

# Data Mining Techniques Contributions to Support Electrical Vehicle Demand Response

João Soares, Sérgio Ramos, *Member, IEEE*, Zita Vale, *Senior Member, IEEE*,  
Hugo Morais, *Member, IEEE*, Pedro Faria, *Student Member, IEEE*

**Abstract** — The introduction of Electric Vehicles (EVs) together with the implementation of smart grids will raise new challenges to power system operators. This paper proposes a demand response program for electric vehicle users which provides the network operator with another useful resource that consists in reducing vehicles charging necessities. This demand response program enables vehicle users to get some profit by agreeing to reduce their travel necessities and minimum battery level requirements on a given period.

To support network operator actions, the amount of demand response usage can be estimated using data mining techniques applied to a database containing a large set of operation scenarios.

The paper includes a case study based on simulated operation scenarios that consider different operation conditions, e.g. available renewable generation, and considering a diversity of distributed resources and electric vehicles with vehicle-to-grid capacity and demand response capacity in a 33 bus distribution network.

**Index Terms** — Classification, Clustering, Data Mining, Demand Response, Electric Vehicle, Mixed Integer Non-Linear Programming (MINLP)

## I. NOMENCLATURE

$\Delta t$	Period $t$ duration (e.g. 15 min., 30 min., 1 hour...)
$\theta_b$	Voltage angle at bus $b$ (rad)
$\theta_k$	Voltage angle at bus $k$ (rad)
$\theta_b^{max}$	Maximum voltage angle at bus $b$ (rad)
$\theta_b^{min}$	Minimum voltage angle at bus $b$ (rad)
$\eta_{c(V)}$	Grid-to-Vehicle Efficiency when the vehicle $V$ is in charge mode
$\eta_{d(V)}$	Vehicle-to-Grid Efficiency when the vehicle $V$ is in discharge mode
$B_{bk}$	Imaginary part of the element in $y_{bk}$ corresponding to the $b$ row and $k$ column
$C_{Charge(V,t)}$	Charge price of vehicle $V$ in period $t$
$C_{DG(DG,t)}$	Generation price of $DG$ unit in period $t$

$C_{Discharge(V,t)}$	Discharge price of vehicle $V$ in period $t$
$C_{EAP(DG,t)}$	Excess available power price of $DG$ unit in period $t$
$C_{NSD(L,t)}$	Non-supplied demand cost of load $L$ in period $t$
$C_{Supplier(S,t)}$	Market energy price of upstream supplier $S$ in period $t$
$E_{BatteryCapacity(V)}$	Battery energy capacity of vehicle $V$
$E_{MinCharge(V,t)}$	Minimum stored energy to be guaranteed at the end of period $t$ , for vehicle $V$
$E_{Stored(V,t)}$	Active energy stored in vehicle $V$ at the end of period $t$
$E_{Trip(V,t)}$	Vehicle $V$ energy consumption in period $t$
$G_{bk}$	Real part of the element in $y_{bk}$ corresponding to the $b$ row and $k$ column
$N_b$	Total number of buses $b$
$N_{DG}^b$	Total number of distributed generators at bus $b$
$N_{DG}$	Total number of distributed generators
$N_L^b$	Total number of loads at bus $b$
$N_L$	Total number of loads
$N_S^b$	Total number of external suppliers at bus $b$
$N_S$	Total number of external suppliers
$N_V^b$	Total number of vehicles at bus $b$
$N_V$	Total number of vehicles
$P_{Charge(V,t)}^b$	Power charge of vehicle $V$ at bus $b$ in period $t$
$P_{Charge(V,t)}$	Power charge of vehicle $V$ in period $t$
$P_{ChargeLimit(V,t)}$	Maximum power charge of vehicle $V$ in period $t$
$P_{DG}^b(DG,t)$	Active power generation of distributed generation unit $DG$ at bus $b$ in period $t$
$P_{DG}(DG,t)$	Active power generation of distributed generation unit $DG$ in period $t$
$P_{DGMaxLimit(DG,t)}$	Maximum active power generation of distributed generator unit $DG$ in period $t$
$P_{DGMinLimit(DG,t)}$	Minimum active power generation of distributed generator unit $DG$ in period $t$
$P_{Discharge(V,t)}^b$	Power discharge of vehicle $V$ at bus $b$ in period $t$

João Soares, Sérgio Ramos, Zita Vale, Hugo Morais, and Pedro Faria are with GECAD – Knowledge Engineering and Decision-Support Research Group of the Electrical Engineering Institute of Porto – Polytechnic Institute of Porto (ISEP/IPP), Rua Dr. António Bernardino de Almeida, 431, 4200-072 Porto, Portugal (e-mail: {japs, scr, zav, hgvm, pnf}@isep.ipp.pt).

This work is supported by FEDER Funds through COMPETE program and by National Funds through FCT under the projects FCOMP-01-0124-FEDER: PESt-OE/EEL/UI0760/2011, PTDC/EEA-EEL/099832/2008, and PTDC/SEN-ENR/099844/2008.

$P_{Discharge(V,t)}$	Power discharge of vehicle $V$ in period $t$
$P_{DischargeLimit(V,t)}$	Maximum power discharge of vehicle $V$ in period $t$
$P_{EAP(DG,t)}^b$	Excess available power by $DG$ unit at bus $b$ in period $t$
$P_{EAP(DG,t)}$	Excess available power by $DG$ unit in period $t$
$P_{Load(L,t)}^b$	Active power demand of load $L$ at bus $b$ in period $t$
$P_{NSD(L,t)}^b$	Non-supplied demand for load $L$ at bus $b$ in period $t$
$P_{NSD(L,t)}$	Non-supplied demand for load $L$ in period $t$
$P_{Supplier(S,t)}^b$	Active power flow in the branch connecting to upstream supplier $S$ at bus $b$ in period $t$
$P_{Supplier(S,t)}$	Active power flow in the branch connecting to upstream supplier $S$ in period $t$
$P_{SupplierLimit(S,t)}$	Maximum active power of upstream supplier $S$ in period $t$
$P_{TripRed(V,t)}^b$	Demand response power reduce of vehicle trip $V$ at bus in period $t$
$P_{TripRed(V,t)}$	Demand response power reduce of vehicle trip $V$ in period $t$
$P_{TripRedMax(V,t)}$	Maximum power to be reduced in trip period $t$ for vehicle $V$
$Q_{DG(DG,t)}^b$	Reactive power generation of distributed generation unit $DG$ at bus $b$ in period $t$
$Q_{DGMaxLimit(DG,t)}$	Maximum reactive power generation of distributed generator unit $DG$ in period $t$
$Q_{DGMinLimit(DG,t)}$	Minimum reactive power generation of distributed generator unit $DG$ in period $t$
$Q_{Load(L,t)}^b$	Reactive power demand of load $L$ at bus $b$ in period $t$
$Q_{Supplier(S,t)}^b$	Reactive power flow in the branch connecting to upstream supplier $S$ at bus $b$ in period $t$
$Q_{SupplierLimit(S,t)}$	Maximum reactive power of upstream supplier $S$ in period $t$
$S_{bk}^{max}$	Maximum apparent power flow established in line that connected bus $b$ and $k$
$T$	Total number of periods
$t_{Last}$	Last connected period of vehicle $V$ before current trip
$V_b$	Voltage magnitude at bus $b$ (rad)
$V_k$	Voltage magnitude at bus $k$ (rad)
$V_b^{max}$	Maximum voltage magnitude at bus $b$
$V_b^{min}$	Minimum voltage magnitude at bus $b$
$X_{(DG,t)}$	Binary variable of unit $DG$ in period $t$
$X_{(V,t)}$	Binary variable of vehicle $V$ related to power discharge in period $t$
$Y_{(V,t)}$	Binary variable of vehicle $V$ related to power charge in period $t$

## II. INTRODUCTION

Electric Vehicles (EVs) are emerging as a reliable alternative solution to the typical internal combustion vehicles, with the advantage of being a good way to reduce CO<sub>2</sub> emissions, as well as to decrease dependence from fossil energy sources [1].

One of the main future technologies to reduce greenhouse gas emissions is the battery powered Electric Vehicle (EV) as well as the Plug-in Hybrid Electric Vehicle (PHEV). [2]

Electric vehicles may be considered as active loads, increasing the demand on the network during charging, and as generators when operating in discharging mode. Therefore, the impact of EVs on the grid, when operating in both modes, needs to be analysed. It is expected that the impact of this new scenario will be significant due to the high energy capacity and mass deployment of EVs in the future [3]. The introduction of renewable resources and the implementation of the concept of smart grids on the distributed power network also adds a degree of complexity in network operation. New approaches to deal with and take full advantages of the new concepts are therefore required [4].

This paper proposes a trip reduce demand response program for EVs. The gist is to provide network operator with another useful resource that consists in reducing vehicles charging necessities for their trips. This demand response program enables vehicle users to get some profit by agreeing to reduce their travel necessities and the minimum battery level requirements.

Fig. 1 presents the proposed framework for the trip reduce demand response program.

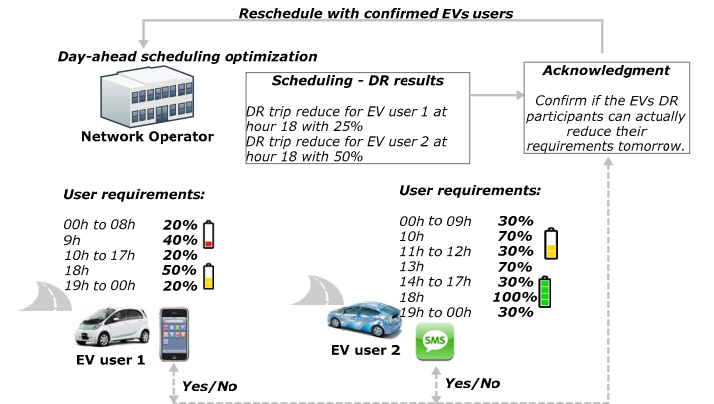


Fig. 1 – Framework of trip reduce DR program

An initial optimization is made assuming that EVs which contracted DR option will participate. With the optimization results it is possible to identify which EVs users are scheduled to participate in the DR event. After that, these EV users can be invited to participate, e.g. through internet application, SMS message. The network operator should wait for a response within a time limit. With the responses of EVs, users the optimization program reschedules the day-ahead problem with the updated information. If additionally EV users are scheduled to participate in the DR program, according to the new optimization results, the operator should follow the same

procedure. The users that do not respond within the time limit are excluded from the DR event. In this work Particle Swarm Optimization (PSO) was used to solve the optimization problem.

In a real environment, a history list of the users that already participated should be maintained and included in the optimization model to enable an attractive commercial model of this DR program and a fair choice of EV users participating in this type of DR programs. It can happen that two users have the same contracted trip reduce DR price; however, user number one might be chosen often due while the other user is never or rarely chosen. If a history list is maintained it is possible to consider this information in order to make the choice of EV users to participate in DR events fairer.

Data Mining (DM) techniques have been used in this work in order to identify trip reduce demand response program profiles for EVs in a given context. Starting from the application of clustering algorithms it was identified several clusters according to the similarity of trip reduce demand response used. Thereafter, a classification model was applied allowing the classification of new trip reduce demand response according to the previous obtained clusters.

After this introduction section, section III briefly reports data mining techniques to support EVs trip reduces demand response programs definition. Section IV explains the mathematical formulation that it was used in the proposed model. In section V a case-study is presented and finally the main conclusions of the paper are presented in Section VI.

### III. DATA MINING TECHNIQUES

Data mining (DM) techniques can be used to extract knowledge from a given database. A clustering process is used to discover groups and identifying interesting distributions in the underlying data [5]. This methodology is based on previous work [6].

#### A. Clustering Process

The clustering process is an unsupervised classification of patterns from observations or data sets, into groups, the so called clusters. Indeed, the clustering analysis is the organization of patterns collection, usually represented as a vector of measurements, or a point in a multidimensional space, into clusters based on similarity [7].

Clustering is used for pattern-analyses such as grouping, decision-making and pattern classification. However, in many such problems, there is little prior information available about the data, and the decision-maker must make as few assumptions about the data as possible. It is under these restrictions that clustering methodology is appropriate for the study of relationships and associations among the data points to make an assessment of their organization [7].

As described in [8], typical pattern clustering activity involves mainly the following steps: pattern representation (optionally including feature extraction and/or selection); definition of a pattern proximity measure appropriate to the data domain; clustering or grouping; data abstraction; and assessment of output.

For cluster analysis there are several clustering methods that can be used. Thus, an important question arises: a decision must be taken in order to choose the clustering method that produces the best data partition. In order to support such decision, indices for measuring the quality of the data partition have been used. To choose the optimal clustering algorithm, mainly two proposed criteria have been followed: compactness, the members of each cluster should be as close to each other as possible and separation, the clusters should be widely spaced from each other.

From the variety of distance measures that have been used the Euclidean distance is emphasize. This distance can often be used to reflect dissimilarity between two patterns, whereas other similarity measures can be used to characterize the conceptual similarity between patterns. Hierarchical clustering algorithms produce a nested series of partitions based on a criterion for merging or splitting clusters based on similarity. Partitional clustering algorithms identify the partition that optimizes a clustering criterion. It is most common to calculate the dissimilarity between two patterns using a distance measure defined on the feature space [7].

In the proposed methodology, different clustering algorithm has been used, the k-means algorithm and the two-step cluster algorithm. In similar works already presented in this context [5], [9], for instance, the algorithm that had the best performance was the k-means and, therefore, was selected for the present paper. The k-means algorithm [10] is the most widely used clustering algorithm. The name represents each of the “k” clusters by the mean or weighted average of its points that is known by the cluster center.

#### B. Classification algorithm

Classification problem is a data supervised classification that the pre-classified data set points are known and the classification algorithm tries to discover a certain rule, which allows mimicking as strictly as possible the observed classification. Typically, the classification task consists in building a classification model that can be applied to unclassified records, seeking to classify them in classes.

The C5.0 algorithm [11] splits the sample based on a criterion that provides the maximum information gain. Each subsample obtained from the previous division is then divided again, based on a different field, and the process is repeated until the subsamples can no longer be divided. Finally, the lowest level splits are re-examined, and those that do not contribute significantly to the value of the model are removed or pruned.

C5.0 produces a decision tree and a rule set. The decision tree is a straightforward description of the splits found by the algorithm. Each node describes a particular subset of the training data. In the present work, the C5.0 algorithm was chosen due to the fact that the results are easy to understand since the rules derived from the model have concrete meaning.

### IV. ENERGY RESOURCE MANAGEMENT PROBLEM

Energy Resource management is an important task that will become crucial in future power systems. The intensification of

the use distributed energy resources contributes to higher problem complexity. The inclusion of electric vehicles with gridable capability introduces new problem variables and constraints. In the formulation proposed in this paper, the main goal is to minimize the operation costs from an aggregator point of view. In the considered context, EVs pay to the network operator to charge their batteries and the operator pays to EV users whenever EV battery energy is used to supply power to the grid. The aggregator also pays to EVs whenever the demand response program proposed in this paper is used.

The mathematical formulation of the optimization problem, including the objective function, is defined as follows:

$$\min f = \sum_{t=1}^T \left[ \begin{aligned} & \sum_{DG=1}^{N_{DG}} P_{DG(DG,t)} \times c_{DG(DG,t)} + \\ & \sum_{S=1}^{N_S} P_{Supplier(S,t)} \times c_{Supplier(S,t)} + \\ & \sum_{V=1}^{N_V} P_{Discharge(V,t)} \times c_{Discharge(V,t)} - \\ & \sum_{V=1}^{N_V} P_{Charge(V,t)} \times c_{Charge(V,t)} + \\ & \sum_{L=1}^{N_L} P_{NSD(L,t)} \times c_{NSD(L,t)} + \\ & \sum_{DG=1}^{N_{DG}} P_{EAP(DG,t)} \times c_{EAP(DG,t)} \\ & + \sum_{V=1}^{N_V} E_{TripRed(V,t)} \times c_{TripRed(V,t)} \end{aligned} \right] \times \Delta t + \quad (1)$$

To improve the solution feasibility the mathematical model includes variables concerning the excess available power ( $P_{EAP(DG,t)}$ ) and non-supplied demand ( $P_{NSD(L,t)}$ ).  $P_{EAP(DG,t)}$  is important because the network operator can establish contracts with uninterruptible generation, for instance, with producers based on renewable energy. In extreme cases, when load is lower than uninterruptible generation the value of  $P_{EAP(DG,t)}$  is different from zero.  $P_{NSD(L,t)}$  is positive when the generation is not enough to satisfy load demand even using demand response.

The minimization of the objective function (1) is subject to the following constraints:

- The network active (2) and reactive (3) power balance, considering power loss, in each period t;

$$\begin{aligned} & \sum_{DG=1}^{N_{DG}} (P_{DG(DG,t)}^b - P_{EAP(DG,t)}^b) + \sum_{S=1}^{N_S} P_{Supplier(S,t)}^b + \sum_{L=1}^{N_L} (P_{NSD(L,t)}^b - P_{Load(L,t)}^b) \\ & + \sum_{V=1}^{N_V} (P_{Discharge(V,t)}^b - P_{Charge(V,t)}^b) = \\ & \sum_{k=1}^{N_B} V_{b(t)} \times V_{k(t)} (G_{bk} \cos(\theta_{b(t)} - \theta_{k(t)}) + B_{bk} \sin(\theta_{b(t)} - \theta_{k(t)})) \\ & \forall t \in \{1, \dots, T\}; k \neq b \end{aligned} \quad (2)$$

$$\begin{aligned} & \sum_{DG=1}^{N_{DG}} Q_{DG(DG,t)}^b + \sum_{S=1}^{N_S} Q_{Supplier(S,t)}^b - \sum_{L=1}^{N_L} Q_{Load(L,t)}^b = \\ & \sum_{k=1}^{N_B} V_{b(t)} \times V_{k(t)} (G_{bk} \sin(\theta_{b(t)} - \theta_{k(t)}) - B_{bk} \cos(\theta_{b(t)} - \theta_{k(t)})) \\ & \forall t \in \{1, \dots, T\}; k \neq b \end{aligned} \quad (3)$$

- Bus voltage magnitude and angle limits;

$$V_b^{min} \leq V_{b(t)} \leq V_b^{max} \quad \forall t \in \{1, \dots, T\} \quad (4)$$

$$\theta_b^{min} \leq \theta_{b(t)} \leq \theta_b^{max} \quad \forall t \in \{1, \dots, T\} \quad (5)$$

- Line thermal limits;

$$\left| V_{b(t)} \times \left[ \left[ (V_{b(t)} - V_{k(t)}) y_{bk} \right] + \left[ V_{b(t)} \times \frac{1}{2} y_{shunt\_b} \right] \right] \right| \leq S_{bk}^{max} \quad \forall t \in \{1, \dots, T\} \quad (6)$$

- Maximum distributed generation limit in each period t;

$$\begin{aligned} P_{DG(DG,t)} & \geq P_{DGMinLimit(DG,t)} \times X_{DG(DG,t)} \\ P_{DG(DG,t)} & \leq P_{DGMaxLimit(DG,t)} \times X_{DG(DG,t)} \end{aligned} \quad (7)$$

$$\forall t \in \{1, \dots, T\}; \forall DG \in \{1, \dots, N_{DG}\}$$

$$\begin{aligned} Q_{DG(DG,t)} & \geq Q_{DGMinLimit(DG,t)} \times X_{DG(DG,t)} \\ Q_{DG(DG,t)} & \leq Q_{DGMaxLimit(DG,t)} \times X_{DG(DG,t)} \end{aligned} \quad (8)$$

$$\forall t \in \{1, \dots, T\}; \forall DG \in \{1, \dots, N_{DG}\}$$

- Upstream supplier maximum limit in each period t;

$$P_{Supplier(S,t)} \leq P_{SupplierLimit(S,t)} \quad \forall t \in \{1, \dots, T\}; \forall S \in \{1, \dots, N_S\} \quad (9)$$

$$Q_{Supplier(S,t)} \leq Q_{SupplierLimit(S,t)} \quad \forall t \in \{1, \dots, T\}; \forall S \in \{1, \dots, N_S\} \quad (10)$$

- Vehicle charge and discharge are not simultaneous;

$$\begin{aligned} X_{(V,t)} + Y_{(V,t)} & \leq 1 \\ \forall t \in \{1, \dots, T\}; \forall V \in \{1, \dots, N_V\}; X_{(V,t)} \text{ and } Y_{(V,t)} & \in \{0, 1\} \end{aligned} \quad (11)$$

- Battery balance for each vehicle. The energy consumption for period t travel has to be considered jointly with the energy remaining from the previous period and the charge/discharge in the period. Efficiency rates are according to [2];

$$\begin{aligned} E_{Stored(V,t)} & = E_{Stored(V,t-1)} + \eta_{c(V)} \times P_{Charge(V,t)} \times \Delta t - \\ & E_{Trip(V,t)} - \frac{1}{\eta_{d(V)}} \times P_{Discharge(V,t)} \times \Delta t \end{aligned} \quad (12)$$

$$\forall t \in \{1, \dots, T\}; \forall V \in \{1, \dots, N_V\};$$

- Discharge limit for each vehicle considering the battery discharge rate;

$$\begin{aligned} P_{Discharge(V,t)} & \leq P_{DischargeLimit(V,t)} \times X_{(V,t)} \\ \forall t \in \{1, \dots, T\}; \forall V \in \{1, \dots, N_V\}; X_{(V,t)} & \in \{0, 1\} \end{aligned} \quad (13)$$

- Charge limit for each vehicle considering the battery charge rate;

$$P_{Charge(V,t)} \leq P_{ChargeLimit(V,t)} \times Y_{(V,t)} \quad (14)$$

$$\forall t \in \{1, \dots, T\}; \forall V \in \{1, \dots, N_V\}; Y_{(V,t)} \in \{0, 1\}$$

- Vehicle battery discharge limit considering the battery balance;

$$\frac{1}{\eta_{d(V)}} \times P_{Discharge(V,t)} \times \Delta t \leq E_{Stored(V,t-1)} \quad (15)$$

$$\forall t \in \{1, \dots, T\}; \forall V \in \{1, \dots, N_V\}; \Delta t = 1;$$

- Vehicle battery charge limit considering the battery capacity and previous charge status;

$$\eta_{c(V)} \times P_{Charge(V,t)} \times \Delta t \leq E_{BatteryCapacity(V)} - E_{Stored(V,t-1)} \quad (16)$$

$$\forall t \in \{1, \dots, T\}; \forall V \in \{1, \dots, N_V\};$$

- Battery capacity limit for each vehicle;

$$E_{Stored(V,t)} \leq E_{BatteryCapacity(V)} \quad (17)$$

$$\forall t \in \{1, \dots, T\}; \forall V \in \{1, \dots, N_V\}$$

- Minimum stored energy to be guaranteed at the end of period  $t$ . This can be seen as a reserve energy (fixed by the EV users) that can be used for a regular travel or an unexpected travel in each period;

$$E_{MinCharge(V,tLast)} \geq E_{Trip(V,t)} \quad (18)$$

$$\forall t \in \{1, \dots, T\}; \forall V \in \{1, \dots, N_V\}$$

$$E_{Stored(V,t)} \geq E_{MinCharge(V,t)} - E_{TripRed(V,t)} \quad (19)$$

$$\forall t \in \{1, \dots, T\}; \forall V \in \{1, \dots, N_V\}$$

$$E_{TripRed(V,t)} \leq E_{TripRedMax(V,t)} \quad (20)$$

$$\forall t \in \{1, \dots, T\}; \forall V \in \{1, \dots, N_V\}$$

## V. CASE STUDY

This section presents a case study that illustrates the application of the proposed methodology. The experiments that will be described in this section were conducted using Clementine version 12 [12]. This is an integrated data mining toolkit, which uses a visual-programming interface, and supports all knowledge data discovery process. MATLAB software has been used to create the scenario database and to implement the Particle Swarm Optimization (PSO) based approach. The basis for this case-study is a distribution network adapted from [13-14] as presented in Fig. 2. A total of 66 DG units and 2000 gridable vehicles were used in the scenario.

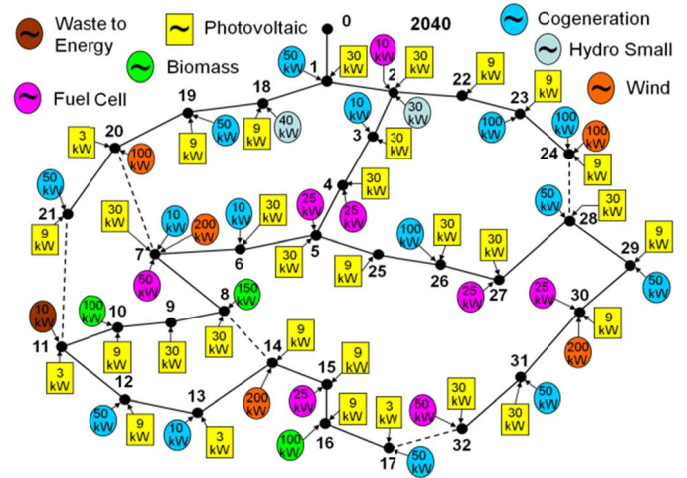


Fig. 2 – 33 bus network distribution system [13-14]

Fig. 3 presents the implemented methodology framework. Starting from an initial case study database [4] some inputs data were modified such as the available Distributed Generation (DG), the price of network suppliers and the base load.

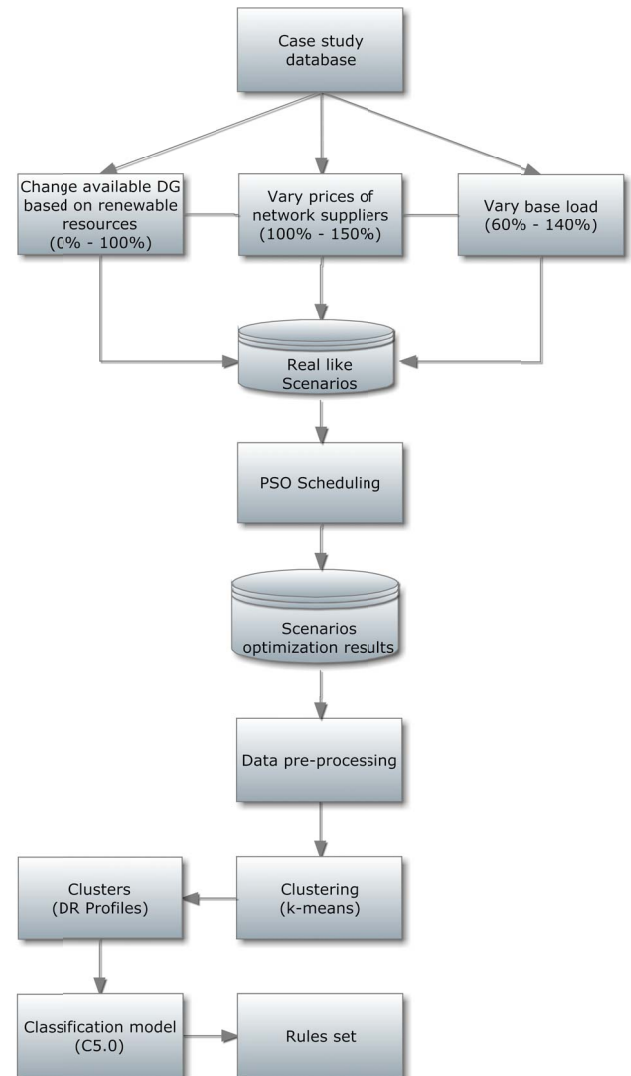


Fig. 3 – Implemented methodology framework

The main goal of these combinations in the input data was to create a range of scenarios. Fig. 2 presents how the 2,079 operation scenarios were obtained for the network presented. Applying particle swarm optimization (PSO) technique to the created scenarios the optimization results were obtained using the mathematical formulation that is presented in the previous sections. The presented formulation results from an improvement of previous works [4] which did not include trip reduce demand response program for electrical vehicles. Besides that, PSO is modified to include a power flow model to handle network constraints. This power flow is based on [15].

The mean execution time for each optimization, using PSO, is 50 seconds, which resulted in a total of 103,950 seconds (29 hours) to compute the 2,079 scenarios.

The parameters used in PSO for the optimization are the following:

- Number of particles was set to 10
- Number of iterations was set to 50
- Maximum and minimum velocities of the particles are automatically adjusted to each scenario using an algorithm taking into account the cost of the network resources and the load context.
- Gaussian mutation of the velocity governing weights is used in PSO core, using 0.80 as mutation factor.

A total of 101,472 PSO variables were used to compute the problem formulation as described in section IV.

Fig. 4 presents an example of the 2,079 optimized scenarios. This case corresponds to a variation of 150% of network supplier prices and 140% of the original base load. The available DG based on renewables such as wind and photovoltaic generation is set to 0%. In this case, the trip reduce demand response occurs mainly between period 9 and period 24. The peak load is 6.10 MW at period 20 as can be

seen in the figure. The power loss in this period is 180 kW. With trip reduce DR program available the PSO scheduled this program in some vehicles and the total reduced energy from vehicle trips amounts to 3.72 MWh of the about 10 MWh expected travel trips requirements. A total charge of 6.23 MWh was dedicated to EVs. The high EVs charge occurred at period 18. 1250 of the 2000 cars were scheduled to discharge at period 9 by PSO. The discharge of the 1250 cars, in this period, presents an average power of 0.51 kW, totaling 637 kW of discharge power. In this same period the amount of trip reduce DR program was scheduled to be 249 kWh with the participation of 130 vehicles.

The data pre-processing phase consisted in converting the optimization results (2,079) to a period basis, resulting in 49,986 operation periods, with the attributes of table II with the exception of the class attribute. After the data pre-processing, a clustering algorithm was used in order to identify similar patterns among trip reduce demand response. It was used the k-means clustering algorithm. In this phase the choice of the number of clusters is extremely important. Typically, the choice of the number of clusters may be supported, by the analysis and evaluation of measurement distances indexes [9]. However, the contribution of expert opinions must also be taken into account. In this case, the number of clusters was chosen to be equal to 10 due to the fact that it allows a reasonable analysis by the network operator. The results of the obtained profiles can be seen in table I.

In order to allow classifying new trip reduce demand response per period a classification model has been implemented

For the classification, rule-based modelling technique C5.0 classification algorithm has been used. The input attributes data sets have been divided to form a training group and a test group.

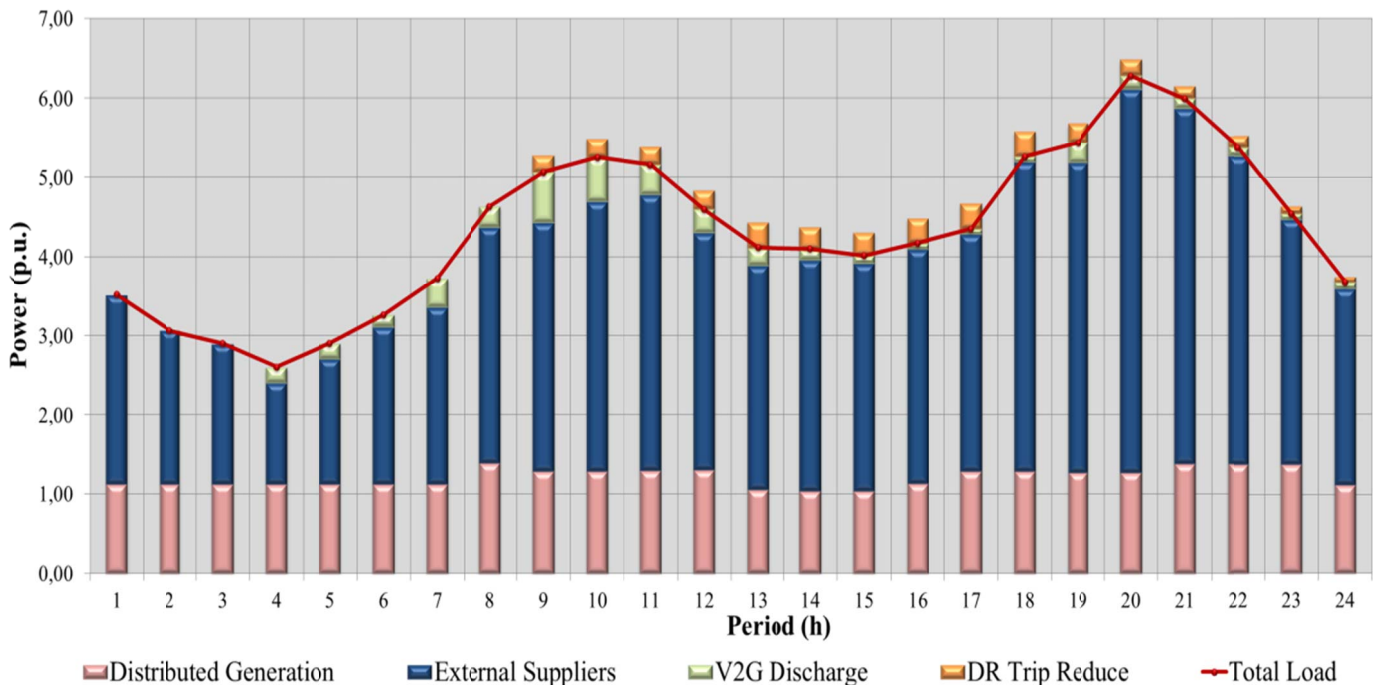


Fig. 4 – Example of scheduling with trip reduce DR available

The separation among test and training classes is to avoid the results being spoiled, i.e. the model error will not be influenced.

The classification model automatically generates the decision tree to provide the rules set.

TABLE I  
OBTAINED CLUSTER FROM THE K-MEANS CLUSTERING ALGORITHM

Cluster	Average capacity usage (kW)	Simulations per cluster
1	0,00	48,204
2	56,29	197
3	104,04	289
4	139,19	265
5	167,77	386
6	205,12	173
7	236,91	208
8	267,63	132
9	310,79	24
10	359,28	18
<b>Total</b>	-	<b>49,986</b>

Table II shows the attributes of the database that were used by the clustering algorithm and by the classification model to generate the rules set. The class attribute is used by the classification model to create the decision tree, i.e. the rule set. After creating the decision tree, the class attribute is not needed to classify new events. In the case of updating trip reduce DR patterns, using the clustering algorithm, the class attribute should be used to create the new updated decision tree and the corresponding rule set.

TABLE II  
ATTRIBUTES OF THE INITIAL DATABASE USED BY THE CLUSTERING ALGORITHM AND CLASSIFICATION MODEL

Variables	Description	Clust.	Class.
PERIOD	Time interval (1-24h)		X
TOTAL_LOAD	Total Load for the given period		X
DG_GEN	Generation from distributed generation for the given period		X
SUPLLIERS_GEN	Generation from suppliers for the given period		X
V2G_DR	Trip reduce demand response program used for the given period	X	
LAST_GEN_PRICE	Last generation price for the given period		X
V2G_LOAD	Total load from EVs for the given period		X
CLASS	Class obtained by clustering		X
<b>Total number of inputs</b>		<b>1</b>	<b>6</b>

As an example, the rule set for class 2 is presented:

<b>Rules for CLASS 2 - contains 3 rule(s)</b>
<i>Rule 1 for CLASS 2</i>
<i>if PERIOD &gt; 15</i>
<i>and PERIOD &lt;= 17</i>
<i>and TOTAL_LOAD &gt; 4,133</i>
<i>and V2G_LOAD &lt;= 0,697</i>

<i>and DG_GEN &lt;= 1,364</i>
<i>and LAST_GEN_PRICE &gt; 0,189</i>
<b>then CLASS 2</b>
<i>Rule 2 for CLASS 2</i>
<i>if PERIOD &gt; 12</i>
<i>and PERIOD &lt;= 13</i>
<i>and LAST_GEN_PRICE &gt; 0,182</i>
<b>then CLASS 2</b>
<i>Rule 3 for CLASS 2</i>
<i>if PERIOD &gt; 15</i>
<i>and PERIOD &lt;= 17</i>
<i>and V2G_LOAD &lt;= 0,697</i>
<i>and LAST_GEN_PRICE &gt; 0,189</i>
<b>then CLASS 2</b>

The classification model generated a rule set with an overall accuracy of 98.18%. Table III summarizes the information concerning the overall accuracy of the used C5.0 algorithm for this case study.

TABLE III  
OVERALL ACCURACY

	Total number	Accuracy (%)
Correct decisions	16,596	98.18
Wrong decisions	308	1.82
Total	16,904	100.00

## VI. CONCLUSIONS

This paper presents a new methodology to support network operators in the definition of the proposed electric vehicles demand response programs. The optimal resource scheduling solutions for the considered scenario set results in a large database from which relevant knowledge is automatically extracted using data mining techniques.

Considering different states of network conditions, 2,079 scenarios have been obtained resulting in 49,986 operation periods. These scenarios have been used to obtain the optimal energy resources scheduling. The optimal solutions considered the trip reduce demand response program for electric vehicles.

The proposed methodology obtained a set of clusters, concerning the values of electric vehicles demand response program by period, from the database scheduling solutions. After this phase, a classification model was applied. A rule set was obtained containing the variables of the studied network and explaining the reasons of a certain demand response usage. This knowledge can be subsequently used by the operator in its daily operation in order to support decisions concerning demand response programs in the context of smart grids paradigm operation.

The application of the proposed methodology enables to estimate the DR usage. The case study produced 10 electric vehicles demand response clusters. The classification model generated a rule set with an overall accuracy of 98.18% to estimate the DR usage.

## REFERENCES

- [1] R. C. Brown and J. D. Gifford, "Four economies of sustainable automotive transportation," *Biofpr: Biofuels, Bioproducts & Biorefining*, vol. 5, pp. 293-304, May-Jun 2011.
- [2] R. Ball, N. Keers, M. Alexander, and E. Bower, "Deliverable D2.1 Modelling Electric Storage Devices for EV," MERGE Project EU, 2010.
- [3] R. C. Green, W. Lingfeng, and M. Alam, "The impact of plug-in hybrid electric vehicles on distribution networks: a review and outlook," in *IEEE/PES Power and Energy Society General Meeting (PES GM)*, pp. 1-8, 2010.
- [4] J. Soares, T. Sousa, H. Morais, Z. Vale, and P. Faria, "An Optimal Scheduling Problem in Distribution Networks Considering V2G," in *IEEE SSCI Symposium on Computational Intelligence Applications in Smart Grid (CIASG) Paris, France, 2011*.
- [5] Z. Vale, H. Morais, S. Ramos, J. Soares, P. Faria, "Using Data Mining Techniques to Support DR Programs Definition in Smart Grids", *IEEE Power and Energy Society General Meeting 2011, Detroit, Michigan, USA, Jul 24 - 29, 2011*.
- [6] S. Ramos, Z. Vale, J. Santana, J. Duarte, "Data Mining Contributions to Characterize MV Consumers and to Improve the Suppliers-Consumers Settlements", *IEEE, Power Engineering Society General Meeting, 2007*, pp. 1-8, 24-28 June, 2007.
- [7] A. K. Jain, M. N. Murty, P. J. Flynn, "Data Clustering: A Review", *ACM Computing Surveys*, Vol. 31, N<sup>o</sup>.3, September, 1999.
- [8] A. K. Jain, R. C. Dubes, "Algorithms for Clustering Data", Prentice-Hall advanced reference series. Prentice-Hall, Inc., Upper Saddle River, NJ, 1988.
- [9] S. Ramos, H. Morais, Z. Vale, P. Faria, J. Soares, "Demand Response Programs Definition Supported by Clustering and Classification Techniques", in *ISAP 2011 – 16th International Conference on Intelligent System Application to Power Systems, IEEE, September 25-28, Hersonissos, Crete, Greece, 2011*.
- [10] J. A. Hartigan, M. A. Wong., "A K-Means Clustering Algorithm", *Applied Statistics* 28:100-108, 1979.
- [11] J. Ferreira, S. Ramos, Z. Vale, J. Soares, "Transmission Expansion Planning supported by Data Mining Based Methodology", *IEEE Intelligent Systems*, Vol. 26, No. 2, March/April, 2011.
- [12] SPSS, "Text Mining for Clementine 12 – User's Guide".
- [13] M. Baran, F. Wu, "Network Reconfiguration in Distribution Systems for Loss Reduction and Load Balancing", *IEEE Transactions on Power Delivery*, Vol. 4, N<sup>o</sup> 2, pp. 1401-1407, 1989
- [14] P. Faria, Z. A. Vale, and J. Ferreira, "Demi - A demand response simulator in the context of intensive use of distributed generation", *IEEE International Conference on Systems Man and Cybernetics (SMC) 2010*, pp. 2025-2032, 2010.
- [15] D. Thukaram, H. M. W. Banda, and J. Jerome, "A robust three phase power flow algorithm for radial distribution systems," *Electric Power Systems Research*, vol. 50, pp. 227-236, Jun 1 1999.

## BIOGRAPHIES



**João Soares** is a BSc graduate in computer science. He received the MSc in Electrical Engineering from the Polytechnic Institute of Porto (ISEP/IPP), Portugal in 2011. He is an Assistant Researcher at GECAD–Knowledge Engineering and Decision-Support Research Center of the Engineering Institute of Porto Polytechnic Institute of Porto (ISEP/IPP) His current research activities are concentrated in future energy distribution networks.



**Sérgio Ramos (M'11 S'07)** is an assistant professor at Polytechnic Institute of Porto. He is a researcher at GECAD – Knowledge Engineering and Decision-Support Research Center and his main research interests include electricity markets operation, and electrical installations. He received his MSc in 2006 graduated in the Instituto Superior Técnico, Lisbon, Portugal. He is currently pursuing the PhD degree in the same institution.



**Zita A. Vale (SM'10 M'93 S'86)** is the director of the Knowledge Engineering and Decision Support Research Group (GECAD) and a professor at the Polytechnic Institute of Porto. She received her diploma in Electrical Engineering in 1986 and her PhD in 1993, both from University of Porto. Her main research interests concern Artificial Intelligence (A.I.) applications to Power System operation and control, Electricity Markets.



**Hugo Morais (M'11 S'08)** received the MSc degree in 2010 from the Polytechnic Institute of Porto, Porto, Portugal. He is a Researcher at GECAD – Knowledge Engineering and Decision-Support Research Center. He is also a PhD student with the University of Trás-os-Montes e Alto Douro, Vila Real, Portugal and his research interests include future power systems.



**Pedro Faria (S'10)** received the MSc degree in Electrical Engineering from the Polytechnic Institute of Porto (ISEP/IPP), Portugal in 2011. Currently, he is an Assistant Researcher at GECAD – Knowledge Engineering and Decision-Support Research Center of ISEP/IPP. His current research activities are focused on future electrical networks with intensive use of distributed generation.