DOI: http://dx.doi.org/10.15446/ing.investig.v38n2.67331

Intelligent system for non-technical losses management in residential users of the electricity sector

Sistema inteligente para la detección de irregularidades en consumidores residenciales de empresas comercializadoras de energía

Miguel A. Uparela[1], Ruben D. Gonzalez[2], Jamer R. Jimenez[3], and Christian G. Quintero[4]

ABSTRACT

The identification of irregular users is an important assignment in the recovery of energy in the distribution sector. This analysis requires low error levels to minimize non-technical electrical losses in power grid. However, the detection of fraudulent users who have billing does not present a generalized methodology. This issue is complex and varies according to the case study. This paper presents a novel methodology to identify residential fraudulent users by using intelligent systems. The proposed intelligent system consists of three fundamental modules. The first module performs the classification of users with similar power consumption curves using self-organizing maps and genetic algorithms. The second module allows carrying out the monthly electricity demand forecasting through of recursive adjustment of ARIMA models. The third module performs the detection of fraudulent users through an artificial neural network for pattern recognition. For the design and validation of the proposed intelligent system, several tests were performed in each developed module. The database used for the design and evaluation of the modules was constructed with data supplied by the energy distribution company of the Colombian Caribbean Region. The results obtained by the proposed intelligent system show a better performance versus the detection rates obtained by the company.

Keywords: non-technical losses, irregular electricity consumption, fraud detection, intelligent systems.

RESUMEN

La identificación de usuarios con consumo fraudulento es una actividad importante en la recuperación de energía en el sector de la distribución. Este análisis requiere bajos niveles de error para minimizar las pérdidas eléctricas no técnicas en la red de distribución. Sin embargo, la detección de usuarios fraudulentos con facturación no tiene una metodología generalizada. Este es un problema complejo y varía de acuerdo con cada caso de estudio. Este artículo presenta una nueva metodología para la identificación inteligente de usuarios fraudulentos residenciales basada en sistemas inteligentes. El sistema inteligente propuesto consiste en tres módulos fundamentales. El primer módulo clasifica a los usuarios con curvas de consumo similares a través de mapas auto-organizativos y algoritmo genéticos. El segundo módulo realiza la predicción de consumos mensuales mediante ajustes recursivos de modelos ARIMA. El tercer módulo es el responsable de llevar a cabo la detección de usuarios irregulares por medio de una red neuronal para reconocimiento de patrones. Para el diseño y validación del sistema inteligente propuesto se realizaron pruebas en cada módulo que lo integra para diferentes tipos de clientes del mercado. La base de datos utilizada para el diseño y evaluación de los módulos fue construida a partir de los datos suministrados por la empresa de distribución de energía de la Costa Caribe Colombiana. Los resultados obtenidos por el sistema inteligente propuesto muestran un mejor desempeño frente a los índices de detección obtenidos por la empresa.

Palabras clave: pérdidas no técnicas, consumo irregular de electricidad, detección de fraudes, sistemas inteligentes.

Received: August 29th_2017
Accepted:

Introduction

The electricity sector has experienced a steady growth in the world. The demand for electricity grows at the same time as societies. The growth of population and the quality of life of people are increasing such demand. This situation generates a dynamic operation in the companies of the electricity sector. Sectors such as rural, industrial, residential, government and commercial can have different curves and peaks of demand. Therefore, the companies of the sector require the application of new technologies.

[1] B.Sc. in Electrical Engineering, Universidad del Norte, Colombia. Affiliation: Department of Electrical and Electronics Engineering, Universidad del Norte, Colombia. E-mail: muparela@uninorte.edu.co.

[2] M.Sc. in Electrical Engineering, Universidad del Norte, Colombia. Affiliation: Department of Electrical and Electronics Engineering, Universidad del Norte, Colombia. E-mail: rdgonzalez@uninorte.edu.co.

[3] M.Sc. in Electronics Engineering, Ph.D. Student, Universidad del Norte, Colombia. Affiliation: Department of Electrical and Electronics Engineering, Universidad del Norte, Colombia. E-mail: jmares@uninorte.edu.co.

[4] Ph.D. in Information Technology, Universitat de Girona, Spain. Affiliation: Department of Electrical and Electronics Engineering, Universidad del Norte, Colombia. E-mail: christianq@uninorte.edu.co

How to cite: Uparela, M., Gonzalez, R., Jimenez, J., Quintero, C. (2018). Intelligent System for non-technical losses management in residential users of the electricity sector. Ingeniería e Investigación, 38(2), 52-60. DOI: 10.15446/ing.investig.v38n2.67331

One of the common problems in this sector has been the electrical losses. Electrical losses can be classified into two groups (Sahoo, Nikovski, Muso, & Tsuru, 2015). On one hand, there are technical losses which usually occur due to the dissipation of energy. These losses are usual in elements of the network such as transmission lines, generators and transformers. On the other hand, non-electrical losses (NTL) are mainly caused by disturbances in measurement equipment, indirect and direct legal connections in the electrical network or human errors in the readings of measurements (Glauner, Meira, State, Valtchev, & Bettinger, 2016).

A type of standard methodology used by electricity companies to detect NTL is based on the study of customers who have null consumption during a certain period. This type of customer whose consumption is null is identified with a possible point of non-technical loss. The problem with this methodology is that this client does not always have an NTL because it can be, i.e. an unidentified unoccupied property (Guerrero *et al*., 2018).

The detection of fraudulent users who have billing is a field in which the implementation of new technologies has been frequent. The detection of irregular users through computational intelligence has been treated by many authors with several approaches. The development of intelligent systems has been an alternative and such systems have included techniques as artificial neural networks (ANNs) (Markoč, Hlupić, & Basch, 2011; Zheng, Yang, Niu, Dai, & Zhou, 2018), principle component analysis (PCA) (Singh, Bose, & Joshi, 2017), fuzzy models (Viegas, & Viera, 2017; Nagi *et al*, 2011), data mining (Chen *et al*., 2014) and support vector machines (SVMs) (Nagi J. *et al*., 2010; Pereira *et al*., 2016).

The use of systems based on neural networks is common in several cases of fraudulent user detection due to the flexibility and capability to associate the features of users with fraudulent patterns. By having real data, an artificial neural network can be trained in a supervised way, achieving suitable results due to learning by error correction (Markoč, Hlupić, & Basch, 2011; Zheng, Yang, Niu, Dai, & Zhou, 2018). However, one of the disadvantages of working with an artificial neural network is the need for real life samples other than the training set to evaluate the network performance.

Some approaches (Viegas, & Viera, 2017; Nagi *et al*, 2011) show how the implementation of a fuzzy cluster allows the classification of fraudulent and non-fraudulent users. These relationships are possible by a classifier called *C-means*. This classifier orders and rates the consumers profile depending on the measures distances in a range of the unit. As a result of the above, the consumers with high probability of being fraudulent or with irregular patterns score higher in the proposed range. In general, features such as average consumption of the last six months, maximum historical consumption, standard deviation of the last six months, total of irregularities found in the last six months and average consumption of customers with the same rate and geographic location in the last six months, they were used to develop the classification by this technique.

The implementation of advance metering infrastructure (AMI) has benefits in the detection of fraudulent users who have billing such as flexibility and adaptability in any electrical system, monitoring data in real time with reduction of electricity costs due to more precise consumption and more accurate location of non-technical losses. This result is achieved through the use of intelligent electronics devices, measures taken through an automated process and advanced measurement systems (Jiang *et al*., 2014; Leite & Mantovani, 2016; Lo, Huang, & Lu, 2012; Su, Lee, & Wen, 2016). The importance of this technology is due to the increase of efficiency of estimation algorithms and classification of users, such as distribution state estimation (DSE), A-Star algorithm and semi-definite relaxation (SDR).

Proposed methodology

The development of the methodology proposed in this work consists of the following fundamental stages:

i) Creation of subsets of users with similar consumption curve profiles (Clustering Module).

ii) Determination of the ARIMA model for the monthly power consumption forecast for each subset of users (Prediction Module).

iii) Determination of fraudulent users (Detection Module).

The proposed block diagram is described in Figure 1.

Creation of subsets of users with similar consumption curve profiles

The historical record of customer consumption was used like the cluster input due to this record is the only necessary feature for the identification of consumption patterns that determine the behavior of the users.

Self-organizing maps were implemented to perform the user classification. This is a special configuration of neural networks in problems of unsupervised learning. The design parameters of the neural network are: number of neurons, network topology, distance function and number of training steps for the initial covering of the input space. The values of the parameters determine the performance and results of the neural network. Therefore, it was necessary to obtain the set of values that will get reliable performance. In response to the above, a multi-objective optimization problem, whose objective functions are to minimize the number of clusters (categories) while grouping performance is maximized, was raised. The first objective function is aimed at reducing runtimes and computational cost while the second is aimed at obtaining the best possible results in the cluster.

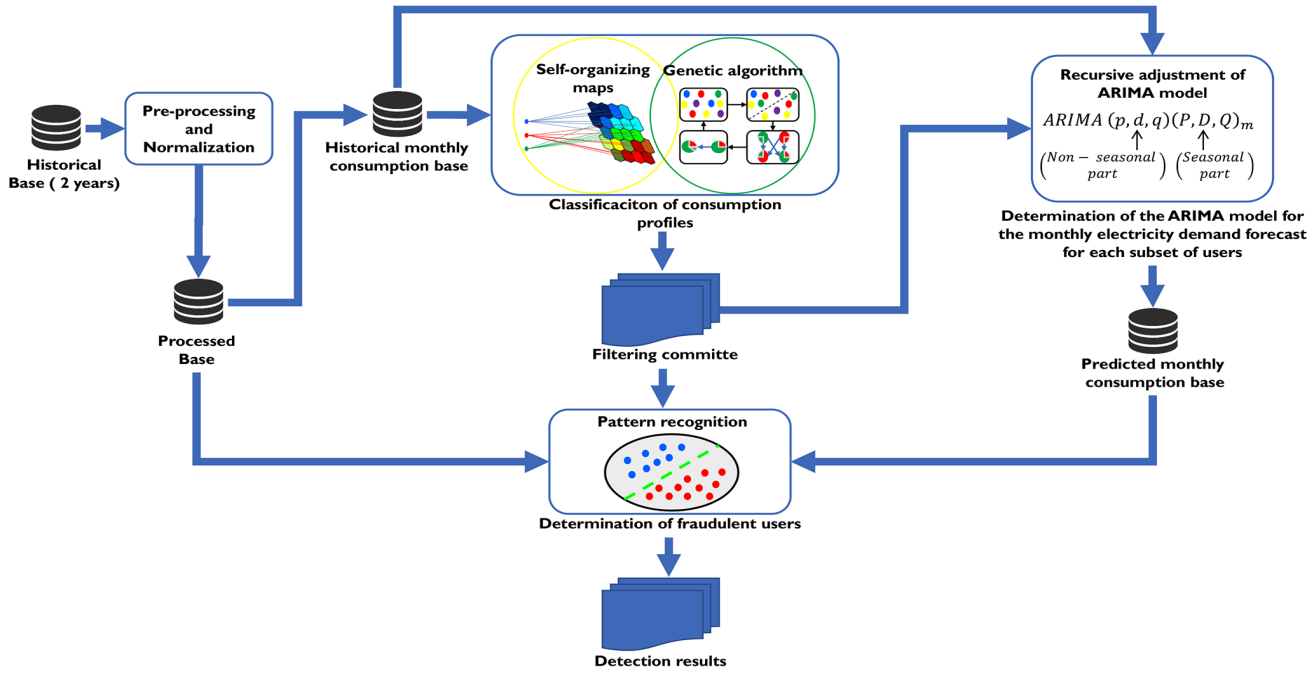

Figure 1. Blocks diagram of the proposed system
Source: Authors

The implemented solution to obtain the optimum values of the parameters of the neural network is a genetic algorithm, which belongs to a search and optimization approach known as meta-heuristic. For each parameter of the neural network, a minimum and maximum value was established. With the minimum and maximum values, two threshold vectors were formed for the search. In addition, it was necessary to include a buffer that stores the already evaluated combinations as good to reduce the number of evaluations. Therefore, if one of the stored combinations is regenerated by a mutation or cross-linking, it is not evaluated because it has already been stored. The final structure of the cluster stage is described in Figure 2, which proved to be a hybrid between neural networks and genetic algorithms.

Determination of the ARIMA model for the monthly electricity demand forecast for each subset of users

The functional structure of the prediction system is mainly based on an integrated autoregressive model of moving average (ARIMA). The system inputs consist of a customer's power consumption history and an indicator of the group to which the user belongs. In order to choose a suitable ARIMA model, the iterative script was developed. In this

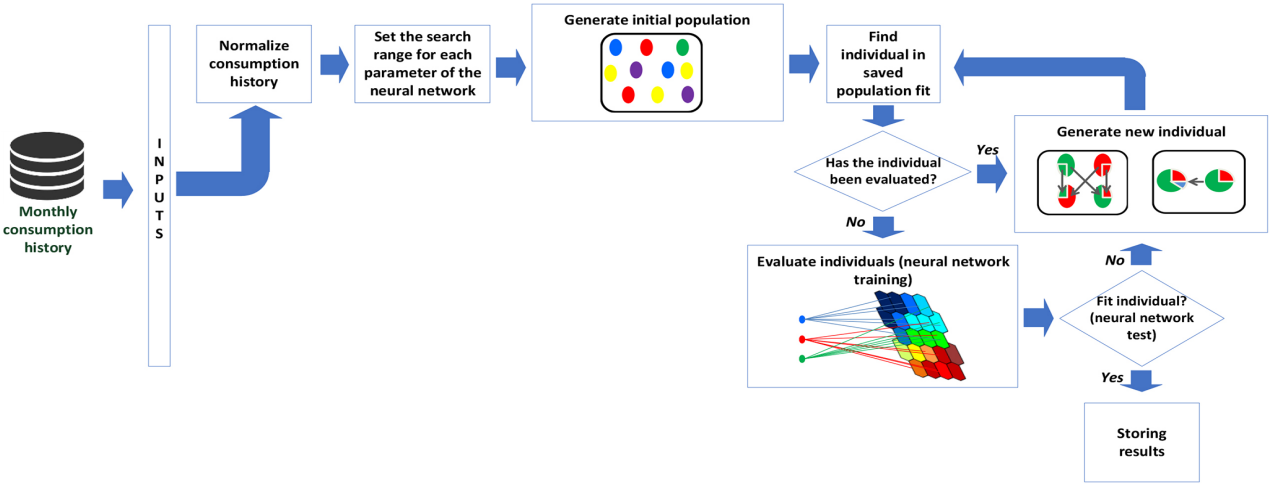

Figure 2. Blocks diagram of the clustering module
Source: Authors

script, the parameters related to the integral and moving average polynomials for the non-seasonal part and the degree of the polynomial of the moving average of the seasonal were varying for several subsets of customers. The parameters with a best performance are chosen for carrying out the power consumption forecasting.

The search for the degrees of polynomials for the seasonal and non-seasonal part is reached through the variation of the degree by means of iterations (Mares, Mercado, & Quintero, 2017). At each iteration, a possible combination is created that gives rise to a model determined by Equation (1).

$$Y_t = -\left(\Delta^d Y_t - Y_t\right) + \phi_0 + \sum_{i=1}^{p} \phi_i \Delta^d Y_{t-i} - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t \quad (1)$$

Where:

d it corresponds to the degree of integration polynomial.

q it corresponds to the degree of the moving average polynomial.

p it corresponds to the degree of the autoregressive polynomial.

ϕ_i it corresponds to the parameters belonging to the autoregressive part of the model.

θ_j it corresponds to the parameters belonging to the moving average part of the model.

ϕ_t it is a constant.

ε_i it corresponds to the error term (also called innovation or stochastic perturbation).

Each combination generates a model which is evaluated considering how well it fits the original series. The objective is to maximize the fit function which represents the *Normalized Root Mean Square Error* (NRMSE) and it is determined by Equation (2).

$$NRMSE = 100\left(1 - \frac{\|y - \hat{y}\|}{\|y - mean(y)\|}\right) \quad (2)$$

Where,

y it corresponds to the original time series.

$\hat{y}$ it corresponds to the set time series.

When a new combination is generated, the performance of the ARIMA model is calculated. The iterations are performed until the first performance is greater than or equal to the established threshold. If no ARIMA model meets the above condition, it selected the model with the best performance among all. Figure 3 shows the proposed monthly load-prediction model.

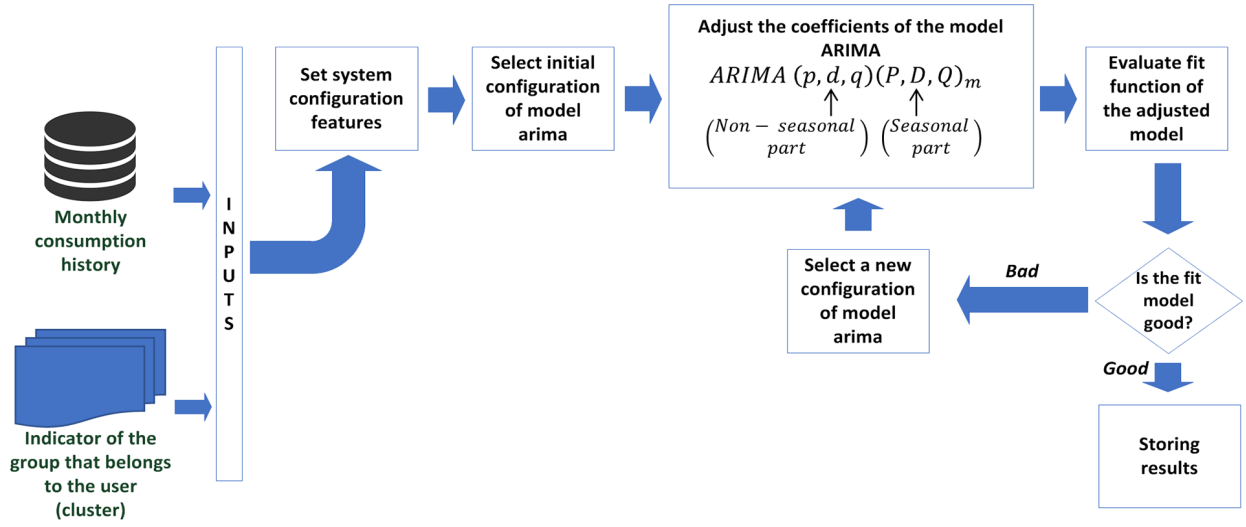

Figure 3. Blocks diagram of prediction module
Source: Authors

Determination of fraudulent users

The results obtained from clustering and prediction stages were used in the analysis to establish the structure of input features of the neural network implemented in the identification of users with irregular electricity consumption.

The approach of the proposed model is based on the classification of users into two classes: fraudulent and non-fraudulent. Therefore, it was necessary to combine user samples with fraudulent and non-fraudulent characteristics into training and testing sets to improve the performance of proposed system. The problem addressed in this work is a supervised classification because there are real inspections'

data to be used for the training. Therefore, a pattern recognition network is used as alternative methodology to solve the addressed problem (Jokar, Arianpoo & Leung, 2016). This is a feed forward network with supervised learning that can be trained to classify inputs according to target classes. Figure 4 shows the methodology proposed for the detection stage.

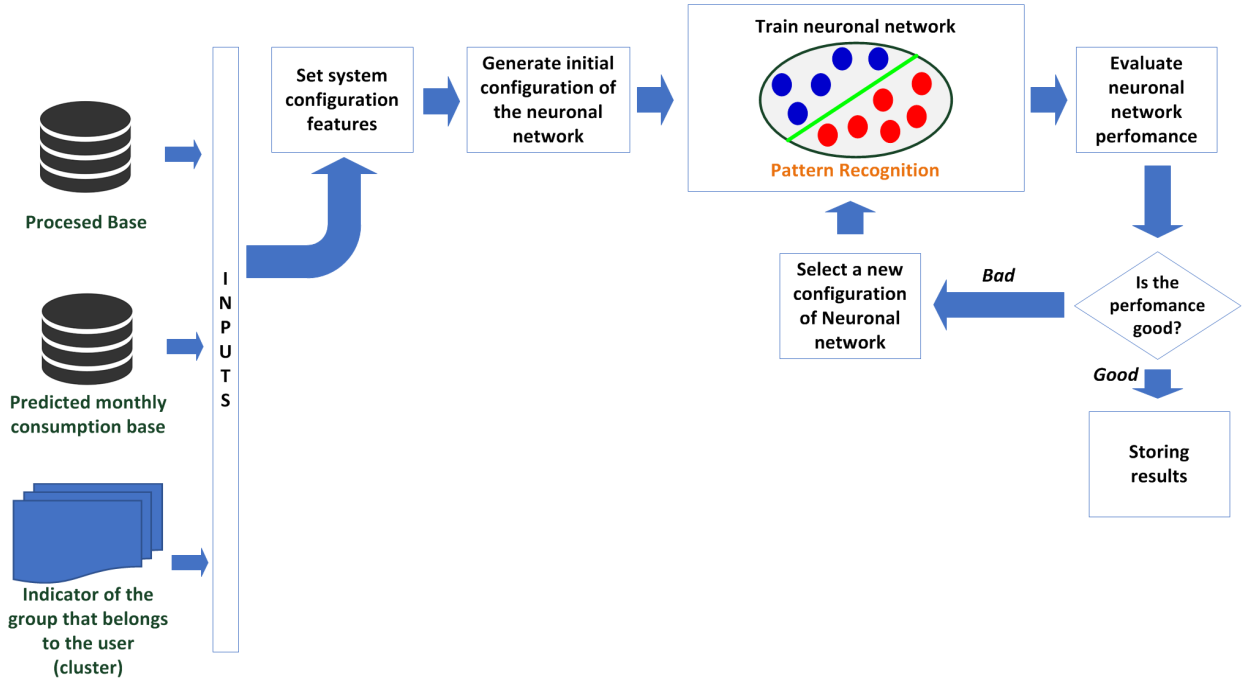

Figure 4. Blocks diagram of detection module
Source: Authors

2. Anomaly reading: it describes irregular situations in user measurements.

3. Power supply state: it refers to the service status of the user, (e.g., suspended, connected and without contract).

4. Type of power supply: it refers to the indicator of user consumption type, (e.g., common building area, direct connection, fixed consumption, normal, historical average, average of the stratum).

5. Overdue billing: it corresponds to the number of invoices not paid by the user.

6. Tariff: it corresponds to the residential tariff class of the user.

7. Type of reading: it describes the type of user's measurement.

8. Cluster indicator: it corresponds to the indicator of the group to which the user belongs.

The features used as inputs of proposed model based on neural networks are described below:

1. Power consumption deviation: the deviation of the energy consumption corresponds to the difference between the actual and predicted consumption in the prediction module.

The data used for the training correspond to the months of July to December of 2015. On the other hand, the data for testing were selected from the months of January to July of 2016.

During the training of the neural network, it was implemented the methodology of (Jimenez, Donado, & Quintero, 2017) to find the best neural network by varying the parameters in the configuration. The modified parameters were:

1. Number of neurons in the hidden layer.

2. Number of hidden layers.

At each iteration, a new neural network was configured and trained with the same training set in all cases. After training, the neural network was validated and the Mean Absolute Percentage Error (MAPE) was calculated. MAPE was adopted as a performance indicator allowing a quantitative comparison of each evaluated configuration. Finally, the best neural network was selected with the least MAPE.

Results

This section presents the results obtained by the clustering, prediction and detection modules that integrate the system proposed in this work.

Clustering module

The clustering module aims to group users with similar behaviors into their monthly power consumption curves. However, fulfilling this goal is a very broad task, so it was necessary to integrate the searching technique of genetic algorithms in order to achieve the least amount of groups with the highest possible quality of clustering.

The execution of the multi-objective search algorithm yielded results in the following set of entries belonging to the Pareto frontier solution (see Table 1).

Table 1. Set of the optimal net parameters

Input	Values				
N° of neurons	4	9	16	25	36
Topology	Hexagonal	Grid	Random	-	-
N° of training steps	50	100	150	-	-
Neuron distance function	Euclidean	Link	Manhattan	Box	-

Source: Authors

The outputs of the neural network for each of the resulting entries sets are shown in Table 2. Some examples of results were obtained from the execution of the neural network in the optimality condition: 6^2 neurons, 100 training steps,

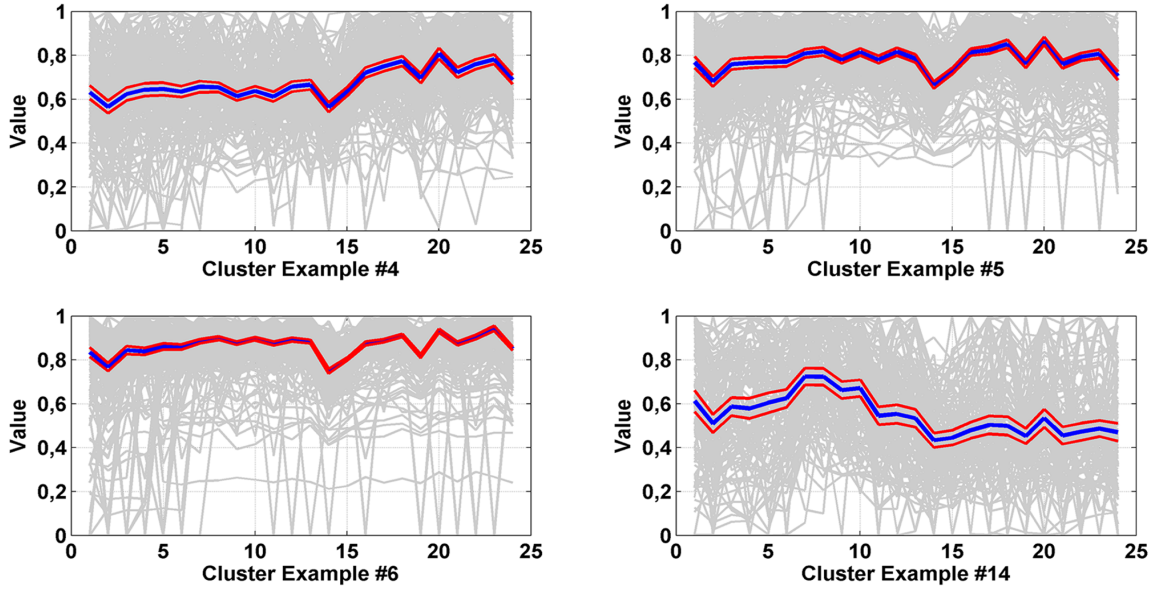

Figure 5. Examples of user profiles (cluster) obtained. The blue curve represents the average of monthly consumption curves. The red curves represent the confidence intervals.
Source: Authors

Hexagonal topology and Euclidean distance function with 24 periods. Figure 5 shows the result of one of the configurations evaluated during the training of the cluster where about 4000 users were used.

Table 2. Neural network performance with optimal entries

Network Entries	
N° of Clusters	**Performance**
4	0,16
9	0,14
16	0,13
25	0,12
36	0,11
49	0,11

Source: Authors

Prediction module

To evaluate the performance of the prediction system, the Normalized Root Mean Square Error (NRMSE) was used as a performance indicator. This indicator seeks to ensure that the initial assumption is met, which consists in finding the best model that fits the time series described by the user's historical power consumption curve. If the model found is able to fit with great accuracy to the time series, it is assumed that the model can correctly forecast the value of the power consumption of the following month. Finally, the absolute mean percentage error (MAPE) was used to figure out the model performance. To test the performance of the prediction module, 1000 users were selected randomly for each test. The selection of users was made from the total of 4000 users used in the clustering stage. Four tests were done and the results are shown in Table 3.

Table 3. Test results of the prediction system

	Results	
N° test	MAPE Mean (%)	NRMSE Mean (%)
1	3,12	86,45
2	3,79	87,89
3	3,93	87,77
4	3,80	87,93

Source: Authors

Figure 6 shows the MAPE cumulative average obtained for each user of the test samples. The average of MAPE is close to the value described in Table 3 and the highest concentration of the MAPE distribution is below average.

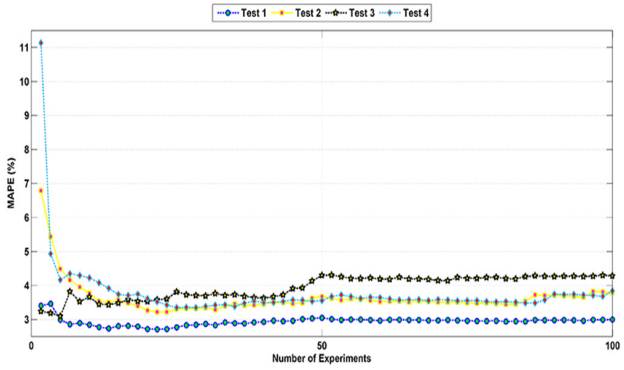

Figure 6. Cumulative average for MAPE.
Source: Authors

Figure 7 shows the NRMSE cumulative average for each user of the test samples. The most users obtained a fit or similarity between the adjusted series and the original series above average.

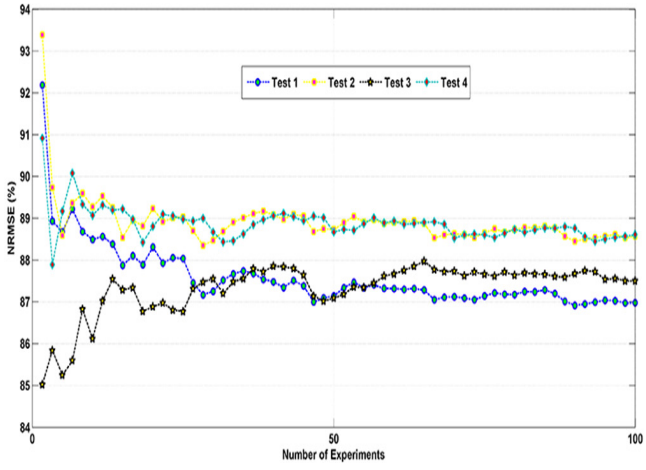

Figure 7. Cumulative average for NRMSE.
Source: Authors

Detection module

The performance of the ANN was figured out through the area under the ROC curve (AUC). This indicator can be interpreted as the probability that a classifier will order or score a positive instance randomly chosen higher than a negative one. In this sense, the performance of the detection module was evaluated according to the experimental design based on observing the behavior of the neural network for different tariffs of the database.

Response variable:

AUC: This value corresponds to the area under the ROC curve. The curve is the representation of the true positive rate (TPR) against the false positive rate (FPR) (Zheng, Yang, Niu, Dai, & Zhou, 2018). The TPR, determined by Equation (3), defines how a classifier is able to detect or classify correctly positives cases of all positive cases available during the test. The FPR, determined by Equation (4), defines how many positive results are incorrect among all the negative cases available during the test.

$$TPR = \frac{\sum True\_positive}{\sum Condition\_positive} \quad (3)$$

$$FPR = \frac{\sum False\_positive}{\sum Condition\_negative} \quad (4)$$

Independent variable:

Tariff: it corresponds to the socio-economic level of a user. This variable is also related to the tariff in kWh on the user's electric power service.

For the proposed experimental design, the following scenarios were defined according to the independent variable selected as shown in Table 4. The selected data were randomly selected in groups of 100 users for each tariff. The number of repetitions used was 100. The data between the months of January and July of 2016 were used to verify the detection module performance.

Table 4. Equivalences of residential stratum

Tariff	Equivalence
Residential stratum 1	Tariff 1
Residential stratum 2	Tariff 2
Residential stratum 3	Tariff 3
Residential stratum 4	Tariff 4
Residential stratum 5	Tariff 5
Residential stratum 6	Tariff 6

Source: Authors

The results achieved are shown in Figure 8, where the accumulated average of the AUC for each type of tariff is displayed. The minimum value reached among all experiments is approximately 88%, a very important fact, since the average success rate of the energy company in the Colombian

Caribbean region is 70%. Therefore, the detection rate of fraudulent customers could be increased by at least 18%.

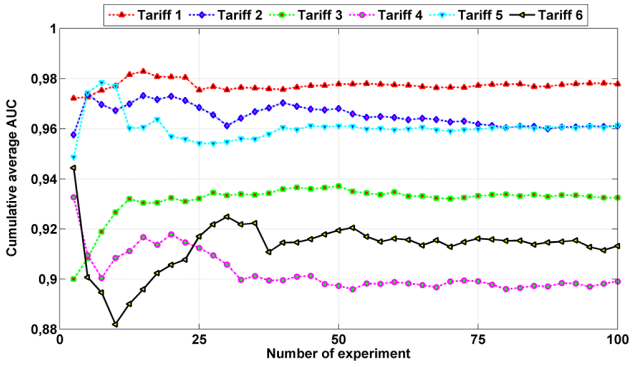

Figure 8. Cumulative average AUC.
Source: Authors

Conclusions

Results in the clustering and prediction stages improved the performance in the detection stage. Figure 8 shows that the proposed methodology allows developing a detection model with an average AUC of 94%. Even from the results, the highest hit rates are concentrated in users of tariffs 1, 2 and 5. This behavior is due to two situations. First, the number of irregular users is greater in tariffs 1 and 2; therefore, they have more samples to define behavior patterns. Second, the number of irregular users is much smaller but the number of non-fraudulent users is greater in tariff 5, therefore, recognition of the patterns of this type of users is made better. In the consumption prediction stage, the proposed model generates adjusted time series very like the original series and for each test the results were very close to each other, i.e., there was not much deviation between the results of each test. Therefore, the proposed model is able to adjust to the real-time series by 87% and predicts values with an average error of 3.5%.

Taking into account the results reached in the detection stage, it is shown that the ANN implemented can work as a reliable tool for the detection of irregular users. However, to obtain suitable results, it is necessary to have a best possible database of samples for the training stage. For the case study, when the neural network is well trained, the results were suitable because the percentage of detection was approximately of 94%. Therefore, the lack of real samples of irregular users can become the greatest disadvantage when implementing an approach such as the proposed in this paper.

As a future work, the authors will continue with the development of an optimization system that provides a suitable route for carrying out recovery billing campaign from the fraudulent users detected.

Acknowledgements

This work was supported by COLCIENCIAS and Universidad del Norte, Barranquilla, Colombia.

References

Chen, H., Fei, X., Wang, S., Lu, X., Jin, G., Li, W., & Wu, X. (2014, November). Energy Consumption Data Based Machine Anomaly Detection. In Advanced Cloud and Big Data (CBD), 2014 Second International Conference on (136-142). IEEE.

Guerrero, J. I., Monedero, I., Biscarri, F., Biscarri, J., Millán, R., & León, C. (2018). Non-Technical Losses Reduction by Improving the Inspections Accuracy in a Power Utility. IEEE Transactions on Power Systems, 33(2), 1209-1218.

Glauner, P., Meira, J., State, R., Valtchev, P., & Bettinger, F. (2016). The challenge of non-technical loss detection using artificial intelligence: A survey. arXiv preprint arXiv:1606.00626.

Jiang, R., Lu, R., Wang, Y., Luo, J., Shen, C., & Shen, X. S. (2014). Energy-theft detection issues for advanced metering infrastructure in smart grid. Tsinghua Science and Technology, 19(2), 105-120.

Leite, J. B., & Mantovani, J. R. S. (2016). Detecting and locating non-technical losses in modern distribution networks. IEEE Transactions on Smart Grid, PP, 1.

Jokar, P., Arianpoo, N., & Leung, V. C. (2016). Electricity theft detection in AMI using customers' consumption patterns. IEEE Transactions on Smart Grid, 7(1), 216-226.

Lo, Y. L., Huang, S. C., & Lu, C. N. (2012, May). Non-technical loss detection using smart distribution network measurement data. In Innovative Smart Grid Technologies-Asia (ISGT Asia), 2012 IEEE (1-5). IEEE.

Mares, J. J., Mercado, K. D., & Quintero, C. G. (2017). A methodology for short-term load forecasting. IEEE Latin America Transactions, 15(3), 400-407.

Markoč, Z., Hlupić, N., & Basch, D. (2011, June). Detection of suspicious patterns of energy consumption using neural network trained by generated samples. In Information Technology Interfaces (ITI), Proceedings of the ITI 2011 33rd International Conference on (551-556). IEEE.

Nagi, J., Yap, K. S., Tiong, S. K., Ahmed, S. K., & Nagi, F. (2011). Improving SVM-based nontechnical loss detection in power utility using the fuzzy inference system. IEEE Transactions on power delivery, 26(2), 1284-1285.

Nagi, J., Yap, K. S., Tiong, S. K., Ahmed, S. K., & Mohamad, M. (2010). Nontechnical loss detection for metered customers in power utility using support vector machines. IEEE transactions on Power Delivery, 25(2), 1162-1171.

Pereira, D. R., Pazoti, M. A., Pereira, L. A., Rodrigues, D., Ramos, C. O., Souza, A. N., & Papa, J. P. (2016). Social-Spider Optimization-based Support Vector Machines applied for energy theft detection. Computers & Electrical Engineering, 49, 25-38.

Sahoo, S., Nikovski, D., Muso, T., & Tsuru, K. (2015, February). Electricity theft detection using smart meter data. In Innovative Smart Grid Technologies Conference (ISGT), 2015 IEEE Power & Energy Society (1-5). IEEE.

Singh, S. K., Bose, R., & Joshi, A. (2017, December). PCA based electricity theft detection in advanced metering infrastructure. In 2017 7th International Conference on Power Systems (ICPS) (441-445). IEEE.

Su, C. L., Lee, W. H., & Wen, C. K. (2016, March). Electricity theft detection in low voltage networks with smart meters using state estimation. In Industrial Technology (ICIT), 2016 IEEE International Conference on (493-498). IEEE.

Viegas, J. L., & Vieira, S. M. (2017, July). Clustering-based novelty detection to uncover electricity theft. In Fuzzy Systems (FUZZ-IEEE), 2017 IEEE International Conference on (1-6). IEEE.

Zheng, Z., Yang, Y., Niu, X., Dai, H. N., & Zhou, Y. (2018). Wide and Deep Convolutional Neural Networks for Electricity-Theft Detection to Secure Smart Grids. IEEE Transactions on Industrial Informatics, 14(4), 1606-1615.