Applied Soft Computing 70 (2018) 195–207

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Hosting capacity of low-voltage grids for distributed generation: Classification by means of machine learning techniques

Sebastian Breker[a], Jan Rentmeister[b], Bernhard Sick[b], Martin Braun[c]

[a] Asset Management of the DSO EnergieNetz Mitte GmbH, Kassel, Germany
[b] Department of Electrical Engineering and Computer Science, Intelligent Embedded Systems, University of Kassel, Germany
[c] Department of Electrical Engineering and Computer Science, Energy Management and Power System Operation, University of Kassel and Fraunhofer Institute for Energy Economics and Energy System Technology (IEE), Germany

ARTICLE INFO

Article history:
Received 17 May 2017
Received in revised form 29 January 2018
Accepted 2 May 2018

Keywords:
Low-voltage grids
Distribution grid planning
Grid classification
Distributed generation
Machine learning
Support vector machines

ABSTRACT

A high amount of installed distributed generators (DG) in low-voltage grids, e.g. photovoltaic generators (PV), may cause serious problems due to overloading of electrical equipment and violation of voltage limits. The assessment of low-voltage grids regarding their hosting capacity for the installation of DG is a difficult task, because grid structures may be diverse and complex. In this article, we classify grids by means of machine learning techniques, in particular support vector machines (SVM). SVM learn to assess grids by means of sample data, that is, grids represented by characteristic features that were assessed by human domain experts (i.e., distribution system operators (DSO) staff). We show that this approach can significantly better reflect domain expert assessments compared to a technique we proposed earlier which is based on a stochastic load flow simulation procedure and a subsequent parametric stochastic model estimation. One key result of this article is that SVM with grid based features significantly outperform SVM using features from load flow simulations regarding the classification accuracy if both are trained with data that were assessed (labeled) by DSO staff. Experiments are based on data for 300 real rural and suburban low-voltage grids.

1. Introduction

Low-voltage grid structures may be very diverse and complex. Hence, the assessment of low-voltage grids regarding their hosting capacity for distributed generators (DG) is a difficult task. In the past decade, with the beginning of the transition of our energy system towards renewable energies, more and more DG—in particular photovoltaic (PV) generators—were connected to low-voltage grids [1]. Without appropriate measures this may cause serious problems in some grids due to overloading of electrical equipment and violation of voltage limits. The (automated) categorization of low-voltage grids regarding their hosting capacity for DG may support the responsible distribution system operator (DSO) in her/his decisions regarding specific enhancement and sustainable development of grid structures. In practice, an automatic classification of low-voltage grids reduces much effort for the DSO because, e.g., exact grid modeling and computation for the installation of DG is only needed for grids with a potentially weaker structure.

In this article, we present an approach for grid classification based on machine learning (ML) techniques, in particular support vector machines (SVM). This work extends own work where low-voltage grids are classified by means of stochastic simulations and the results are compared to domain expert assessments [2]. We will show here that SVM are able to remarkably better reflect the domain expert assessments than this simulation based approach. With regard to the German energy legal framework, the evaluation of grid strength becomes an important part for future grid planning. In Germany, private investors can build distributed generators, e.g., photovoltaic generators, on their roofs and obtain funding. The DSO is committed to integrate these distributed generators into the grid and to realize grid enhancement if needed.

The use of ML techniques requires sample data from which these techniques learn to cope with the classification problem at hand. For a supervised training of a classifier, the sample data must consist of sample inputs together with the corresponding sample outputs. The trained classifier can then be applied to new data, i.e., it is able to classify grids that were not part of the training data set (generalization). Basically, three main questions have to be answered:

(1) Which ML paradigm (e.g., SVM or other classifier types) shall be used for the classification problem?

E-mail addresses: Sebastian.Breker@EnergieNetz-Mitte.de (S. Breker), jan.rentmeister@gmail.com (J. Rentmeister), bsick@uni-kassel.de (B. Sick), martin.braun@uni-kassel.de (M. Braun).

https://doi.org/10.1016/j.asoc.2018.05.007
1568-4946/

(2) Which input information (i.e., the features that represent the characteristics of low-voltage grids) shall be provided to the classifier?
(3) Which output categories (classes that are meaningful to the problem) shall be defined and how can output sample data (class labels) be gathered for the available input (e.g., from DSO staff)?

We will address these questions and many more that arise from a practical perspective, e.g.: How many grids are needed to train a classifier? Which input information to the classifiers describing characteristics of a grid is most important? How many human domain experts are needed to provide class labels? How can the training algorithm of SVM be easily parametrized? We will try to answer these questions by means of many experiments based on data for 300 real rural and suburban low-voltage grids with corresponding class labels provided by five domain experts from a DSO.

In a nutshell, the highlights of this article are:

- the classification of electrical grids with regard to their hosting capacity,
- the use of support vector machines (SVM) in power systems applications,
- a new kind of decision support for distribution system operators,
- evaluation with data from 300 real grids and five domain experts, and
- the result that SVM outperform other machine learning and conventional techniques.

The remainder of the article is organized as follows: Section 2 gives a summary of the state-of-the-art regarding the application of ML techniques in the context of energy systems. Section 3 describes the methodical foundations of our approach: the ML classifiers we use, the data considered as possible inputs, and the way we gathered output data (assessment of grids by human domain experts) to train the classifiers in a supervised way. Section 4 presents the results of numerous experiments focusing not only on classification accuracy but also on practical applicability of the proposed approach. In Section 5, the key findings are summarized and future research activities are sketched.

2. State-of-the-art

In [2], we already analyzed the state-of-the-art in low-voltage grid classification in detail. Grid classification is often conducted with conventional classification techniques rather than machine learning (ML) methods. Thus, we focus on ML applications in the field of power systems here.

In the field of power systems, the application of ML techniques is often related to smart grids [3–5]. In [3], a framework for a future integration of ML techniques into smart grids is presented. The focus is on an enhancement of data processing as well as a better reliability of the grid while the amount of distributed generators (DG) and the number of electric vehicles increase. A similar prospect to the potential of machine learning techniques in smart grids can be found in [4]. Some methods are presented and their potentials are investigated by means of simulation results. Additional application fields for ML techniques in future grids are proposed in [5,6]. Furthermore, approaches for dynamic pricing in smart grids are investigated in [7].

An important aspect of future grid operation is the classification of system states and failures. The classification and location of faults on transmission lines is addressed using support vector machines (SVM) and a wavelet transform in [8]. Zhang et al. present an approach to make grid protection more intelligent and adaptive by an online classification of grid operation conditions using SVM [9]. In [10], a new method for grid state classification based on supervised learning is proposed which applies adaptive boosting as well as classification and regression trees (CART). Two new approaches to detect and classify grid failures on selective lines are developed in [11]. This is done with the help of data acquisition in the transient phase. Haruna et al. use ML techniques for load shedding [12].

Several applications of ML techniques in the field of power systems address forecasting applications. Forecasting power generation and/or load are important for network operation, trading, and network planning. An overview of load forecasting techniques is given in [13,14]. Examples for different approaches can be found in [15] (short-term load forecasting using support vector regression), [16] (hyper-spherical ARTMAP for load forecasting), or [17] (deep learning techniques). Power generation forecasting for wind farms or photovoltaic systems is also done with ML techniques, e.g., deep learning or ensemble techniques [18,19]. The problem of electricity price forecasting is solved, e.g., in [20] by means of dynamic choice artificial neural networks.

The potential to enhance grid planning and operation with ML techniques is shown in the following articles: Stark and Krost develop a design tool for micro grids with the help of artificial neural networks, an expert system, and a fuzzy system [21]. An approach to estimate the failure probability of lines in a city grid is developed in [22]. For this purpose, the lines are ordered by an SVM using aggregate physical data and time series with regard to the line operation. The validation is done with data from a real grid in New York. This approach is further improved by Rudin et al. who develop a general process based on the previously gathered results which allows failure prediction and preventive maintenance in grid operation [23].

Niknam et al. propose the use of particle swarm optimization (PSO) algorithms to optimize the operation management of fuel cell power plants in the distribution network [24]. PSO is also suggested by Aghaei et al. in [25] for distribution expansion planning. Here, several aspects such as costs, active power losses, or voltage stability are considered. The approach is extended further in [26].

The application of ML techniques, especially SVM, as proposed in this article aims on an assessment of real grids similar to the methods presented in [23,22]. In contrast to [23,22], we do not only consider single grid equipments, e.g. lines, but assess the whole structure of a low-voltage grid with regard to the hosting capacity for distributed generators (DG) by an ordinal classification. For this purpose, we use grid information, information derived from an analysis of the graph underlying a grid, and information obtained by means of stochastic simulations, as presented in [2]. In our scope was the provision of a methodology with high practical relevance regarding its application to a large amount of real low-voltage grids. Thus, our work is also strongly related to and can be seen as a refinement of the investigations done by Kerber and Witzmann in [27] and Dickert et al. in [28]. They present approaches for the distinction of distribution grids by an analysis of grid parameters which are of practical relevance (e.g., distribution of rated transformer power). In contrast to our approach they do not classify the grids according to their hosting capacity for DG, but investigate only the parameters of a lot of real grid structures.

Altogether, the way we use ML techniques to assess low-voltage grids can be regarded as new.

3. Methodical foundations

In this Section, we first describe the ML technique used to classify the grids. Then we outline the (possible) inputs of this classifier