Expert Systems With Applications 45 (2016) 322–330

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Novelty detection and multi-class classification in power distribution voltage waveforms

André Eugênio Lazzaretti[a,*], David Martinus Johannes Tax[b], Hugo Vieira Neto[c], Vitor Hugo Ferreira[d]

[a] Institute of Technology for Development (LACTEC), Avenida Comendador Franco 1341, Curitiba, PR, 80215-090, Brazil
[b] Pattern Recognition Laboratory, Delft University of Technology, Mekelweg 4, Delft, 2628CD The Netherlands
[c] Federal University of Technology – Paraná (UTFPR), Avenida Sete de Setembro 3165, Curitiba, PR, 80230-901, Brazil
[d] Federal University Fluminense (UFF), Rua Passo da Pátria 156, Niterói, RJ, 24210-240, Brazil

ARTICLE INFO

Keywords:
Novelty detection
New class identification
Open set recognition
Smart grids
Waveform classification

ABSTRACT

The automatic analysis of electrical waveforms is a recurring subject in the power system sector worldwide. In this sense, the idea of this paper is to present an original approach for automatic classification of voltage waveforms in electrical distribution networks. It includes both the classification of the waveforms in multiple known classes, and the detection of new waveforms (novelties) that are not available during the training stage. The classification method, based on the Support Vector Data Description (SVDD), has a suitable formulation for this task, because it is capable of fitting a model on a relatively small set of examples, which may also include negative examples (patterns from other known classes or even novelties), with maximal margin separation. The results obtained on both simulated and real world data demonstrate the ability of the method to identify novelties and to classify known examples correctly. The method finds application in the mitigation process of emergencies normally performed by power utilities' maintenance and protection engineers, which requires fast and accurate event cause identification.

1. 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 impacts regarding equipment failure or consumer damage (Mahela, Shaik, & Gupta, 2015). Therefore, the correct identification of such events is highly desirable, with the following events being of particular importance: short-circuits, lightning strikes, switching transients along the distribution line, and switching events in distributed power generation (Fig. 1).

Variations in voltage waveforms are of great concern for power distribution utilities, especially voltage interruptions and their duration, number of outages, voltage levels, frequency deviations, transients, and harmonic contents. Several countries specify standards that indicate the expected quality of service for distribution networks and exceeding standard service limits can incur in fines for power utilities, which are imposed by regulatory agencies.

With these restrictions in mind, power utilities have been taking a series of actions to enable broad monitoring of their distribution networks, as well as the identification and classification of waveform variations, which may help in mitigation processes, maintenance, and fault characterization, building up a support system for decision making (Dahal, Abuomar, King, & Madani, 2015; Faisal, Mohamed, Shareef, & Hussain, 2011; Kulkarni, Lee, Allen, Santoso, & Short, 2010; Pires, Amaral, & Martins, 2011).

A common characteristic of most classifiers already investigated in the literature for automatic waveform analysis in electric power systems is the use of supervised learning and multi-class classification models (Biswal, Biswal, Mishra, Member, & Jalaja, 2014; De Yong, Bhowmik, & Magnago, 2015; Demir, 2010; Ekici, 2009; Ferreira, Seixas, & Cerqueira, 2015; Kulkarni & Santoso, 2010; Sánchez, Montoya, Manzano-Agugliaro, & Gil, 2013). However, existing methods do not consider the possibility of new, previously unseen events occurring, which is originally considered in the present work. In the context of this paper, waveforms can be classified as novelties (or previously unknown) due to the following:

- **Lack of previous records to characterize and classify the events:** in this case, experts may know about the existence of some events, but there are few or even no records to incorporate

* Corresponding author. Tel.: +554196579732.
E-mail addresses: lazzaretti@lactec.org.br, andrelzt@yahoo.com.br, andrelazzaretti152@hotmail.com (A.E. Lazzaretti), D.M.J.Tax@tudelft.nl (D.M.J. Tax), hvieir@utfpr.edu.br (H. Vieira Neto), vitor@vm.uff.br (V.H. Ferreira).

http://dx.doi.org/10.1016/j.eswa.2015.09.048
0957-4174/

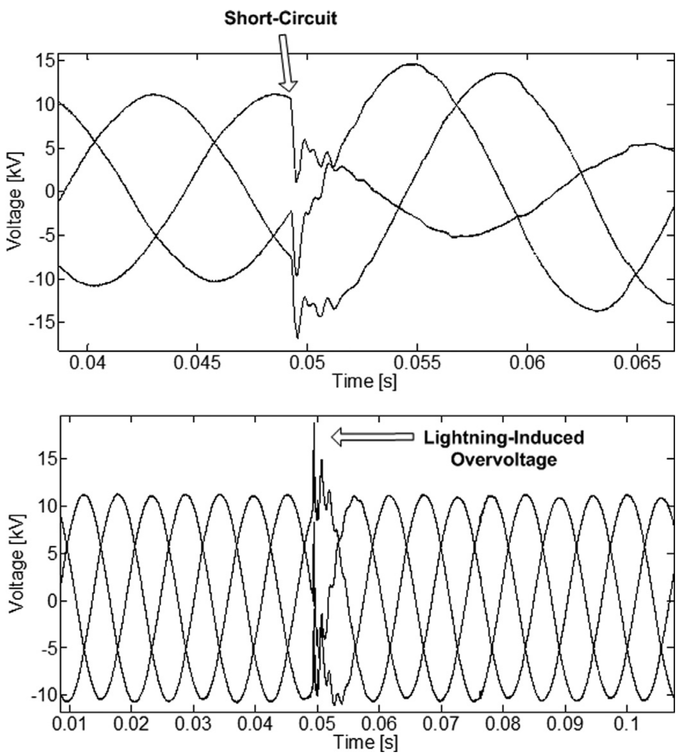

Fig. 1. Examples of voltage waveform transients. The graphs show two different real events (short-circuit and lightning-related) in a power distribution line (Lazzaretti et al., 2011). The start of each event is indicated with arrows. Before the occurrence of each event, the distribution system was in a normal condition.

their features in the model. It is known that certain types of events have very complex models, such as lightning-induced overvoltages (Akbari et al., 2013), which limits their correct classification in some cases (Lazzaretti et al., 2011);

- **Complex waveform interpretation:** there may be events that result in difficult identification and characterization, e.g. simultaneous or subsequent events;
- **Lack of knowledge about the occurrence of the events:** with the latest changes being made in the power distribution sector, it is expected that there may be a set of currently unknown transient events, mainly related to the interconnection of distributed generation units in the system.

Hence, the first contribution of this work concerns a framework for automatic waveform recognition, including multi-class classification of known events and new class identification – novelty detection – that is suitable for the analysis of oscillographic records in power systems. To the extent of the authors' knowledge, the proposed approach is original in this application domain. The kind of classification in focus can be viewed as an open set recognition problem (Scheirer, de Rezende Rocha, Sapkota, & Boult, 2013), in which complete knowledge of the dataset is unavailable at training phase, and new (unknown) classes may appear during testing (Fig. 2).

There are different approaches to build classifiers under the open set assumption. Most of them are based on multi-class classification and novelty detection (Cabral & Oliveira, 2014; Mu & Nandi, 2009; Zhou, Hao, Ning, Yang, & Li, 2015). Novelty detection is applied to identify unknown classes, whereas multi-class classification is used to discriminate among multiple known classes.

A first approach for multi-class classification with novelty detection is based on the use of one-class classifiers (Tax, 2001), one for each class known *a priori* (Hao, Chiang, & Lin, 2007; Miller & Browning, 2003; Mu & Nandi, 2009; Wu & Ye, 2009). In this approach, an input pattern is first checked by multiple one-class classifiers and, if it does not fit any of the modeled classes, it is considered to be a novelty. By combining one-class classifiers, it is also possible to define a multi-class classifier that expands as new classes are identified, as presented in Tax and Duin (2008). In that work, a new heuristic for one-class classifier output normalization was presented, so that one-class classifiers can be combined regardless of their output characteristics (probabilistic or distance-based).

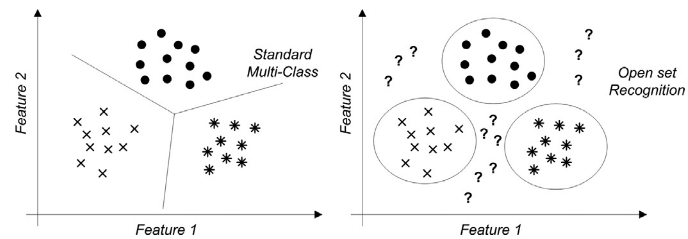

Fig. 2. Open set recognition. In standard multi-class classification (left), all the training and test examples come from known classes. In open set recognition (right), there are multiple known classes and (possibly many) unknown classes.

Novelty detection with multiple known classes is also widely applied in data stream classification (Al-Khateeb et al., 2012; Silva et al., 2013). A data stream can be defined as a sequence of patterns that are continuously produced according to a probability distribution, which can change in the course of time. In a data stream scenario, new classes may appear and previously known classes may change. A recent algorithm that addresses novelty detection in data streams multi-class problems by using a *k-means* algorithm to include new classes and classify known objects is presented in Faria, Gama, and Carvalho (2013). It should be noted that it is possible to modify (extend) known classes using that approach.

Another recent classifier that was proposed to perform this general task is based on a null space method for novelty detection (Bodesheim, Freytag, Rodner, Kemmler, & Denzler, 2013). The idea in this case is to map all training samples of one class to a single point in a null space kernel representation. The authors also present the extension of that approach for a multi-class problem, treating multiple known classes and novelty detection jointly, within a single model. During the test stage, the classification in the null space is distance-based, following a similar procedure to that presented in Hao et al. (2007).

As discussed so far, there are different approaches to perform multi-class classification along with novelty detection. Regarding applications in power systems, initial results with simulated voltage waveforms were presented in Lazzaretti, Ferreira, Vieira Neto, Toledo, and Pinto (2013b), using standard one-class classifiers. The general conclusion of that work is that superior overall performances can be achieved when one-class classifiers were applied for each known class of the problem during the training phase. Also, boundary and reconstruction methods achieved better overall performances when compared to density-based methods, since poorly sampled classes normally occur in waveform datasets. However, the resulting overall performance, even for boundary and reconstruction methods, was relatively far from the desired result, which should be over 80% considering practical applications.

One way to overcome those limitations is to use the model presented in Hao et al. (2007), Wu and Ye (2009), whose formulation has the desired features for the main goal of the present work: closed boundary assumptions for each class with maximal margin separation between classes. Additionally, the proposed formulation is suitable for unbalanced datasets with just a few examples per class, since it is based on SVDD. It is noteworthy that the model presented in Scheirer et al. (2013) is similar to the SVDD-based model used in Hao et al. (2007), Wu and Ye (2009), which reinforces the choice for an open set recognition approach. In this sense, the second contribution of this work is to indicate and demonstrate