18th International Conference on Renewable Energies and Power Quality (ICREPQ'20)
Granada (Spain), 1st to 2nd April 2020
Renewable Energy and Power Quality Journal (RE&PQJ)
ISSN 2172-038 X, Volume No.18, June 2020

Application of Neural Networks to Determine the Customer Connectivity based on Smart Meters

A. Gastalver-Rubio[1], and R. Carmona-Pardo[1]

[1] Ingelectus
Calle Leonardo da Vinci 18, 41092 Sevilla (Spain)
Phone: +0034 954 460446, e-mail: agastalver@ingelectus.com, rcarmona@ingelectus.com

Abstract. Along with the massive installation of Smart Meters in the distribution grid, new applications, such as state estimation, have been developed in order to improve the operation of the electrical network. Such applications require a faithful knowledge of the network topology, specifically the feeder and phase where the customers are connected to. Classical solutions for this complex combinatorial problem usually fail in such mission. Fortunately, with the development of artificial intelligence techniques, such as machine learning through neural networks, it can be solved efficiently. This works shows the results of applying, to different currently-operating distribution grids, artificial neural networks which discover the customer connectivity to the network using smart meter measurements.

Key words. Artificial Intelligence, Neural Networks, Machine Learning, Distribution Grid, Smart Grids, Network Connectivity

1. Introduction

The installation of new measurements devices in the low voltage distribution grid, including the Smart Meters in every consumer and distributed generation, and the secondary substation meters, is continuously providing a large quantity of electrical data. However, distribution system operators keep this information in databases which are mainly used for billing purposes and rarely leveraged for extracting new knowledge. Although the computer memory and storage is highly efficient, taking advantage of the measurement data can benefit not only the electric market, but also improve the planning and operation of the low voltage distribution grid. For this reason, the research and development of new technologies capable generate knowledge by exploiting low voltage data has become a high-interest matter within the research centers and the energy distributors.

This topic has also raised the interest of the governmental associations, who have created different programs, for financing related research and development projects, in example, the FEDER Interconnecta program. The latest includes multiple initiatives, such as the MONICA project [1],[2], and PASTORA project [3], who have become pioneers in the application of state estimation in medium voltage and low voltage distribution networks, not only being advantageous for the distribution system operator but also increasing the state of the art of the line of research [4]. Others have being studying the leverage of this data and results to the discovering of non-technical losses [5], and to the optimization of the network losses and unbalances by an unbalance operation of the distributed generation [6].

Nevertheless, these researches and applications often have to deal with an obstacle when being applied in an operational environment: the definition of the network topology. Database errors and misinformation, uncertainty in the network connectivity, and the constant change of the distribution grid, are the main source of errors when applying the previous technologies or when calculating energy balances. Confidence in the connectivity of consumers and distributed generation to the three-phase four-wire low voltage network is an essential base to sustain the development of new applications.

In such mission, previous works have been proposed to solve this problem by using different sources of information. Some uses geographical information [7] to correctly associate consumers to a feeder or secondary substation. Others improved this methodology by including the voltage measurements [8]. However these methods ignored the connectivity to a electrical phase and failed when the geographical information system was missing or uncompleted.

In current works at industrial environments, the usage of current measurements differences [9] is proposed to verify the network connectivity, in both, to feeder and phase. However, it requires the installation of an specific manufacturer device in the secondary substation, and the

update of the Smart Meter firmware, which could become a drawback to some distribution system operators.
Along with the advance in artificial intelligence technologies, such as deep neural networks [10], new implementations can provide an statistical leverage of the historical and continuously generated data, in order to solve the connectivity problem whilst providing new features to the solution, such as three-phase consumers characterization or non-technical losses detection. This works shows the results of applying this emerging technology to discover consumer connectivity in a real operating environment, Enel's Smartcity Málaga.

2. Modelling and Methodology

This section shows the elementary foundations of the proposed connectivity modelling and methodology, and clarifies the performance indicators that have been applied in the benchmarking to validate the technology.

The following design is defined by the constraint that Ingelectus focus on the development of a tool capable of solving the problem while taking into account that the technology must be: applicable to all currently-operating distribution grids with a minimum implementation cost, provide as much information as possible whilst adapting to the constant change of the network, capable of leveraging previously generated historical data to solve the problem in early stages of its implementation, and independent of the meters' manufacturers.

A. *Measurement Dataset*

Although big distributors may have a significantly wider range of electrical data within their SCADA systems, both small and big distributors are using the energy measurements in order to bill their consumers. Bearing the stated constraints, the usage of active energy measurements and historical data is taking as an input for the methodology used in the first implementation. This work also shows an second implementation, including reactive energy measurements, which improves the results of the solution.

In order to the methodology to detect connectivity to a phase and feeder, energy measurements from the secondary substation are necessary. Fortunately, along with the increase in distributed generation and flexible grids, this is becoming common within the Spanish distributors.

B. *Connectivity Model*

The work assumes that the topology of the network, electrical lines and nodes, is uncertain or unavailable. Thus, for each consumer and feeder or phase, a connection probability, or feed, is assigned and will be expected to be corrected by the input data.

The output of the algorithm is a classification based in a probability of the consumers within the stated network.

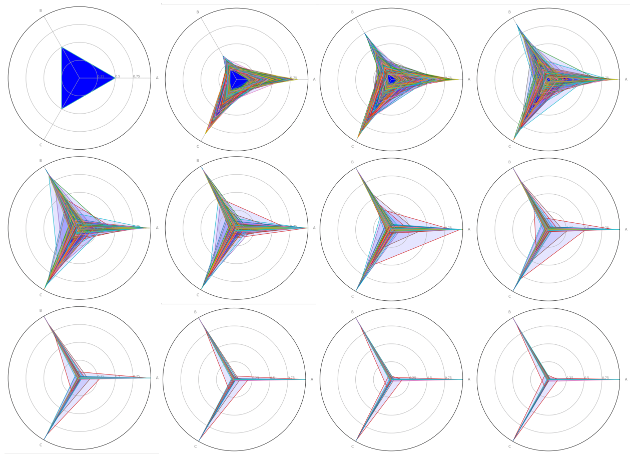

Fig. 1. Representation of iteration learning of the Ingelectus' neural network classification algorithm in a 400 single-phase consumers network.

In the Fig. 1, the Ingelectus learning algorithm, based in neural networks, corrects the connection probability, normalizes and classifies the consumers depending on their nature:

1) *Single-phase consumers (1-P)*: a list of single-phase consumers with its feeder and phase is stated as an output.
2) *Three-phase consumers (3-P):* a list of three-phase consumers along with three percentage values are returned per client, indicating the most probable consumption characterization.
3) *Non-consumers (NC):* a list of consumers which were included as an input but that are statically not belonging to the feeders or phases included.

These outputs are often useful for applying energy balances which provides a better understanding of where the technical and non-technical losses resides.

C. *Considered limitations*

The objective of the algorithm is to solve the problem in a industrially relevant and operation environment. Thus, different limitations must be considered for the data used for learning and classification. The most common errors when dealing with real operational data are: time synchronization, measurement deviation error, non-technical losses or fraud connections, and misplacing or lacking of information.

The proposed algorithm solves the time synchronization and measurement deviation error by assuming a Gaussian distribution in the provided measurements. However, the latest limitations can affect the success of the results. The decrease in performance of these limitations are studied in the results sections of this paper.

D. *Key Performance Indicators*

As a classification problem, it is important to define those metrics used to evaluate the performance of the algorithm.

These metrics are the same to phase identification as well as feeder identification. The Key Performance Indicators (KPI) used for that purpose are those that fulfil a confusion matrix. Using this matrix applied to the problem solved in this paper, the two main conditions are:

4) *Positive condition (P)*: Smart Meter identified, it belongs to a phase or feeder
5) *Negative condition (N):* Smart Meter not identified in reality, it does not belong to any phase or feeder

Once defined the conditions, hereunder it is detailed the four different outcomes depicted in the confusion matrix.

1) *True positive (TP)*: Smart Meter identified by algorithm matching with reality.
2) *False positive (FP):* Smart Meter identified by algorithm not matching with reality.
3) *False negative (FN):* Smart Meter not identified by algorithm while it is identified in reality.
4) *True negative (TN)*: Smart Meter not identified by algorithm, matching with reality.

For this study, success is defined as guessing properly whether a Smart Meter belongs or not to a phase or feeder, hence: **TP + TN.**

3. Benchmarking Results

This section is devoted to show and discuss the results obtained in the different test sets that have been carried out. Every test has been done using active and reactive energy from real clients and header sensors installed in some Secondary Substations (SS) contained in the scope of the Smartcity Málaga Living Lab from Endesa. In order to understand the insights provided by the benchmarking, hereunder every test is explained.

A. *Montecarlo Simulations*

Since the beginning of these studies, the main objective was to develop a tool capable of adapting to every situation that may occur during its implementation and could affect to its performance. Over the course of the tests, two drawbacks have been faced:

1) *Communication problems*: since Smart Meters (SM) where installed in field, utilities have to cope with the challenge of communicating with their whole SM fleet. This involves a lack in the energy curves of a part of their clients.
2) *Initial clients connection uncertainty*: there is no guarantee that the SMs whose energy curves are inputs to the algorithm are actually connected to the Secondary Substation.

Both drawbacks have been considered as a non-technical losses parameter.
In order to simulate these kind of situations and study the performance of the algorithm, forty different scenarios has been used with Montecarlo's algorithm. The idea generating these scenarios was to increase, randomly, the percentage of non-technical losses from 0% to 100%.
The following image compares the evolution of the identification algorithm's performance (red points) to the success obtained choosing the element to which a Smart Meter is connected randomly (green points).

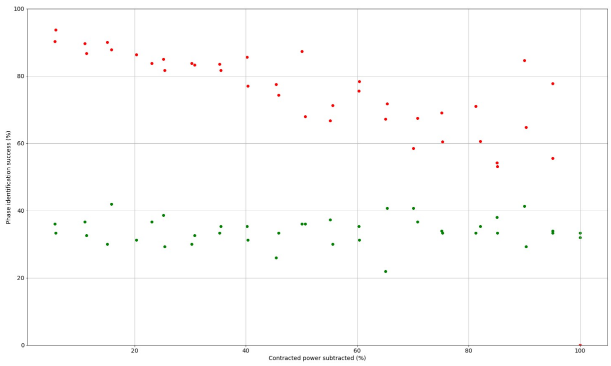

Fig. 2. Comparison between Ingelectus' algorithm success and random selection success.

As it can be seen in Fig. 2, even having a 40% of non-technical losses, the success of the algorithm is greater than 80%.

B. *Benchmark 1: Proof-of-Concept*

The very first Proof-of-Concept (PoC) in which the algorithm was displayed is *PASTORA* project (*Preventive Analysis of Smart grid with real Time Operation and Renewable Assets Operation)*. In this PoC, the algorithm was applied in a particular Secondary Substation formed by 5 feeders and 121 clients downstream connected, particularly 86 single-phase meters and 35 three-phase meters, with a total of 0.77 MW of contracted power.

Using the data from these Smart Meters to find the phase and feeder identification in a completely real situation and, then, compare with the original connections was the main idea of this POC. For that purpose, some field inspections were carried out in order to guarantee that the original identification to which the algorithm's results will be compared is correct.

The characteristic of the implementation is using active and reactive energy curves available for a period of 1.5 months and deal with a 50% of non-technical losses. In this case, the main objective was to demonstrate how active (P) and reactive (Q) energy curves support phase and feeder identification working in conjunction rather than input them independently.

Table I. – POC results

Input data	Phase identification success (%)	Feeder identification success (%)
P + Q	79.70	66.00
P	9.50	20.20
Q	45.23	9.20

Table I shows how using both sources of data as input strongly benefits the results.

C. *Benchmark 2: Operational Environment Validation*

The Operational Environment Validation of Ingelectus' tool has also been displayed in *PASTORA* project but, this time, the target was to test how it behaves being implemented in a few different kind of SS within the scope of the project, varying the amount of clients connected downstream and using a month of data of only active energy curves. The main characteristics of each SS are detailed in Table I.

Table II. – Operational Environment Validation scope

SS	Number of clients	Contracted Power
SS_1	235	1.03 MW
SS_2	410	1.42 MW
SS_3	548	3.75 MW
SS_4	631	2.14 MW
SS_5	840	2.75 MW
SS_6	1045	6.78 MW

Other important aspect about this second benchmark is that, in addition to complexity that involves identify phase and feeder of a SS consisting of up to 1045 clients, it has been also simulated the two drawbacks previously introduced by generating 28 validation scenarios, focusing on two aspects:

6) *Clients that actually belongs to the SS (D1).* Every scenario misses a particular percentage of energy curves from the SS's actual clients. This percentage goes from 0% to 30%, with a step of 5%. Depending the size of the SS, every percentage will imply a different number of clients.
7) *Clients that belongs to adjacent SS (D2).* Every scenario includes a particular percentage of energy curves from the clients that actually belongs to an adjacent SS. This percentage goes from 0% to 7.5%, with a step of 2.5%.

Hereunder it is presented the evolution of the phase identification's success for all the SS, varying D1 from 0% to 30% (represented in X-axis as amount of clients) and fixing D2 to 0%.

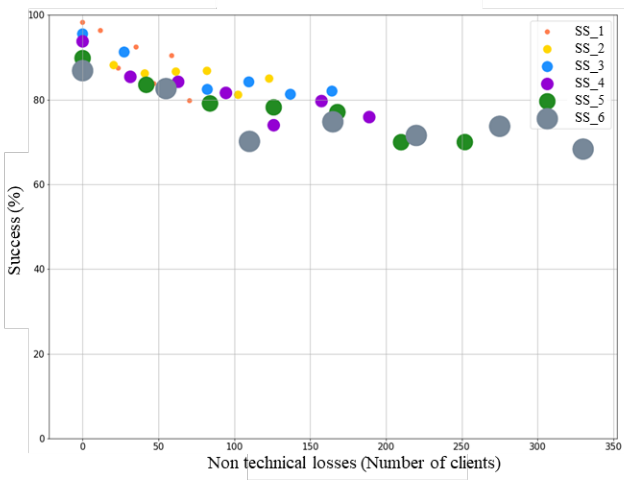

Fig. 3. Evolution of phase identification success for all SS varying D1 and fixing D2.

As seen in Fig. 3, a great performance is obtained for all SS despite facing critical such critical situations. When all clients' curves input the algorithm (D1 equals to 0%), the success for all SS is almost over 90% (98.25%, 95.74%, 95.54%, 93.77%, 89.90% and 86.84%, for each SS respectively). As D1 starts growing, success tendency becomes more or less linearly decreasing, reaching 70% of success in the worst scenario for SS_6. It must be highlighted that under 100 clients of D1, success remains over 80%.

4. Conclusion and Future Work

This paper has explored the benefits of applying neural networks to the leverage of low voltage distribution grid active and reactive energy measurements to obtain the network connectivity of actual operating networks. For doing so, Ingelectus proposed an innovative algorithm which provides differentiation between connections and non-connections, feeder and phase connectivity, and three-phase consumption characterization, considering the possible lack and uncertainty of topology and measurement data, and fraud connections. The paper has quantified the performance of the proposed strategy a different technology readiness level (TRL): simulations, industrially relevant and operational environments.

The results shows that the proposed technology performs significantly well as a solution to the problem, obtaining a mean of 90% of success (98.25%, 95.74%, 95.54%, 93.77%, 89.90% and 86.84%) in real operational distribution grid, and, showing greater than 80% of success when the non-technical losses and/or missing data are under 40%. This clearly proves the reliability of the technology for solving the connectivity problem. In addition, the technology also provides: information about the three-phase consumption characterization, is capable of detecting and fix changes in topology with the upcoming information, and is totally independent of the source of measurements data, hence any manufacturer or version of the meter is available to the distribution system operator.

This technology may be used as an input to other solutions, such as state estimation to improve the certainty of the connectivity of the network, or energy balances, to detect non-technical losses. The latest provides an improvement in the new upcoming measurements, hence allowing the algorithm to perform better.

Finally, it is important to remark that this algorithm can be extended with multiple inputs like power, current and voltage measurements, the topology of the grid, and the geographical information. In the opinion of the authors, this work may have promising future extensions, whilst being able to be currently applied in an operational environment.

Acknowledgement

The authors of this research work wish to acknowledge the PASTORA consortium and SmartCity Málaga for its contributions in the network measurement datasets. The project PASTORA is financed by CDTI under the FEDER Innterconecta 2018 program.

References

[1] S. Carillo-Aparicio, D. Morales-Wagner, P. L. Cruz-Romero, "*Advanced monitoring and control in MV and LV*," in ICREPQ'17, 2017.

[2] A. Gómez-Exposito, D. Morales-Wagner, S. Carillo-Aparicio, A. De la Villa-Jaen, E. Romero-Ramos, A. Gastalver-Rubio, et. al., "*Herramientas avanzadas para monitorización y operación de redes MT/BT*," in IV Congreso Smart Grids, Madrid, Dic. 2017.

[3] A. Gómez-Expósito, D. Morales-Wagner, A. Gastalver-Rubio et al., "*Herramientas avanzadas para monitorización y operación de redes MT/BT*," in V Congreso Smart Grids, Madrid, Dic. 2018.

[4] A. Gómez-Expósito et al., "*City-Friendly Smart Network Technologies and Infrastructures: The Spanish Experience*," in Proceedings of the IEEE, vol. 106, no. 4, pp. 626-660, Apr. 2018.

[5] M. M. Buzau, J. Tejedor-Aguilera, P. Cruz-Romero and A. Gómez-Expósito, "*Detection of Non-Technical Losses Using Smart Meter Data and Supervised Learning*," in IEEE Transactions on Smart Grid, vol. 10, no. 3, pp. 2661-2670, May. 2019.

[6] A. Gastalver-Rubio, E. Romero-Ramos and J. M. Maza-Ortega, "*Improving the Performance of Low Voltage Networks by an Optimized Unbalance Operation of Three-Phase Distributed Generators*," in IEEE Access, vol. 7, pp. 177504-177516, 2019.

[7] M. A. Yehia, E. E. Matar, N. Y. Hobeila and A. D. Avedikian, "*A heuristic algorithm for electric distribution networks optimal feeder configuration using geographic information system*," in IEEE Transactions on Power Systems, vol. 17, no. 4, pp. 1232-1237, Nov. 2002.

[8] W. Luan, J. Peng, M. Maras, J. Lo and B. Harapnuk, "*Smart Meter Data Analytics for Distribution Network Connectivity Verification*," in IEEE Transactions on Smart Grid, vol. 6, no. 4, pp. 1964-1971, Jul. 2015.

[9] H. Zubia, "*Analítica avanzada para la simulación de flujos de potencia en red de baja tensión (BT)*," in VI Congreso Smart Grids, Madrid, Dic. 2019.

[10] I. Goodfellow, Y. Bengio, and A. Courville, "*Deep Learning*," The MIT Press, 2017.