Data Mining for Prosumers Aggregation considering the Self-Generation

Catarina Ribeiro[1,2], Tiago Pinto[1], Zita Vale[1], José Baptista[2]

[1]GECAD – Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development, Institute of Engineering, Polytechnic of Porto (ISEP/IPP), Portugal
[2]UTAD – Universidade de Trás-os-Montes e Alto-Douro
{acrib, tmcfp, zav}@isep.ipp.pt, baptista@utad.pt

Abstract. Several challenges arrive with electrical power restructuring, liberalized electricity markets emerge, aiming to improve the system's efficiency while offering new economic solutions. Privatization and liberalization of previously nationally owned systems are examples of the transformations that have been applied. Microgrids and smart grids emerge and new business models able to cope with new opportunities start being developed. New types of players appear, allowing aggregating a diversity of entities, e.g. generation, storage, electric vehicles, and consumers, Virtual Power Players (VPPs) are a new type of player that allows aggregating a diversity of players to facilitate their participation in the electricity markets. A major task of VPPs is the remuneration of generation and services (maintenance, market operation costs and energy reserves), as well as charging energy consumption. The paper proposes a normalization method that supports a clustering methodology for the remuneration and tariffs definition. This model uses a clustering algorithm, applied on normalized load values, the value of the micro production, generated in the bus associated to the same load, was subtracted from the value of the consumption of that load. This calculation is performed in a real smart grid on buses with associated micro production. This allows the creation of 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.

1 Introduction

Power sector has been completely revolutionized by the emergence of liberalized electricity markets (EM) aiming to improve the system's efficiency while offering economic solutions. Characterized by an increase in competition and changes in participant entities, potential benefits will depend on the efficient operation in the market and in bilateral contract negotiation and remuneration of aggregated players. In consequence of these structural changes, has been a gradual decentralization of decision making and a redistribution of responsibilities for different players [1]. Important developments concerning electricity market players modelling and simulation including decision-support capabilities can be widely found in the literature [2-3]. Subsystems of the main network are evolving into a reality. The coordination of

The present work has been developed under the EUREKA - ITEA2 Project FUSE-IT (ITEA-13023), Project GREEDI (ANI|P2020 17822), and has received funding from FEDER Funds through COMPETE program and from National Funds through FCT under the project UID/EEA/00760/2013.

these entities is a challenge that requires the implementation of distributed intelligence, potentiating the concept of Smart Grid (SG) [4, 5]. However, EM and SG are not converging towards common goals and technical and economic relationships are addressed in an over simplistic way. Operation methods and EM models do not take all the advantage of installed DG, yielding to inefficient resource management that should be overcome by adequate optimization methods [6]. The aggregating strategies of players, allow them to gain 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], heterogeneous entities that aggregate different types of resources where each aggregated player has its individual goals. The VPP should conciliate all players in a common strategy, while enabling each player to pursue its own objectives [8]. There are some simulators in literature that enable modeling VPPs aggregation and resource management process. One relevant system in this domain is MASGriP (Multi-Agent Smart Grid Simulation Platform) [9], manages and controls the most relevant players acting in a SG environment. MASGriP is connected to MASCEM (Multi-Agent Simulator for Competitive Electricity Markets) [2,7], thus providing the means to perform joint simulations. A decision support system has been integrated in MASCEM, in order to allow players to automatically adapt their strategic behavior according to the operation context and with their own goals. This platform is Adaptive Learning Strategic Bidding System (ABidS) [10], and it provides agents with the ability of analyzing contexts of negotiation. This paper proposes a data mining methodology, based on the application of a clustering algorithm, applied on normalized load values, which groups the typical load profile of the consumers of a SG according to their similarity for the remuneration and tariffs definition from VPPs. The value of the micro production, generated in the bus associated to the same load, was subtracted from the value of the consumption of that load. This allows the creation of sub-groups of data according to their correlations. The clustering process is evaluated so that the number of data subgroups that brings the most added value for the decision making process is found,

2 RemT – Remuneration and tariff decision support tool for electricity markets

The Remuneration and Tariff Mechanism (RemT) [11, 12] 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. In the scope of MASCEM, VPPs 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 establishment of remuneration and tariffs is based on the identification of players' types and development of contract models for each player type. This considers players with a diversity of resources and requirements, playing several distinct roles (consumers, producers and can be responsible for one or several electrical vehicles). The players modelling takes into account the operation and market context. The terms for new contracts and best strategies for each context are determined by means of machine learning methods and data-mining algorithms. The definition process is presented in Figure 1.

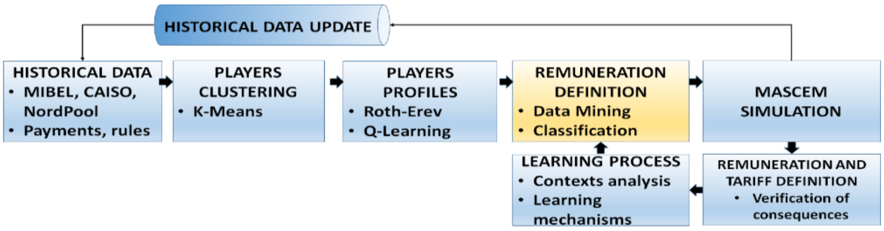

Figure 1. RemT definition process.

A. Clustering approach

There are many clustering algorithms in literature and unfortunately, no single algorithm that can by itself, discover all sorts of cluster shapes and structure[13]. K-means[14], has been used, as it proves to be a robust model for distinct applications: it minimizes distance from each point to the centre of the respective cluster, where μ_i is the mean of points in C_i, *i.e.* the cluster *centroid* as defined in (1).

$$\min \sum_{i=1}^{k} \sum_{x \in c_i} \left\| x - \mu_i \right\|^2 \tag{1}$$

To determine the quality of the division of players into different clusters the clusters validity indices MIA and CDI[15] were used, as formalized in (2) and (3).

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

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

Where *d* represents the Euclidian distance between two points, and R is the representative load profile of all consumers. This indices represent distances, the smaller (or greater) is the MIA and CDI value, it indicates more (or less) compact clusters. To facilitate the analysis of results, we will consider that the higher the value of the MIA and CDI, the larger the error associated with this cluster.

B. Normalization method considering prosumers self-generation

Analysing the results of previous work [11], 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. *L* is the value of load.

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

$$ML_c = \max(L_c), \forall c \in co \tag{5}$$

Where *N* is the common normalized load, for each consumer *c*, for each hour *h, a*nd *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 difference normalization process is introduced. In different process the value of the micro production *P*, generated in the bus associated to a load, was subtracted from the value of the consumption of that load. This calculation is performed for loads from 1 to 17, loads which are on buses with associated micro production, for each specific load in the 24 periods. It is formalized in (6) and (7).

$$\mathrm{T_{c,h}} = \left|\mathrm{L_{c,h}} - \mathrm{P_{c,h}}\right|, \forall \mathrm{c} \in \mathrm{co} \tag{6}$$

Where T corresponds to the load where consumption subtracted by micro production, for each consumer c, for each hour h.

$$SN_{c,h} = \frac{T_{c,h}}{SML_h}, \forall c \in co \tag{7}$$

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 values of production in each bus were taken on site, in the real distribution network [16]. This method aims to combine the advantages of both previous approaches (using non-normalized data, and regular normalization subtracting the value of production of the different loads), 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. 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. This approach ensures that data are also normalized in a range between 0 and 1, but without losing information related to differences between amounts of consumption among consumers. Two independent variables are subtracted to accentuate the difference between classes of loads. This is because, apparently, local production depends on the class of loads, see [16], residential houses produce more throughout the year than they consume whereas it is the opposite for commercial building, and residential building produce as much as they consume. This supports the use of the proposed method. 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 have normalized values with smaller values, proportional to the difference between the consumption quantities 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.

4 Case study

This case study intends to show the adequacy of the proposed 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 [16]. 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 commercial1) 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 2 which presents the load profiles of consumers that have been grouped in cluster 1 and in cluster 2 using the non-normalized data.

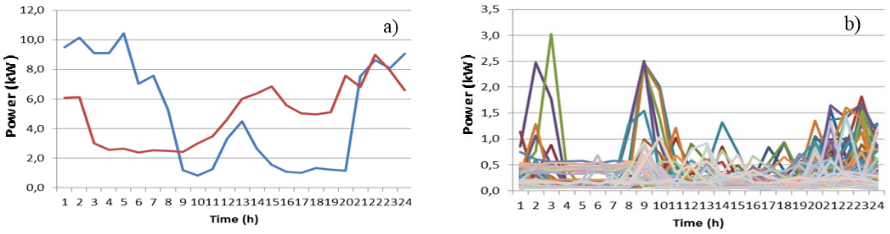

Figure 2. Consumption profile of loads allocated to: a) cluster 1; b) cluster 2

From Figure 2 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. 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 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 of a larger number is not reflected by a significant gain in clustering error. Analyzing the results with 2 clusters, the separation is not as clear as it was with non-normalized values. The two commercial buildings(load 1 and 2), were aggregated in different clusters, with residential consumers. However, the clustering process with normalized values has better results from the load profile separation stand point. Figure 3 presents the allocation of the consumers considering normalized and 2 clusters.

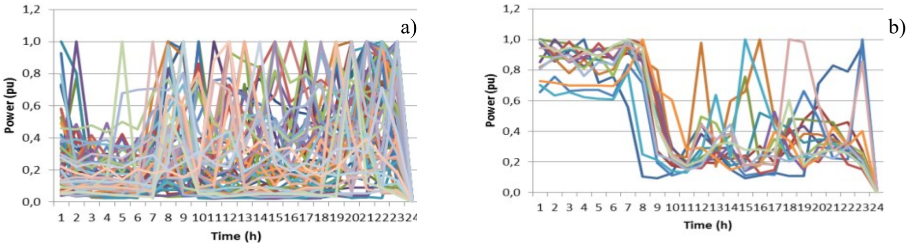

Figure 3. Consumption profile of loads represented in: a) cluster 1, b) cluster 2

From Figure 3 is visible that although the consumer types cannot be separated correctly with this approach as occurs when using non-normalized data (Figure 2), 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. *Proposed normalization process*

In this process, the value of the micro production, generated in the bus associated to the same load, was subtracted from the value of the consumption of that load. This calculation is performed for loads from 1 to 17, these are the loads, which are on buses with associated micro production. MIA and CDI are used to analyze the clustering error, they are presented in Figure 4.

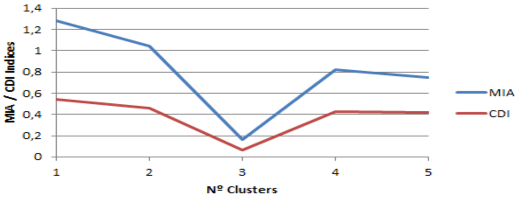

Figure 4. MIA and CDI results for difference normalization values for 24 hour period.

In difference normalization process, the best clustering results are achieved with the use of 3 clusters, as the clustering error is minimal, similar to what happened in previous cases. Using 2 clusters, a clear separation of residential houses and buildings from commercial buildings is visible. Also, 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. When considering the grouping process with 3 clusters, the difference is that there is an even better separation of consumers types, commercial buildings were allocated to cluster 1, residential houses to cluster 2 and residential building to cluster 3. In this case, the two types of commercial buildings stayed in the same cluster, although they present very different load profiles.

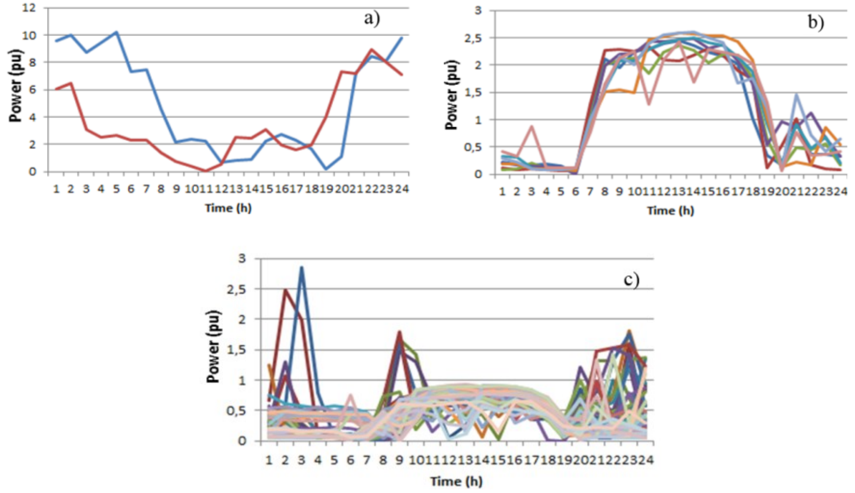

Figure 5. Consumption profile of load allocated to: a) cluster 1; b) cluster 2; c) cluster 3

The proposed difference 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 clearer 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

EMs are experiencing profound transformations. This case study demonstrated the usefulness and advantage of data mining methodologies, based on the application of clustering process to group typical load profiles of consumers according to their similarity to allow proposing specific consumption tariffs to each group, so that consumers load profile is taken into account to meet the objectives of the SG aggregator. This work allows 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 difference normalization method has been introduced. The results of the presented case study, based on real consumption data, show that the difference 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 EM 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 Int. Systems, 18,6,54-60, 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," IEEE Trans. on Smart Grid, 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. Oliveira, P. et.al., "MASGriP - A Multi-Agent Smart Grid Simulation Plataform," IEEE 2012 - Power and Energy Society General Meeting, San Diego, USA, 2012, pp. 1-10
10. Pinto, T., et.al., “Adaptive Learning in Agents Behaviour : a Framework for Electricity Markets Simulation,” Integr. Comput. Aided. Eng., vol. 21, no. 4, pp. 399–415, 2014
11. C. Ribeiro., et al., “Data Mining approach for Decision Support in real data based Smart Grid scenario” IATEM, 2015
12. Ribeiro C., et al., “ Intelligent Remuneration and Tariffs in for Virtual Power Players”, IEEE PowerTech (POWERTECH) Grenoble, France, 16-20 June, 2013
13. Anil K. Jain et. al., (1999) “Data Clustering: A Review.” ACM Computing Surveys, 31 (3). pp. 264-323.
14. Anil K. Jain, “Data Clustering: 50 years beyond K-Means”. Pattern Recognition Letters, Elsevier, Vol. 31, Issue 8, pp.651-666, June 2010.
15. 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.
16. 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