2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

Fault Recognition in Smart Grids by a One-Class Classification Approach

Enrico De Santis, Lorenzo Livi, Fabio Massimo Frattale Mascioli, Alireza Sadeghian, and Antonello Rizzi

***Abstract*— Due to the intrinsic complexity of real-world power distribution lines, which are highly non-linear and time-varying systems, modeling and predicting a general fault instance is a very challenging task. Power outages can be experienced as a consequence of a multitude of causes, such as damage of some physical components or grid overloads. Smart grids are equipped with sensors that enable continuous monitoring of the grid status, hence allowing the realization of control systems related to different optimization tasks, which can be effectively faced by Computational Intelligence techniques. This paper deals with the problem of faults modeling and recognition in a real-world smart grid, located in the city of Rome, Italy. It is proposed a suitable classication system able to recognize faults on medium voltage feeders. Due to the nature of the available data, the one-class classication framework is adopted. Experiments are presented and discussed considering a three-year period of measurements of fault events gathered by ACEA Distribuzione S.p.A., the company that manages the smart grid system under analysis. Results demonstrate the effectiveness and validity of our approach.**

I. INTRODUCTION

THERE are many definitions for a Smart Grid (SG). The SG European Technology Platform defines a SG as an "electricity network that can intelligently integrate the actions of all the connected users – generators, consumers and those that do both, in order to efficiently deliver sustainable economic and secure electricity supply" [1]. A SG employs innovative products and services together with intelligent monitoring, control, communication, and self-healing technologies in order to:

- better facilitate the connection and operation of generators of all sizes and technologies;
- allow consumers to play an active part in optimizing the operation of the system;
- significantly reduce the environmental impact of the whole electricity supply system;
- preserve or improve the level of system reliability, quality, and security;
- efficiently maintain and improve the existing services.

SGs can be considered as an "evolution" rather than a "revolution" of the existing energy networks [2]. The evolution is leaded by the symbiotic exchange between power grid technologies and the Information and Communication Technologies (ICT). ICT provide instruments, such as *Smart Sensors* (SS) to monitor the network status, wired and wireless communication network to collect and transport data, and powerful computational architectures to define control actions. A SG can be framed, both in the applicative context and in a theoretical framework, as a complex non-linear and time-varying system [3], [4], where heterogeneous elements, including environmental factors, are extremely interconnected through the exchange of both energy and information.

Computational Intelligence (CI) techniques offer a solid framework for "injecting intelligence" into the power network [5], providing to the system the capability of monitoring, decision making, and adaptation. Well-known CI techniques adopted in the SG context include approximate dynamic programming [6], neural networks and fuzzy inference systems for prediction and control [7], swarm intelligence and evolutionary computation for optimization [8], [9].

An important key issue in SGs is the Decision Support System (DSS), which is an expert system that provides decision support for the commanding and dispatching system of the power grid. It is usually meant for providing forecasting, early warnings on malfunctions, and also for autonomous decisions. Such a system analyzes the risk for damage of crucial equipments, assesses the power grid security, forecasts and provides warnings about the magnitude and location of possible faults, and timely broadcasts the early-warning signals through suitable communication networks [4]. The information provided by the DSS can be used for Condition Based Maintenance (CBM) in the power grid [10]. CBM is defined as "a philosophy that posits repair or replacement decisions on the current or future condition of assets". The objective of CBM is thus to minimize the total cost of inspection and repair by collecting and interpreting (heterogeneous) data related to the operating condition of critical components.

Collecting heterogeneous measurements is of paramount importance. As an instance, the available measurements can be used for dealing with various important pattern recognition and data mining problems on SGs, such as event classification [11]. On the basis of the specific data type, different problem types could be formulated. In [12] authors have established a relationship between environmental features and fault causes. A fault cause classifier based on the linear discriminant analysis (LDA) is propose in [13]. Information regarding weather conditions, longitude-latitude information, and measurements of physical quantities (e.g., currents and voltages) related to the power grid have been taken into account. In [14], the authors proposed a system

Enrico De Santis, F.M. Frattale Mascioli and Antonello Rizzi are with the Department of Information Engineering, Electronics, and Telecommunications, "Sapienza" University of Rome, Via Eudossiana 18, 00184 Rome. (email: {enrico.desantis, antonello.rizzi}@uniroma1.it, mascioli@infocom.uniroma1.it). Lorenzo Livi and Alireza Sadeghian are with the Dept. of Computer Science, Ryerson University, Toronto, ON M5B 2K3, Canada. (email: llivi@scs.ryerson.ca, asadeghi@ryerson.ca).

978-1-4799-1484-5/14/$31.00 ©2014 IEEE

based on LDA, which processes Phasor Measurement Unit data, with the aim of recognizing and locating faults on power lines. As concerns fault diagnosis in power grids, in [15] is proposed a Support Vector Machine (SVM) based method to perform the recognition of faults related to high-voltage transmission lines. The One-Class Quarter-Sphere SVM algorithm is proposed for faults classification in the power grid [16]. The reported experimental evaluation is, however, performed on synthetically generated data only. In [17] is proposed a genetic algorithm for the optimization of a neurofuzzy system used to predict faults in cables and accessories.

This paper presents a study on the problem of modeling and recognizing fault instances in the real-world SG system of ACEA. The considered SG feeds the entire city of Rome, Italy. The problem is faced by a combined approach involving dissimilarity measures and the so-called one-class classification paradigm [18]. In Sec. II it is described the considered SG system and related data. Sec. III provides the details of the designed recognition system. Results of tests are shown in Sec. IV, and finally in Sec. V conclusions are drawn.

II. The Considered Smart Grid System

The ACEA power grid is the electrical distribution grid of Rome, the capital city of Italy. It is constituted of backbones of uniform section exerting radially with the possibility of counter-supply if a branch is out of order. Each backbone of the power grid is supplied by two distinct Primary Stations (PS) and each half-line is protected against faults through the breakers. Along each line, there is a breaking point whose position is chosen with respect to the preassigned constraint of the electric current that flows in each half-line. The Medium Voltage (MV) power grid consists in lines (MV feeders) in which the nominal voltage is 20 kV, with the presence of few "legacy" lines that still work at 8.4 kV. The MV part of the network covers 10,490 km, while the Low Voltage (LV) section covers 11,120 km. Cables can be on air or underground and their sections can vary along the backbone with the presence of bottlenecks. The MV section has 1,565 lines in service and it is supplied with 76 PSs, while LV section is supplied with 13.292 Secondary Stations (SS).

This study deals with the problem of modeling and recognizing a particular type of MV grid fault, which is commonly termed as Localized Fault (LF) [19]. Before providing a precise definition of a LF, it is important to discriminate, according to the CEI 5160 normative [20], among *outages* and *faults*. An outage (i.e., an interruption of the service) is the condition in which the voltage on the access point to the electrical energy of a user is less than 5% of the declared voltage on all phases of supply [19]. Three types of outage are considered according to their duration:

- long, if the duration is more than three minutes (long outages);
- short, if the duration is more than one second and less than three minutes (short outages);
- transient, if the duration does not exceed one second (transient outages).

A fault instead is related to the failure of the electrical insulation (e.g., cables insulation) that compromises the correct functioning of (part of) the grid. Therefore, a LF is effectively a fault in which a physical element of the grid is permanently damaged causing long outages.

III. The Proposed One-class Classification System for Smart Grid Fault Detection

A. Representation of Fault Patterns

Instances of Fault Patterns (FP) describing LFs occurred in the SG have been elaborated from a historical database provided by ACEA. The considered period spans across 2009–2012. Faults on MV feeders are characterized by heterogeneous data, including weather conditions, spatio-temporal data (i.e., longitude-latitude pairs and time), physical data related to the state of power grid and its electric equipments (e.g., measured currents and voltages), and finally meteorological data. As a consequence, a FP is effectively defined by features of various types, containing categorical (nominal) data, quantitative data (i.e., data belonging to a normed space), and also Time Series (TS) describing the sequence of short outages occurred before a LF. A detailed description of the considered features characterizing a FP is provided in Tab. I.

B. Data Preprocessing

Data normalization is a well-known important aspect in pattern analysis, which becomes even more crucial when processing patterns characterized by many heterogeneous features. Data have been normalized using the affine normalization technique:

$$v = \frac{c - m}{(M - m)} \in [0, 1]. \tag{1}$$

where c is the original (non-normalized) value; m and M are, respectively, the minimum and maximum values that the considered feature can assume.

1) Temporal Data: The "Day start" and "Time Start" features (Tab. I) have been encoded as integer values. The former ranges in $\{0, 1, ..., 364\}$, and for the latter in $\{0, 1, ..., 1439\}$, which corresponds to the number of minutes in a year. Normalization of those data follows straightforwardly.

2) Spatial Data: Three types of information regarding the geographical position of a LF are available: the absolute position of the PS feeding the element where the LF has occurred, and the absolute position of the two SSs delimiting the section of power line where the revealing system detected the LF. The original coordinates of the geographical position of the LF have been expressed in WGS84 (decimal degrees), the same that it is used in the GPS geolocalization system. Those three features (the PS and the two SS positions) have been reduced to two features: (i) the distance between the SS location ("Primary Station fault distance") and (ii) the

TABLE I: List of the considered features describing a FP instance.

Feature	Data type and features space label	Description
Day start	Quantitative (Integer) $\mathcal{F}_D$	Day in which the LF was detected
Time start	Quantitative (Integer) $\mathcal{F}_T$	Time stamp (minutes) in which the LF was detected
Primary Station (PS) code **Protection tripped** **Voltage line** **Location element** **Material**	Categorical (String) $\mathcal{F}_C$	Unique backbone identifier Type of intervention of the protective device Nominal voltage of the backbone Element positioning (aerial or underground) Constituent material element (CU, AL)
Primary station fault distance	Quantitative (Real) $\mathcal{F}_1^Q$	Distance between the primary station and the geographical location of the LF
Median point	Quantitative (Real) $\mathcal{F}_2^Q$	Fault location calculated as median point between two secondary stations
# Secondary Stations (SS)	Quantitative (Integer) $\mathcal{F}_3^Q$	Number of out of service secondary stations after the LF
Current out of bounds	Quantitative (Integer) $\mathcal{F}_4^Q$	The maximum operating current of the backbone is less than or equal to 60% of the threshold "out of bounds", typically established at 90% of capacity
Max. temperature	Quantitative (Real) $\mathcal{F}_5^Q$	Maximum registered temperature
Min. temperature	Quantitative (Real) $\mathcal{F}_6^Q$	Minimum registered temperature
Delta temperature	Quantitative (Real) $\mathcal{F}_7^Q$	Difference between the maximum and minimum temperature
Rain	Quantitative (Real) $\mathcal{F}_8^Q$	Millimeters of rainfall in a period of 24 hours preceding the LF
Cable section	Quantitative (Real) $\mathcal{F}^S$	Section of the cable, if applicable
Backbone Electric Current	Quantitative (Real) $\mathcal{F}^{EC}$	Extracted feature from Time Series of electric current values that flows in a given backbone of the considered power grid
Interruption (breaker)	TS (Integer) $\mathcal{F}_1^{TS}$	Sequence of opening events of the *breakers* in the primary station
Petersen alarms	TS (Integer) $\mathcal{F}_2^{TS}$	Sequence of alarms detected by the device called "Petersen's coil" due to loss of electrical insulation on the power line
Saving intervention	TS (Integer) $\mathcal{F}_3^{TS}$	Sequences of decisive interventions of the Petersen's coil which have prevented the LF

middle point among the two SSs ("Median point"); see Fig. 1.

The maximum spatial resolution of the geographical localization of the LF is therefore defined by the two SS positions. The preprocessing of original data has been operated considering that the output power line from the PS has a radial structure.

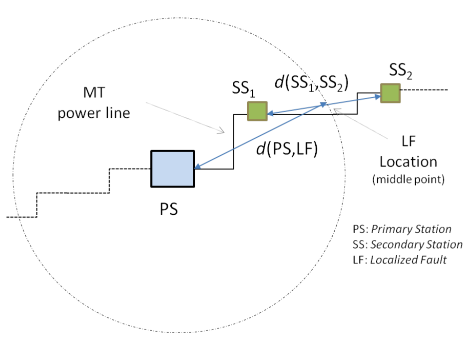


Fig. 1: Radial structure of the output power line from the PS and distance calculation in order to reduce the number of features.

The distance between two geographical locations is calculated through the Vincenty's algorithm [21]. The normalization process of the position data is based on the calculation of the minimum size rectangle that includes all PSs and SSs. Hence, applying the affine normalization (1), the spatial positions of the LFs result normalized in $[0, 1]$. The affine normalization is applied also for the distance values among PSs and the positions of the LFs.

3) Physical Data: The data describing the physical power grid is defined by both categorical and quantitative information. As concerns categorical data ("Location element"), the analyzed dataset has few missing values – less than 5%. However, the missing values of a feature have been substituted with the most frequent category for that feature. The normalization of quantitative data (i.e., "# Secondary station", "Current out of bounds", and "Cable Section") is implemented by means of (1).

It is well known that a possible cause of faults in distribution systems affecting cables and joints is the abrupt change in the current loads, more than the actual amount of electrical power flowing in the devices. To define a feature taking this effect into account, current measures sampled every 10 minutes have been considered in a main window of 24 hours before the LF occurrence. This time window is divided in two non overlapping sub-windows, w1 and w2, each of 12 hours. The feature "Backbone Electric Current" is finally computed as the absolute difference between the average of current values in each sub-window. That values are normalized with respect to the available minimum and maximum values recorded in the considered backbone.

4) Meteorological Data: The meteorological data are acquired by suitable stations located in different areas of Rome. The "Rain" feature is calculated as the average millimeters of rain observed in the 24 hours antecedent the LF.

5) Short Outages Sequences (SOSs): Short Outages Events are automatically registered by the protection systems as soon as they occur (asynchronous recording); as a consequence, representing this information as an ordered set of events in time gives rise to a sparse record (the time series is not uniformly sampled). The considered TS events are the Interruption (breaker), the Petersen Alarms, and the Saving Intervention (see Tab. I) occurred in a time window of three months before the LF. The short outages events are represented as variable-length sequences, which contain the temporal distances (expressed in seconds) from the LF event, where it is positioned the reference time (see Fig. 2). A SOS S^i of K outage events is defined as follows:

$$S^i = \left[\xi_1^i, \xi_2^i, ..., \xi_{K_{i(n)}}^i\right], \tag{2}$$

where ξ is the temporal distance from the LF event considered as the origin, $i \in \{1, 2, 3\}$ is an integer code distinguishing the three aforementioned types of outages, and $K_{i(n)}$ is the number of events for the i-th type of outage (that depends on the n-th pattern).

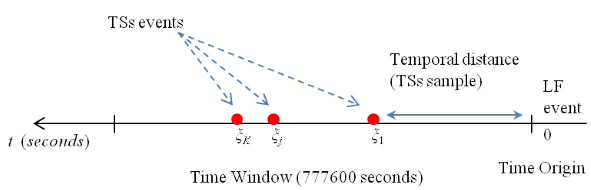


Fig. 2: Representation of SOSs of outages recorded before the occurrence of a LF.

Normalization is computed in two steps: (i) the integer values are first transformed in real values in the range $[0, 1]$ by means of (1) where M here is the maximum length of the considered time window in seconds, (ii) let us consider a given dissimilarity measure between SOSs (see Sec. III-C.1.c). First the dissimilarity matrix D, i.e a matrix containing the pairwise dissimilarity values between SOSs, is computed considering all the short outages sequences available. Hence a dissimilarity measure between any given pair of SOSs is normalized dividing its value by the maximum entry in D.

C. The Proposed One-class Classifier

As a consequence of the difficulty of modeling useful instances of non-faults in the considered SG, we designed an OCC for the purpose of recognizing LFs only. Such a goal is implemented by building a one-class classifier relying on clustering techniques. The idea of using a cluster for modeling a region of the idealized "fault space", $\mathcal{F}$, containing target patterns denoting LFs, is reasonable and also intuitive. The underlying assumption is that similar status of the SG have similar chances of generating a LF, assumption that it is reflected by the cluster model.

Given a dataset of FPs, it is partitioned in k (disjoint) clusters, where each cluster contains faults having similar features. As a consequence, the most important component of the OCC system is the core dissimilarity measure $d : \mathcal{F} \times \mathcal{F} \rightarrow \mathbb{R}^+$, which assigns dissimilarity values to a pair of FPs. The partition, as well as other parameters that will be described in the following, constitute the model of the OCC.

A (one-class) classification problem instance is defined as a triple of disjoint sets, namely training set ($\mathcal{S}_{tr}$), validation set ($\mathcal{S}_{vs}$), and test set ($\mathcal{S}_{ts}$), all containing FP instances. Given a specific parameters setting, a classification model instance is synthesized on $\mathcal{S}_{tr}$ and it is validated on $\mathcal{S}_{vs}$. Finally, performance measures are computed on $\mathcal{S}_{ts}$. The OCC system parameters defining the model are optimized by means of a standard Genetic Algorithm (GA), which is guided by a suitable performance measure that includes as the most important factor the classification accuracy achieved on $\mathcal{S}_{vs}$.

1) The Dissimilarity Measure Among FPs: A FP $x \in \mathcal{S}$ is described as:

$$x = \{\mathbf{F}_1, \mathbf{F}_2, ..., \mathbf{F}_m\}, \tag{3}$$

where the l-th feature, $\mathbf{F}_l$, $1 \leq l \leq m$, lies in its specific feature space $\mathcal{F}_l$. Hence, each x lies on the m-fold product feature space $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2 \times ... \times \mathcal{F}_m$.

Given two FPs $x, y \in \mathcal{S} \subset \mathcal{F}$, the general formulation of the proposed weighted dissimilarity measure reads as:

$$d(x, y; \mathbf{W}) = \|x \ominus y; \mathbf{W}\| = \sqrt{(x \ominus y)^T \mathbf{W}^T \mathbf{W} (x \ominus y)}, \tag{4}$$

where T denote the transpose operator. The $\ominus$ operator represents a generic dissimilarity measure among features. In this paper, $\mathbf{W}$ is restricted to be a $m \times m$ diagonal matrix, whose diagonal is used as a weight vector $\underline{\mathbf{w}} = (w_1, w_2, ..., w_m) \in [0, 1]^m$. This results in the following expression:

$$d(x, y; \underline{\mathbf{w}}) = \sqrt{\sum_{j=1}^{m} (w_j\,(x_j \ominus y_j))^2} = \sqrt{\sum_{j=1}^{m} (w_j \times d_j)^2}, \tag{5}$$

which corresponds to the weighted l_2 norm of the vector of dissimilarity values ($d_j = x_j \ominus y_j$).

The weights $\underline{\mathbf{w}}$ are optimized during the training phase of the OCC by means of a GA. The dissimilarity measure (5) allows to compute the overall dissimilarity value between two FPs by combining different dedicated dissimilarity measures for each specific data type.

The following paragraphs describes how it is computed the dissimilarity between two FP components according to their type. The specific nature of the feature, $\mathcal{F}_i$, will be specified using the same notation of Tab. I.

a) Categorical Data: Categorical attributes, also referred to as nominal attributes, are attributes without a semantically valid ordering (see Tab. I for the data treated as nominal). Let $\mathcal{F}^c = \{\eta_1, \eta_2, ..., \eta_n\}$ be the set of all the categorical features of the entire data set, each described by d categorical attributes: $\nu_1, \nu_2, ..., \nu_d$. Let us define the domain of the attribute ν_j, $\mathrm{DOM}(\nu_j) = \{A_{j_1}, A_{j_2}, ..., A_{j_{n(j)}}\}$, where A_{j_l} $(1 \leqslant l \leqslant n(j))$ is the set of possible values for the categorical attribute ν_j, and $n(j)$ its cardinality. Let us consider the well-known *simple matching* distance, defined as follows:

$$\delta(x,y) = \begin{cases} 0 & x = y, \\ 1 & x \neq y. \end{cases} \tag{6}$$

Let x^c and y^c be the projections on the categorical feature space $\mathcal{F}^c$ of two generic patterns x, y. The dissimilarity measure between the two categorical objects described by d categorical attributes is implemented as:

$$d^c(x^c, y^c) = \frac{1}{d}\sum_{j=1}^{d} \delta(x^c_j, y^c_j). \tag{7}$$

b) Quantitative Data: As concerns the quantitative data (see Tab. I) we distinguish between (i) "Normal" quantitative data and (ii) "Special" quantitative data. The former type includes both numerical and integer values (normalized in $[0, 1]$) and the generic difference operator $\ominus$ is implemented by the absolute difference: $d^N = |x - y|$.

As concerns integer values that rely to temporal information, such as the day in which the LF happened and the timestamp within that day, it is defined a particular dissimilarity measure implementing $\ominus$, called *circular difference*. Given an ordered set of integer numbers $\{0, 1, ..., a\}$, the circular difference among any x, y in this set is defined as:

$$d^{CD}(x, y; a) = \min(|x - y|, a - |x - y|). \tag{8}$$

For "Day start" and "Time start", which are referred to the $\mathcal{F}^D$ and $\mathcal{F}^T$ features subspaces, the maximum value for a in (8) is 364 and 1439, respectively. The implementation of the circular difference is designed to avoid that pairs of close days or timestamps give raise to high values of the dissimilarity function.

"Special" quantitative data are normalized in the range $[0, 1]$, but can assume also a special symbol ϵ, indicating the "not applicable" condition. It is the case for the "Cable section" feature, since for LFs not related to cables is undefined. The dissimilarity measure $d^S : \{[0, 1] \cup \epsilon\} \times \{[0, 1] \cup \epsilon\} \to [0, 1]$ is defined as follows. Given two special quantitative values $x, y \in \mathcal{F}^S$, we have:

$$d^S = \begin{cases} |x - y| & (x \neq \epsilon \wedge y \neq \epsilon), \\ 1 & (x = \epsilon \vee y = \epsilon), \\ 0 & (x = \epsilon \wedge y = \epsilon). \end{cases} \tag{9}$$

c) Time Series Data: Dynamic time warping (DTW) is a well-known technique to find an optimal alignment between two sequences of variable length. The use of the DTW as dissimilarity measure for sequences of generic objects is increased considerably in many applications, such as biology, finance, multimedia, and image analysis [22], [23]. An in-depth description of DTW can be found in [24].

Following the notation introduced in Sec. III-B.5, the data set consists in three types of TSs, $S^i, (i = \{1, 2, 3\})$. Each one represents a vector belonging to the TSs feature vector subspace, $\mathcal{F}^{TS}_i$ $(i = 1, 2, 3)$, and thus we compute a DTW dissimilarity measure between them, separately.

2) Model Definition and Testing of the Classifier: The most important part of the OCC model is the partition P of $\mathcal{S}_{tr}$, which is obtained through a clustering algorithm – see Fig. 3 for an overview.

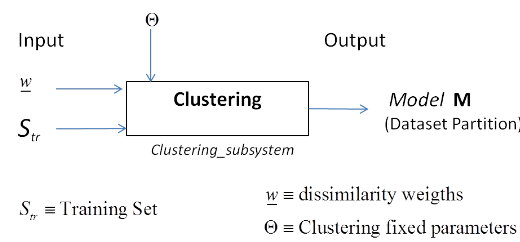


Fig. 3: Model synthesis of the OCC performed by a cluster analysis of the training set $\mathcal{S}_{tr}$.

A hard partition of order k is a collection of k disjoint and non-empty clusters, $P = \{\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_k\}$. Each cluster $\mathcal{C}_i \in P$ is synthetically described by a *representative* element, denoted as $c_i = R(\mathcal{C}_i)$; accordingly, with $R(P) = \{c_1, c_2, ..., c_k\}$ it is denoted the set of representatives of the partition P. The representative of a cluster $\mathcal{C}$ is computed in two ways that will be compared. The first one relies on the computation of the component-wise mean for the numerical values and mode for categorical values [25], while the second one relies on the MinSOD computation [26], that is, the element $\nu \in \mathcal{C}$ that minimizes the sum of distances:

$$\nu = \arg\min_{F_j \in \mathcal{C}} \sum_{i=1}^{|\mathcal{C}|} d(F_j, F_i). \tag{10}$$

A cluster representative c_i can be considered as a prototype of a *typical fault scenario* individuated in $\mathcal{S}_{tr}$.

3) The Classifier Decision Rule: The information provided by the cluster $\mathcal{C}_i$ as a whole is useful to conceive a region of the pattern space "around" c_i, which describes similar fault scenarios. By defining $\delta(\mathcal{C}_i) \geq 0$ as a measure of "cluster extent", it is constructed the decision region associated to the cluster $\mathcal{C}_j$, used to implement the classification rule. The cluster extent can be computed as the average/maximum intra-cluster dissimilarity values or by considering their standard deviation, for instance. However, in addition to $\delta(\mathcal{C}_i)$ it is considered also a tolerance parameter, $\sigma_i \geq 0$, aiming to extend the decision region. The decision region of the cluster $\mathcal{C}_i$ is thus defined by the quantity $B(\mathcal{C}_i) = \delta(\mathcal{C}_i) + \sigma_i$. This choice is motivated by the difficulty of defining precise decision regions in an OCC problem, due to the absence of patterns describing meaningful non-target instances.

Fig. 4 provides a schematic overview of a cluster model and its use in the process of classifying a test pattern $\bar{x}$. The classification rule for a test pattern $\bar{x}$ operates in two stages. First, the nearest cluster representative $c_i^* \in R(P)$ is individuated according to the following expression:

$$c_i^* = \arg\min_{c_j \in R(P)} d(\bar{x}, c_j). \tag{11}$$

The second step consists in comparing the dissimilarity value $d(\bar{x}, c_i^*)$ with $B(\mathcal{C}_i)$. To this end, it is defined a binary-valued function $f(\cdot)$ that performs hard classification:

$$f(x) = \begin{cases} 1 & \text{if } d(\bar{x}, c_j^*) \leq B(\mathcal{C}_i), \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

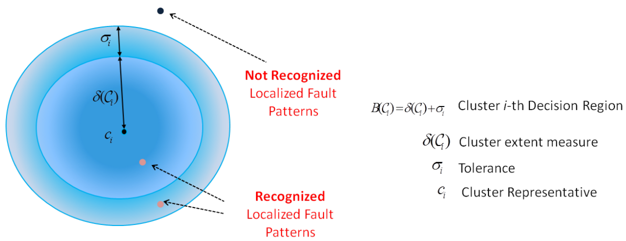


Fig. 4: Cluster decision region and its parameters.

4) *Training of the Classifier:* It is proposed a learning strategy to synthesize the OCC model that is based on the well-known k-means [27]. The number of clusters is fixed a priori by setting an integer valued parameter, k. The dissimilarity measure described in Sec. III-C.1 depends on a vector of weights $\underline{\mathbf{w}}$; moreover, the cluster model (III-C.2) is based on the thresholds σ_i. Setting those parameters is of utmost importance, and of course it has a significant influence on the results produced by k-means. Therefore, such parameters $p_j = [\underline{\mathbf{w}}_j, \underline{\sigma}_j]$ are optimized by means of a cross-validation technique. The GA minimizes the following objective function:

$$f(p_j) = \alpha ER(\mathcal{S}_{vs}) + (1-\alpha) \sum_{i=1}^{k} \sigma_i. \tag{13}$$

In (13) $ER(\mathcal{S}_{vs})$ is the recognition rate on $\mathcal{S}_{vs}$ (i.e., the fraction of misclassified validation patterns). The GA is in charge to find the parameters setting, p_j, that minimizes the l_1 norm of the tolerances, while at the same time minimizing the number of errors. Fig. 5 shows a diagram describing the optimization stage.

As concerns the learning phase, it is well-known that the k-means algorithm is sensible to the adopted cluster initialization strategy; in particular, here it is used the randomized initialization. As a consequence, the current version of the OCC takes as external parameter the k value and a classification model is synthesized for each k in a given user-defined range $k_{\min}, k_{\max}$. For each k in this range, three models are synthesized, considering different random initializations of cluster's representatives. The fitness associated to a candidate solution, p_j, is hence the average of the objective function values (13) computed with respect to each model. In the test phase, the models are used in "a majority" voting scheme to reach a final decision (hard decision).

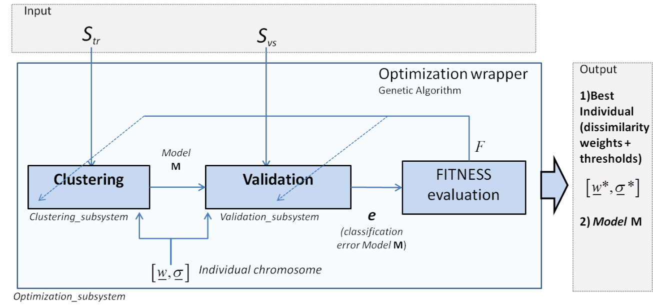


Fig. 5: Block diagram depicting the optimized classification model synthesis.

It is worth pointing out that it has been conceived also a soft decision scheme, which provides a way to give also a measure of reliability of the classification. Such a perspective is not exploited in this paper.

IV. Experimental Evaluation

In this section are described the experiments performed on synthetic and ACEA datasets (see Tab. II).

A. *System Benchmarking on Synthetic Data*

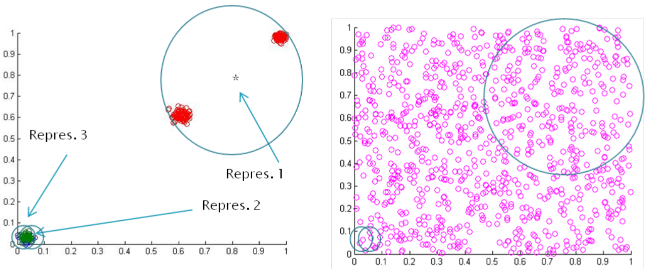
To first benchmark the OCC in a controlled setting, it has been prepared a synthetic dataset characterized by five numerical features. The experiments are conducted in two stages. The first stage ("true test") is based on training the classifier on a dataset with three well-formed clusters, characterized by a normal distribution of known mean and variance. The test phase is performed with a dataset that presents similar statistics. The second stage ("random sampling") is based on testing the model learned in the first stage on a dataset that is generated by a uniform random sampling of the whole domain space. The experiments highlight the intrinsic difficulties in evaluating the system where the behavior of power grid status is modeled only with positive instances (i.e., faults), trying to estimate the classifier capability in recognizing non-fault patterns, in order to minimize false alarms.

In Fig. 6 is represented the situation (a) in which the clustering model obtained with $k = 3$ reaches a suboptimal solution, and (b) the comparison between the extent of the learned decision regions and the "random sampling" distribution tested in the second stage. Choosing the representative in the middle, a cluster is split in two parts, while the other decision region is forced to embed the other two clusters. Consequently, the associated decision region is very large. In this situation, (Mod. N°3) during the "random sampling" test, about 15% of test patterns fall in the overall decision region. Conversely, the other learned models have an optimal partition (they do not split the clusters) achieving a low classification rate: 0.4% for the Mod. N°2 and and 0.2% for the Mod. N°1.

The majority voting scheme proves here its robustness: the result of the final classification rate over the "random

TABLE II: Description of the considered datasets and OCC setting.

	$\#S_{tr}$	$\#S_{vs}$	$\# S_{ts}$	Representative initialization	α	Cluster extent
Benchmark dataset	300	300	300	random	0.5	mean radius
ACEA S.p.A. dataset	532	470	178	random	0.2	mean radius



(a) "true test" dataset. (b) "random sampling" dataset.

Fig. 6: Scatter plot of the partition computed on the synthetic dataset (Mod. N°3) with $k = 3$ in the case of suboptimal clustering (the representative c_j is the mean for numerical values).

sample" dataset is 0.4%, showing the capability of the system in finding small and suitable decision regions.

Tab. III shows the performance comparison between two variants of the OCC system, relying on different ways to represent clusters (i.e., average–median and the MinSOD; see III-C.2) As concerns the use of MinSOD as representative, increasing the number of clusters (see Tab. III) the classification rate on the "random sample" is stable and low as we expect, and on the "true test" it denotes good performances.

B. Tests on the ACEA Data

As regards the tests performed on the ACEA dataset, we set the α weight of fitness function to 0.2, giving more importance to the minimization of the tolerances of the decision regions compared to the recognition rate of true LF. The cluster extent measure is computed as the average dissimilarity among the patterns and the representative of the cluster, while the threshold σ can be different for each cluster. The search range for the k parameter is $k_{\min} = 2$, $k_{\max} = 15$.

In Fig. 7 is shown the normalized classification rate for each of the three models versus the number of clusters (i.e., the k value). The best results, in term of classification rate, are achieved with $k^* = 7$ – see Tab. IV. For this k value, each partition shows good compactness and separability values, which are measured with the Davies–Bouldin index [28]. It is important underlying also that the tolerances are very small, denoting that the extent of the LF decision region is just a small fraction of the whole input domain. The second best result (data not shown) is obtained with a model formed by $k = 10$ clusters with 91.3% of test set patterns correctly classified. In this last case, performance degrades for what concerns compactness and separability (the lower, the better), where the value of the Davies–Bouldin index is 34.700, 92.094, and 95.922 for the three models (Mod. N°1, 2, and 3, respectively).

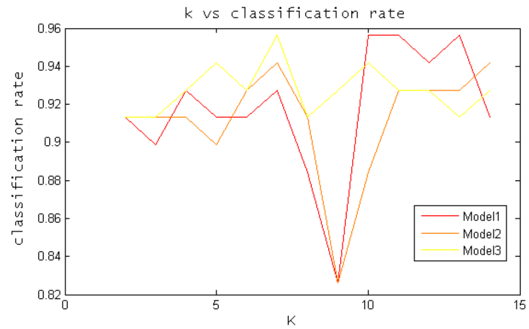


Fig. 7: Classification rate achieved during the OCC test phase (ACEA dataset).

V. Conclusions

In this study, it has been faced the problem of modeling and recognizing fault instances in a real-world SG. The considered SG feds the entire city of Rome, Italy. This work is not an end-point but rather a starting point for the development of a complex system for the management and control of faults in the ACEA power grid. From the computational viewpoint, it has been faced the problem by following the one-class classification framework. In fact, pattern instances denoting conditions of "non-fault" are not available for the dataset at hand. The designed one-class classifier has been conceived by clustering techniques, modeling the whole fault decision region as the union of elementary decision regions, each one defined by a cluster. Since the considered SG data is characterized by very heterogeneous features, it has been defined a suitable dissimilarity measure, which is in charge of providing effective dissimilarity values among the input fault patterns. Experiments have been carried out on both synthetic (controlled) and real-world data (ACEA). Results on synthetic data prove the effectiveness of the system in defining appropriate and compact decision regions. Test set results achieved on the ACEA dataset show an interesting generalization capability of the classification system in a real-world situation.

The designed classifier is modular, with the possibility to employ several different dissimilarity measures and clustering algorithms. This represents a next step in which we will evaluate a faster clustering algorithm which does not require the a priori definition of the partition order. So far, classification is performed by means of Boolean decision functions. An important improvement concerns the possibility to define a measure of classification reliability. Each decision region will be equipped with a suitable "membership function"

TABLE III: Comparison among MinSOD and mean cluster representative on the benchmarking dataset.

	MinSOD		mean representative	
k	Random Sampling	True $\mathbf{S_{ts}}$	Random Sampling	True $\mathbf{S_{ts}}$
2	0.9 %	61.3 %	11.2 %	100 %
3	0.1 %	97.6 %	0.4 %	99.6 %
4	0.1 %	98.3 %	0.7 %	100 %
5	0.1 %	96.6 %	1.2 %	99.8 %
6	0.1 %	97.0 %	1.3 %	100 %

TABLE IV: Summary test set results achieved with best k^* value for the ACEA dataset.

$k^* = 7$	**Mod. 1**	**Mod. 2**	**Mod. 3**
σ^*	0.0801, 0.0703, 0.0571, 0.0662, 0.0948, 0.0877, 0.0993		
$\delta(\mathcal{C}_i)$	0.2287, 0.2361, 0.2169, 0.2128, 0.2432, 0.2532, 0.2287	0.1884, 0.2106, 0.2268, 0.1998, 0.1982, 0.2325, 0.2749	0.2353, 0.2122, 0.2222, 0.2257, 0.2244, 0.2263, 0.2327
Davies–Bouldin index	21.684	17.333	16.648
Classification rate	92.7 %	94.2 %	95.6 %
Voting result	95,6 %		

that will be used to provide the user with an additional information regarding the decision. We stress the importance of this fact, which assumes paramount importance in the particular setting of one-class classification, where non-target pattern instances are not available during the training stage. Finally, future works will focus on the design of a non-fault patterns generator (possibly not trivial), exploiting the available data, in order to better evaluate the possibility to employ such a system in a CBM procedure.

References

[1] "The SmartGrids european technology platform | SmartGrids." [Online]. Available: http://www.smartgrids.eu/ETPSmartGrids

[2] "International energy outlook 2011 - energy information administration." [Online]. Available: http://www.eia.gov/forecasts/ieo/index.cfm

[3] G. K. Venayagamoorthy, "Potentials and promises of computational intelligence for smart grids," in *Power Energy Society General Meeting, 2009. PES '09. IEEE*, Jul. 2009, pp. 1–6.

[4] S. Mei, M. Cao, and X. Zhang, *Power grid complexity*. Springer, 2011.

[5] G. K. Venayagamoorthy, "Dynamic, stochastic, computational, and scalable technologies for smart grids," *IEEE Computational Intelligence Magazine*, vol. 6, no. 3, pp. 22–35, 2011.

[6] R. Anderson, A. Boulanger, W. Powell, and W. Scott, "Adaptive stochastic control for the smart grid," *Proceedings of the IEEE*, vol. 99, no. 6, pp. 1098–1115, 2011.

[7] A. Anvari Moghaddam and A. R. Seifi, "Study of forecasting renewable energies in smart grids using linear predictive filters and neural networks," *Renewable Power Generation, IET*, vol. 5, no. 6, pp. 470–480, 2011.

[8] E. De Santis, A. Rizzi, A. Sadeghian, and F. M. Frattale Mascioli, "Genetic optimization of a fuzzy control system for energy flow management in micro-grids," in *2013 Joint IFSA World Congress and NAFIPS Annual Meeting*. IEEE, 2013, pp. 418–423.

[9] G. L. Storti, F. Possemato, M. Paschero, A. Rizzi, and F. M. Frattale Mascioli, "Optimal distribution feeders configuration for active power losses minimization by genetic algorithms," in *2013 Joint IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS)*, 2013, pp. 407–412.

[10] D. Raheja, J. Llinas, R. Nagi, and C. Romanowski, "Data fusion/data mining-based architecture for condition-based maintenance," *International Journal of Production Research*, vol. 44, no. 14, pp. 2869–2887, Jul. 2006.

[11] M. Afzal and V. Pothamsetty, "Analytics for distributed smart grid sensing," in *Innovative Smart Grid Technologies (ISGT), 2012 IEEE PES*, 2012, pp. 1–7.

[12] S. D. Guikema, R. A. Davidson, and L. Haibin, "Statistical models of the effects of tree trimming on power system outages," *IEEE Transactions on Power Delivery*, vol. 21, no. 3, pp. 1549–1557, 2006.

[13] Y. Cai and M.-Y. Chow, "Exploratory analysis of massive data for distribution fault diagnosis in smart grids," in *Power Energy Society General Meeting, 2009. PES '09. IEEE*, 2009, pp. 1–6.

[14] Y. Zhang, Y. Liu, X. Wang, and Z. Wang, "Fault pattern recognition in power system engineering," in *Industrial Mechatronics and Automation, 2009. ICIMA 2009. International Conference on*, 2009, pp. 109–112.

[15] Z. Wang and P. Zhao, "Fault location recognition in transmission lines based on support vector machines," in *Computer Science and Information Technology, 2009. ICCSIT 2009. 2nd IEEE International Conference on*, 2009, pp. 401–404.

[16] N. Shahid, S. Aleem, I. Naqvi, and N. Zaffar, "Support vector machine based fault detection amp; classification in smart grids," in *Globecom Workshops (GC Wkshps), 2012 IEEE*, 2012, pp. 1526–1531.

[17] A. Rizzi, F. M. F. Mascioli, F. Baldini, C. Mazzetti, and R. Bartnikas, "Genetic optimization of a PD diagnostic system for cable accessories," *IEEE Transactions on Power Delivery*, vol. 24, no. 3, pp. 1728–1738, July 2009.

[18] S. S. Khan and M. G. Madden, "A survey of recent trends in one class classification," in *Artificial Intelligence and Cognitive Science*, ser. Lecture Notes in Computer Science, L. Coyle and J. Freyne, Eds. Springer Berlin Heidelberg, 2010, vol. 6206, pp. 188–197.

[19] "La road map per la continuit del servizio di distrbuzione elettrica di roma, acea distribuzione spa, aracne editrice, citt ducale (ri)," 2010.

[20] (Jun.) Cei - comitato elettrotecnico italiano @ONLINE. [Online]. Available: http://www.ceiweb.it/it/

[21] T. Vincenty, "Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations," *Survey Review*, vol. XXII, Apr. 1975.

[22] A. P. Shanker and A. Rajagopalan, "Off-line signature verification using DTW," *Pattern Recognition Letters*, vol. 28, no. 12, pp. 1407 – 1414, 2007.

[23] A. Rizzi, L. Livi, H. Tahayori, and A. Sadeghian, "Matching general type-2 fuzzy sets by comparing the vertical slices," in *2013 Joint IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS)*, 2013, pp. 866–871.

[24] M. Müller, "Dynamic time warping," in *Information Retrieval for Music and Motion*. Springer Berlin Heidelberg, 2007, pp. 69–84.

[25] Z. Huang, "Extensions to the k-means algorithm for clustering large data sets with categorical values," 1998.

[26] G. Del Vescovo, L. Livi, F. M. Frattale Mascioli, and A. Rizzi, "On the Problem of Modeling Structured Data with the MinSOD Representative," *International Journal of Computer Theory and Engineering*, vol. 6, no. 1, pp. 9–14, 2014.

[27] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, Jun. 2010.

[28] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224–227, 1979.