Income Classification of Residential Consumers Through Intelligent Techniques and Load Curves

Danilo Gastaldello and Haroldo Amaral
Dept of Electrical Engineering
Polytechnic School, University of São Paulo - USP
São Paulo, Brazil
danilo.gastaldello@usp.br, agaelema@globo.com

André Nunes de Souza and Caio C. Oba Ramos
Department of Electrical Engineering
Univ. Estadual Paulista - UNESP
Bauru, Brazil
andrejau@feb.unesp.br, caioramos@gmail.com

Abstract—**The internal power consumption of a country has always been related to its Gross Domestic Product (GDP) - the consumption increases when GDP grows. However, this relationship is changing, as it is being directly affected by the oil problems. The scarcity of power and geopolitical and environmental problems have led the most developed countries to look for ways to optimize production and to spend less energy. In Brazil, the dependence on oil is less than in most countries because its energy is linked to hydropower. The residential segment, with 22%, is the second in energy demand, which it justifies our study. The residential energy consumption is mainly based on two factors: (i) equipment ownership and the (ii) consumer habits. The paper presents a study using intelligent techniques to classify residential consumers with different salary ranges through load curves.**

Index Terms— **Energy consumption, Intelligent systems, Load Curve, Residential Consumer.**

I. Introduction

The electricity consumption of households can reach up to 50% of the available energy in electricity markets of developed countries. However, the management of this type of consumption becomes very challenging due to the variation of each consumer profile, which is linked to people's habits and each specific equipment used, which is related to the culture, climate, economic conditions, education, seasons of the year, days of the week, and holidays [1].

The economic factor is related to family income, which directly influences the possession of equipment and consequently the demand for energy. However, this relationship is not so simple to be understood, because the choice of the investment is individual, ie, is linked to the needs and the power of free will of each person. So, to establish a relationship between the salary range of a family and its use can be difficult.

The concern with the increase of energy consumption and the difficulty to implement new generation plants leads to the need for several studies in order to improve the energy consumption efficiency [2]. Methodologies are being developed to encourage consumers to use energy more rationally. Additionally, the smart grid concept contributes to data communication, improved resources, and network management, including the advantages of joining consumption and distributed generation [3-5].

In Brazil, the state of São Paulo has great economic power and large population, mainly in relation to the consumption of electricity. With just over 42 million habitants and density of 170 habitants per km2 in 2013, the state of São Paulo consumed approximately a total of 240,083 GWh of electricity, being 63,946 GWh consumed by residential customers (26.63%) [6-8]. For example, in 2010, the european electricity consumption has been increasing since 1990 and the residential consumers were responsible for 29.70% of the total electricity consumption in EU-27 [9].

In the context of load curves study, what concerns the resources management in order to manage the final load curve, the recent and relevant literature has been focused on themes as the distributed generation, demand response [10-12] and electric vehicles integration [13]. While several works present techniques and methods to address the load curve data cleaning and treatment [14], others focus on the load curve definition for residential consumers [15]. However, the natural, socioeconomic and cultural factors, behind the load curve, have not been taken into account.

Given this context, to explore and experience the energy consumers is key to the delivery of good energy demand planning, control and power, favoring the market. The goal is to create a set of residential customers data through an algorithm that takes into account social, economic and cultural aspects, then perform an assessment of clever techniques for classification by salary range, the residential consumer data set. Initially statistics of the population will be used for the creation of load curves construction algorithm, however, one sees in the future see if there are common features in the world that relate to the residential consumer using the form electricity.

After this introduction section, in Section II presents a brief theoretical introduction of the intelligent techniques used, in Section III explains the proposed methodology. In Section IV, we show the application and the results of the proposed methodology into a case study. The main conclusions of the paper are presented in Section V.

978-1-4673-6605-2/15/$31.00 ©2015 IEEE

II. Intelligent Techniques Applied

In this section, we present a brief theoretical introduction on intelligent techniques used in this work (SVM – Support Vector Machine, ANN - Artificial Neural Networks and OPF - Optimum Path Forest).

A. SVM

A support vector machine is a concept in computer science for a set of supervised learning methods that analyse the data and recognize patterns, used for classification and regression analysis.

Thus, standard SVM consists to find a hyperplane that maps the data of the space from one dimension to another larger so that the samples of data become linearly separable using a kernel function. New samples are then mapped in the same space and predicted as belonging to a category based on which side of the space they are placed [16].

B. ANN

Artificial neural networks are a method to solve problems by simulating the human brain, including their behavior, i.e., learning, making mistakes and making discoveries. Are computational techniques which have a model inspired by the neural structure of intelligent organisms and acquire knowledge through experience.

Neural networks have nodes or processing units. Each unit has links to other units, in which receive and send signals. Each unit can have local memory. These units are the simulation of neurons, receiving and relaying information. Add to the inputs and returns an output, if it is greater than the value of the sum.

A neural network may have one or multiple layers. For example with three layers could be input layer, wherein the receiving unit patterns; the middle layer, which is made processing and feature extraction; and the output layer, and has concluded that the outcome. The greater the number of layers, the better the learning ability.

The input layer must have a special unit referred to as bias, used to increase the degrees of freedom, allowing for a better adaptation, by the neural network, knowledge supplied to it [17].

C. OPF

Recently, novel frameworks for graph-based classifiers that reduce the pattern recognition problem as an optimum path forest computation (OPF) in the feature space induced by a graph were presented.

These kinds of classifiers do not interpret the classification task as a hyperplanes optimization problem, but as a combinatorial optimum-path computation from some key samples (prototypes) to the remaining nodes. Each prototype becomes a root from its optimum-path tree and each node is classified according to its strongly connected prototype that defines a discrete optimal partition (influence region) of the feature space.

The OPF-based classifiers have some advantages with respect to the aforementioned classifiers, which allow the development of real time applications for fraud detection in electricity systems [18]:

- one of them is free of parameters,
- they do not assume any shape/separability of the feature space,
- run training phase faster.

II. Methodology

The methodology is presented and used for the development of the algorithm to construct the load curves of residential consumers and case study for classification in salary ranges from residential customers through intelligent techniques.

A. Algorithm for database construction

We analyze the data presented by National Program for Energy Conservation (PROCEL) on the energy consumption habits of residential consumers, as well as socioeconomic data of families and the possession of the residences equipment, presented by Brazilian Institute of Geography and Statistics (IBGE) in the 2010 Demographic Census. This way, it is possible to develop the algorithm that constructs the residential load curves. The algorithm use as variables number of people, ages of family components, regions the country, economy factor (family income), size of residence, season and day of the week.

On this way, we can study the influence of social, cultural and economic characteristics on the load curves. These seven variables have relationships with each other and were called active variables, because they influence directly in possession of equipment and / or in consumer habits, with more or less intensity.

The first step of the algorithm is to determine which equipment composes the load of residence, defining the weights in the raffles realized for the variables that influence each equipment of possession (region, economy, number of people, age range and size residence).

After selecting the equipment that composes the house, the algorithm start the process of daily and monthly usage of each device, being that all seven variables influence on consumer habits. However, we have been decided to make some assumptions at this moment, because in order to determine the person's habit is not a trivial task.

The main reference is related to the average curves practiced by Brazilian Electricity Regulatory Agency (ANEEL), they indicate to the algorithm the required average standard of load curves [19].

Finally, the Fig. 1 shows a diagram to represent the entire load curves construction program.

We have been added more 13 variables taken from the load curves generated, which are conclusive or apparent variables as installed capacity, consumption, consumption class, average demand, maximum demand, load factor, demand factor, number of peaks, number of peaks above the average, number of valleys, number of valleys above the average, total points above the average and day of the month with the highest consumption.

All these variables are important because they characterize each type of consumer and are used to assist intelligent systems, defining energy consumption patterns and classifying consumers according to consumption.

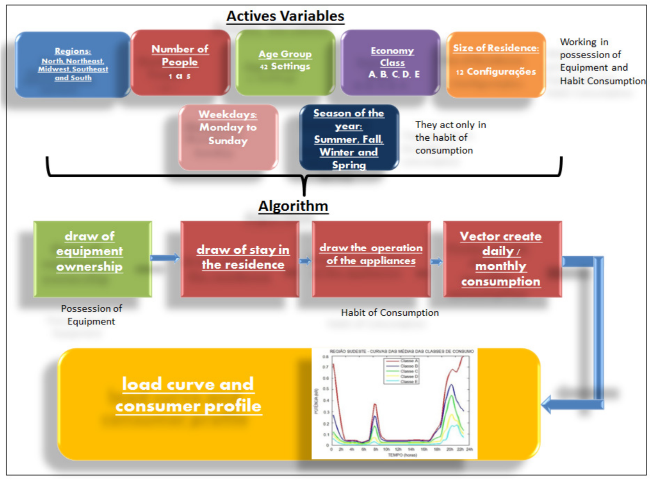

Figure 1. Schematic of Algorithm for Load Curves Construction.

B. Intelligent techniques for income classification

A study of the intelligent techniques application was performed to obtain the income classification of a family consumer. We have chosen three techniques for classification: ANN, OPF and SVM. These techniques are compared in terms of accuracy and execution time for classification of the same data set.

The algorithm constructs the dataset, which is composed by 16 features and consists of 12,600 residential consumers profiles.

Two rounds of experiments have been performed: the first that evaluates the accuracy of classification techniques on all 16 variables. In the last round, it was decided to remove the consumption class from the dataset. Thus, it can be seen if the consumption class will influence the family income classification.

The dataset was divided using 50% of samples for training phase and 50% for the test phase in both rounds of experiments.

The settings for the classification techniques are as follows:

- • OPF: There are no parameters to be chosen. We used the LibOPF, which is a library for the design of optimum-path forest-based classifiers [20].
- • ANN: Multilayer perceptron with one hidden layer of 10 neurons, tansig activation function in all layers and Levenberg-Maquardt training algorithm. We used Matlab (Neural Network Toolbox).
- • SVM: One against all strategy, C-SVC type, with C = 1, RBF kernel function and gamma = 0.2. We used the last version of the LIBSVM for Matlab [21].

Intelligent techniques have been used to classify the consumers in relation to family income. The Table I shows that there are 5 different classes, which are related to family income given in Reais.

TABLE I. FAMILY INCOME CLASSES [7]

Class	Minimum Wage (MW)	Family Income (R$)
A	Above 20	14,500.00 or more
B	10 to 20	7,250.00 to 14,499.99
C	4 to 10	2,900.00 to 7,249.99
D	2 to 4	1,450.00 to 2,899.99
E	Up to 2	Up to 1,449.99

The criterion for the division of consumers amount of income is applied by IBGE, other criteria have been developed and are used by Brazilian Association of Research Companies (ABEP), which take into account equipment ownership characteristics and educational levels of the heads of families. We opted for the IBGE criteria because presents economic aspects, allowing the study with relation to GDP.

The intelligent techniques that are being compared are widely used in research classification, but it is known that each of the techniques can be configured and adjusted empirically to obtain better results, but the principle is the best set to perform a further study depth beyond studying it better configuration of input specifications, as this get all the data suggested is rather costly.

III. RESULTS AND DISCUSSION

The methodology is applied to generate daily load curves according with social, economic and cultural aspects for residential consumers in Southeast of Brazil. The intelligent techniques have been applied for two rounds proposed in this work, with or without the family income economic factor.

A. Results of the proposed methodology

According with results through application of methodology, Fig. 2 presents an example of daily load curves generated for a particular consumer by the algorithm in Southeast of Brazil.

The most important aspect of the figure is the peak demand during the period of large consumption (between 18 and 20h). It also presented a mean curve of all curves obtained.

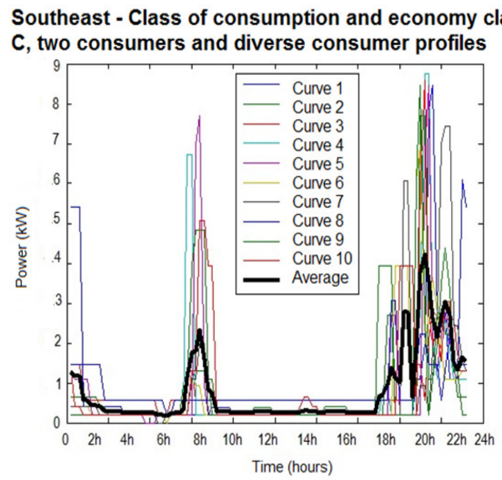

Figure 2. Random curves and mean curve in the Southeast Region for a particular consumer profile.

This algorithm allows us to evaluate either daily curve, such as monthly curves of energy consumption. In addition, it is possible to do statistics analysis of the influence of each features or variables in the consumer unit. However, this algorithm has been used to create a consistent dataset for this work, containing necessary information to be used in classification study through intelligent techniques. The average curve presented compared with actual measurements of load curves in the southeast are very similar, but a more critical analysis will be performed to validate correctly the proposed algorithm for database creation.

B. *Discussion of the results*

Tables I and II present the results of accuracy for consumption class classification and the execution time (in seconds) through intelligent techniques for the two rounds proposed.

TABLE II. ACCURACY FOR FAMILY INCOME CLASSIFICATION WITH ALL 16 FEATURES

Technique	Accuracy (%)	Execution time (s)
OPF	56.08 ±0.0	25.41
SVM	19.95±0.0	648.70
ANN	54.47±0.63	58.24

TABLE III. ACCURACY FOR FAMILY INCOME CLASSIFICATION WITH 15 FEATURES – WITHOUT CONSUMPTION CLASS

Technique	Accuracy (%)	Execution time (s)
OPF	56.10±0.0	24.99
SVM	19.95±0.0	643.71
ANN	54.93±0.89	45.90

We can see from the results given in Tables I and II that both OPF as ANN have similar accuracy with a slight advantage to the OPF. Furthermore, OPF completes the task around two times faster than ANN.

SVM showed a poor performance than other techniques with low accuracy and high execution time. We can also observe that the presence of the consumption class variable has not taken significant changes to the results, that is, this variable does not improve results.

IV. CONCLUSIONS

A set of residential consumer data has been created through an algorithm that takes into account social, economic and cultural aspects. This work performed an assessment of intelligent techniques to family income classification of this dataset.

We conducted experiments to evaluate the accuracy of ANN, SVM and OPF. In this sense RNA and OPF outperformed SVM, but OPF is faster than ANN. In experiments, we observed that the presence of the consumption class variable is not important to the family income classification. The results are very interesting if we consider the effectiveness and efficiency of the OPF classifier.

The next steps of the investigation will be conducted to improve the algorithm to validate load curves with real curves comparisons and study on the influence of each feature in the load curve, demonstrating its importance through evolutionary algorithms for optimization. This type of study will help the power distribution planning, demand response study and the implementation of smart grid.

V. REFERENCES

[1] Empresa de Pesquisa Energética (Brasil), "Balanço Energético Nacional 2013: Ano base 2012 / Empresa de Pesquisa Energética", Rio de Janeiro, 2013.

[2] A.G. Siraki, P. Pillay, P. Angers, "Full Load Efficiency Estimation of Refurbished Induction Machines From No-Load Testing", IEEE Transactions on Energy Conversion, vol. 28, no. 2, pp. 317-326, June 2013.

[3] K. M. Santos, A. N. Souza, D. S. Gastaldello, "Desenvolvimento de um Algoritmo para Determinar Curvas de Carga a partir da Posse e dos Hábitos de Consumo para Consumidores: Residenciais, Comerciais e Industriais", SBSE – Simpósio Brasileiro de Sistemas Elétricos, Goiânia, 2012.

[4] N. Kagan, C. C. B. Oliveira, E. J. Robba, "Introdução aos Sistemas de Distribuição de Energia Elétrica", 1ª Edição, São Paulo: Edgard Blucher, 2005.

[5] P. Faria, J. Soares, Z. Vale, H. Morais, T. Sousa, "Modified Particle Swarm Optimization Applied to Integrated Demand Response and DG Resources Scheduling", IEEE Transactions on Smart Grid, vol. 4, no. 1, pp. 606-616, March 2013.

[6] J.A. Jardini, C.M.V. Tahan, S.U. Ahn, S.L.S. Cabral, Determination of the Typical Daily Load Curve for Residential Area Based on Field Measurements. In IEEE Proceedings Supplements, T&D, Chicago, 1994.

[7] PROCEL, "Avaliação do Mercado de Eficiência Energética no Brasil: Classe Residencial, Relatório Brasil", Julho, 2007.

[8] Energy Balance of the State of São Paulo – 2014, Governo do Estado de São Paulo, Secretaria de Energia, 2014.

[9] European Environment Agency, "Final electricity consumption by sector, EU-27", 2013.

[10] EPE – Empresa de Pesquisa Energética , "Plano Nacional de Energia – Cenário Econômico 2050", Rio de Janeiro, 2014.

[11] ABRADEE – Associação Brasileira de Distribuidores de Energia, "Bandeiras Tarifárias", 2014.

[12] H. Morais, T. Sousa, Z. Vale, P. Faria, "Evaluation of the electric vehicle impact in the power demand curve in a smart grid environment", Energy Conversion and Management, vol. 82, pp. 268-282, June 2014.

[13] T. Zhao, Y. Peng, A. Nehorai, "An Optimal and Distributed Demand Response Strategy With Electric Vehicles in the Smart Grid", IEEE Transactions on Smart Grid, vol. 5, no. 2, pp. 861-869, March, 2014.

[14] G. Mateos, G. B. Giannakis, "Robust Nonparametric Regression via Sparsity Control With Application to Load Curve Data Cleansing", IEEE Transactions on Signal Processing, vol. 60, no. 4, pp. 1571-1584, April 2012.

[15] M. Chaouch, "Clustering-Based Improvement of Nonparametric Functional Time Series Forecasting: Application to Intra-Day Household-Level Load Curves", IEEE Transactions on Smart Grid, vol. 5, no. 1, pp. 411-419, January 2014.

[16] C. Cortes and V. Vapnik, "Support-Vector Networks," in Machine Learning. Kluwer Academic Publishers, 1995, pp. 273–297

[17] S. Haykin, "Neural networks – A comprehensive foundation", New York: Macmillan, 1994.

[18] J.P. Papa, A.X. Falcão, and C.T.N. Suzuki. Supervised pattern classification based on optimum-path forest". International Journal of Imaging Systems and Technology, 19(2):120–131, 2009.

[19] ANEEL - AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA. Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional– PRODIST, versão de 2008.

[20] J. P. Papa, C. T. N. Suzuki, A. X. Falcão, LibOPF: A library for the design of optimum-path forest classifiers", software version 2.1 (2014).

[21] Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines". ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27, 2011.