Probabilistic Model for Predicting the Probability of Failure of High Voltage Circuit Breakers

Miguel Moreira da Silva, Jorge Casaca (REN) and Ricardo Azevedo

Abstract - The new challenges faced by electricity transmission system operators (TSOs) are characterized by the large-scale introduction of renewable energy sources, ensuring quality of service and cost-efficiency of investments. On the other hand, the TSO assets begin to age, which introduces additional risks to the reliability of the equipment as well as increased maintenance costs. In this sense, the monitoring of the electric network assets has become increasingly important over the last years.

Asset condition-based maintenance is intended to be the approach to implement. The circuit breakers (CBs) are responsible for the protection of the network and its elements, thus they acquire high importance in this context. The development of failure probability indicators for circuit breakers allows a broader view of the equipment and optimized budgeting for asset replacement and maintenance.

The development of a failure probability index (IPF), coupled with new forecasting methods based on artificial intelligence, allows the prioritization of maintenance actions, bearing in mind the CBs condition. The creation of a learning model, based on a failure probability index and a set of circuit breakers attributes, allowed simulating several scenarios in a multi-year perspective. The final results highlight the importance of the continuous circuit breaker data storage in order to improve the learning.

Keywords - Asset Management, Circuit Breakers, Probability of Failure Index, Learning Model, Artificial Intelligence

I. INTRODUCTION

Mobilized by the strategic efficiency associated with the operational activities of the TSO and the adequacy of the condition of the assets to the performance of the National Transmission Network (RNT), the projects for replacement and modernization of assets fall within the specific optimization investments to increase the lifetime of the assets [1].

The new approach of asset management taken by the TSOs, allow to understand the increasing importance of data analysis. An asset management based on the condition of the equipment reveals to be the best approach comparing with a preventive management based on time.

In this way the implementation of strategies to predict the condition of an asset allows to optimize maintenance costs. The creation of condition indexes and probability failure indexes, reveals to be the way to understand which assets are about to fail [1]. The analysis intends to answer to the follow questions:

- What are the key attributes that will allow to realizing a learning model to estimate a failure probability index of a circuit breaker, considering all the parameters available and susceptible to analyze?
- How should the various attributes evolve over the long term?
- What is the evolution of the IPF of the long-term assets in the various proposed models?
- What is the IPF of a circuit breaker in the various proposed models?
- Which models best fit the perception of a specialist?

II. METHODOLOGY

The methodology used is divided in two different parts:

The first part intends to:

- Select the IPF attributes. (Section A)
- Create a data set. (Section B)
- Normalize the data set, according to the selected attributes. (Section C)
- Examination of a specialist in MAT circuit-breakers by obtaining a qualitative assessment of the condition of the assets. (Section D, E, F, G)
- Application of learning models to a hypothetical dataset, considering the evaluation of a specialist. (Section H, I, J)

The second part intends to:

- Extrapolate the selected data set to 5 years for 3 different scenarios. (Section K)
- Normalize extrapolated data. (Section L)
- Estimate the 5-year IPF in the 3 scenarios, using the learning models obtained previously. (Section M)

A. Attribute Selection to Calculate IPF
Initially was set a list of attributes related to circuit breakers that can be considered for the model. The grouped attributes are shown on the next table.

<table>
<thead>
<tr>
<th>Group</th>
<th>Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables related to time</td>
<td>Age, Operating age, Year of the last fault</td>
</tr>
<tr>
<td>Static variables</td>
<td>Short-circuit current, Rated current, Application type, Drive system, Type of technology, Operating voltage</td>
</tr>
<tr>
<td>Dynamic variables</td>
<td>Number of defects, Number of maneuvers, Maximum operating power, Maximum operating current, Cumulative cut-off current, Cumulative power, Percentile 75 of operating current, MTBF</td>
</tr>
</tbody>
</table>

Table 1 – Grouped attributes.

Based on this Table 1, it is possible to select the most suitable attributes for the predictive model. The criteria used to make this selection was:

- Mutual exclusivity;
- Technical reasons;
- Dimensional reduction;
- Inconsistency in dataset.

Attributes related to time were selected based on mutual exclusivity, where the attribute that makes more sense to consider was the age of the circuit breaker. All the others were rejected.

On the static variables due to the fact that these attributes do not change in time, it was only considered the type of application. The selection of only one attribute in this category, allow to reduce the number of attributes.

The dynamic variables where selected using the inconsistency in dataset and technical reasons criteria. In this way, the maximum power and cumulative power have presented missing values in the dataset and so they were discarded. The maximum operating current was discarded too because some circuit breakers present high value comparing with the percentile 75 of current. The considered final attributes are:

- Age;
- Type of application;
- Number of maneuvers;
- Cumulative cut-off current;
- Percentile 75 of operating current;
- MTBF.

B. Data Set Creation

A hypothetic dataset was created with 1270 circuit breakers and the respective attribute values.

C. Dataset Normalization

The normalization process is necessary to create uniformity on dataset. To do so, a limit for the normalization $\bar{x}$ is defined and a set of rules too:

$$[\bar{x} \in \mathbb{R} | 0 \leq \bar{x} \leq 3]$$

If $x_{\text{min}} \leq x < \mu$, then $0 \leq \bar{x} < 1$

If $\mu \leq x \leq \mu + \sigma$, then $1 \leq \bar{x} \leq 2$

If $\mu + \sigma < x \leq x_{\text{max}}$, then $2 < \bar{x} \leq 3$

Where:

- $x$ is the value to be normalized.
- $\mu$ is mean value of a dataset in a attribute.
- $\sigma$ is the standard deviation of the dataset in a given attribute.
- $x_{\text{min}}$ is the minimum value registered in the dataset in a given attribute.
- $x_{\text{max}}$ is the minimum value registered in the dataset in a given attribute.

An example of an attribute normalization is shown on the next figure.

![Attribute distribution](image)

![Distribution of the normalized attribute](image)

Figure 1 – Age attribute normalization.

\(^1\) Mean time before failure. It is calculated with the ratio between age and the number of defects.
D. Expert Evaluation

After the full dataset normalization, an expert IPF is needed to make the predictive model. This results of the analysis of the full dataset normalized by a circuit break expert.

\[ IPF_{Exp} \in \mathbb{N} : IPF_{Exp} = \{1, 2, 3\} \] (3)

The next table show an example of 3 circuit breaks with the normalized attribute and IPF evaluation.

<table>
<thead>
<tr>
<th>Age</th>
<th>Type of application</th>
<th>Number of operations</th>
<th>Cumulative cut-off current</th>
<th>Percentile 75 of current</th>
<th>MTBF</th>
<th>IPF_{Exp}</th>
</tr>
</thead>
<tbody>
<tr>
<td>2,46</td>
<td>0</td>
<td>0,29</td>
<td>0,27</td>
<td>0,40</td>
<td>2,07</td>
<td>2</td>
</tr>
<tr>
<td>1,69</td>
<td>0</td>
<td>1,39</td>
<td>0,15</td>
<td>0,33</td>
<td>1,39</td>
<td>1</td>
</tr>
<tr>
<td>2,02</td>
<td>0</td>
<td>2,24</td>
<td>1,12</td>
<td>1,32</td>
<td>1,29</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 2 – Normalized attributes and expert IPF.

E. Expert IPF Analysis

In order to understand if the dimension of the expert IFP (grouped in 3 levels) is suitable to the dataset, an analysis with the k-means [2] algorithm is performed.

Applying the Elbow method, that performers the sum of the distance to the respective centroid and the calculation of the Euclidian distance, allows to calculate the best number of groups. The Elbow method formula for \( k \) clusters [3]:

\[ WCSS(k) = \sum_{i=1}^{k} \sum_{j=1}^{x_{i \text{ cluster } j}} \| x_i - \bar{x}_j \|^2 \] (4)

Where:

- \( \bar{x}_j \) mean sample value on cluster \( j \).

The optimal \( k \) clusters value corresponds where the \( k \) presents a maximum value and for analyzed dataset the best fit is when \( k = 4 \).

F. Correlation Between Expert IPF and Attributes

The analysis of correlation gives the perception of how the attributes are related with each other and with the expert IPF.

The attributes higher correlated with the expert IPF are the age, MTBF, cumulative cut-off current and the percentile 75 current.

G. Cumulative variance Between Expert IPF and Attributes

The variance measures the squared deviation of a variable to its mean and the cumulative value of it with the attributes higher correlated with expert IPF, demonstrate the influence of attributes on the expert IPF.
H. Creating the Learning Model

In order to create the prediction model, the dataset is split in training set (67% of data set) and test set (33% of data set) [4].

After that, it is applied the learning algorithms on the training set data: logistic regression [5], nearest neighbors [2], linear support vector machine [2], gradient boosting classifier [6] [7], decision tree [2], random forest [8], neural networks [2] [9] and naive Bayes [2].

I. Predictive Model Performance

The performance of predictive models can be measure with a train set indicator, test set indicator, time set indicator and the mean squared error.

The train set indicator is defined by [2]:

\[ P_{Train}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^{n} 1 \times (y_i = \hat{y}_i) \] (5)

Where:
- \( y_i \) is the train value.
- \( \hat{y}_i \) value result of model.

The test set indicator is defined by [2]:

\[ P_{Test}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^{n} 1 \times (y_i = \hat{y}_i) \] (6)

Where:
- \( y_i \) is the test value.
- \( \hat{y}_i \) value result of model.

The train time indicator is defined by:

\[ TT = T_f - T_i \] (7)

Where:
- \( T_i \) is the initial time.
- \( T_f \) is the final time.

The mean square error is defined by [2]:

\[ MSE(\hat{f}) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{f}(x_i))^2 \] (8)

Where:
- \( y_i \) is the known test value.
- \( \hat{f}(x_i) \) predicted value by the model in \( x_i \) test.

Applying these metrics to the predicted models is possible to verify the model that makes the best fit between the dataset attributes and the expert IPF. On Table 3 it is possible to verify the sorted algorithms that have a better fit.

<table>
<thead>
<tr>
<th>Model</th>
<th>Train set indicator</th>
<th>Test set indicator</th>
<th>Train time</th>
<th>Mean square error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gradient boosting classifier</td>
<td>1,0</td>
<td>0,955</td>
<td>2,845</td>
<td>0,045</td>
</tr>
<tr>
<td>Random forest</td>
<td>1,0</td>
<td>0,950</td>
<td>2,652</td>
<td>0,050</td>
</tr>
<tr>
<td>Linear SVM</td>
<td>0,942</td>
<td>0,924</td>
<td>0,034</td>
<td>0,076</td>
</tr>
<tr>
<td>Decision tree</td>
<td>1,0</td>
<td>0,924</td>
<td>0,002</td>
<td>0,076</td>
</tr>
<tr>
<td>Nearest neighbors</td>
<td>0,946</td>
<td>0,917</td>
<td>0,002</td>
<td>0,083</td>
</tr>
<tr>
<td>Neural net</td>
<td>0,926</td>
<td>0,914</td>
<td>1,031</td>
<td>0,086</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>0,929</td>
<td>0,910</td>
<td>0,011</td>
<td>0,090</td>
</tr>
<tr>
<td>Naive Bayes</td>
<td>0,904</td>
<td>0,883</td>
<td>0,002</td>
<td>0,117</td>
</tr>
</tbody>
</table>

Table 3 – Models performance

J. Model ROC Curve

The ROC curve is a two-dimensional graphic that makes the relation between the false positive rate and the true positive rate of a binary classification system [2].

The calculation of the area under the ROC curve allows to understand the predictive model performance. The area value is between 0 and 1. A higher area value indicates a better performance [2].
Figure 5 – Gradient boosting classifier ROC curve vs naive Bayes

**K. Dataset Extrapolation**

The dataset extrapolation in time is needed in order to apply the predictive models. A set of rules are implemented for each attribute to extrapolate data in time.

**Age:**

\[ Age(t) = Age_{t=0} + t \]  

**Type of application:**

\[ TA(t) = TA_{t=0} \]  

**Number of operations:**

\[ VM(t) = VM_{\#} \times t \]

Where:

\[ VM \] the annual average growth rate of the number of operations

Cumulative cut off current and percentile 75 of current:

\[ CCA(t) = CCA_{t=0} \times (1 + 0.015)^t \]  

\[ P75C(t) = P75C_{t=0} \times (1 + 0.015)^t \]

**MTBF:**

\[ MTBF(t) = \frac{t}{MTBF_{t=0}} \times 0.6 + MTBF_{\#} \times 0.4 \]

Where:

\[ MTBF_{\#} \] the annual average of MTBF for a certain operating voltage of the circuit-breaker.

**L. New Dataset Normalization**

The new dataset normalization is done using the same rules defined on the section B.

<table>
<thead>
<tr>
<th>Model</th>
<th>Mean ROC Area</th>
<th>ROC Area class 0</th>
<th>ROC Area class 1</th>
<th>ROC Area class 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gradient boosting classifier</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>1.0</td>
</tr>
<tr>
<td>Random forest</td>
<td>1.0</td>
<td>0.99</td>
<td>0.99</td>
<td>1.0</td>
</tr>
<tr>
<td>Linear SVM</td>
<td>0.99</td>
<td>0.98</td>
<td>0.98</td>
<td>1.0</td>
</tr>
<tr>
<td>Decision tree</td>
<td>0.95</td>
<td>0.93</td>
<td>0.91</td>
<td>0.89</td>
</tr>
<tr>
<td>Nearest neighbors</td>
<td>0.99</td>
<td>0.97</td>
<td>0.97</td>
<td>1.0</td>
</tr>
<tr>
<td>Neural net</td>
<td>0.99</td>
<td>0.98</td>
<td>0.98</td>
<td>1.0</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>0.99</td>
<td>0.97</td>
<td>0.96</td>
<td>0.99</td>
</tr>
<tr>
<td>Naive Bayes</td>
<td>0.98</td>
<td>0.98</td>
<td>0.97</td>
<td>1.0</td>
</tr>
</tbody>
</table>

**Table 4 – ROC area analysis.**

<table>
<thead>
<tr>
<th>Age</th>
<th>Type of application</th>
<th>Number of operations</th>
<th>Cumulative cut-off current</th>
<th>Percentile 75 of current</th>
<th>MTBF</th>
<th>IPF_{Exp}</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2,46</td>
<td>0</td>
<td>0.29</td>
<td>0.27</td>
<td>0.40</td>
<td>1.91</td>
</tr>
<tr>
<td>24</td>
<td>1,69</td>
<td>0</td>
<td>1.11</td>
<td>0.12</td>
<td>0.29</td>
<td>0.28</td>
</tr>
<tr>
<td>118</td>
<td>2,02</td>
<td>0</td>
<td>2.20</td>
<td>1.12</td>
<td>1.32</td>
<td>0.69</td>
</tr>
</tbody>
</table>

**Table 5 – New dataset normalized. Extrapolated to 5 years.**

**M. Application of Predictive Models**

In order to apply the models, it was set 3 scenarios with different purposes.

The first scenario corresponds to a simple situation, where all the attributes evaluate according the rules set on point J.

The second scenario pretends to simulate a replacement of the old circuit breakers for new circuit breakers. It was considered that the 20% older are replaced. The rules applied to these are:

**Age:**

\[ Age(t) \in N : Age(t) = [0,1] \]

**Type of application:**

\[ TA(t) = TA_{t=0} \]

**Number of operations:**

\[ VM(t) = 0 \]

**Cumulative cut off current:**

\[ CCA(t) = 0 \]
Percentile 75 of current:

\[ P_{75C}(t) = P_{75C_0} \times (1 + 0.015)^t \]  \hfill (19)

MTBF:

\[ MTBF(t) = \overline{MTBF}_{ak} \times 0.4 \]  \hfill (20)

Where:

- \( \overline{MTBF}_{ak} \) the annual average of MTBF for a certain operating voltage of the circuit-breaker.

The third scenario’s goal is to recreate aging effect on the older circuit breakers. On the 20% older circuit breakers it was applied the follow rules:

Age:

\[ Age(t) \in \mathbb{N} : Age(t) = \{39, 40, 41, 42\} \]  \hfill (21)

All of the rest attributes will have specific values to force aging on the circuit breakers. These values were identified on circuit breakers with the expert IPF equals to 3.

III. CASE STUDY

The analysis of the predicted scenarios could be done in 3 parts for each algorithm:

- A comparative analysis of the frequency levels of predicted IPF in \( t=0 \) and \( t=5 \) years.
- The changes between levels of predicted IPF in \( t=0 \) and \( t=5 \) years.
- A comparative analysis between every algorithm for the mean probability value of a predicted IPF to be in level 3 for \( t=0 \) and \( t=5 \) years.

The scenario 1, where it is considered that the attributes only suffer an extrapolation in time (\( t=5 \) years), the IPF level 1 tends to move to level 2 and the level 2 tends to move for level 3. These movements are not too expressive due to the fact that it was only made a shift right in time for attributes.

The mean value for a circuit breaker to be on level 3 of the predicted IPF increase between \( t=0 \) and \( t=5 \). As shown on the next figure for all algorithms there is a higher mean probability of a circuit breaker to be in level 3.

The scenario 2, is supposed to simulate a replacement of old circuit breakers with new ones. From the analysis of the frequency levels, it is possible to notice that level 3 has less circuit breakers comparing \( t=5 \) and \( t=0 \). In this way the natural evolution demonstrates that the IPF tends to balance between level 1 and 2.
The mean probability value tends to decrease in $t=5$ in relation with $t=0$ because the new circuit breakers added are new.

The scenario 3 pretends to simulate forced aging on circuit breakers. Analyzing the frequency levels, it is possible to verify that the level 3 of IPF suffers a sharp growth and the level 2 reduce his number. Some part of level 1 and 2 circuit breakers moves to level 3.
In general, the mean probability value increases a lot for a circuit breaker to be un level 3. This is a predictable behavior because the 20% of the older circuit breakers in $t=5$, have specific value to simulate premature aging.

In conclusion it is possible to make a good approximation between the dataset attributes and the expert IPF. Anyway, to make predictions in time is necessary to have a large database for circuit breakers that allows improving the study of all the attributes in time and consequently creating a better data extrapolation in time.

In all the presented scenarios, it was assumed that the number of circuit breakers do not change in time. The expansion of the circuit breaker on the grid could be an interesting analysis that could give another perspective to the level divisions.

Other approaches to the problem, like considering sound, vibration and environment attributes could also help to improve the predative models.

REFERENCES

[4] R. M. S. Kevin K Dobbin, "Optimally splitting cases for training and testing high dimensional classifiers".

[7] A. AlexeyNatekin, "Gradientboostingmachines, atutorial".


