A New Approach for Event Classification and Novelty Detection in Power Distribution Networks

André E. Lazzaretti*, Vitor H. Ferreira†, Hugo Vieira Neto‡, Luiz F. R. B. Toledo* and Cleverson L. S. Pinto§
*Institute of Technology for Development (LACTEC), Comendador Franco 1341, Curitiba-PR, Brazil
†Federal University Fluminense (UFF), Passo da Pátria 156, Niterói-RJ, Brazil
‡Federal University of Technology – Paraná (UTFPR), Sete de Setembro 3165, Curitiba-PR, Brazil
§Energy Company of Paraná (COPEL Distribuição S.A.), José Izidoro Biazetto 158, Curitiba-PR, Brazil

***Abstract*—This paper presents a new approach for automatic oscillography classification in distribution networks, including the detection of patterns not initially presented to the classifier during training, which are defined as novelties. We performed experiments with coupled novelty detection and multi-class classification, and also in separate stages, using the following classifiers: Gaussian Mixture Models (GMM), K-means clustering (KM), K-nearest neighbors (KNN), Parzen Windows (PW), Support Vector Data Description (SVDD), and multi-class classification based on Support Vector Machines (SVM). Preliminary results for simulated data in the Alternative Transient Program (ATP) demonstrate the ability of the method to identify new classes of events in a dynamic learning environment. This work was partially supported by COPEL within the Research and Development Program of the Brazilian Electrical Energy Agency (ANEEL).**

***Index Terms*—Multi-class Classification, Support Vector Machines, Novelty Detection, Automatic Waveform Analysis.**

I. Introduction

Several events are responsible for changes in voltage and current waveforms in electrical power systems. In the particular case of voltage waveforms (oscillographic records) in a power distribution system, there is a range of events with relevant impact regarding equipment failure or consumer damage [1]. These events involve changes in the waveforms, whose correct identification is desirable – in particular, the following events can be highlighted: short-circuits, lightning, switching transients, and the start of heavy-duty engines.

For power distribution utilities, voltage waveform variations cause increasing concern with supply disruption and their duration, number of outages, voltage levels, frequency deviations, transients, and harmonic contents. In several countries there are standards that specify the expected quality of service for distribution networks. The extrapolation of product limits can incur in fines for power utilities, imposed by regulatory agencies.

With these restrictions in mind, power utilities have been taking a series of measures to enable broad monitoring of their distribution networks, as the identification and classification of waveform variations may help in mitigation processes, maintenance, and fault characterization, constituting a support system for decision making.

A point in common for most classifiers developed for automatic analysis of oscillographies in electric power systems is the use of supervised learning for multi-class classification models [2], [3]. However, in this work we take into account the possibility of new, previously unseen events happening, without being considered in the classifier that was initially defined. This is an important consideration for events that do not have records stored for modeling, such as data resulting from transient maneuvers in distributed generation units in a new smart distribution network.

In this context, it is important that the model is able to identify events that do not fit the classes known *a priori*, i.e. the classes used for training the classifier. There are some approaches available in the literature that use this idea, mainly in computer vision applications [4]. To the extent of our knowledge, none of these approaches has been applied for automatic analysis of oscillographies, indicating that the novelty detection stage suggested in this work is an important contribution to the field.

A first approach for multi-class classification and novelty detection is based on the use of one-class classifiers, one for each class known *a priori* [5]. In this approach, a new pattern is first checked by each one-class classifier and, if it does not fit any of the modeled classes, it is considered a novelty. There are cases in which patterns fit more than one class model, requiring a similarity measure to define which of them is more likely to be the class of the pattern being checked. A second approach available in the literature use a broader modeling method, performing multi-class classification and novelty detection altogether, enabling the inclusion of new classes and, therefore, making continuous learning possible. For example, in [4], a Gaussian Mixture Model is used to identify new galaxies from outer space images. A very similar method was proposed in [6], where a K-means clustering that allowed on-line addition of new classes to the classifier was used.

By combining one-class classifiers [7], it is also possible to construct a multi-class classifier that expands as new classes are identified, as shown in [8]. In that work, a model for one-class classifier output normalization was presented, so that these classifiers can be combined regardless of their output characteristics (probabilistic or distance-based).

The methods proposed in [4], [6], [8] form the basis for multi-class and novelty detection for event classification in the present paper, which we use to promote a more robust and adaptive classification process to the existing dynamics in event analysis for distribution networks.

II. Theoretical Aspects

Since the one-class classifiers used in this paper (GMM, KM, KNN, PW, and SVDD) are broadly explored in the literature, we emphasize here the theoretical aspects on our multi-class classification and novelty detection method, especially because of its autonomous features during the training stage. Further details on the use of one-class classification in the novelty detection context can be found in [7].

A. One-class Classifier

A one-class classifier, also known as novelty detector or novelty filter, is defined as a classifier based on previously known patterns, which are arranged as a cluster that allows the identification of new patterns that are not present in the originally defined dataset [7] – these previously unseen patterns are defined as novelties.

There are three different approaches to construct one-class classifiers [7]. The first group is based on models based that estimate the probability density function of input patterns. From this function, it is possible to establish if a new pattern is an outlier or not, based on its probability value. In this approach, we highlight parametric estimators based on Gaussian Mixture Models and nonparametric estimators based on Parzen Windows [7].The second group comprises models with imposed boundaries upon the data set, assuming an unknown distribution. Therefore, a boundary optimization problem is solved in order to represent the data. In this approach, we highlight the KNN and SVDD methods. And the third group of models is based on clustering methods. From this kind of representation, it is possible to define if a given input pattern is a novelty or not, based on the distance from the input pattern to previously defined clusters. Within this group, K-means clustering is one of the the most used methods [6].

It is possible to apply one-class classifiers to a multi-class problem [9]. There are basically two approaches to accomplish this task. The first one is to perform multi-class classification and novelty detection in two different stages [7]. First, during the training phase, a multi-class classifier is constructed using the known classes. After that, all known classes are grouped and defined as a normal class using an one-class classifier. In the application phase, new input patterns are fed to the trained one-class classifier – if the pattern belongs to the normal class (one of the known classes), its corresponding class is defined via the multi-class classifier; otherwise, the new pattern is classified as a novelty. In this paper, this approach will be called *Independent Multi-Class Classification and Novelty Detection* (I-MCCND).

The second approach consists in an one-class classifier applied for each known class of the problem during the training phase [5]. In this approach, it is possible to obtain a closed boundary for each class, allowing joint multi-class classification and novelty detection. In the application phase, a verification of whether the input pattern fit into one of the defined one-class models is done. At the end of this process, some patterns will fit into only one model, while others will fit into more than one. A third set of patterns will not belong to any of the adjusted models, being characterized as novelties.

For patterns that fit into more than one model, especially those located in regions of overlapped classes, it is necessary some post-processing to confirm to which class these patterns belong. The post-processing is based on a similarity analysis between a given pattern and each model involved. For density-based models (GMM, KNN, and PW), the similarity measure is based on probabilities – the pattern is assigned to the class that yields the highest *a posteriori* probability.

Usual K-means clustering algorithms adopt the Euclidean distance as the similarity measure between an input pattern and the center of each cluster, so that the pattern is assigned to the nearest cluster. Similarly, the SVDD model takes into account the distance between the pattern and the centers of hyperspheres in the feature space. In this paper, this approach will be called *Coupled Multi-Class Classification and Novelty Detection* (C-MCCND).

B. Multi-class Classifier

Support Vector Machines were originally developed to solve classification problems using the concept of an optimum separation hyperplane, which maximizes the separation margin ρ between classes. The motivation for maximizing ρ is based on a complexity measurement known as the Vapnik-Chervonenkis (VC) dimension [10], whose upper limit is inversely proportional to ρ. Mathematically, the output of an SVM can be expressed as:

$$f(\boldsymbol{x}, \boldsymbol{W}, b) = \mathrm{sgn}[\boldsymbol{W}^T \boldsymbol{\Phi}(\boldsymbol{x}) + b], \tag{1}$$

where $\boldsymbol{\Phi}(\boldsymbol{x}) : \mathbb{R}^n \rightarrow \mathbb{R}^N$ is a nonlinear input mapping in feature space, $\boldsymbol{W}$ is the parameter set of the model, b is the bias and sgn$(\cdot)$ is the sign function. Maximization of the separation margin ρ can be formulated via the following restrict optimization problem:

$$\min_{\boldsymbol{W}, b, \boldsymbol{\xi}} E_s(\boldsymbol{W}) = \frac{1}{2} \boldsymbol{W}^T \boldsymbol{W} + C \sum_{i=1}^{N} \xi_i \tag{2}$$

s.t.

$$\begin{cases} d_i[\boldsymbol{W}^T \boldsymbol{\Phi}(\boldsymbol{x}) + b] \geq 1 - \xi_i \\ \xi_i \geq 0 \end{cases} , \; i = 1, 2, \ldots, N$$

In equation (2), the first term of the objective function is responsible for complexity control of the model by means of the maximization of ρ. The second term relates to classification errors for the dataset, as for correctly classified data ξ_i is equal to zero. The hyperparameter C is responsible for the balance between model complexity and goodness of fit to the training data, and therefore is denominated regularization parameter [10].

The quadratic optimization problem in equation (2) can be solved using the Lagrange multipliers method, whose dual formulation is given by:

$$\max_{\boldsymbol{\alpha}} \Psi(\boldsymbol{\alpha}) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} d_i d_j K(\boldsymbol{x}_i, \boldsymbol{x}_j) \alpha_i \alpha_j \tag{3}$$

s.t.

$$\begin{cases} 0 \le \alpha_i \le C \\ \sum\limits_{i=1}^{N} \alpha_i d_i = 0 \end{cases}, i = 1, 2, \ldots, N,$$

where $\boldsymbol{\alpha}$ represents the set of Lagrange multipliers and $K(\boldsymbol{x}_i, \boldsymbol{x}_j)$ the dot product kernel in feature space, as follows:

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = [\boldsymbol{\Phi}(\boldsymbol{x_i})]^T \boldsymbol{\Phi}(\boldsymbol{x_j}). \tag{4}$$

There are several types of kernel $K(\boldsymbol{x}_i, \boldsymbol{x}_j)$, which must abide to the conditions of Mercer's theorem. In this work, we use the Gaussian kernel given by:

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \mathrm{e}^{\left[-\sum\limits_{l=1}^{N} \sigma_i^2 (x_{il} - x_{jl})^2\right]}, \tag{5}$$

where σ_i^2 are the kernel hyperparameters.

At the optimum of equation (3) not all α_i^* are nonzero. The vectors for which α_i^* are different than zero are the so called support vectors, which define the decision surface of the SVM as follows:

$$f(\boldsymbol{x}, \boldsymbol{W}, b) = \operatorname{sgn}\left[\sum_{i=1}^{N_S} \alpha_i d_i K(\boldsymbol{x}_i, \boldsymbol{x}) + b\right], \tag{6}$$

where N_S is the number of support vectors.

Despite being concerned with the complexity control in their formulation and yielding the model structure as a subproduct of the training algorithm through the number of support vectors, the SVM have some hyperparameters that must be specified by the user, such as the regularization constant C and the kernel hyperparameters σ_i^2. These values are commonly selected via cross-validation and in this work we selected them by means of minimizing the upper limit of the estimated generalization error in a leave-one-out approach. This resampling method yields an almost non-biased estimate for the generalization ability of the model [11], but is computationally intensive. On the other hand, the upper limit used in this work was analytically developed in [11] and is conceptually founded on the span of support vectors. To minimize this limit, we use genetic algorithms, whose optimum reflects an estimate of C and σ_i. Therefore, it is not necessary to choose parameters for training the model in our approach, making it autonomous in the sense of parameter choice.

III. Method

A. Simulated Data

In order to test and validate the segmentation models, we decided to perform simulations in the Alternative Transient Program (ATP) environment [12]. In the basic ATP model used for the generation of simulated events, we have considered the basic elements of a distribution substation from COPEL (power distribution company in the state of Paraná, Brazil), as well as the structures needed for the simulation of the events [13]. These elements are: substation transformer, capacitor bank, grounding transformer, four bar feeders, and the equivalent of the electric circuit up to the substation transformer. Only one feeder was modeled, being segmented in several parts, with different cable types. For other feeders, only the equivalent load was used. All parameters used in the model were based on real data obtained from COPEL.

We decided to simulate events that can actually occur in the feeder under analysis, and events that occur in the electrical system upstream the chosen distribution substation – 29 classes of events were considered, including:

- Single-line-to-ground faults (in each phase)
- two-phase-line-to-ground faults (in each pair of phases)
- three-phase-line-to-ground faults
- two-phase faults (in each pair of phases)
- three-phase faults
- single phase feeder switch-off (in each phase)
- two-phase feeder switch-off (in each pair of phases)
- three-phase feeder switch-off
- single phase feeder reclosing (in each phase)
- two-phase feeder reclosing (in each pair of phases)
- three-phase feeder reclosing
- capacitor bank switching
- start of heavy-duty engines
- single-line-to-ground faults and two-phase-line-to-ground faults in the electrical system before the chosen distribution substation

For each isolated event, which was logged by an oscillographer, we varied the instant and location of occurrence, the fault resistance, and the measurement site. Varying these parameters, we obtained 18 events per class for short-circuits and ten events per class for switching events. From a total of 621 events, 414 were used for training and 207 for testing the classifiers. In addition, we also included simultaneous and subsequent events, with an occurrence interval of 1/4 of a cycle between each event – the following sequences were adopted:

- Feeder switch-off (single or two phase) followed by feeder reclosing (two or three phase)
- Short-circuit (single, two, or three phase) followed by feeder reclosing (single, two, or three phase)
- Feeder switch-off (single or two phase) followed by feeder switch-off (two or three phase)
- Short-circuit (single or two phase) followed by short-circuit (two or three phase)
- Capacitor bank switching followed by short-circuit (single, two, or three phase)
- Capacitor bank switching simultaneously to a short-circuit (single, two, or three phase)
- Feeder reclosing simultaneously to a short-circuit (single, two, or three phase)

In this way, we obtained 128 subsequent and 60 simultaneous events. It is worth mentioning that the subsequent and simultaneous events will be considered here as part of the class defined as novelty.

B. Multi-Class Classification and Novelty Detection

The method applied for model training is summarized in figure 1. Initially, samples from all 29 classes were randomly combined to define the classes known *a priori* and the classes

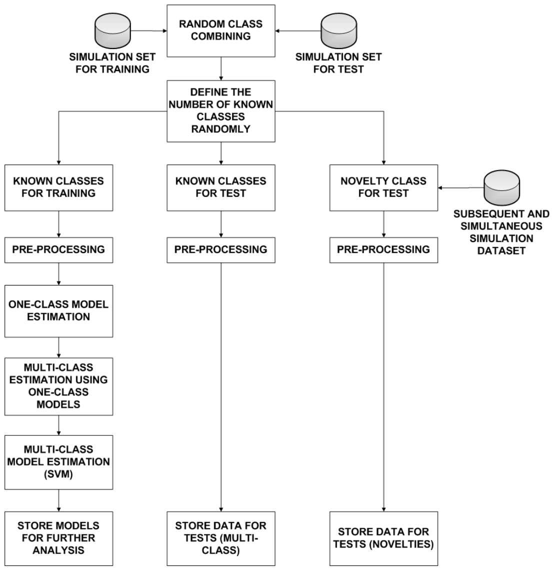

Fig. 1. Training method.

that are to be identified as novelties. In this process, the number of classes that form the basis of supervised learning were also chosen randomly – the minimum number of known classes used in this work was defined as three. The process shown in figure 1 was repeated 100 times, making it possible to obtain an average classification performance for different numbers of classes known *a priori*.

Three sets are extracted from the group of known classes. The first is the training set, which is used for model estimation with 414 events generated for this purpose. The second is the test set, which is used in the test stage, both for novelty detection and multi-class classification, with the 207 remaining events. The third group of classes is related to expected novelties – this set is based on the classes that were not selected for supervised learning, along with the set of simultaneous and subsequent events.

After defining the training and testing sets, we applied a pre-processing stage to the data – each waveform of each phase of the signal was decomposed using a Discrete Wavelet Packet Transform (DWPT) in four levels (16 sub-bands at a sampling frequency of 15360Hz) using the Daubechies-8 mother wavelet [14]. After decomposing the input signals, we computed the energy contents of each DWPT level. The energy contents was computed before and during the event, i.e. for each DWPT level we obtained the energy ratio between the cycle where the event occurred and the cycle immediately before the occurrence, extracting signal features as shown in [13].

Once the pre-processing was performed, the matrices corresponding to the test and novelty sets were stored for the test stage. The training set was used to estimate all models in the I-MCCND and C-MCCND approaches, such as: one-class classifier for all classes, one-class classifier for each class, and the multi-class classifier using SVM.

After model estimation, we assessed the performance of

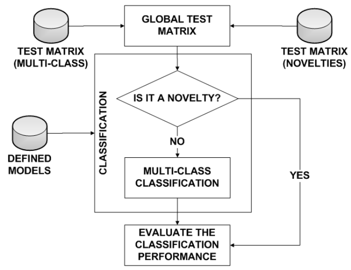

Fig. 2. Test method.

each proposed method, comprising the two different approaches – I-MCCND and C-MCCND. For each experiment run, the test matrix (multi-class) and the novelty matrix were merged in order to be used for performance assessment. The resulting matrix is called global matrix and is composed by all extracted features during the pre-processing stage and the class label index K of each pattern, while patterns in the novelty class are labeled with the index $K + 1$.

In the I-MCCND approach, novelty detection and multi-class classification are performed independently. Initially, it is verified if a given pattern belongs to the set of known classes, via an one-class classifier that groups all known classes. Then, if the pattern is considered as known beforehand, a multi-class classifier is used to define to which class the pattern belongs. This approach is summarized in figure 2, where the classification block is segmented in two stages. In the C-MCCND approach, one-class classifiers are used for each of the known classes and, therefore, it is possible to implement the classification block in figure 2 in a single stage.

For overall performance analysis in both I-MCCND and C-MCCND approaches, one can construct a confusion matrix with K classes known *a priori* and the novelty class $K + 1$. Using this matrix, one can compute the average performance for all classes, including the novelty class. The performance reflects the ability of the method to identify novelties and perform the multi-class classification.

IV. Results

A. I-MCCND Approach

Table I presents the results obtained for the I-MCCND approach, using the PW, GMM, KM, KNN, and one-class SVDD classifiers for novelty detection, followed by a multi-class SVM classifier. The results shown refer to 100 different arrangements of known classes selected as described previously.

TABLE I
GLOBAL AVERAGE PERFORMANCE FOR THE I-MCCND APPROACH.

Method	Average Performance(%) $\pm$ Std. Dev.
PW	20.69 $\pm$ 7.70
GMM	57.50 $\pm$ 15.70
KM	59.09 $\pm$ 12.17
KNN	60.27 $\pm$ 13.53
SVDD	36.72 $\pm$ 17.82

In table I, one can observe that the average performance for all methods is below 61% of accuracy. The best performances were obtained for GMM, KM, and KNN methods. However, some arrangements of known classes presented a far superior overall performance when compared to the average performance. We observed that performance was maximized when the group of known classes was formed by classes of the same nature, i.e. short-circuits – there was an overall performance of 82.52%, 91.37%, and 80.15% for the GMM, KM, and KNN methods, respectively, in the particular case of the group of known classes formed by single-line-to-ground faults, two-phase-line-to-ground faults, three-phase-line-to-ground faults, two-phase faults, three-phase faults; and the novelty class formed by all remaining classes, along with subsequent and simultaneous events.

B. C-MCCND Approach

For the C-MCCND approach, we have used the same 100 different arrangements of known classes used previously for the I-MCCND approach. Thus, it is possible to perform a fair comparison between these two approaches. The overall performance of the C-MCCND method is shown in table II, where one can notice a significant improvement for KM, KNN, and SVDD methods. For the GMM method, the overall performance was worse than the performance shown in table I.

TABLE II
GLOBAL AVERAGE PERFORMANCE FOR THE C-MCCND APPROACH.

Method	Average Performance(%) $\pm$ Std. Dev.
PW	21.12 $\pm$ 3.78
GMM	36.30 $\pm$ 5.86
KM	71.09 $\pm$ 7.71
KNN	68.64 $\pm$ 7.99
SVDD	51.12 $\pm$ 12.28

The same observation drawn for the I-MCCND approach, regarding grouping of known classes, also applies to the C-MCCND. In the latter case, a global result of 86.69% was obtained for the GMM classifier, considering the same class arrangement previously discussed for the former case.

When comparing the results in table II with those in table I, it is observed that the C-MCCND approach yielded performances that are superior to the I-MCCND approach in most cases. KM and KNN methods presented better overall results in both proposed approaches – the overall accuracy of 71.09% yielded by the KM method in the C-MCCND approach can be highlighted as the best.

V. CONCLUSIONS

In this paper we proposed a novelty detection stage operating with a multi-class classifier for event classification in power distribution networks using voltage oscillographies as input data. To the extent of our knowledge, this approach is original – especially when classification of oscillographic records in power networks is considered. Experiments were performed using the novelty detection and multi-class classification engaged in independent stages, using GMM, PW, KM, KNN, and SVDD one-class classifiers, followed by an SVM multi-class classifier. The whole process was assessed using simulated data in the Alternative Transient Program.

When we consider the overall performance of the novelty detector and multi-class classifier, it is possible to obtain an average performance of 71.09% for different classes known beforehand. For a specific group of classes regarding events of the same nature (short-circuits), we observed 91.37% of accuracy. The number of classes and arrangements used in this work result in large confusion matrices that could not be presented due to space restrictions, but we summarize them stating that the classification performances for short-circuits was always superior to the ones for other classes.

Once novelty detection is considered original in oscillographic record classification, we were not able to establish a comparison to current methods. However, we conclude that the preliminary performances obtained indicate a real possibility of including aspects related to dynamic learning, i.e. new class identification, in oscillographic record classification.

REFERENCES

[1] M. Bollen and I. Gu, *Signal Processing of Power Quality Disturbances*. John Wiley & Sons Inc., 2006.

[2] O. Gerek, D. Ece, and a. Barkana, "Covariance Analysis of Voltage Waveform Signature for Power-Quality Event Classification," *IEEE Transactions on Power Delivery*, vol. 21, pp. 2022–2031, 2006.

[3] J. Faiz, S. Lotfi-fard, and S. H. Shahri, "Prony-Based Optimal Bayes Fault Classification of Overcurrent Protection," *IEEE Transactions on Power Delivery*, vol. 22, pp. 1326–1334, 2007.

[4] D. Miller and J. Browning, "A mixture model and EM-based algorithm for class discovery, robust classification, and outlier rejection in mixed labeled/unlabeled data sets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 1468–1483, 2003.

[5] F. Jinglong, W. Wanliang, W. Xingqi, L. Zhe, and I. Introduction, "A SVDD Method Based on Maximum Distance between Two Centers of Spheres," *Chinese Journal of Electronics*, vol. 21, pp. 107–111, 2012.

[6] E. Spinosa and A. de Carvalho, "Combining one-class classifiers for robust novelty detection in gene expression data," in *Advances in Bioinformatics and Computational Biology*. Springer Berlin / Heidelberg, 2005, vol. 3594, pp. 54–64.

[7] D. M. J. Tax and R. P. W. Duin, "Combining One-Class Classifiers," in *Second International Workshop on Multiple Classifier Systems*, 2001, pp. 299–308.

[8] D. Tax and R. Duin, "Growing a multi-class classifier with a reject option," *Pattern Recognition Letters*, vol. 29, pp. 1565–1570, 2008.

[9] P.-Y. Hao, J.-H. Chiang, and Y.-H. Lin, "A new maximal-margin spherical-structured multi-class support vector machine," *Applied Intelligence*, vol. 30, pp. 98–111, 2007.

[10] V. N. Vapnik, *Statistical Learning Theory*. New York: Wiley-Interscience, 1998.

[11] V. Vapnik and O. Chapelle, "Bounds on error expectation for support vector machines," *Neural Computation*, vol. 12, no. 9, pp. 2013–2036, September 2000.

[12] EMTP, *Alternative Transients Program (ATP) Rule Book*, Can Am EMTP User's Group, 1995.

[13] A. E. Lazzaretti, V. H. Ferreira, H. Vieira Neto, R. J. Riella, and J. Omori, "Classification of events in distribution networks using autonomous neural models," in *15th International Conference on Intelligent System Applications to Power Systems*, 2009.

[14] I. Daubechies, *Ten Lectures on Wavelets*. SIAM, Philadelphia, 1992.