.

We carried out the calculation of the proximity among the objects of each cluster to verify their similarity. To facilitate the observation and analysis, these indices were normalized in a range between 0 and 1. It is noteworthy that the lower the index value (closer to 0), the more similar the objects. The threshold adopted to perform the analysis was 0.25, i.e., pairs of objects that had indices equal to or less than 0.25, were considered similar.

By applying the analysis of similarity to the objects of clusters represented by their respective vectors and comparing them to the indices obtained by applying Manhattan Distance and Euclidian Distance, we observed that for cluster 2, both measures present an index less than or equal to 0.25 (Tables 10 and 12). Regarding cluster 9, there is a certain distance between some pairs of objects such as those identified as C25 and C35. The index calculated based on the Euclidian Distance was 0.48 for the aforementioned pair of objects and 0.33 for the same pair considering the index obtained by the Manhattan Distance calculation (Tables 11 and 13).

Although there were occurrences of weak similarity between some objects that are grouped in the same cluster, it is clear that by using the Manhattan Distance measure, in only 15.8% of the identified clusters, there are objects with little similarity. Thus, the results provided by the neural network actually lead to similarity among objects.

After obtaining the clusters by the ARTa module, which characterize the consumer profiles in the universe analyzed, we proceeded to the data processing (ARTb module) concerning the ownership and consumption habits of electronic devices together with the result obtained by the ARTa module. Two of the three approaches (relating to the input vector content for this module) were performed and analyzed. In

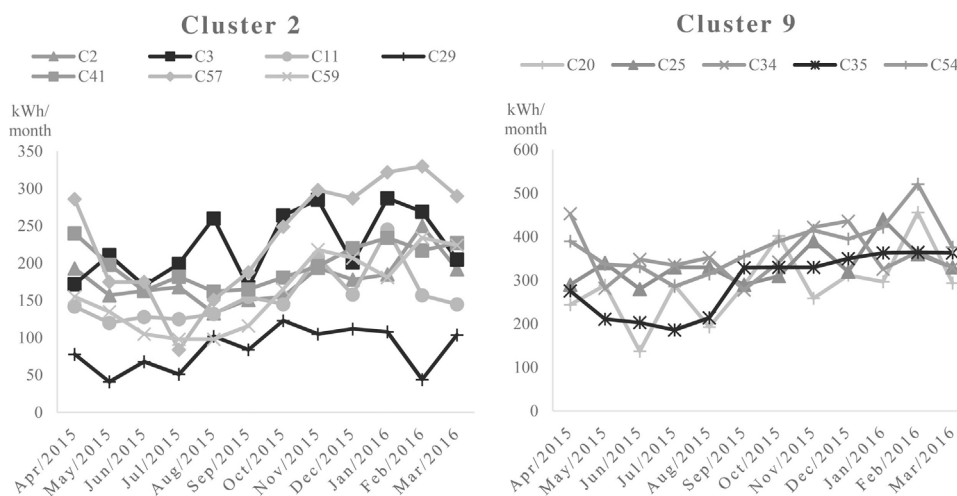


Fig. 3. Monthly electricity consumption profile (in kWh) in clusters 2 and 9.

Table 10
Indices obtained by the Euclidian Distance calculation for the objects of cluster 2.

Cluster	C2	C3	C11	C29	C41	C57	C59
C2	0.00	0.18	0.13	0.22	0.13	0.18	0.18
C3	0.18	0.00	0.22	0.13	0.13	0.00	0.18
C11	0.13	0.22	0.00	0.25	0.18	0.22	0.13
C29	0.22	0.13	0.25	0.00	0.18	0.13	0.22
C41	0.13	0.13	0.18	0.18	0.00	0.13	0.13
C57	0.18	0.00	0.22	0.13	0.13	0.00	0.18
C59	0.18	0.18	0.13	0.22	0.13	0.18	0.00

Table 11
Indices obtained by the Euclidian Distance calculation for the objects of cluster 9.

Cluster	C20	C25	C34	C35	C54
C20	0.00	0.40	0.13	0.25	0.40
C25	0.40	0.00	0.28	0.48	0.18
C34	0.13	0.28	0.00	0.28	0.28
C35	0.25	0.48	0.28	0.00	0.48
C54	0.40	0.18	0.28	0.48	0.00

Table 12
Indices obtained by the Manhattan Distance calculation for the objects of cluster 2.

Cluster	C2	C3	C11	C29	C41	C57	C59
C2	0.00	0.11	0.06	0.17	0.06	0.11	0.11
C3	0.11	0.00	0.17	0.06	0.06	0.00	0.11
C11	0.06	0.17	0.00	0.22	0.11	0.17	0.06
C29	0.17	0.06	0.22	0.00	0.11	0.06	0.17
C41	0.06	0.06	0.11	0.11	0.00	0.06	0.06
C57	0.11	0.00	0.17	0.06	0.06	0.00	0.11
C59	0.11	0.11	0.06	0.17	0.06	0.11	0.00

Table 13
Indices obtained by the Manhattan Distance calculation for the objects of cluster 9.

Cluster	C20	C25	C34	C35	C54
C20	0.00	0.22	0.06	0.11	0.22
C25	0.22	0.00	0.17	0.33	0.11
C34	0.06	0.17	0.00	0.17	0.17
C35	0.11	0.33	0.17	0.00	0.33
C54	0.22	0.11	0.17	0.33	0.00

Approach I, 15 clusters were obtained and are distributed as shown in Fig. 4. In Approach II, there were only 4 clusters obtained, which have the concentration of consumers shown in Fig. 5. In Approach III, 7 clusters were obtained, and the consumers are distributed as shown in Fig. 6.

Tables 14 and 15 shows the results of the application of the ARTab neural system and shows the association of some of the variables with the profiles that were determined by the clustering process in Approach I and Approach III, respectively.

In Approach II, it is observed a higher concentration of consumers in only one of the clusters: the cluster identified by the number 1. At first, this result does not seem to be useful from the point of view of identifying distinct groups of consumers so that they can treat them also distinctly as the models suggest to determine actions in Response Demand programs. By applying the analysis of similarity between the pairs of objects, we observed a high rate of occurrence of dissimilarity that leads to the conclusion that these clusters are not significant.

There are several household profiles in Brazil. However, there is no intention, in this paper, that the proposed methodology meets such broad range of consumer profiles, although this may happen in future

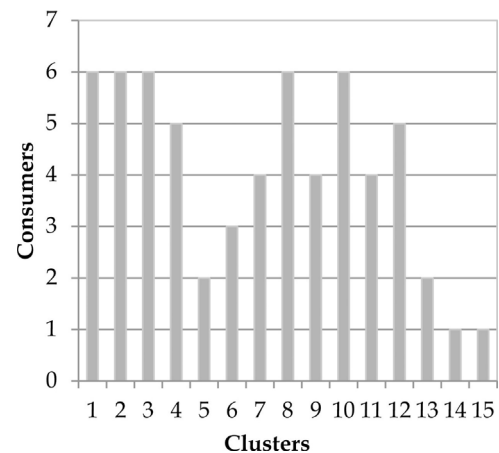


Fig. 4. Distribution of consumers by clusters in Approach I.

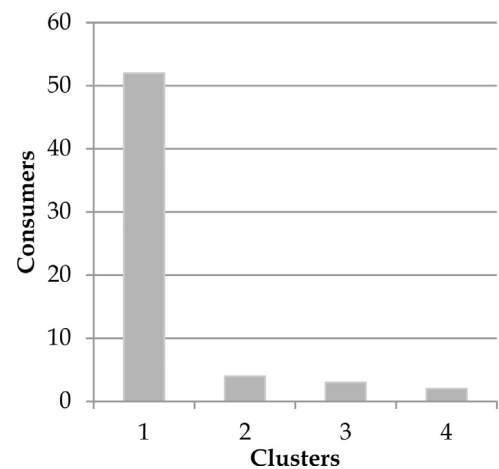


Fig. 5. Distribution of consumers by clusters in Approach II.

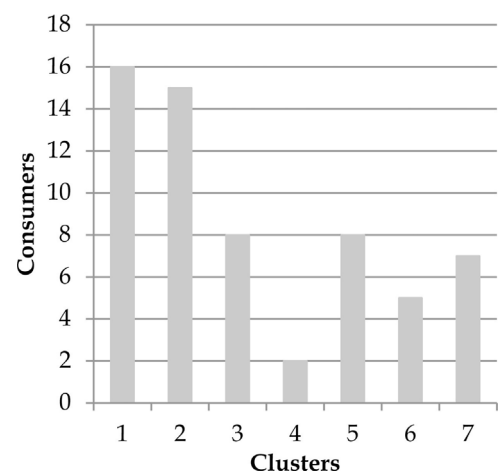


Fig. 6. Distribution of consumers by clusters in Approach III.

approaches based on the use of the Fuzzy-ARTMAP neural network with continuous-learning (Marchiori et al., 2011).

5. Conclusions

Initially, based on the availability of a set of information collected from volunteer consumers through the use of the proposed procedure, we draw important information aiming to gain a better perception of behavioral electricity consumption. This methodology was developed

Table 14
Interpretation of part of the clusters obtained by the implementation of the ARTab neural network in Approach I.

Cluster	Consumer identification	Monthly electricity consumption profile	Features observed
1	2, 6, 17, 40, 53, 60 (Total: 6)		<ul style="list-style-type: none">• 83.3% of households are medium sized• 66.7% belong to Class 2 power consumption• The highest consumption occurs in November (14% above the annual average)• 83.3% of households have children, and in 66.7% of households, there are residents who remain at home during the day• 50% belong to salary range 2• Do not have a freezer• 16.7% have a dishwasher and air conditioning• Do not have a clothes dryer• 83.3% use electric shower• 50% of households are medium sized• 50% belong to Class 3 power consumption• January and November are the months with more intensive use of electricity, at 38.4% and 25% above the annual average, respectively• 50% of households have children, and in 83.3% of households, there are residents who remain at home during the day• 50% belong to salary range 3• Do not have a freezer, dishwasher or clothes dryer• 83.3% have air conditioning• 83.3% use an electric shower• 50% of households are medium sized, and 50% are small• 50% belong to Class 2 power consumption• January, February and November are the months of increased consumption, at 24.5%, 17.8% and 13.2% higher than the annual average, respectively• 16.7% of households have children, and in 66.7% of households, there are residents who remain at home during the day• 66.7% belong to salary range 3 or 4• Do not have a freezer or electric oven• 16.7% have dishwasher and clothes dryer• 66.7% have air conditioning• 83.3% use an electric shower• 66.7% of households are small• 66.7% belong to Class 1 or 2 power consumption• The highest consumption occurs in October (10.6% above the annual average)• 16.7% of households have children, and in 50% of households, there are residents who remain at home during the day• 50% belong to salary ranges below 2• Do not have a freezer• 16.7% have a dishwasher, and 33.3% have air conditioning• Do not have a clothes dryer or electric oven• 100% use an electric shower• 66.7% of households are small• 66.7% belong to Class 2 power consumption.• The highest consumption occurs in March (18.4% above the annual average)• 50% of households have children, and in 50% of households, there are residents who remain at home during the day• 50% belong to salary range 4• 100% have a freezer• Do not have a dishwasher, clothes dryer or electric oven• 50% have air conditioning• 83.3% use an electric shower
2	3, 10, 16, 33, 48, 51 (Total: 6)		
3	5, 11, 13, 14, 27, 57 (Total: 6)		
8	7, 9, 12, 18, 30, 37 (Total: 6)		
10	38, 41, 43, 47, 49, 52 (Total: 6)		

and based on a neural network from the ART family. It is a concept with great application potential because it is stable and plastic and has a low computational cost for implementation. By using it, it was possible to carry out this work by considering a small set of data and generating some knowledge. As new information becomes available, by means of incremental incorporation, the neural network will become increasingly “smarter”, similar to what happens with humans. These details are different compared with other methods of clustering. Through the clustering of consumers by the similarity of their profiles and the comprehension of their consumption behavior, new programs to improve the habits of consuming electricity may be proposed. Thus, the consumer would have access to a sampling of their consumption. By knowing the similar consumers, better negotiation strategies with electricity suppliers are offered, especially in the case of a smart grid, in which consumers will be free to choose their suppliers. The main contribution of this paper is identifying electricity consumer’s clusters with the purpose of obtaining a better consumption behavioral perception by

using a neural network system composed of Fuzzy ART modules. This work, in addition to studying residential consumers’ behavior from an electrical point of view, will also help improve the forecasting of multimodal loads (Nose-Filho, Lotufo, & Minussi, 2011).

Funding

This research has not received any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgements

The authors would like to thank Bruna E. S. Guerra for the contribution providing language help on translating this article and to American Journal Experts on reviewing it. The authors would also like to thank the citizens that have contributed for the completion of the household survey questionnaire.

Table 15

Interpretation of part of the clusters obtained by the implementation of the ARTab neural network in Approach III.

Cluster	Consumers identification	Monthly electricity consumption profile	Features observed
1	2, 3, 4, 6, 11, 17, 25, 26, 35, 40, 46, 51, 53, 57, 60, 61 (total: 16)		<ul style="list-style-type: none"> 56.3% of households are medium sized, and 18.7% are large 62.5% belong to Class 2 power consumption Higher consumption is between November and January, at 19.2%, 9.5% and 21.3% higher than the annual average, respectively 68.7% of households have children, and in 68.7% of households, there are residents who remain at home during the day 62.5% belong to salary ranges above 3 Do not have a freezer Less than 12.5% have a clothes dryer and dishwasher 25% have an electric oven 81% use an electric shower 56.3% have air conditioning
2	5, 7, 15, 16, 18, 22, 27, 29, 33, 39, 42, 44, 50, 55, 58 (total: 15)		<ul style="list-style-type: none"> 73.3% of households are medium sized, and 25% are small 46.7% belong to Class 3 power consumption The highest consumption occurs in January (31% above the annual average) 66.7% of households do not have children. However, in 66.7% of households, there are residents who remain at home throughout the day 66.7% belong to salary range 3 or 4 Do not have a freezer 33.3% have a clothes dryer, 20% have a dishwasher, and 40% have an electric oven 66.7% use an electric shower 86.7% have air conditioning
3	8, 9, 12, 13, 28, 36, 37, 56 (total: 8)		<ul style="list-style-type: none"> 37.5% of households are small, and 37.5% are medium sized 37.5% belong to Class 3 power consumption The highest consumption occurs in November (21.7% above the annual average) 75% of households do not have children, and in 50% of households, there are residents who remain at home during the day 87.5% belong to salary range 4 or 5 37.5% have a freezer Less than 25% have a clothes dryer, dishwasher and electric oven 100% use an electric shower 75% have air conditioning
5	20, 23, 30, 38, 41, 49, 54, 59 (total: 8)		<ul style="list-style-type: none"> 87.5% of households are small 50% belong to Class 3, and 37.5% belong to Class 2 power consumption The highest consumption occurs in February (21.5% above the annual average) 75% of households do not have children, and in 62.5% of households, there are residents who remain at home during the day 50% belong to salary range 3 62.5% have a freezer Do not have an electric oven or dishwasher 87.5% use an electric shower 50% have air conditioning

Appendix A. – Fuzzy ART Neural Network Algorithm

Fuzzy ART neural network algorithm consists basically in the following steps (Carpenter et al., 1991a, 1991b).

Step 1: Input Data

Input data are allocated in the vector $\mathbf{a} = [a_1 a_2 a_3 \dots a_M]$. This vector is normalized in order to achieve the criterion of fuzzy logic (Zadeh, 1965) such that the normalized vector components pertain to the interval [0,1]:

$$\bar{\mathbf{a}} = \mathbf{a}/|\mathbf{a}| \quad (A1)$$

where

$\bar{\mathbf{a}}$: input vector normalized;

$$|\mathbf{a}| = \sum_{i=1}^M \bar{a}_i$$

$$\bar{a}_i \geq 0.$$

Step 2: Input Vector Encoding

The complement coding is realized to obtain a vector of length M :

$$\bar{\mathbf{a}}^c = [\bar{a}_1^c \bar{a}_2^c \bar{a}_3^c \dots \bar{a}_M^c] \quad (A2)$$

where

$\bar{\mathbf{a}}^c$: complementary vector normalized;

$$\bar{a}_i^c = 1 - \bar{a}_i.$$

That way, the dimension of the vector I is $2M$. Thus, the vector is defined by:

$$I = [\bar{a}\bar{a}^c] \quad (A3)$$

where

$$|I| = M.$$

Step 3: Activity Vector

The activity vector on recognition layer (RL) is defined by $y = [y_1 y_2 y_3 \dots y_N]$, where N is the category number formed in the RL according to Eq. (A4):

- $y_j = 1$, if the neuron j of RL is active;
 $y_j = 0$, if the neuron j of RL is inactive. (A4)

Step 4: Neural Network Parameters

Parameters used in the processing of the fuzzy ART neural network are:

1. Choose parameter : $\alpha > 0$;
2. Training rate : $\beta \in [0,1]$;
3. Vigilance parameter : $\rho \in [0,1]$.

Step 5: Weight Initialization

Initially, all weights have a value equal to 1:

$$w_{j,1}(0) = w_{j,2}(0) = \dots = w_{j,2M}(0) = 1 \quad (A5)$$

It is indicating that there is no active category (class).

Step 6: Choose Class

Taking input vector I on competition layer (CL), for each node j in RL, the choice function T_j can be calculated by:

$$T_j = |I \wedge w_j| / (\alpha + |w_j|) \quad (A6)$$

where

\wedge : fuzzy AND operator being defined by:

$$(I \wedge w)_i = \min(I_i, w_i) \quad (A7)$$

The category is chosen when the node J is active:

$$J = \arg\{\max(T_j)\}, j = 1, 2, \dots, N. \quad (A8)$$

Using Eq. (A8), if there is more than one active category (T_j with identical values), the chosen category will be the one with a lower value.

Step 7: Resonance or Reset

Resonance occurs if the vigilance criterion Eq. (A9) is satisfied if:

$$|I \wedge w_j| / (|I|) \geq \rho \quad (A9)$$

If the criterion, defined by Eq. (A9), is not satisfied, another candidate (process set to reset) must be chosen. In the reset agent, the RL node J is excluded from the search process (Eq. (A6)), i.e., $T_j = 0$. Then, a new class is selected (by Eq. (A8)) for the process of resonance. This procedure should be repeated until a specific category satisfies the criterion given by Eq. (A9).

Step 8: Weight Adapting (training)

Once settled the state of resonance, i.e., node J has been confirmed as the winner neuron associated weights should be adjusted using Eq. (A10):

$$w_j^{\text{new}} = \beta(I \wedge w_j^{\text{old}}) + (1 - \beta)w_j^{\text{old}} \quad (A10)$$

where

J : Active category;

w_j^{new} : Updated weight vector;

w_j^{old} : Vector weights for the previous update.

It is emphasized that if $\beta = 1$, there is the rapid training.

References

- Albert, A., & Rajagopal, R. (2013). Smart meter driven segmentation: What your consumption says about you. *IEEE Transactions on Power Systems*, 28(No. 4), 4019–4030. <http://dx.doi.org/10.1109/TPWRS.2013.2266122>.
- Angelos, E. W. S., Saavedra, O. R., Cortés, O. A. C., & Souza, A. N. (2011). Detection and identification of abnormalities in customer consumptions in power distribution systems. *IEEE Transactions on Power Delivery*, 26(No. 4), 2436–2442. <http://dx.doi.org/10.1109/TPWRD.2011.2161621>.
- Aydinalp, M., Ugursal, V. I., & Fung, A. S. (2002). Modeling of appliance, lighting, and space-cooling energy consumption in the residential sector using neural network. *Applied Energy*, 71(No. 2), 87–110. [http://dx.doi.org/10.1016/S0306-2619\(01\)00049-6](http://dx.doi.org/10.1016/S0306-2619(01)00049-6).
- Aydinalp, M., Ugursal, V. I., & Fung, A. S. (2004). Modeling of the space and domestic hot water heating energy-consumption in the residential sector using neural networks. *Applied Energy*, 79(No. 2), 159–178. <http://dx.doi.org/10.1016/j.apenergy.2003.12.006>.
- Aydinalp-Koksall, M., & Ugursal, V. I. (2008). Comparison of neural network, conditional demand analysis, and engineering approaches for modeling end-use energy consumption in the residential sector. *Applied Energy*, 85(No. 4), 271–296. <http://dx.doi.org/10.1016/j.apenergy.2006.09.012>.
- Babu, T., Murthy, T., & Sivaiah, B. (2017). Detecting unusual customer consumption profiles in power distribution systems ?APSPDCL. *Proc. of IEEE international conference on computational intelligence and computing research (ICCIC)*<http://dx.doi.org/10.1109/ICCIC.2013.6724264>.
- Beckel, C., Sadamori, L., Staake, T., & Santini, S. (2014). Revealing household characteristics from smart meter data. *Energy*, 78, 397–410. <http://dx.doi.org/10.1016/j.energy.2014.10.025>.
- Benitez, I., Quijano, A., Diez, J. L., & Delgado, I. (2014). Dynamic clustering segmentation applied to load profiles of energy consumption from Spanish customers. *International Journal of Electrical and Power Energy Systems*, 55, 437–448. <http://dx.doi.org/10.1016/j.ijepes.2013.09.022>.
- Cao, H., Beckel, C., & Staake, T. (2013). Are domestic load profiles stable over time? An attempt to identify target households for demand side management campaigns. *Proc. of 39th annual conference of the IEEE industrial electronics society (IECON)*<http://dx.doi.org/10.1109/IECON.2013.6699900>.
- Carpenter, G. A., Grossberg, S., & Rosen, D. B. (1991a). Fuzzy ART: fast stable learning

- and categorization of analog patterns by an adaptive resonance system. *Neural Network*, 6(No. 4), 759–771. [http://dx.doi.org/10.1016/0893-6080\(91\)90056-B](http://dx.doi.org/10.1016/0893-6080(91)90056-B).
- Carpenter, G. A., Grossberg, S., & Roden, D. B. (1991b). Fuzzy ART: an adaptive resonance algorithm for rapid, stable classification of analog patterns. *Conference on neural networks Seattle*. <http://dx.doi.org/10.1109/IJCNN.1991.155368>.
- Chicco, G. (2012). Overview and performance assessment of the clustering methods for electrical load pattern grouping. *Energy*, 42(No. 1), 68–80. <http://dx.doi.org/10.1016/j.energy.2011.12.031>.
- Dent, I., Craig, T., & Aickelin, U. R. T. (2012). Finding the creatures of habit: clustering households based on their flexibility in using electricity. *Proc. of 3rd annual digital economy all hands conference, digital futures 2012*.
- Energy Research Company (EPE) (2015a). *Demand projection for electricity for the next 10 years (2015–2024). Feb-2015*. Available: http://www.epe.gov.br/mercado/Documents/S%C3%A9rie%20Estudos%20de%20Energia/20091222_2.pdf (In portuguese) [Accessed 07 April 2016].
- Energy Research Company (EPE) (2015b). *Energy demand 2050. Ago-2015*. [Available <http://www.epe.gov.br/Estudos/Documents/DEA%2013-14%20Demanda%20de%20Energia%202050.pdf> (In portuguese) [Accessed 07 April 2016].
- Fan, J., & Borlase, S. (2009). The evolution of distribution. *IEEE Power and Energy Magazine*, 7(No. 2), 63–68. <http://dx.doi.org/10.1109/MPE.2008.931392>.
- Flath, C., Nicolay, D., Conte, T., & Dinther, C. F. N. L. (2012). Cluster analysis of smart metering data: An implementation in practice. *Business & Information Systems Engineering*, 4(No. 1), 31–39. <http://dx.doi.org/10.1007/s12599-011-0201-5>.
- Haben, S., Singleton, C., & Grindrod, P. (2016). Analysis and clustering of residential customers energy behavioral demand using smart meter data. *IEEE Transactions on Smart Grid*, 7(No. 1), 136–144. <http://dx.doi.org/10.1109/TSG.2015.2409786>.
- Jain, A. K., & Dubes, R. C. (1988). *Algorithms for clustering data, upper saddle river*. NJ USA: Prentice Hall.
- Jain, A. K., Murty, M., & Flynn, P. (1999). Data clustering: A review. *ACM Computing Surveys*, 31(No. 3), 264–323. <http://dx.doi.org/10.1145/331499.331504>.
- Jokar, P., Arianpoo, N., & Leung, V. C. M. (2015). Electricity theft detection in AMI using customers' consumption patterns. *IEEE Transactions on Smart Grids*, 7(No. 1), 216–226. <http://dx.doi.org/10.1109/TSG.2015.2425222>.
- Kwac, J., Flora, J., & Rajagopal, R. (2014). Household energy consumption segmentation using hourly data. *IEEE Transactions on Smart Grid*, 5(No. 1), 420–430. <http://dx.doi.org/10.1109/TSG.2013.2278477>.
- Marchiori, S. C., Silveira, M. C. G., Lotufo, A. D. P., Minussi, C. R., & Lopes, M. L. M. (2011). Neural network based on adaptive resonance theory with continuous training for multi-configuration transient stability analysis of electric power systems. *Applied Soft Computing*, 11, 706–715. <http://dx.doi.org/10.1016/j.asoc.2009.12.032>.
- Martinho, V. R. C., Nunes, C., & Minussi, C. R. (2013). A new method for prediction of school dropout risk group using fuzzy-ARTMAP neural network. *IEEE 25th international conference on tools with artificial intelligence* (pp. 159–166). <http://dx.doi.org/10.1109/ICTAI.2013.33>.
- McLoughlin, F., Duffy, A., & Conlon, M. (2012). Characterizing domestic electricity consumption patterns by dwelling and occupant socio-economic variables: An Irish case study. *Energy and Buildings*, 48, 240–248. <http://dx.doi.org/10.1016/j.enbuild.2012.01.037>.
- McLoughlin, F., Duffy, A., & Conlon, M. (2015). A clustering approach to domestic electricity load profile characterization using smart metering data. *Applied Energy*, Vol 141, 190–199. <http://dx.doi.org/10.1016/j.apenergy.2014.12.039>.
- Nose-Filho, K., Lotufo, A. D. P., & Minussi, C. R. (2011). Short-Term multinodal load forecasting using a modified general regression neural network. *IEEE Transaction on Power Delivery*, 26(No. 4), 2862–2869. <http://dx.doi.org/10.1109/TPWRD.2011.2166566>.
- ÓDoherty, J., Lyons, S., & Tol, R. (2008). Energy-using appliances and energy-saving features: Determinants of ownership in Ireland. *Applied Energy*, 85(No. 7), 650–662. <http://dx.doi.org/10.1016/j.apenergy.2008.01.001>.
- Räsänen, T., & Kolehmainen, M. (2009). *Feature-based clustering for electricity use time series data. Adaptive and natural computing algorithms (LNCS 5495)*, Berlin Germany http://dx.doi.org/10.1007/978-3-642-04921-7_41.
- Stephen, B., & Galloway, S. J. (2012). Domestic load characterization through smart meter advance stratification. *IEEE Transactions on Smart Grid*, 3(No. 3), 1571–1572. <http://dx.doi.org/10.1109/TSG.2012.2198314>.
- Stephen, B., Mutanen, A., Galloway, S. J., Burt, G., & Järventausta, P. (2014). Enhanced load profiling for residential network customers. *IEEE Transactions on Power Delivery*, 29(No.1), 88–96. <http://dx.doi.org/10.1109/TPWRD.2013.2287032>.
- The World Bank (IBRD-IDA) (2014). *Electric power consumption (kWh per capita)*. Available: <http://data.worldbank.org/indicator/EG.USE.ELEC.KH.PC?locations=BR> [Accessed 07 April 2016].
- Witten, I., & Frank, E. H. M. (2011). *Data Mining: Practical machine learning tools and techniques*. Burlington USA: Morgan Kauffman.
- Yohanis, Y. G., Mondol, J. D., Wright, A., & Norton, B. (2008). Real-life energy use in the UK: How occupancy and dwelling characteristics affect domestic electricity use. *Energy and Buildings*, 40(No. 6), 1053–1059. <http://dx.doi.org/10.1016/j.enbuild.2007.09.001>.
- Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(No. 3), 338–353. [http://dx.doi.org/10.1016/S0019-9958\(65\)90241-X](http://dx.doi.org/10.1016/S0019-9958(65)90241-X).