

Customized Normalization Method to enhance the Clustering process of Consumption Profiles

Catarina Ribeiro, Tiago Pinto, Zita Vale

GECAD – Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development, Institute of Engineering, Polytechnic of Porto (ISEP/IPP), Portugal
{acrib, tmcfp, zav}@isep.ipp.pt

Abstract. The restructuring of electricity markets brought many changes to markets operation. To overcome these new challenges, the study of electricity markets operation has been gaining an increasing importance. With the emergence of microgrids and smart grids, new business models able to cope with new opportunities are being developed. New types of players are also emerging, allowing aggregating a diversity of entities, e.g. generation, storage, electric vehicles, and consumers. The virtual power player (VPP) facilitates their participation in the electricity markets and provides a set of new services promoting generation and consumption efficiency, while improving players' benefits. The contribution of this paper is a customized normalization method that supports a clustering methodology for the remuneration and tariffs definition from VPPs. To implement fair and strategic remuneration and tariff methodologies, this model uses a clustering algorithm, applied on normalized load values, which creates sub-groups of data according to their correlations. The clustering process is evaluated so that the number of data sub-groups that brings the most added value for the decision making process is found, according to players characteristics. The proposed clustering methodology has been tested in a real distribution network with 30 bus, including residential and commercial consumers, photovoltaic generation and storage.

Keywords: Clustering, Consumption profiles, Dynamic Tariffs

1 Introduction

The restructuring of the electricity sector is characterized by an enormous increase in competition and profound changes in the participant entities. The restructuring was performed so that the competitiveness could be increased, but it also had exponential implications in markets complexity and unpredictability [1]. Potential benefits will depend on the efficient operation in the market and, on the other hand, in the remuneration of aggregated players. Important developments concerning electricity market players modelling and simulation including decision-support capabilities can be widely found in the literature [2-3].

Much like electricity markets, subsystems of the main network are rapidly evolving into a reality, coordinating these entities is a huge challenge that requires the implementation of distributed intelligence, potentiating the concept of Smart Grid (SG) [4, 5]. However, the two concepts are not converging towards common goals and technical and economic relationships are addressed in an over simplistic way. Present operation methods and Electricity markets models do not take full advantage

The present work was done and funded in the scope of the following projects: People Programme of the European Union's Seventh Framework Programme FP7/2007-2013/ project ELECON, REA grant agreement No 318912; EUREKA - ITEA2 Project M2MGrids with project number 13011

of installed DG, yielding to inefficient resource management that should be overcome by adequate optimization methods [6]. Player aggregating strategies allows players gaining technical and commercial advantages, individuals can achieve higher profits due to specific advantages of a mix of technologies to overcome disadvantages of some technologies. The aggregation of players gives rise to the concept of Virtual Power Player (VPP) [7]. VPPs aggregate different types of resources. Each aggregated player has its individual goals; VPPs should conciliate all players in a common strategy able to enable each player to pursue its own objectives [8].

This paper proposes a data mining methodology, based on the application of a clustering process, which groups the typical load profile of the consumers of a Smart Grid according to their similarity. The separation of consumers in different groups allows proposing specific consumption tariffs to each group, so that consumers' load profile is taken into account to meet the objectives of the Smart grid aggregator. This methodology is tested using a smart grid that includes several real consumers of different types (residential and commerce). This work thus proposes a customized normalization method to treat data before it is used by the clustering process.

2 Smart Grid and Electricity Markets simulation

Many works have been developed using simulators to model the complex interaction between electricity market players. Successful examples sustain the fact that a multi-agent system (MAS) with adequate simulation abilities, is the best approach for simulating electricity markets [9-11]. The Multi-Agent Simulator for Competitive Markets (MASCEM) is a platform that simulates several electricity market types, while providing decision support to players' actions [9]. This type of simulators are able to represent market mechanisms and players' interactions. However, for them to be valuable decision support tools in foreseeing market behaviour, they need to be used in testing adequate and realistic scenarios. Real data analysis by means of a knowledge discovery process will be a crucial step forward to assure that MASCEM agents exhibit adequate profiles and strategies.

Multi-Agent Smart Grid simulation Platform (MASGriP) [10] simulates, manages and controls the most important players acting in a Smart Grid environment. This system includes simulated players, which interact with agents that control real hardware. The considered players include operators, and energy resources, such as several types of consumers, producers, electric vehicles, among other. Aggregators are also considered, namely: VPPs and Curtailment Service Providers (CSP) [12]. These players introduce a higher level of complexity to the management of the system. Joint simulations of MASCEM and MASGriP enable a simulation environment that includes the participation of Smart Grid players in electricity markets, or even internal Smart Grid markets, using complex markets models provided by MASCEM.

3 Decision support tool for electricity markets remuneration and tariff definition

The Remuneration and Tariff Mechanism (RemT) [13-14] is a decision support mechanism that is being developed to support the VPP actions in the definition of the best tariff and remuneration to apply to each of the aggregated players, regarding the VPP objectives and the individual goal of each aggregated player. VPPs in the scope of MASCEM use RemT to remunerate aggregated players, according to the

results obtained in the electricity market, the penalties for breach of contract, contracts established to guarantee reserve, demand response programs and incomes of aggregated consumers. The definition of remuneration and tariffs is based on the identification of players' types and on the development of contract models for each player type. This considers players with a diversity of resources and requirements, playing several distinct roles (a player can be a consumer, a producer and can be responsible for one or several V2G). The terms for new contracts and the best strategies for each context are determined by means of machine learning methods and data-mining algorithms.

A. Clustering approach

A wide variety of clustering algorithms can be found in the literature and unfortunately, there is no single algorithm that can by itself, discover all sorts of cluster shapes and structure [15]. K-means [16], has been used, as it proves to be a robust model for distinct applications: K-means minimizes the distance from each point to the centre of the respective cluster, as defined in (1).

$$\min \sum_{i=1}^k \sum_{x \in c_i} \|x - \mu_i\|^2 \quad (1)$$

where μ_i is the mean of points in C_i , *i.e.* the cluster *centroid*. To determine the quality of the division of players into different clusters the clusters validity indices MIA and CDI [17] have been used, as formalized in (2) and (3) respectively.

$$MIA = \sqrt{\frac{1}{K} \sum_{k=1}^K d^2(x^{(k)}, \mu^{(k)})} \quad (2)$$

$$CDI = \sqrt{\frac{\frac{1}{K} \sum_{k=1}^K \left[\frac{1}{2n^{(k)}} \sum_{n=1}^{n^{(k)}} d^2(x^{(m)}, \mu^{(k)}) \right]}{\sqrt{\frac{1}{2K} \sum_{k=1}^K d^2(x^{(k)}, R)}}} \quad (3)$$

where d represents the Euclidian distance between two points, and R is the representative load profile of all consumers.

B. Customized Normalization

Analysing the results of previous work [13], is possible to conclude that aggregation strategies have very good results, and are very useful, because they provide a good separation according to what is intended. The non-normalization grouping process has led to a clear separation between different consumers types, as it considers the absolute consumption amounts in the clustering process. The normalized data, used as formalized in (4) and (5), reveals a separation through consumption profiles, although it is not able to consider the differences in consumption quantity.

$$N_{c,h} = \frac{L_{c,h}}{ML_c}, \forall c \in co \quad (4)$$

$$ML_c = \max(L_c), \forall c \in co \quad (5)$$

where N is the common normalized load, for each consumer c , for each hour h , and co is the set of all considered consumers. ML is the largest consumption value, of the consumer c , considering all hours.

To improve the results achieved in the previous works, the customized normalization process is introduced. This method normalizes data using each consumers' load value at each period divided by the largest recorded value of all loads in all periods, it is formalized in (6) and (7).

$$SN_{c,h} = \frac{L_{c,h}}{SML_h}, \forall c \in co \quad (6)$$

$$SML_h = \max(L_{co,h}), \forall c \in co \quad (7)$$

where SN is the load with a different normalization process, for each consumer c , for each hour h . SML is the largest consumption value recorded for all consumers at the time h .

The proposed customized normalization method aims to combine the advantages of both previous approaches (using non-normalized data, and regular normalization), so as to achieve consumer groups that capture both differences in the quantities of consumption, and also the trends of consumer profiles along the hours. The clustering process takes into account the tendency of the consumption values trough the time, regardless of its absolute amount. This separation is very important, according to different consumers' types and profiles, it works as a base for personalized and dynamic consumption tariff definition. Using this approach ensures that the data are also normalized in a range between 0 and 1, but without losing information related to differences between amounts of consumption among consumers. While using the regular normalization, the value 1 is attributed to the greater consumption value of each consumer (thus both consumers with large and small values will always have one value of 1 in a certain hour), using the customized normalization method, only the largest consumer of all, will have a value of 1. The smaller consumers will have normalized values with smaller values, proportional to the difference between the quantities of consumption of that consumer and the largest consumer in each hour. Thus normalization is still made between 0 and 1, but there is visible difference between higher and lower consumption among different consumers, and the evolution of consumption of each consumer profile is also captured.

4 Case study

This case study intends to show the adequacy of the proposed customized normalization clustering methodology to solve the problem of remuneration of players with heterogeneous characteristics and behaviors. In order to test the adequacy of the method, a clustering algorithm has been applied, concerning the consumption data of a total of 82 consumers (8 residential houses, 8 residential buildings with 72 loads, and 2 commercial buildings). Data has been collected from a real distribution network throughout one year. The Smart grid accommodates distributed generation (photovoltaic and wind based generation) and storage units, which are integrated in

the consumption buildings. The accommodated photovoltaic generation, wind based generation and storage units are related to the building installed consumption power, according to the current legislation in Portugal. Further details on the considered distributed network can be seen in [18].

The K-means algorithm has been used to perform the clustering process using non-normalized values of load (section A), and also normalized values, using both the regular normalization method (section B) and the proposed customized normalization method (section C).

A. Non-normalized data

The clustering process is performed for different numbers of clusters, in order to enable grouping consumers according to the similarity of their consumption profiles, in order to support the definition of specific tariffs that are suited for each of the consumer groups. From [13] it has been concluded that, by analyzing MIA and CDI results from the clustering of non-normalized data, the best clustering results are achieved with the use of 3 clusters, as the clustering error is minimal. When using 2 clusters, a clear separation of residential houses and buildings from commercial buildings is visible. It is also visible that the two commercial buildings (corresponding to loads 1 and 2) have been allocated to cluster 1, and the rest of the loads, corresponding to residential consumers, have been aggregated in cluster 2. This can be observed in Figure 1 which presents the load profiles of consumers that have been grouped in cluster 1 and in cluster 2 using the non-normalized data.

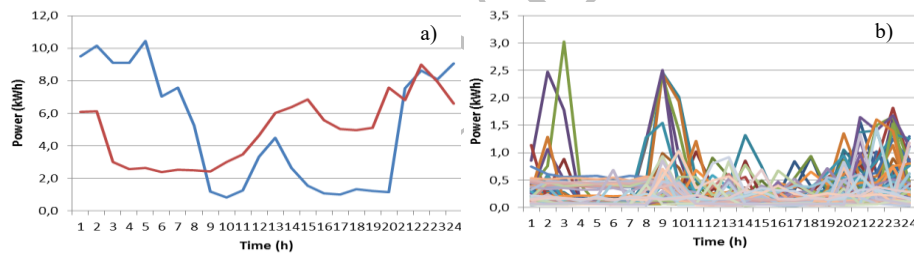


Figure 1. Consumption profile of loads allocated to: a) cluster 1; b) cluster 2

From Figure 1 it is visible that cluster 1 includes the two commercial buildings, with very distinct load profiles, and cluster 2 includes all the residential buildings and houses. When considering the grouping process with 3 clusters, the difference is that there is still a separation from residential houses and buildings to the commerce. However, in this case the two types of commercial buildings are also separated, as they present very different load profiles.

B. Normalized data

In the second clustering process, regular normalized data were used. The normalization was made considering each type of consumer. The value of load corresponding to each period was divided for the maximum value register in that specific load in the 24 periods. When using normalized values, a more accentuated descent of the clustering error values is visible. The descent in the error value is however, stable from the start, which hardens the identification of the optimal number of clusters that should be used. For this reason it is not advantageous to use more than 2 or 3 clusters, since the use a larger number is not reflected by a significant gain in clustering error.

By analysing the results of the clustering process with 2 clusters, it can be seen that the separation is not as clear as it was with non-normalized values. The two commercial buildings corresponding to load 1 and 2, were aggregated in different clusters, together with several residential consumers. However, the clustering process with normalized values has better results from the load profile separation stand point, as can be seen from Figure 2, which presents the allocation of the consumers to the different clusters, when considering normalized data and 2 clusters.

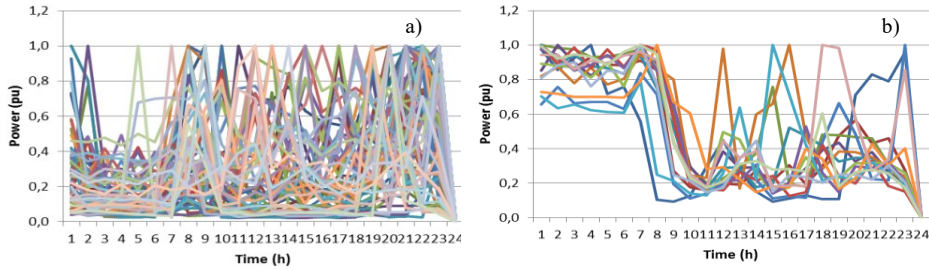


Figure 2. Consumption profile of loads represented in: a) cluster 1, b) cluster 2

From Figure 2 it is visible that although the consumer types cannot be separated correctly with this approach as occurs when using non-normalized data (Figure 1), the separation of the load profiles is more evident in this case, since profiles are grouped independently from the gross amount of consumption itself.

C. Customized Normalization

The clustering process is performed for different numbers of clusters (from 2 to 6). MIA and CDI are used to analyse the clustering error. Figure 3 presents the comparison of the MIA and CDI values that are achieved when using from 2 to 6 clusters, with each of the three considered methods: non-normalized data, normalized data using the regular method, and customized normalization method.

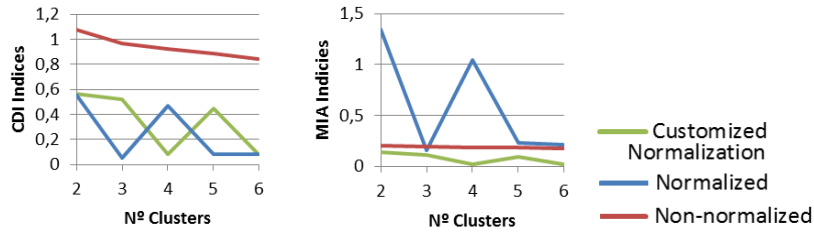


Figure 3. CDI and MIA results for different numbers of clusters, using non-normalized, normalized and customized normalization

In Figure 3 it is possible to see that the proposed method has a lower error when compared with the previous methods. The best clustering results are achieved with the use of 4 clusters, as the clustering error is minimal. With the use of 4 clusters, the two commercial buildings are separated into a different cluster each, the third cluster allocates some of the residential buildings that have similar load profiles, and the rest of the loads, corresponding to residential houses and some residential buildings, have been aggregated in the final cluster. Figure 4 represents the load profile of the

different consumer types considering 3 clusters. Figure 4 a) represents the consumption profiles of the two commercial buildings, which, as it is possible to see, have very different consumption profiles, especially during the night. This is why they were allocated into different clusters when 4 clusters are considered. Figure 4 b) and c) represent the consumption profiles of the loads allocated to the other two clusters.

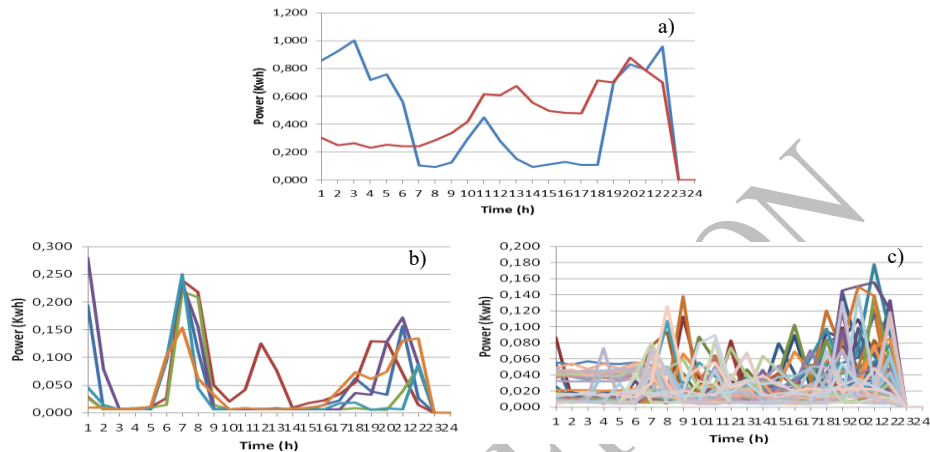


Figure 4. Consumption profile of: a) commercial buildings, b) cluster 2; c) cluster 3

From Figure 4 the separation of the load profiles is evident. It is also visible that a separation taking into account the gross amount of consumption of each consumer has been accomplished, as commercial consumers, which present much higher consumption values, have been separated from the residential consumers.

The proposed customized normalization brings, therefore, clear advantages to the RemT tariff definition process. It enables to clearly identify different consumers, taking into account their consumption tendency and amount, therefore breaking the way for an objective and fare definition of dynamic electricity tariffs, which can suitably fit each of the identified groups, i.e. consumers with similar consumption tendencies, taking into account their dimension. The new type of normalization, when compared with the previous normalization types, allows an even more clear separation of consumer types, which is evident from the load profile graphs that show the separation into different clusters, and also by the MIA and CDI values, which show that the proposed method achieves smaller clustering error values than the other methods.

6 Conclusions

Electricity markets are experiencing profound transformations. Currently there is a gap in what concerns VPP aggregated players' tariffs and remuneration. In order to overcome this problem, developing appropriate methods is essential. This paper presents the development of a tool that provides a decision support for VPP definition of best tariff and remuneration to apply to each aggregated player, RemT. To develop RemT a clustering methodology that uses different data normalization methods was presented, and a new customized normalization method has been introduced.

The results of the presented case study, based on real consumption data, show that the customized normalization method combines the advantages of both previous

approaches, so as to achieve more consumer groups that capture both differences in the quantities of consumption, as well as the trends of consumer profiles along hours. This is crucial, according to different consumers' types and profiles, as it works as a basis for personalized and dynamic consumption tariff definition. Thus normalization is the same made between 0 and 1, but there is visible difference between higher and lower consumption among different consumers, and the evolution of consumption of each consumer profile is also captured. RemT mechanism is evolving to become a crucial tool to go a step forward in electricity markets simulation, by enabling a fair and dynamic means to define electricity tariffs for different types of consumers.

References

1. L. Meeus, et al., "Development of the Internal Electricity Market in Europe", *The Electricity Journal*, vol. 18, no. 6, pp. 25-35, 2005
2. I. Praça, C. Ramos, Z. Vale, M. Cordeiro, "MASCEM: A Multi-Agent System that Simulates Competitive Electricity Markets", *IEEE Intelligent Systems*, 18, 6, pp. 54-60, 2003
3. V. Koritarov, "Real-World Market Representation with Agents: Modeling the Electricity Market as a Complex Adaptive System with an Agent-Based Approach", *IEEE Power & Energy magazine*, pp. 39-46, 2004
4. M. Shahidehpour, et al., "Market Operations in Electric Power Systems: Forecasting, Scheduling, and Risk Management", Wiley-IEEE Press, pp. 233-274, 2002
5. Blumsack S and Fernandez A. "Ready or not, here comes the smart grid!" *Energy*. 2012; 37(1):61-8
6. Sousa, T., et al., "Intelligent Energy Resource Management Considering Vehicle-to-Grid: A Simulated Annealing Approach," *Smart Grid*, *IEEE Trans. on*, 3, 535-542, 2012
7. Z. Vale, T. Pinto, I. Praça, H. Morais, "MASCEM - Electricity markets simulation with strategically acting players", *IEEE Intelligent Systems*, vol. 26, n. 2, Special Issue on AI in Power Systems and Energy Markets, 2011
8. T. Pinto, et al., "Multi-Agent Based Electricity Market Simulator With VPP: Conceptual and Implementation Issues", 2009 IEEE PES General Meeting, 2009
9. T. Pinto, et al., "Strategic Bidding Methodology for Electricity Markets using Adaptive Learning", *Modern Approaches in Applied Intelligence*, vol. 6704, pp.490-500, 2011
10. Oliveira, P. et al., "MASGrid - A Multi-Agent Smart Grid Simulation Platform," *IEEE Power and Energy Society General Meeting*, San Diego, California USA, 2012, pp. 1-10
11. Z. Vale, et al., "Mascem - electricity markets simulation with strategic agents," *IEEE Intelligent Systems*, Special Issue on Agents and Markets, vol. 26, no. 2, pp. 9-17, 2011
12. C. Kieny et al., "On the concept and the interest of Virtual Power plant: some results from the European project FENIX", 2009 IEEE Power and Energy Society General Meeting, Calgary, Canada, 26-30, July 2009
13. C. Ribeiro., et al., "Data Mining approach for Decision Support in real data based Smart Grid scenario" *IATEM*, 2015
14. Ribeiro C., et al., " Intelligent Remuneration and Tariffs in for Virtual Power Players", *IEEE PowerTech (POWERTECH) Grenoble*, France, 16-20 June, 2013
15. Anil K. Jain et. al., (1999) "Data Clustering: A Review." *ACM Computing Surveys*, 31 (3), pp. 264-323.
16. Anil K. Jain, "Data Clustering: 50 years beyond K-Means". *Pattern Recognition Letters*, Elsevier, Vol. 31, Issue 8, pp.651-666, June 2010.
17. Chicco et al., "Support Vector Clustering of Electrical Load Pattern Data". *IEEE Transactions on Power Systems*, vol.24, no.3, pp.1619-1628, August 2009.
18. Canizes B. et. al., "Resource Scheduling in Residential Microgrids Considering Energy Selling to External Players", *Power Systems Conference (PSC 2015)*, South Carolina, USA, 10-13 March, 2015