Engineering Applications of Artificial Intelligence 48 (2016) 134–139

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Fuzzy-rough imbalanced learning for the diagnosis of High Voltage Circuit Breaker maintenance: The SMOTE-FRST-2T algorithm

E. Ramentol [a], I. Gondres [b], S. Lajes [b], R. Bello [c], Y. Caballero [a], C. Cornelis [d,*], F. Herrera [d,e]

[a] Department of Computer Science, University of Camaguey, Cuba
[b] Department of Electrical Engineering, University of Camaguey, Cuba
[c] Department of Computer Science, Universidad Central de Las Villas, Cuba
[d] Department of Computer Science and AI, University of Granada, Spain
[e] Faculty of Computing and Information Technology - North Jeddah, King Abdulaziz University, Saudi Arabia

ARTICLE INFO

Article history:
Received 10 September 2013
Received in revised form
11 October 2015
Accepted 20 October 2015
Available online 21 November 2015

Keywords:
High Voltage Circuit Breaker (HVCB)
Imbalanced learning
Fuzzy rough set theory
Resampling methods

ABSTRACT

For any electric power system, it is crucial to guarantee a reliable performance of its High Voltage Circuit Breaker (HCVB). Determining when the HCVB needs maintenance is an important and non-trivial problem, since these devices are used over extensive periods of time. In this paper, we propose the use of data mining techniques in order to predict the need of maintenance. In the corresponding data, one class (minority, or positive class) is significantly less represented than the other (majority, or negative class). For this reason, we introduce a new imbalanced learning preprocessing algorithm, called SMOTE-FRST-2T. It combines the well-known Synthetic Minority Oversampling Technique (SMOTE) with a strategy of instance selection based on fuzzy rough set theory (FRST), using two different thresholds for cleaning synthetic minority instances introduced by SMOTE, as well as real majority instances. Our experimental analysis shows that we obtain better results than a range of state-of-the-art algorithms.

1. Introduction

A High Voltage Circuit Breaker (HVCB) is a device designed to open or close an electric circuit. It should be able to open circuits that operate on a wide range of capacities, varying from capacitive currents of a few hundred Amperes to inductive currents of many kA. This is the main reason why it is crucial for any electric power system to ensure a reliable performance of its breakers (ANSI, 2000).

Maintaining the HVCB is a very important task to improve its operational reliability. This maintenance is performed in time intervals dictated by the manufacturer. Nevertheless, very often these intervals are not workable in practice (AREVA, 2005). This difference between the predicted and real times occurs because there are some important variables that are not taken into account in the prediction (Lindquist et al., 2008; Rudd et al., 2011; Fan and Xiaoguang, 2012; Fnineche and Aitken, 2012; Runde et al., 2012).

Therefore, we need a reliable tool to predict when maintenance is necessary for an HVCB, avoiding undesired electric system faults.

* Corresponding author.
E-mail addresses: enislay@gmail.com (E. Ramentol), igondrest@gmail.com (I. Gondres), santiago.lajes@reduc.edu.cu (S. Lajes), rbellop@uclv.edu.cu (R. Bello), yailec@yahoo.com (Y. Caballero), chriscornelis@ugr.es (C. Cornelis), herrera@decsai.ugr.es (F. Herrera).

http://dx.doi.org/10.1016/j.engappai.2015.10.009
0952-1976/

This problem can be considered as a classification task since given some input variables, the system should decide on two possible outputs: "maintenance needed" (positive class) or "maintenance not needed" (negative class).

To obtain the data for this problem, several measurements over different HVCBs have been taken and for each of them a team of experts evaluates whether maintenance is necessary or not. In the majority of cases, the answer is "maintenance not needed". In other words, the class distribution in the resulting dataset is imbalanced; this characteristic is well-known and can be solved by imbalanced classification techniques (He and García, 2009; López et al., 2013; Sun et al., 2009).

A successful strategy to tackle imbalanced classification uses resampling methods that preprocess the data prior to classification (García et al., 2009). Many state-of-the-art resampling methods are based on the Synthetic Minority Oversampling Technique (SMOTE, (Chawla et al., 2002), an oversampling method that creates artificial minority class examples by interpolating between real minority examples and their nearest neighbors. SMOTE is often used in conjunction with a data cleaning method that eliminates examples (artificial minority ones, or majority ones) that are considered harmful for classification. Prominent methods include SMOTE-Tomek links and SMOTE-ENN (Batista et al., 2004), Borderline-SMOTE1 and Borderline-SMOTE2 (Han et al., 2005), Safe-Level-SMOTE (Bunkhumpornpat et al., 2009), SPIDER2 (Napierala et al., 2010), SMOTE-RSB* (Ramentol et al., 2012a) and SMOTE-FRST (Ramentol et al., 2012b).

In this paper, we propose a modification of the SMOTE-FRST algorithm in order to improve it for solving the HVCB maintenance problem. SMOTE-FRST uses fuzzy rough set theory (FRST, Dubois and Prade, 1990) in order to remove data points (instances) that do not sufficiently belong to the fuzzy rough positive region. An important drawback of the method is that to use a high threshold for instance removal usually ends up in eliminating too many original majority examples, causing a reduction in classification performance. On the other hand, choosing the threshold too low undermines the method's data cleaning purpose. In this paper, we deal with this dilemma by cleaning/reducing the training data using a double threshold for eliminating original majority data on the one hand, and synthetic minority data on the other hand. The resulting method is called SMOTE-FRST-2T.

We set up an experimental study to compare our proposal with SMOTE-FRST, as well as with the eight previously mentioned resampling algorithms. As we will see, SMOTE-FRST-2T outperforms all selected methods, demonstrating its competitiveness, and in particular is able to strike a balance between a low number of false negatives (that is, a failure to predict maintenance in the HVCB when it is actually necessary) and false positives (predicting an unnecessary maintenance step).

The remainder of the paper is structured as follows. In Section 2, we recall the HVCB maintenance problem and introduce the dataset used in our research. In Section 3, we present the details of our new proposal, while in Section 4 it is evaluated experimentally. Finally, in Section 5 we conclude.

2. Diagnosis of High Voltage Circuit Breaker maintenance

In this section, we describe in detail the problem of HCVB maintenance. HVCBs are mechanical switching devices that carry and disrupt electrical current in a circuit. Circuit breakers must function in normal and abnormal conditions, and must accommodate short circuits and outages. Circuit breakers are used with switching generators, power stations, cable feeders, transformers, and overhead lines in power distribution systems (Garzon, 2002).

2.1. The HVCB problem

The primary functions of a High Voltage Circuit Breaker include carrying rated current at rated voltage and power frequency when in closed position; interrupting rated currents at rated voltage and power frequency on command; and maintaining rated dielectric (power frequency and impulse) withstand levels when in open position. Sometimes the breaker may not open or close on command, allowing the fault to exist for a longer duration than the system can sustain while functioning normally (Garzon, 2002). Unless a breaker failure initiate action is taken, faults of breakers can lead to undesired changes in system functioning that may result in the system going into an abnormal state, potentially causing major system-wide power outage. This is the reason why it is so important to timely predict when maintenance is necessary for an HVCB.

To decide whether an HVCB needs maintenance is not obvious in many cases for they are used, open or close, for extensive periods of time. The need for a correct prediction of their performance grows in time with the expansion of the transmission systems since they transport more energy in wider regions. Hence preventive or time based maintenance is used. This means that maintenance to the HVCB is scheduled regularly on preset time-slots independent of its state. Nevertheless with the development of the technologies, new approaches for this scheduling have also been derived (ANSI, 2000). Predictive maintenance bases the decision on the inspection of the equipment on regular time-slots. It includes the objective (with the adequate tools) and subjective (with human senses) inspection as well as the reparation of the problem (potential fault). The goal is to accurately predict the condition of the breaker without opening it for inspection, augmenting its efficiency and significantly dropping its maintenance cost. This technique is normally performed by tests, statistical analysis or condition monitoring.

The effectiveness of the predictive maintenance depends on the accuracy of the analysis of the visual revision, tests and statistics to determine the level of the damage. In practice, there are several variables to be included in such an analysis which is often affected by the expertise of the specialist. In this paper, we use data mining techniques to perform such a prediction and the task is divided into the following steps:

1. Making-up the data-set, i.e.,
   (a) determining the variables;
   (b) measurements of each variable;
   (c) labeling each observation as maintenance needed or not by a human expert.
2. Determining the data-set characteristics given the amount of instances on each class.
3. Choosing the classifier to be used in the application.
4. Choosing the preprocessing techniques to be used in the application.
5. Evaluating the performance of the system in an experimental study.

2.2. Construction of the dataset

The variables used in our data were previously determined from international studies, rules and procedures according to approaches of several specialists of the electric company. The Delphi method was used to consult several experts to validate the feasibility and relevance of the variables used in the training set. This consult was applied to 35 renowned specialists, including 25 locals and 10 foreigners. For creating the dataset, the variables chosen were those that had more effect on the class. This procedure resulted in 17 variables that demonstrated to have a high impact on the decision associated to maintenance. These variables are shown in Table 1.

After deciding on the variables, we proceed with the measurements. In this process, experts provided important information related to the state of the breakers. All measurements were

Table 1
Descriptive variables.

Variable names	Type
Insulation chamber	Real
Insulation support	Real
Total insulation	Real
Contact resistance	Real
Gas pressure SF6	Real
Resistance coils	Real
Number of operations	Integer
Principal terminals	Nominal {1–10}
Porcelain	Nominal {1–10}
Temperature-compensated	Nominal {1–10}
Cabinets	Nominal {1–10}
Grounding	Nominal {1–10}
Connections	Nominal {1–10}
General control	Nominal {1–10}
Operational criticality	Nominal {high, normal, low}
Electrical wear	Nominal {high, normal, low}
Breaker wear	Nominal {high, normal, low}