



HAL
open science

A New Hybrid Approach for Identifying Obsolescence Features: Applied to Railway Signaling Infrastructure

Elie Saad, Mariem Besbes, Marc Zolghadri, Victor Czmil, Vincent Bourgeois

► To cite this version:

Elie Saad, Mariem Besbes, Marc Zolghadri, Victor Czmil, Vincent Bourgeois. A New Hybrid Approach for Identifying Obsolescence Features: Applied to Railway Signaling Infrastructure. 2024. hal-04456703

HAL Id: hal-04456703

<https://hal.science/hal-04456703>

Preprint submitted on 26 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

A New Hybrid Approach for Identifying Obsolescence Features: Applied to Railway Signaling Infrastructure

Elie Saad^{1,2}, Mariem Besbes¹, Vincent Bourgeois², Victor Czml², and Marc Zolghadri¹

¹ Institut Supérieur de Mécanique de Paris, Laboratoire Quartz (EA7393), 3 Rue Fernand Hainaut, 93400 Saint-Ouen-sur-Seine, France,

{`elie.saad`, `mariem.besbes`, `marc.zolghadri`}@isae-supmeca.fr,

² SNCF Réseau, Département de Signalisation Ferroviaire, 6 Av. François Mitterrand, 93210 Saint-Denis, France,

{`elie.saad`, `vincent.bourgeois1`, `victor.czml`}@reseau.sncf.fr

Abstract. Electrical component obsolescence poses a major issue especially within systems with large life cycles. Thus, finding the optimal management solution for each obsolescence case is as crucial as knowing what to consider when faced with an obsolescence case. In this paper, a novel hybrid approach for identifying features affecting electrical component obsolescence management is introduced, which combines features engineering techniques and expert knowledge. The method then uses machine learning to predict obsolescence resolution techniques in order to find the optimal resolution. The motivation behind this research is driven by the imperative need for SNCF RESEAU to optimally address and mitigate the challenges posed by electrical component obsolescence in railway infrastructure.

Keywords: component obsolescence, obsolescence management, feature selection, feature extraction, machine learning, hybrid approach

1 Introduction

Obsolescence, pertains to the process of becoming non-available, non-adequate, and/or non-suitable [Zolghadri et al., 2021]. Obsolescence poses a problem as it renders previously functional components obsolete, leading to waste, inefficiency, and the need for constant adaptation. One of the major causes of obsolescence is the difference between the component life cycle (≈ 18 months or less) and the systems within which it is contained (30 years or more) [Renaud et al., 2018], e.g., planes, trains, railway infrastructures, etc.

The difference between the component and system life-cycles triggers component obsolescence that requires a response, sometimes incurring costs rarely acceptable. For instance, the decades-old railway infrastructure of SNCF RESEAU constantly grapples with issues related to the obsolescence of electrical components, resulting in significant and unplanned annual expenditures. For that

reason, multiple approaches have been proposed for obsolescence management strategy selection focused on minimizing the cost [Porter, 1998, Zheng et al., 2015, Singh and Sandborn, 2006]. But various features beyond cost, including the number of suppliers, time constraints, solution reliability, and multiple stakeholder opinions from design, operations, manufacturing, sales, and service, can significantly influence the determination of the optimal decision for obsolescence management [Pingle, 2015].

The contribution of this work is twofold: proposing an approach for deriving the set of features affecting the decision-making process for obsolescence management, and proposing an analysis method for determining the contribution of each feature within the decision-making process through the use of machine learning. The remainder of this paper is separated as follows: starting with the dataset formation and methodology description in Section 2, followed by the presentation of the results in Section 3, and finally ending with a brief conclusion and discussion in Section 4.

2 Contribution

The section starts off by tackling the dataset formation methodology from the process described in Figure 1 (shown in blue), afterwards the models used are discussed in Section 2.2 (shown in red), and finally the analysis method is tackled in Section 2.3 (shown in green).

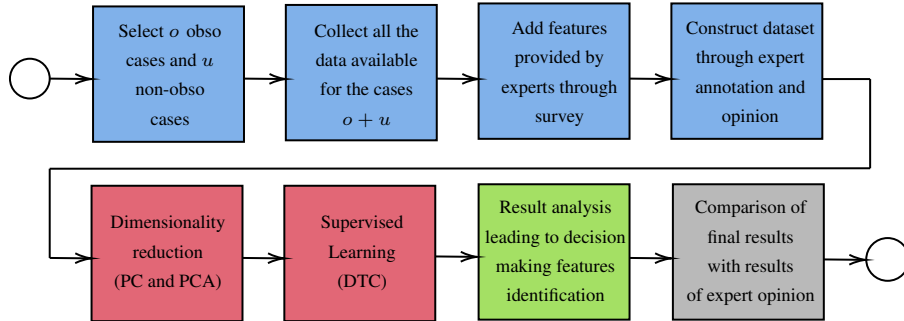


Fig. 1: Methodological Approach Process

2.1 Dataset Formation Methodology

Data Acquisition A survey has been conducted regarding the features affecting obsolescence management solution selection of electrical products, e.g., blinkers, power supplies, etc. The survey was sent out to the obsolescence management team of SNCF RESEAU, who monitor, detect, and decide on the appropriate resolution for each obsolescence case. These personnel come from various backgrounds and domains of work, including the supply chain, technical

management, technicians, engineering, and purchasing departments. For each obsolescence case, a set of features is recorded, but the knowledge is held by the experts of the team.

The dataset was constructed from numerous internal databases within the company following the formation of the set of features obtained from the survey conducted, the literature, and an internal company data visualization interface. The quality of the data varies depending on the database source. The data regarding the features referring to the technical properties of the various units and components are not always digitized, and thus the knowledge of the personnel as well as the archives of the company was accessed to be filled.

Dataset Formalization The dataset is formalized in what follows. Given the dataset of obsolescence and non-obsolescence cases \mathcal{D} such that $\mathcal{D} = \{(x_i, y_i) \in \mathcal{X} \times \mathcal{Y} | i = 1, \dots, N_{o+u}\}$, where each case x_i is a D -dimensional vector comprised of D features $x_i \in \mathbb{R}^D$, and each class label, i.e., obsolescence management solution (e.g. replacement, repairing, etc. [Rust et al., 2022]), is represented by an integer $y_i \in \mathcal{Y} \equiv \{1, \dots, N\}$ resulting in N distinct classes. Let $\mathcal{X} \equiv \mathbb{R}^D$ for generality. The set \mathcal{X} is divided into two subsets: the obsolete set containing obsolete units, and the non-obsolete set containing non-obsolete units. The obsolete set is given by $\mathcal{O} = \{(x_i^o, y_i^o) \in \mathcal{X} \times \mathcal{Y} | i = 1, \dots, N_o\}$, and the non-obsolete set is given by $\mathcal{U} = \{(x_i^u, y_i^u) \in \mathcal{X} \times \mathcal{Y} | i = 1, \dots, N_u\}$. The obsolete and non-obsolete sets are disjoint $\mathcal{O} \cap \mathcal{U} = \emptyset$. Each of the sets above are represented by their corresponding matrices where \mathcal{O} , \mathcal{U} , and \mathcal{X} are represented by the matrices $O \in \mathbb{R}^{o,D}$, $U \in \mathbb{R}^{u,D}$, and $X \in \mathbb{R}^{o+u,D}$ respectively such that $X = \begin{pmatrix} O \\ U \end{pmatrix}$

The dataset \mathcal{D} is composed of $o = 316$ obsolete instances, and $u = 10$ non-obsolete instances. The instances are labeled with one of the following classes of obsolescence remediation: existing stock (7 instances), Last-time Buy Order (30 instances), minor redesign (28 instances), major redesign (124 instances), and substitution (126 instances). The labeling process was conducted through expert annotation.

2.2 Model Methodology

As a preliminary process, and in order to combat the restrictive nature of binary values, all the binary fields are normalized on a scale from $[1, 5]$ following the normal distribution as follows $1 + 4 \frac{(x - \bar{x})}{\sigma}$ where x is the row feature vector and σ is the standard deviation.

Step 1: PC. Highly correlated features pertain to variables that exhibit a strong linear relationship with one another, i.e., they convey nearly identical information. Consequently, this phenomenon can give rise to a predicament referred to as multicollinearity [Montgomery et al., 2021], wherein the independent impact of each variable on the target variable becomes arduous to ascertain. Thus, to address this problem, the Pearson Correlation (PC) coefficient is calculated for

X^T [Mei et al., 2022]. Any vector with a correlation above an arbitrarily selected α is removed following the backward elimination method [Kotu and Deshpande, 2014] from the matrix X^T , i.e., to reduce redundancy, thus reducing its dimensionality $X \in \mathbb{R}^{o+u,h}$ such that $h \leq D$.

Step 2: PCA. When confronted with a large number of features and a small dataset, the model may inadvertently fit the noise within the data rather than discerning the underlying patterns. Consequently, the ability of the model to generalize effectively diminishes [Goodfellow et al., 2016]. By reducing the number of features the complexity of the model diminishes as well, thereby reducing the likelihood of overfitting the training data which leads to better generalization and higher accuracy on novel data [Sammut and Webb, 2011]. Ergo, the Principal Component Analysis (PCA) algorithm is used which computes the orthonormal ℓ -dimensional subspace $X^* \in \mathbb{R}^{o+u,\ell}$ of X forming \mathcal{X}^* , and by consequence \mathcal{D}^* such that $\ell < h$.

Step 3: DTC. The next and final step is to harness the knowledge that the experts possess through learning a function f that takes in the expert provided instances and maps them to the expert classified class labels $f : \mathcal{X}^* \rightarrow \mathcal{Y}$. The aim of this step is to mimic the reasoning behind the decisions made by the experts. For the reason of ease of interpretability, the Decision Tree Classifier (DTC) algorithm is used to learn the decision conditions of each class $y_i \in \mathcal{Y}$, i.e., obsolescence management solution.

2.3 Analysis Methodology

The first step of the analysis involves analyzing the mean decrease in Gini impurity within the learned DTC model f , which can be used to determine the importance of each feature: the higher the mean decrease in impurity, the more important the feature is considered [Sandri and Zuccolotto, 2010]. The second step is to determine the features that contribute the most to the formation of the orthonormal subspace X^* . This is done by taking the sum of the highest ℓ elements of the vectors of the column space of X^* , where each vector corresponds to a feature present in X [Goodfellow et al., 2016].

3 Experimentation and Discussion

The number of initial features $D = 16$ was reduced to $\ell = 7$ principal components. The correlation threshold, set at $\alpha = 0.15$, was determined by identifying the optimal value within the range $\alpha \in [-1, 1]$ that maximizes the average accuracy of the DCT model while minimizing the dimensionality of \mathcal{D} , resulting in $D = 12$. Similarly, the number of principal components $\ell = 7$ was determined using a similar optimization method and objective. The optimal parameters were

derived by selecting the values corresponding to the model with the highest accuracy among the 500 tests conducted.

Afterwards, an additional 1,000 tests were conducted on the proposed model using the obsolete set \mathcal{O} , which is shuffled before every test to reduce bias and evaluate its generalization. The maximum and minimum accuracy values of DCT are reported as 61.90% and 19.05%, respectively. The standard deviation, arithmetic mean, and geometric mean of the accuracy measurements are 5.95%, 37.48%, and 37.00%, respectively. These values suggest a balanced spread, indicating that the accuracy measurements are not skewed and are symmetric. This implies a degree of inconsistency in the model’s behavior, as illustrated in Figure 2a, but it is not considered erratic. The top-performing model attained a 40% accuracy on the non-obsolete set \mathcal{U} suggesting a potential application for monitoring using the model.

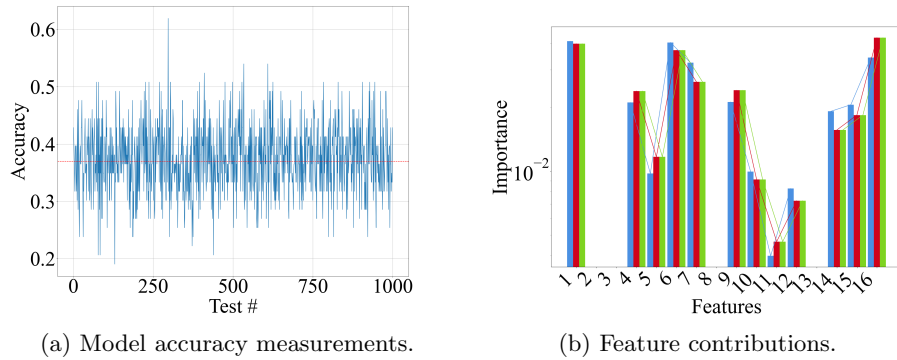


Fig. 2: Obtained results displaying feature contributions on a logarithmic scale, following the methodology outlined in Section 2.3. Blue bars indicate mean contributions, red bars show values from the best model, and green bars represent contributions from the top-performing model on non-obsolete instances.

The sorted list of highest average feature contributions aligns closely with the model achieving the highest accuracy on both \mathcal{O} and \mathcal{U} . In both lists, the top three elements include “repair complexity,” “several components are involved,” and “supplied as new, no longer supply, no longer newly supply, repairable only.” However, it contradicts the list derived from the combined options of experts “existing stock quantity,” “existing substitution product,” and “security block/product”), highlighting a lack of consensus on the subject. For detailed lists, refer to the GitHub repository: <https://github.com/Inars/A-Hybrid-Approach-for-Identifying-Obsolescence-Features>.

The results shown display a discrepancy within the decision making process of the SNCF RESEAU obsolescence management team. They emphasize the need of a unified decision making model that attempts to deduce consensus by

optimally finding a resolution to a given obsolescence case using the features that are most relevant.

4 Conclusion and Perspective

This paper presents a hybrid methodology designed to identify feature contributions that influence decision-making in obsolescence management resolutions. The proposed method highlights the divergence of expert opinions on the subject. During the dataset formation process, a noticeable discrepancy in the number of class instances was observed, necessitating a substantial increase in instances. Two key research areas emerge: data generation techniques to address data scarcity and k-shot learning models for resolving class imbalances.

References

- [Goodfellow et al., 2016] Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press.
- [Kotu and Deshpande, 2014] Kotu, V. and Deshpande, B. (2014). *Predictive analytics and data mining: concepts and practice with rapidminer*. Morgan Kaufmann.
- [Mei et al., 2022] Mei, K., Tan, M., Yang, Z., and Shi, S. (2022). Modeling of feature selection based on random forest algorithm and pearson correlation coefficient. In *Journal of Physics: Conference Series*, volume 2219, page 012046. IOP Publishing.
- [Montgomery et al., 2021] Montgomery, D. C., Peck, E. A., and Vining, G. G. (2021). *Introduction to linear regression analysis*. John Wiley & Sons.
- [Pingle, 2015] Pingle, P. (2015). *Selection of obsolescence resolution strategy based on a multi criteria decision model*. PhD thesis, Iowa State University.
- [Porter, 1998] Porter, G. Z. (1998). *An economic method for evaluating electronic component obsolescence solutions*. Boeing.
- [Renaud et al., 2018] Renaud, G., Abiven, Y.-M., Ta, F., Tran, Q., and Zhang, S. (2018). MACUP (Material for data ACquisition - UPgrade): Project Focusing on DAQ Hardware Architecture Upgrades for SOLEIL. In *Proc. of International Conference on Accelerator and Large Experimental Control Systems (ICALEPCS'17), Barcelona, Spain, 8-13 October 2017*, pages 1330–1334.
- [Rust et al., 2022] Rust, R. M., Elshennawy, A., and Rabelo, L. (2022). A literature review on mitigation strategies for electrical component obsolescence in military-based systems. *South African Journal of Industrial Engineering*, 33(1):25–38.
- [Sammut and Webb, 2011] Sammut, C. and Webb, G. I. (2011). *Encyclopedia of machine learning*. Springer Science & Business Media.
- [Sandri and Zuccolotto, 2010] Sandri, M. and Zuccolotto, P. (2010). Analysis and correction of bias in total decrease in node impurity measures for tree-based algorithms. *Statistics and Computing*, 20:393–407.
- [Singh and Sandborn, 2006] Singh, P. and Sandborn, P. (2006). Obsolescence driven design refresh planning for sustainment-dominated systems. *The Engineering Economist*, 51(2):115–139.
- [Zheng et al., 2015] Zheng, L., Terpenney, J., and Sandborn, P. (2015). Design refresh planning models for managing obsolescence. *Iie Transactions*, 47(12):1407–1423.
- [Zolghadri et al., 2021] Zolghadri, M., Addouche, S.-A., Baron, C., Soltan, A., and Boissie, K. (2021). Obsolescence, rarefaction and their propagation. *Research in Engineering Design*, 32(4):451–468.