

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/344556520>

A Methodology for the Identification of the Critical Components of the Electrical Distribution Network of CERN's Large Hadron Collider

Conference Paper · November 2020

DOI: 10.3850/978-981-14-8593-0_5680-cd

CITATIONS

0

READS

95

6 authors, including:



Ahmed Shokry

École Polytechnique

47 PUBLICATIONS 164 CITATIONS

[SEE PROFILE](#)



Piero Baraldi

Politecnico di Milano

250 PUBLICATIONS 3,723 CITATIONS

[SEE PROFILE](#)



Andrea Castellano

Politecnico di Milano

4 PUBLICATIONS 6 CITATIONS

[SEE PROFILE](#)



Ugo Gentile

Nestlé S.A.

28 PUBLICATIONS 156 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Uncertainty analysis and belief reliability theory [View project](#)



Quantitative methods of uncertainty representation and modelling in risk analysis for decision-making practice [View project](#)

A Methodology for the Identification of the Critical Components of the Electrical Distribution Network of CERN's Large Hadron Collider

Ahmed Shokry

Energy Department, Politecnico di Milano, Via Lambruschini 4, 20156, Milan, Italy

Piero Baraldi

Energy Department, Politecnico di Milano, Via Lambruschini 4, 20156, Milan, Italy.

Email: piero.baraldi@polimi.it

Andrea Castellano

Energy Department, Politecnico di Milano, Via Lambruschini 4, 20156, Milan, Italy

Ugo Gentile

Engineering Department, CERN, 1211 Geneva 23, Switzerland

Luigi Serio

Engineering Department, CERN, 1211 Geneva 23, Switzerland

Enrico Zio

Energy Department, Politecnico di Milano, Via Lambruschini 4, 20156, Milan, Italy

MINES ParisTech, PSL Research University, CRC, Sophia Antipolis, France

Eminent Scholar, Department of Nuclear Engineering, Kyung Hee University, South Korea

We consider the propagation of disturbances in the electrical distribution network of CERN's (*European Organization for Nuclear Research*) Large Hadron Collider (LHC), which is a complex system made of a large number of mutually interconnected and interdependent components. The objective of this work is the identification of the components most critical in the determination of the system operating/failed state. Given the complexity of the system, the identification is to be based on the operational data collected from the monitoring systems. The critical components are sought as those whose condition monitoring signals are most correlated to the system operating/failed state. The method of the identification is based on the use of the Relief feature selection technique to rank the monitoring signals according to their importance with respect to the classification of the system operating/failed state. The criticality of the identified signals and the corresponding components has been validated by the system technical responsables.

Keywords: Critical components, Relief feature selection, Filter methods, Classification, Large Hadron Collider.

1. Introduction

The main goal of CERN is to design, construct and operate technical infrastructures and particle accelerators to provide proton beams with given characteristic to experimental facilities for high-energy physics research (CERN, 2017). Figure 1 shows the LINAC2, Proton Synchrotron Booster (PSB), Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS) accelerators, which are used to accelerate and inject the beams particles into the Large Hadron Collider (LHC) where they

are accelerated up to 7 TeV (Brüning, et al., 2012; CERN, 2017).

The LHC is the world's largest and highest-energy particle collider. It consists of a 27 km ring of superconducting magnets and infrastructures extending over the Swiss and French borders. The beams travel in opposite directions in separate beam pipes, which are kept at ultra-high vacuum. The particles are guided and deflected around the accelerator ring by a strong magnetic field maintained by superconducting electromagnets, whose superconducting state requires the magnets

Proceedings of the 30th European Safety and Reliability Conference and the 15th Probabilistic Safety Assessment and Management Conference

Edited by Piero Baraldi, Francesco Di Maio and Enrico Zio

Copyright © ESREL2020-PSAM15 Organizers. *Published by Research Publishing, Singapore.*

ISBN/DOI: 978-981-14-8593-0

to be cooled down at $-271.3\text{ }^{\circ}\text{C}$ (Brüning, et al., 2012; Kahle, 2014; Nielsen & Serio, 2016).

caused by short-circuits or surges (Kahle, 2014) inside or outside CERN electrical network. In the

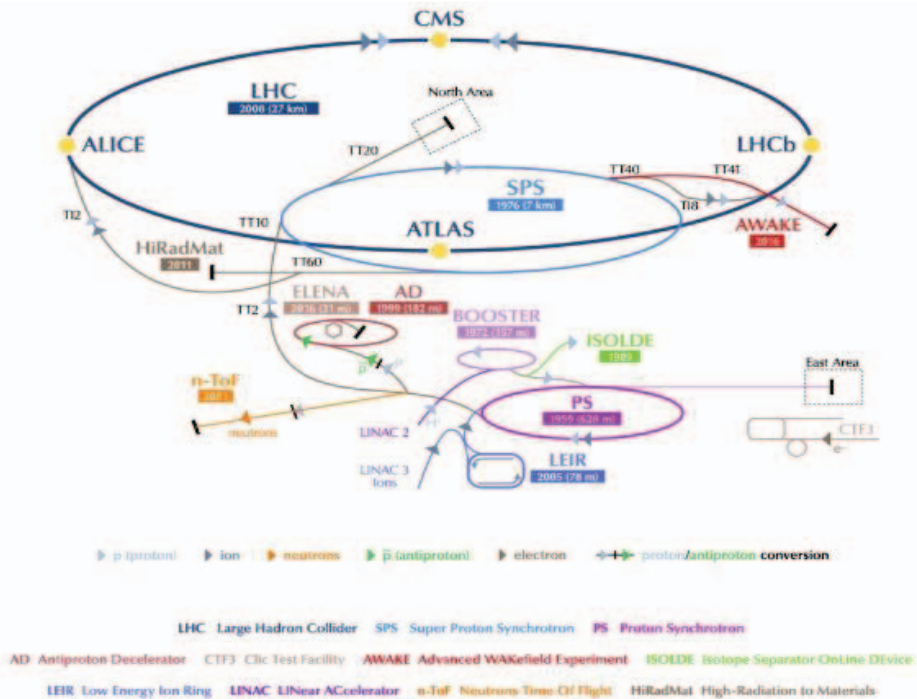


Figure 1. CERN accelerators and experimental facilities layout (CERN, 2017).

The functionality and availability of the LHC is conditioned by the safe and reliable operation of different large-scale complex systems made by tens of thousands of interconnected and interdependent components (Nielsen & Serio, 2016; Todd, et al., 2017). For example, the electrical network system provides power to all facilities, components and machines in CERN, the cryogenic system is responsible for the refrigeration of the superconducting electromagnets, and the cooling and ventilation system operates all the air conditioning plants, pumping stations and fluid distribution systems across all CERN facilities. In this work, we consider the electrical network system, whose malfunctions directly impact the LHC availability. For example, the analysis of the LHC operation during 2016 has shown that a significant amount of the overall LHC downtime has been caused by electrical disturbances (Nielsen & Serio, 2016; Todd, et al., 2017). An electrical disturbance is a voltage fluctuation

latter case, disturbances can propagate from the external power supply lines (French EDF/RTE 400 kV line and Swiss SIG/EOS 130 kV line) to the downstream distribution lines inside CERN campus, in the following sequence: 66 kV, 18 kV, 3.3 kV, 400V lines up to the power converters level. As a consequence, alarms are raised, and occasionally interlocks are triggered causing the particle beams to be stopped preventively (an operation called beam dump in technical jargon) to avoid any risk of damage of the LHC components (Brüning, et al., 2012; Kahle, 2014; Nielsen & Serio, 2016).

Since the electrical disturbances propagate following unknown mechanisms, which are influenced by the configuration and operational state of the network components (e.g. healthy, degraded, failed, under maintenance), the identification of the electrical network components that are critical for the disturbances propagation is of paramount importance to detect the system bottlenecks to which consolidation

efforts should be allocated (Johansson, et al., 2013).

Due to the complexity of the system topology, its geographic distribution and the differences in its components technologies and functionalities, traditional reliability and risk analysis methods for the computation of the components importance measures, based on the knowledge of the system functional logic, cannot be applied (Borgonovo & Apostolakis, 2001; Zio, et al., 2007). Then, in this work we propose a novel data-driven method for the identification of critical components of CERN's LHC electrical distribution network. The method is based on the analysis of hundreds of monitoring signals collected from the large number of components in the network over long periods of time and under different operational states of the LHC. The method relies on the selection of the subset of signals most relevant and informative with respect to the LHC operating/failed state by using the Relief feature selection technique.

Relief, which is a filter feature selection method, directly computes from the data a numerical index that evaluates the feature (i.e. signal) importance or relevance with respect to the observed output (the LHC state) (Urbanowicz, et al., 2018). The Relief algorithm has been chosen due to its relatively small computational requirements, tolerance to noise and robustness to feature interactions (Chandrashekar & Sahin, 2014). At the end of the procedure, the most informative signal subset is analyzed to identify the critical components.

The rest of the paper is organized as follows. Section 2 presents the case study and the involved dataset, and formalizes the objective of the work. Section 3 describes the proposed methodology for the identification of the critical signals using the Relief algorithm. Section 4 shows the application results and their analysis. Finally, Section 5 concludes the work.

2. Case study

The LHC electrical network is geographically decomposed into eight zones, which are referred to as LHC Points 1, 2, 3, 4, 5, 6, 7 and 8, and is functionally decomposed into the LHC power

distribution at 400 kV, 66 kV, 18 kV, 3.3 kV and 400 V levels. Starting from the top-hierarchy, 400kV line, to the bottom-hierarchy, 400V line, the number of components/units and parameters to be monitored explodes exponentially from hundreds to tens of thousands.

This work considers a dataset, D , including a number of $N = 260$ monitoring signals, whose values have been collected during the year 2016 by sensors installed on distribution switchboards and breaker units at the 66 kV and 18 kV lines of the LHC electrical network. The signals are recorded in correspondence of the times of occurrence of electrical disturbances, which are detected by oscilloscopes installed in specific locations.

We exploit the information content of this dataset D containing historical instances, $(\mathbf{y}(k), x(k))$, $k = 1, \dots, K = 3303$, where $\mathbf{y}(k) = (y_1(k), y_2(k), \dots, y_j(k), \dots, y_N(k))$ represents the vector of N signal values collected at the past time instant k and $x(k) \in \{0,1\}$ is the corresponding operating/failed system state. Since the failure state is a rare event, the number of positive instances (i.e. $x(k) = 1$) is much smaller than that of negative instances (i.e., $x(k) = 0$). In the considered dataset, only 43 instances have caused the beam dump, which is considered as the system failed state.

The objective of the work is the identification of the subset of signals critical for the state of the system, $\mathbf{y}^* = (y_1^*, y_2^*, \dots, y_l^*, \dots, y_q^*)$, with $q < N$ and $y_i^* \in \{y_1, \dots, y_N\}$ for any $l = 1, \dots, q$. The critical components are, then, identified as those whose states are monitored by the selected subset of signals, \mathbf{y}^* .

3. Methodology

We consider a subset of signals \mathbf{y}_1 more relevant than a subset of signals \mathbf{y}_2 with respect to the system operating/failed state, x , if the classification performance of an empirical classifier $G^{*1}(\mathbf{y}_1): x^{CTI} = G^{*1}(\mathbf{y}_1)$, built using the historical instances $(\mathbf{y}_1(k), x(k))$, is superior to that of a classifier $G^{*2}(\mathbf{y}_2): x = G^{*2}(\mathbf{y}_2)$, built using the historical instances $(\mathbf{y}_2(k), x(k))$, $k = 1, \dots, K$. We, then, define a selection problem, that has the objective of finding the subset of signals \mathbf{y}^* which allows building the best empirical system state classifier G^* .

3.1 Relief

The Relief algorithm is a filter feature selection method that provides a proxy statistic for each feature (also called feature weight), which acts as a measure of the relevance of the feature with respect to a target concept (Urbanowicz, et al., 2018). The algorithm is capable of handling binary classification, multi-class classification, and regression problems, including discrete and continuous features (Urbanowicz, et al., 2018).

Relief assigns low score to features that have different values for neighboring instances belonging to the same class and high scores to features that have different values for neighboring instances belonging to different classes (Urbanowicz, et al., 2018). Consider the dataset D containing instances $[\mathbf{y}(k), x(k)], k = 1, \dots, K$, $\mathbf{y}(k) = (y_1(k), y_2(k), \dots, y_j(k), \dots, y_N(k))$, $x(k) \in \{0,1\}$. The Relief algorithm first initializes the weights of all features (i.e. signals) to zero, $W = [w_1, w_2, \dots, w_j, \dots, w_N] = 0$. Then, the weights are updated through $\mathcal{V} \leq K$ iterations. In the generic v -th iteration, $v \leq \mathcal{V}$, an instance, \mathbf{y}_v , is randomly selected from the dataset, D , without replacement and the features weights are updated considering the difference between their values at this instance (\mathbf{y}_v) and their values at the Q nearest neighbor instances, $\mathbf{y}_q^v, q = 1, 2, \dots, Q$, of each class:

$$\begin{aligned}
 &W^v \\
 &= W^{v-1} - \sum_{q=1}^Q \frac{\text{diff}(\mathbf{y}_v, \mathbf{y}_q^v)}{\mathcal{V} \cdot Q} \\
 &\quad \quad \quad x(\mathbf{y}_q^v) = x(\mathbf{y}_v) \\
 &+ \sum_{x(\mathbf{y}_q^v) \neq x(\mathbf{y}_v)} \frac{\left[\frac{p(\mathbf{y}_q^v)}{1 - p(\mathbf{y}_v)} \cdot \sum_{q=1}^Q \text{diff}(\mathbf{y}_v, \mathbf{y}_q^v) \right]}{\mathcal{V} \cdot Q}
 \end{aligned}$$

where $p(\mathbf{y}_q^v)$ is the prior probability of the class to which \mathbf{y}_q^v belongs and $p(\mathbf{y}_v)$ is the prior probability of the class to which \mathbf{y}_v belongs, which can be estimated from the training data, and $\text{diff}(\mathbf{y}_v, \mathbf{y}_q^v) = |\mathbf{y}_v - \mathbf{y}_q^v| / \max(\mathbf{y}) - \min(\mathbf{y})$. Finally, the obtained weights are normalized between -1 and +1, and the normalized weight vector, $W =$

$[w_1, w_2, \dots, w_j, \dots, w_N]$, is used to rank the N features according to their informativeness or relevance. For further details on the Relief algorithm, the reader is invited to refer to (Urbanowicz, et al., 2018).

The final outcomes of the Relief algorithm are the weights that express the relative importance or relevance of the signals with respect to the observed phenomena or output. The task of selecting a subset of the top-ranked signals can be performed by setting a minimum threshold on the signals' weights or a maximum threshold on the number of top-ranked signals, (John, et al., 1994; Langley, 1994). The threshold values are dependent on the objective of the subsequent task (i.e. constructing a predictive model whose inputs could be a subset of the top-ranked signals or analyzing the system behavior with respect to a subset of the top-ranked signals).

4. Results

The Matlab function “*relieff*” with $\mathcal{V} = K$ and $Q = 7$ is used to rank the available signals. Figure 1 shows the ranking and the associated weights of the 260 signals. Signals measured on the 66 kV and 18 kV lines are highlighted by different colors.

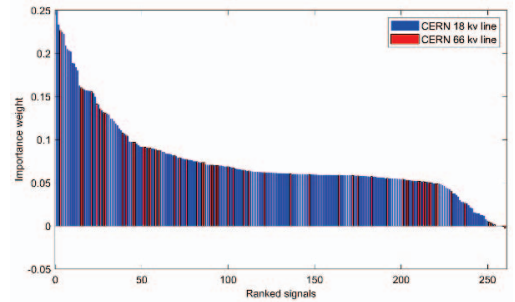


Figure 1. Relief weights and corresponding ranking of the monitoring signals.

Notice that 9 of the 10 top-ranked signals are taken from the 18 kV line and only one is selected from the 66 kV line. This is due to the fact that signals from the LHC 66 kV line are less relevant than the ones from the 18 kV line with respect to the LHC state in case of electrical disturbances, since the 18 kV line is at the bottom of the network hierarchy.

filter feature selection approach based on the Relief technique is used to rank the monitoring signals with respect to the LHC state. The system critical components are identified as the ones associated to the subset of the top-ranked signals.

Table 1. List of the ten top-ranked signals by the Relief.

Signal Rank	Weight Value	Network Line	Signal Type	Component Number	LHC Geographic Zone
1	0.249	18 kV	power	205	Point 5
2	0.233	66 kV	power	212	Point 5
3	0.227	18 kV	power	100	Point 4
4	0.224	18 kV	power	302	Point 1
5	0.222	18 kV	power	203	Point 1
6	0.209	18 kV	power	213	Point 5
7	0.205	18 kV	power	101	Point 4
8	0.203	18 kV	power	213	Point5
9	0.202	18 kV	power	205	Point 1
10	0.189	18 kV	power	203	Point 1

Table 1 reports the list of the 10 top-ranked signals and the corresponding weights. The analysis reveals that:

- six components from the LHC 18 kV line and one component of the LHC 66kV line are identified as the most critical, among tens of components that are related to all the 260 signals collected from both lines,
- most of these critical components are located in the geographical areas of the LHC corresponding to Points 1,4 and 5, and
- all the selected signals are power signals.

5. Conclusions

This work presents a novel data-driven method for the identification of critical components. It has been tested and evaluated through its application to the CERN LHC electrical network, based on the analysis of hundreds of monitoring signals recorded over long periods of time and under different operational settings. The problem is addressed as a feature selection problem with the objective of identifying the subset of signals most informative with respect to the system state. A

The methodology application has allowed identifying as most critical specific components, geographic areas (points 1 and 5), type of signals (power) and even hierarchical level (18 kV) whose criticality is confirmed by the equipment responsables.

References

- Borgonovo, E. & Apostolakis, G. E., (2001). A New Importance Measure for Risk-Informed Decision Making. *Reliability Engineering & System Safety*, 72, 193-212.
- Brüning, O., Burkhardt, H. & Myers, S., (2012). The Large Hadron Collider. *Progress in Particle and Nuclear Physics*, 67, 705-734.
- CERN, 2017. The LHC Guide. Geneva: CERN Document Server.
- Chandrashekar, G. & Sahin, F., (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, 40,16-28.
- Johansson, J., Hassel, H. & Zio, E., (2013). Reliability and vulnerability analyses of critical infrastructures: Comparing two approaches in the context of power systems. *Reliability Engineering & System Safety*,120, 27-38.

- John, G., Kohavi, R. & Pfleger, K., (1994). Irrelevant features and the subset selection problem. In: Machine Learning Proceedings 1994. New Brunswick: Proceedings 11th International Conference on Machine Learning, 121-129.
- Kahle, K., (2014). Power Converters and Power Quality. Baden, Switzerland, Proceedings of the CAS-CERN Accelerator School: Power Converters, 7-14.
- Langley, P., (1994). Selection of relevant features in machine learning. California: Proceedings of the AAAI Fall symposium on relevance, 184, 245-271.
- Nielsen, J. & Serio, L., (2016). Technical Services: Unavailability Root Causes, Strategy and Limitations. Evian Les Bains, France, Proceedings 7th Evian Workshop on LHC beam operation, 59-62.
- Todd, B., Ponce, L., Apollonio, A. & Walsh, D. J., (2017). LHC Availability 2017: Standard Proton Physics. Geneva, CERN-ACC-NOTE-2017-0063.
- Urbanowicz, R. J. et al., (2018). Relief-Based Feature Selection: Introduction and Review. Journal of Biomedical Informatics, 85, 189-203.
- Zio, E., Marella, M. & Podofillini, L., (2007). Importance measures-based prioritization for improving the performance of multi-state systems: application to the railway industry. Reliability Engineering & System Safety, 92, 1303-1314.