

Allocation and Sizing of Distributed Generation with Data Mining Code of Repositories - DAMICORE

**Gesiel Rios Lopes, Alessandro Wilk S. Almeida,
Claudio F. Motta Toledo, Alexandre C. B. Delbem**

Institute of Mathematical and Computer Sciences
University of São Paulo (ICMC-USP)
São Carlos-SP, Brazil

{gesielrios, alessandrowilk}@usp.br, {claudio, acbd}@icmc.usp.br

ABSTRACT

The power system networks have undergone significant changes in the last decades, where the increase in distributed generation represents enormous technical challenges. The distribution networks are traditionally radial, where the devices and procedures were designed according to this topology. Thus, the benefits brought by the allotted generation depend on changes in radial system, where the location and size of the generating units will have a direct impact. The correct re-design of the network can lead to reduction in energy losses and increase in the reliability levels for the whole system. In this paper, we present a research methodology for allocation and sizing of distributed generators. The methodology combines Data Mining Code of Repositories (DAMICORE) with the resolution of Mixed Integer Linear Programming (MILP) models. The results achieved indicate that the proposed method has potential to become a useful tool for decision-making on distributed power system design.

KEYWORDS. Allocation, Sizing, Distributed Generation, Data Mining and DAMICORE.

1. Introduction

Electric Power Systems has undergone significant transformations in the last decades where decentralized power systems have become a viable option as reported by Adefarati e Bansal [2016]. The distributed generation (DG) systems provide a source of active electric power which is connected directly to the distribution network as defined by Ackermann et al. [2001]. The benefits brought by distributed generation depend of the location and size of the generating units, where the correct sizing of such units will lead to the reduction of energy losses and increase of reliability levels for the entire power system.

The increase of the distributed generation represents enormous technical challenges for distribution networks that are traditionally designed following a radial topology. The insertion of DG brings different operating conditions for this type of network topology such as power flow inversion, fluctuations in voltage profile, increase of fault levels, among others. In this sense, the development of new computational methodologies and tools are necessary in the face of this distributed scenario, aiming to support decisions about generation units placement and their power sizing.

The world today is overloaded with information and data in the most different fields, the volume of these data are growing daily. In the last decades, the society has been inundated with data and information of the most varied types. The large volume, the speed of availability and varieties of such data drew attention of large corporations and state departments. They have spend a large amount of money to collect terabytes of data. However, all these investments do not mean that the relevant information hidden within the data repositories will be properly gathered (Bibri [2018]).

In this context, Data Mining techniques have been developed and applied by companies to uncover patterns and trends from their available databases. Data mining is a knowledge discovery process as stated by Ponniah [2001]. The volume growth of data and databases, concerning everyday activities, brings data mining as the new business technologies.

Moreover, data mining helps to understand the essence of the data by unveiling patterns and trends in the data that we never knew existed. In Bigus [1996], data mining is defined as the efficient discovery of valuable, non-obvious information from a large collection of data, where its core relies in the automated discovery of new facts and the relationship between the data.

Mathematical Programming (MP) is a technique that belongs to Operations Research field and it is widely applied to military, business and industry problems, among others. MP has helped the process of making decision in power system problems as reported by Lee et al. [2018] or optimize the structure of DG systems as described in Twaha e Ramli [2018].

The present paper introduces a methodology that combines data mining and mathematical programming to make decisions on power network system with distributed generation. The goal is to find the best allocation and sizing for distributed generators. The data mining method applied is called Data Mining Code of Repositories (DAMICORE), which integrates three algorithms to find relationships between data. The method can be employed over data type of any nature by using compression of the samples (files) to define a phylogenetic tree. The main idea is to identify similarities between the data from the relationships established by the phylogenetic tree.

A mixed-integer linear programming (MILP) model is proposed to describe the power distribution system network problem with distributed generation. Thus, data of the distribution system are processed by MILP model such as: active power DG, minimal voltage, objective function value, active power losses value, minimal voltage and normally open switches. The values returned by the MILP will feed DAMICORE that returns the possible allocation of the DG units for the decision-maker.

The paper is organized as follows: Section II presents the related works. Section III introduces the formulation for our problem. Section IV describes the proposed method, while Section V shows how DAMICORE was applied to solve the problem of allocation and sizing of distributed generators. Section VI presents the results and section VII has our conclusions.

2. Related works

The presence of the distributed generation can provide several benefits to the electrical system such as to avoid new investments in the network expansion, improvement in voltage profile, diversity of the energy matrix and low environmental impact. However, the benefits brought by the distributed generation depend on the placement and size of the generating units. The insertion of DG in radial distribution system brings benefits such as: reduction of line losses, voltage profile improvement, better power quality and system reliability.

Several techniques have been used for allocation and sizing of distributed generators in distribution systems: Genetic Algorithm (GA), Evolutionary Programming (EP), Particle Swarm Optimization (PSO), Artificial Bee Colony (ABC) Algorithm, Dynamic programming, Ant Colony Search Algorithm, Tabu Search Algorithm (TS), Teaching and Learning Based Optimization (TLBO), Firefly Algorithm, Analytical Methods, among others as reviewed by Twaha e Ramli [2018].

In Griffin et al. [2000], the authors proposed an algorithm to determine the optimal allocation of distributed generation constrained by minimization of system losses. A methodology to evaluate the impacts of distributed generation on electric losses, is proposed in Borges e Falcao [2003] for distribution systems, regarding the allocation, sizing, reliability and voltage profile. The methodology is based on a power flow method to evaluate electric losses and voltage profile, while an analytic method was used to calculate the reliability indices to handle multiple distributed generations.

Genetic Algorithm and Optimal Power Flow are introduced in Mardaneh e Gharehpetian [2004] to find optimal location and sizing of distributed generation. The idea is minimizing the active and reactive power cost for distributed generation units. In C.Wang e Nehrir [Nov. 2004], the authors present a method to determine the optimal placement for a distributed generators with unity power factor in radial systems. The goal is to minimize power loss of the whole network system.

The authors in Kashem et al. [2006] proposed sensitivity analysis to determine the optimal size and allocation point for distributed generation, when power losses in the radial systems must be minimized. To obtain the optimal size of distributed generation, techniques were developed taking into account constant impedance and constant load.

A comparison between several heuristic approaches namely, Multiobjective Genetic Algorithm (MOGA), Particle Swarm Optimization (PSO) and Biogeography Based Optimization (BBO) is done in Hazem et al. [December, 2017]. The methods have to find the optimal allocation, sizing and number of DG units within a distribution system.

In our paper, we apply the data mining technique, named DAMICORE, to find the allocation for distributed generators. DAMICORE was applied by Sanches et al. [2011] to find similar software code, where the method needs to handle codes with different types of representations. A distance matrix is used by DAMICORE to return relevant software similarities. DAMICORE was also employed by de Abreu Silva et al. [2014] to define the configuration of an heterogeneous multi-core processor (HMP) and the related cache sizes. This data mining approach evaluates the real energy consumption of the HMPs and searches for similarities among their configurations.

A hybrid genetic algorithm is introduced by de Souza e de Albuquerque [2006] to plan trajectory and control a robot navigation. The genetic algorithm defines the actions to be executed by the robot, while the Normalized Compression Distance (NCD) algorithm, applied by DAMICORE,

controls the robot by identifying its current state. The most related work applying DAMICORE is described in Reche et al. [2019], where a decision support tool for fault location in radial distribution systems is introduced. DAMICORE and a fault location method are combined within a cloud computing environment. Complex data are provided from two different scenarios and the proposed method is able to satisfactorily predict faults in the radial distribution system.

3. Problem Description

The problem approached here is the minimization of losses in radial distribution systems, taking into account the current injection from the distributed generation (DG).

This problem is described as a mixed-integer linear programming model, where expression 1 represents the objective function that minimizes the active losses in the power distribution system.

$$\min \sum_{\langle i,j \rangle \in arcs} (I_{Ri,j}^2 + I_{Ii,j}^2) R_{i,j}; \quad (1)$$

For the resistance (model parameter) through the branch i, j , we have $arcs = \{ \langle i, j \rangle \mid i, j \in N, R_{i,j} \geq 0.001 \}$, with the variables $I_{Ri,j}$ as the real part of the flow of current in the branch i, j and $I_{Ii,j}$ as the imaginary part of the flow of current in the branch i, j .

Constraint (2) assures the radial topology for the power system

$$\sum_{\langle i,j \rangle \in arcs} x_{i,j} = Nnos - |F|; \quad (2)$$

where, $x_{i,j}$ is the binary variable that defines if there is an active connection (closed state of switch), $x_{i,j} = 1$, between nodes i and j , $x_{i,j} = 0$; otherwise. Parameters $Nnos$ and F are the number of nodes and set of feeders, respectively.

Constraints (3) and (4) state that the real and imaginary parts of the demand will be satisfied by the flow of current

$$\sum_{\langle i,j \rangle \in arcs} I_{Rj,i} - I_{Ri,j} = D_{Ri}, \quad \forall i \in N \setminus F \quad (3)$$

$$\sum_{\langle i,j \rangle \in arcs} I_{Ij,i} - I_{Ii,j} = D_{Ii}, \quad \forall i \in N \setminus F \quad (4)$$

where, parameter D_{Ri} is the demand of active power and D_{Ii} is the demand of reactive power in node i .

Constraints (5) and (6) describe the Kirchhoff law for nodes where DG units were allocated.

$$\sum_{\langle i,j \rangle \in arcs} I_{Rj,i} + \frac{P_{DG_i}}{V_{base}} - I_{Ri,j} = D_{Ri}, \quad \forall i \in G \quad (5)$$

$$\sum_{\langle i,j \rangle \in arcs} I_{Ij,i} + \frac{Q_{DG_i}}{V_{base}} - I_{Ii,j} = D_{Ii}, \quad \forall i \in G \quad (6)$$

P_{DG_i} is the active power parameter of DG unit at node i and Q_{DG_i} is the reactive power of DG unit at node i .

Constraints (7) activate only one arc direction, allowing to change the current flow during the system reconfiguration.

$$x_{i,j} + x_{j,i} \leq 1, \quad \forall \langle i, j \rangle \in arcs; \quad (7)$$

Constraints (8) together with constraint (2) will not allow to disconnect nodes in the network power system.

$$\sum_{\langle j,i \rangle \in arcs} x_{j,i} \leq 1, \quad \forall i \in N \setminus F. \quad (8)$$

Constraints (9) will allow or not the flow of current through the switch $\langle i, j \rangle$.

$$I_{Ri,j} \leq 10000 \cdot x_{i,j}, \quad \forall \langle i, j \rangle \in arcs; \quad (9)$$

Constraints (10) define a lower bound for the variable v_{Ri} (voltage value at node i).

$$v_{min} \leq v_{Ri}, \quad \forall i \in N \setminus F. \quad (10)$$

Constraints (11) describe the voltage drop at the current direction from active switches.

$$v_{Ri} - v_{Rj} \leq R_{i,j} \cdot M_{i,j} \cdot x_{i,j} + 10.000(1 - x_{i,j}), \forall \langle i, j \rangle \in arcs \quad (11)$$

4. Data Mining of Code Repositories - DAMICORE

Data Mining of Code Repositories (DAMICORE) is a data mining technique proposed by Sanches et al. [2011] which combines three algorithms:

- Normalized Compression Distance (NCD) from Theory of Information, Cilibrasi [2005].
- Neighbor Joining (NJ) from Phylogenetics, Saitou e Nei [1987].
- FastNewman (FN) from Complex Networks, Newman [2004].

The basic principle of DAMICORE is to identify similarities between any type of data: numeric, images, texts, codes, etc. In order to apply DAMICORE, it is required that all the data must have the same nature and they must be separated into files. The method will group the similarities found within clusters. We summarize next the main aspects of these techniques.

1. *Normalized Compression Distance* (NCD) is a distance metric that can be used for several types of data, even when one does not know characteristics that can be used to represent properly the data to be analyzed, as reported in Sanches et al. [2011]. In Cilibrasi [2005], the authors proposed the NCD metric based on the approximation complexity of *Kolmogorov* as follows:

$$NCD(x; y) = \frac{C(xy) - \min\{C(x), C(y)\}}{\max\{C(x), C(y)\}} \quad (12)$$

where $C(xy)$ denotes the compression size for the concatenation of x and y , $C(x)$ denotes the compression size for the concatenation of x and $C(y)$ denote the compression size for the concatenation of y , with x and y being the two objects to be compared.

2. *Neighbor Joining* (NJ) is a construction method for phylogenetic trees. The principle of NJ is to identify pairs of objects that minimize the total branch length at each stage. These pairs of objects are called neighbors. The tree is built through an iterative process where all data (leaf nodes) are connected through one interior node within the tree, Saitou e Nei [1987]. Neighbor Joining (NJ) method is a relatively simple technique and computationally efficient.
3. *Fast Newman* (FN) is a technique used for detection of communities in complex networks. The algorithm can deal with large-scale networks finding the main representative communities of a system, Newman [2004]. Complex Networks is today a research area that presents models more effective for representation and complex modelling, attracting increasing attention from several fields of knowledge as reported in Newman [2006].

5. Allocation and Sizing of Distributed Generation via Damicore

The main objective of the proposed methodology is to define the allocation and sizing of distributed generators (DGs), where we combine data mining with mathematical programming techniques. There are many approaches applied through the data mining phases, and the authors in Rezende et al. [2003b] divide the data mining process into three stages: pre-processing, pattern extraction and post-processing. The authors also include a phase before the pre-processing that is domain knowledge, and another phase after post-processing that addresses the use of knowledge. As described in Rezende et al. [2003a], we add the domain knowledge phase before pre-processing, and the acquired knowledge phase after post-processing.

In the domain knowledge phase, it is defined the objectives to be achieved. In our problem, the objective is to minimize the power system losses, based on data gathered from DGs units previously allocated in the system network. The pre-processing phase extracts relevant characteristics to generate samples (data) that will be sent to Damicore. In this phase, each node of a benchmark power system is chosen as a possible location for the DG unit, except by the feeder one. Moreover, for each node where a DG unit is allocated, a configuration of the radial distribution system is generated aiming to minimize losses. The system configuration is achieved by solving the related MILP model described in section 3. The solutions found by the MILP will tell us the open-close switches in the systems that will minimize the power system losses as well as the DG allocation. The values returned by such solutions are employed to choose the best data sets for extracting of patterns.

In the patterns extraction phase, all data returned by the MILP solutions are processed simultaneously by Damicore to look for patterns of allocation and sizing of DG. The main data attributes are placed in a standard format such as every file has the same number of bytes and attributes. The chosen attributes to be sent to DAMICORE technique are: node where the DG unit is allocated, active power DG (kW), minimal voltage (V), node location of minimal voltage and active power loss (kW). The normally open switches (tie switches) are also input data for DAMICORE. Redundant or irrelevant attributes were removed since they do not contribute to the accuracy of the methodology, or they could decrease the accuracy of the method. These irrelevant attributes are infeasible solutions, active power below 2 kW and open switches near the substation. Table 1 illustrates the selection of attributes.

Where,

- Node = Candidate nodes to allocate DG units

Table 1: Class of attribute for Data Mining

Conf.	Node	Pdg (Kw)	V_{min}	N_{vmin}	P_{Loss}
DG25-07	25	793.57	11.898	31	115.98
DG6-09	6	854.98	11.871	31	118.08
DG15-05	15	614.30	11.893	30	102.62
DG30-08	30	825.04	11.738	12	113.95

- Pdg = Active Power of DG (kW);
- V_{min} = Minimal voltage of DG (V);
- N_{vmin} = Node minimal voltage;
- P_{Loss} = Active power loss (kW).

candidate nodes to allocate DG units

Damicore generates a phylogenetic tree by clustering samples based on their characteristic. The extraction phase produces the initial model of knowledge. The phylogenetic tree is a graphic representation, where each inner node (small node) represents a level of similarity for an attribute, each path represent a relationship and leaf nodes stands for the class attribute.

The post-processing stage is assessed the extracted initial model and it is inferred if the knowledge extracted is relevant, Piatetsky-Shapiro e Matheus [1994]. Finally, we can exploit knowledge to define the best allocation and sizing for distributed generators.

Algorithm 1 was developed in order to support the decision-making process about where to allocate generators as well as to define their power injection amount. First, the number of nodes

Algorithm 1 Allocating generators

Require: $\overline{N}$, min , max , $systemInstance^0$

Ensure: $damicoreReport$

```

for all  $i \in \overline{N}$  do
   $count \leftarrow 0$ 
  while  $(count++) < 10$  do
     $systemInstance \leftarrow systemInstance^0$ 
     $p_{inj} \in [min, max]$ 
     $systemInstance \leftarrow (i, p_{inj})$ 
     $DATASET \leftarrow ModelSolution(systemInstance)$ 
  end while
end for
 $damicoreReport \leftarrow Damicore(DATASET)$ 

```

$\overline{N}$ without feeders in the network system is provided as well as the min and max values for power injection from a DG unit. The $systemInstance^0$ provides all the other data related to the bus system under consideration. Next, a DG unit is place at each node $i \in \overline{N}$ and 10 possible power injection values, $p_{inj} \in [min, max]$, are associated with such generator.

For each combination of (i, p_{inj}) , the $systemInstance$ is updated with the generator position i and power injection p_{inj} . The exact solution of the $systemInstance$ is achieved, by using the mathematical model and the Branch & Cut algorithm from CPLEX. The solutions will

feed the *DATASET* as previously described. Finally, *damicoreReport* has all the information extracted using *Damicore(DATASET)*.

6. Computational Results

All experiments were evaluated through a benchmark distribution system from literature. The test system 33-bus is a hypothetical system proposed by Baran e Wu [April 1989], as shown in Figure 6, with 33 bus, base voltage 12.66 kV, 37 branches, 32 switches (normally closed) and 5 tie switches (normally open). The switches normally open are numbered from s33 to s37, the switches normally closed are numbered from s1 to s32. The total substation loads for the base configuration are 5084.26 kW e 2547.32 kVar, the total real and reactive power loads on the system are 3715 kW e 2315 kVar. The initial power loss of this system is 202.67 kW and after reconfiguration (optimal configuration) the active power loss of 139.55 kW is reached, which means a loss reduction of 31.14 %.

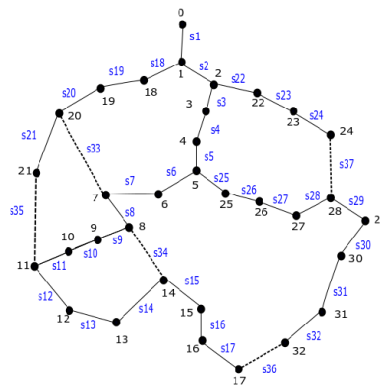


Figure 1: Initial Configuration of the 33-bus System

We assume three hypothesis when allocating DGs:

- DGs is used for supplying active power only;
- DGs are assigned to all nodes;
- DGs are not connected to feeders.

Thus, the 33-bus radial distribution system is used here to test the proposed method when defining the allocation and sizing of DG units. Figure 2(a) illustrates a possible configuration for the 33-bus system, when a DG is allocated at node s15. This configuration is achieved by solving the MILP describe in section 3, where the red arrows indicate a unidirectional power flow from the source(transmission system) to the loads.

As previously explained in Algorithm 1, the MILP will return a configuration for DG assigned to each one of the nodes, except by the feeder, in the 33-bus system. Next, values from these solutions are organized in the class attribute, as illustrate in Table 1, and send to DAMICORE that will generate one phylogenetic tree. Figure 2(b) shows the phylogenetic tree returned by DAMICORE for the 33-bus system.

In the phylogenetic tree, the nodes with same color present a strong similarity. DAMICORE found seventeen clusters and the phylogenetic tree has various levels of similarity. The

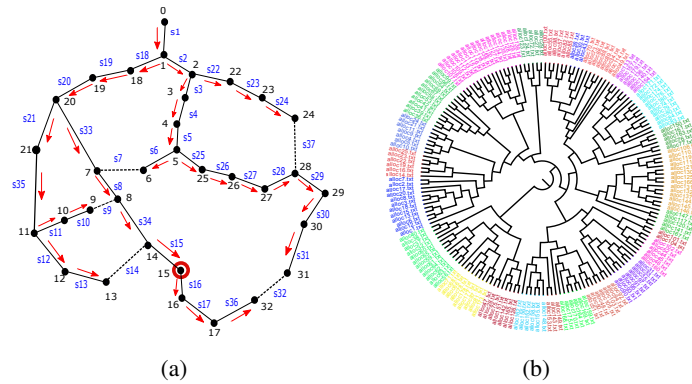


Figure 2: (a) Single Allocation Distribution Generated 33 – bus System and (b) Clusters returned by DAMICORE.

class attribute files (e.g, alloc151.txt), processed by DAMICORE, can be seen as leaf nodes in the phylogenetic tree. Thus, Figure 2(b) indicates the initial level of correlation between these files.

Next, after analysis of the extracted patterns, it was returned four groups of patterns from the phylogenetic tree, based on allocation and sizing of DG. There are also several subgroup with high correlations levels for values of minimal voltage, node minimal voltage and power loss. The green pattern stands for allocations in the nodes from 1 to 5. Pattern pink has DG units attached to nodes 22-32, 16 and 17. Pattern yellow returned allocation in nodes 6, 7, 18, 19 and 20. Pattern blue has allocation in the node between 8-15 and node 21. Figure 3(a) shows the phylogenetic tree updated after patterns extraction of the groups.

Figure 3(b) illustrates the branches with the respective best allocation nodes of DG. Thus, based on the pattern described in each branch, it is possible to identify the respective nodes within the 33-bus system. For instance, if a decision-maker decides to allocate a DG unit, following the yellow pattern for minimal voltage, node minimal voltage and power loss, the candidate nodes are 6, 7 and 18-20.

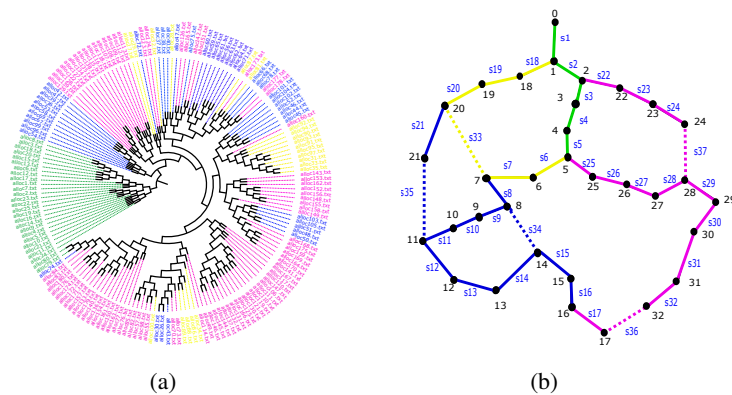


Figure 3: (a) Interpretation patterns by clusters and (b) Interpretation system by pattern.

Figure 4(a) illustrates voltage profile improvement for each pattern returned by DAMICORE. while If we decide to evaluate the overall results, note also which this improvement occur in the cluster pink. Moreover, this improvement can be achieved by strategically placing and sizing

DG along the network, as showed in figure 4(b).

The system voltage profile is shown in Figure 9 for different simulated configurations, based on the four patters returned by DAMICORE. The minimum voltage of the initial system configuration occurs on bus 17 with 11.55V. There is an improvement of the voltage profile in all cluster patterns found by DAMICORE. For instance, let's take the minimum voltage values at bus 8. These values happen when the DG is allocated at nodes 29, 31 and 32 with minimum voltage increased by 2.89 %, 2.14 % and 1.87 % respectively. The only exception occurs at bus 25, when the DG is allocated at bus 15 with a voltage worsening around 4.45 %.

Figure 4(b) shows the distribution of active power losses considering all simulated configurations. The distribution indicates that the smallest losses occur below 100 Kw when the active power is between 500Kw and 2000Kw.

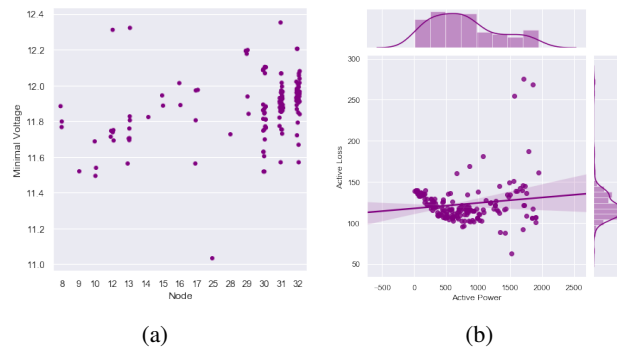


Figure 4: (a) Interpretation voltage profile improvement and (b) Interpretation active power loss.

Table 2 has the best results found by DAMICORE when compared against the active power loss of the initial configuration: 202.67 kW.

Table 2: Best pattern configuration DAMICORE.

Cluster	Node	P _{dg} (Kw)	V _{min}	N _{umin}	P _{Loss}
Green	5	754.31	12.015	31	95.50
Pink	23	1704.17	12.055	32	92.25
	24	1422.34	12.067	32	87.66
	24	1341.79	12.080	32	88.35
Blue	8	777.89	12.086	30	96.10
	10	1542.82	12.089	30	62.75
Yellow	7	524.44	11.929	31	103.80

The Green pattern extracted by DAMICORE presents a reduction of 52.87 % when the DG unit is allocated at bus 5. In the Pink pattern, the best DG allocation points for such reduction are buses 23 and 24. The bus 23 returns a reduction in P_{loss} that reaches 54.48 %, while bus 24 obtains 56.74 % and 56.40 % for 1422.34Kw and 1341.79Kw values of P_{dg} , respectively. In the Blue pattern, the best DG allocation happens at bus 8 and 10 with P_{loss} reduction reaching 52.58 % and 69.00 %, respectively.

7. Conclusions

In this work, a method combining data mining and mathematical programming was proposed to help the decision-making process on power network system with distributed generation.

The potential of the Data Mining Code of Repositories (DAMICORE) technique was investigated to find the better sizing and better location of distributed generators in radial distribution systems. DAMICORE is feed by data from solutions found when solving different configurations of the distributed power system. These configurations are solved by using a Mixed Integer Linear Programming (MILP) model. DAMICORE presented the ability to extract relevant patterns for allocation and sizing of distributed generators, by reducing significantly the power system losses of a benchmark system. As future work, other benchmark systems will be solved and the patters returned by DAMICORE will be compared against other data mining techniques.

Acknowledgments

The authors of this work would like to thank the CAPES (PROEX), CNPQ (grant 312605/2018-8), Center for Artificial Intelligence (C4A-IUSP) and the support from the São Paulo Research Foundation (FAPESP grant 2019/07665-4, 2020/16578-5) and from the IBM Corporation. We would like also to thank the CEPID-CeMEAI/ICMC-USP (CEPID, FAPESP grant 2013/07375-0)). We also thank ICMC-USP and LCR for offering the necessary infrastructure for this study.

References

- Ackermann, T., Göran, A., e Lennart, S. (2001). Distributed generation: a definition. *Electric power systems research*, 57(3):195–204.
- Adefarati, T. e Bansal, R. C. (2016). Integration of renewable distributed generators into the distribution system: a review. *IET Renewable Power Generation*, 10:873–884.
- Baran, M. E. e Wu, F. (April 1989). Network reconfiguration in distribution system for loss reduction and load balancing. *IEEE Trans. Power Del*, 4(2):1401–1407.
- Bibri, S. E. (2018). Systems thinking and complexity science and the relevance of big data analytics, intelligence functions, and simulation models. In *Smart Sustainable Cities of the Future*, p. 297–369. Springer.
- Bigus, J. P. (1996). *Data Mining with Neural Networks: Solving Businees Problem-from Application Development to Decision Support*. Mcgraww-Hill.
- Borges, C. L. T. e Falcao, D. M. (2003). Impact of distributed generation allocation and sizing on reliability, losses and voltage profile. in *Proc.IEEE Power Tech Conf., Bologna, Italy*, 2:1–5.
- Cilibrasi, P., R.and Vitany (2005). Clustering by compression. *IEEE Transactions on Information Theory*, 51(4):1523–1545.
- C.Wang e Nehrir, M. H. (Nov. 2004). Analytical approaches for optimal placement of distributed generation sources in power systems. *IEEE Trans Power Syst.*, 19(4):2068–2076.
- de Abreu Silva, B., Cuminato, L. A., Delbem, A. C., Diniz, P. C., e Bonato, V. (2014). Application-oriented cache memory configuration for energy efficiency in multi-cores. *IET Computers & Digital Techniques*, 9(1):73–81.
- de Souza, B. A. e de Albuquerque, J. M. C. (2006). Optimal placement of distributed generators networks using evolutionary programming. in *Proc. Transm. Distrib. Conf. Expo.: Latin Amer.*, p. 1–6.

- Griffin, T., Tomsovic, K., Secrest, D., e Law, A. (2000). Placement of dispersed generations for reduced losses. *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*.
- Hazem, N., Elshahed, M. A., e Osman, Z. H. (December, 2017). Optimal placement of dispatchable and non-dispatchable distributed generation of different technologie. *Nineteenth International Middle East Power Systems Conference (MEPCON)*, p. 1023–1030.
- Kashem, M. A., Le, A. D., Negnevitsky, M., e Ledwich, G. (2006). Distributed generation for minimization of power losses in distribution systems. in *Proc. IEEE Power Eng. Soc. Gen. Meeting*, p. 1–8.
- Lee, J.-Y., Aviso, K. B., e Tan, R. R. (2018). Optimal sizing and design of hybrid power systems. *ACS Sustainable Chemistry & Engineering*, 6(2):2482–2490.
- Mardaneh, M. e Gharehpetian, G. B. (2004). Siting and sizing of dg units using ga and opf based technique. in *Proc. IEEE Region 10 TENCON Conf.*
- Newman, M. E. J. (2004). Fast algorithm for detecting community structure in networks. *Physical Review*, 69(6):066133.
- Newman, M. E. (2006). Finding community structure in networks using the eigenvectors of matrices. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 74(3):1–19.
- Piatetsky-Shapiro, G. e Matheus, C. (1994). he interestingness of deviations. *International Conference on Knowledge Discovery and Data Mining*, p. 23–36.
- Ponniah, P. (2001). *Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals*. John Wiley Sons, Inc., 1nd ed edition.
- Reche, E. A., de Sousa, J. V., Coury, D. V., e Fernandes, R. C. A. S. (2019). Data mining-based method to reduce multiple estimation for fault location in radial distribution systems. *IEEE Transactions on Smart Grid*, 10(4):3612–3619.
- Rezende, S. O., Pugliesi, J. B., Melanda, E. A., e Paula, M. F. (2003a). *Intelligent Systems: Fundamentals and Applications*. Manole.
- Rezende, S. O., Pugliesi, J. B., Melanda, E. A., e Paula, M. F., editors (2003b). *Mineração de dados. Sistemas inteligentes: fundamentos e aplicações*. Editora Manole.
- Saitou, N. e Nei, M. (1987). The neighbor-joining method : A new method for reconstructing phylogenetic trees. *Molecular biology and evolution*, 4(4):406—425.
- Sanches, A., Cardoso, J. M., e Delbem, A. C. (2011). Identifying merge-beneficial software kernels for hardware implementation. In *2011 International Conference on Reconfigurable Computing and FPGAs*, p. 74–79. IEEE.
- Twaha, S. e Ramli, M. A. (2018). A review of optimization approaches for hybrid distributed energy generation systems: Off-grid and grid-connected systems. *Sustainable Cities and Society*, 41:320–331.