Data-Driven Identification of Critical Components in Complex Technical Infrastructures Using Bayesian Additive Regression Trees

Xuefei Lu

Department of Energy, Politecnico di Milano, Italy. E-mail: xuefei.lu@polimi.it

Federico Antonello

Department of Energy, Politecnico di Milano, Italy. E-mail: federico.antonello@polimi.it

Piero Baraldi

Department of Energy, Politecnico di Milano, Italy. E-mail: piero.baraldi@polimi.it

Enrico Zio

Department of Energy, Politecnico di Milano, Italy.
MINES ParisTech, PSL Research University, CRC, Sophia Antipolis, France.
Eminant Scholar, Department of Nuclear Engineering, College of Engineering, Kyung Hee University, Republic of Korea.
E-mail: enrico.zio@polimi.it

Complex technical infrastructures are systems-of-systems characterized by hierarchical structures, made by thousands of interconnected components performing different functions associated to various domains. Given the difficulty of deriving their functional logic using traditional risk and reliability analysis methods, we address the problem of critical component identification from an innovative perspective, which exploits the large amount of available monitored data of operation. Specifically, we develop a data-driven framework of analysis which employs Bayesian additive regression trees and validate it on a synthetic case study, which mimics the complexity of a complex technical infrastructure.

Keywords: Critical Components Identification, Bayesian Additive Regression Trees, Complex Technical Infrastructure, Feature Selection.

1. Introduction

Complex Technical Infrastructures (CTIs) are large-scale systems of systems consisting of numerous mutually interconnected components. The various CTI systems perform different functions, use technologies from various domains, and are typically designed and built independently (Boardman and Sauser, 2006; Keating et al., 2008; Eusgeld et al., 2011; Zio, 2016). The identification of critical components in a CTI has become a priority for improving CTI reliability and availability, and reducing maintenance and operation costs. The traditional risk and reliability analysis approach for the identification of critical components is based on the use of Importance Measures (IMs), which quantify the contribution of components to a measure of system performance, such as, system reliability, unreliability, unavailability or risk. The computation of IMs requires the knowledge of the functional logic of the system in the form of a structure function, which is typically not known for CTIs due to their complexity and continuous transformations.

On the other hand, recent developments in sensors, signal processing systems and machine learning have opened up opportunities for analyzing the large amount of data available to support cost-effective and robust decision-making for design, operation and maintenance. In this context, the objective of the present work is to define an innovative data-driven framework of analysis for the identification of critical components based on the use of the substantial operational data collected from the CTI systems.

Specifically, the identification of the CTI critical components is formulated as a feature selection problem and addressed using Bayesian Additive Regression Trees (BART) (Bleich et al., 2014), which is a Bayesian non-parametric regression approach based on an ensemble of decision trees. The critical

Proceedings of the 29th European Safety and Reliability Conference.
Edited by Michael Beer and Enrico Zio

Published by Research Publishing, Singapore.
ISBN: 978-981-11-2724-3; doi:10.3850/978-981-11-2724-3_0714-cd

components are identified considering the posterior inclusion frequencies of the corresponding signals. The defined method is validated on a synthetic case study, which mimics the complexity of a real CTI.

The remainder of the work is as follows. Section 2 presents the problem setting. Section 3 reviews the feature selection methods, whereas Section 4 illustrates the BART method for critical components identification. Section 5 presents the results of its application to a synthetic case study. Section 6 summarizes the main findings of the work.

2. Problem Statement

We consider a CTI made by p components C_j, whose degradation and failure process is monitored by measuring signals $X_j \in \mathcal{R}, j = 1 \dots p$. The set of all monitoring signals are referred to as $\mathbf{X} = (X_1, \dots X_p) \in \mathcal{R}^p$ and the overall CTI safe(0)/failure(1) state as $Y \in \mathcal{Y}$ with $\mathcal{Y} = \{0, 1\}$. A large amount of data $\mathcal{D} = \{(\mathbf{x}^i, y^i)\}_{i=1}^n$ are collected during the CTI operation, containing the measurements $\mathbf{x}^i = (x_1^i, \dots, x_p^i)$ of p signals and the corresponding safe(0)/failure(1) states y^i of the CTI at n time instants.

The objective of the present work is to identify the CTI critical components $\mathbf{C}^\star = (C_{r_1}, \dots, C_{r_q}), 1 \leq r_1 \leq \dots \leq r_q \leq p$ to the CTI safe/failure state. To this aim, we consider the use of feature selection techniques for the identification of the subset $\mathbf{X}^\star = (X_{r_1}, \dots, X_{r_q})$ of the relevant monitoring signals.

3. Feature Selection Techniques

In general, feature selection techniques have the objective of identifying the subset of signals (features) $\mathbf{X}^\star$, which allows maximizing the classification accuracy of a learning machine performing the mapping $g : \mathcal{X} \to \mathcal{Y}$ (Genuer et al., 2010, 2015; Bolón-Canedo and Alonso-Betanzos, 2019).

Feature selection approaches fall into the three categories of wrapper, embedded and filter methods. Wrapper methods select an optimal subset of features using the learning machine itself, i.e., the learning machine is wrapped within the search algorithm which aims at identifying the feature subset providing the 'best' classification performance. Filter methods rank the features according to their statistical association (e.g., mutual information) with the response.

Embedded methods perform feature selection as part of the learning machine training. They select signals using importance indicators obtained during the training procedure, such as the node importance in decision trees (Blum and Langley, 1997; Guyon et al., 2005), and the regression coefficients in the Least Absolute Shrinkage and Selection Operator (LASSO) (Tibshirani, 1996).

As the feature selection literature is vast and rapidly growing, we refer to the works of Chandrashekar and Sahin (2014); Salcedo-Sanz et al. (2018); Stetco et al. (2019); Bolón-Canedo and Alonso-Betanzos (2019) for broader views.

4. Critical Components Identification Based on BART

The conjecture behind the use of feature selection for critical component identification is that if a signal is needed for the classification of the system failure/safe state, the component monitored by the signal is critical. In this work, we investigate the potentiality of an embedded feature selection method based on BART for critical components identification.

We consider the problem of building a learning machine $g : \mathcal{X} \to \mathcal{Y}$, which minimizes the classification error $P(Y \neq g(\mathbf{X}))$. According to Chipman et al. (2012), ensemble of regression trees are more capable of capturing interactions and non-linearities, as well as additive effects, than single trees. BART is a full Bayesian approach based on an ensemble of trees. Specifically, the BART model for classification assumes a probit transformation of a regression tree:

$$P(Y = 1|\mathbf{X}) = \Phi\left(\sum_{t=1}^{m} \mathcal{T}_t(\mathbf{X})\right), \qquad (1)$$

where Φ denotes the standard normal cumulative density function, $\mathcal{T}_t$'s are distinct binary regression trees. The prediction is given by the sum of m leaf values when recursing down all m trees.

BART considers a set of priors to provide regularization by preventing the domination of any single tree. Specifically, priors are assigned to the tree structure and the leaf parameters. The tree structure prior controls the size and shape of $\mathcal{T}_t$ through the probability of splitting a nonterminal node of certain depth. Usually the depth of $\mathcal{T}_t$ is kept small, e.g. less than 5. The splitting rule consists of two steps: 1) random selection of a feature to be split according to a probability distribution (e.g. discrete uniform or Bernoulli distribution), and 2) random choice of the splitting value among a set of values according to a uniform distribution. The prior on each leaf parameter follows a conjugate normal distribution $\mathcal{N}\left(\mu_\mu, \sigma_\mu^2\right)$, such that the

induced prior on $\mathbb{E}[Y|\mathbf{x}]$ is $\mathcal{N}\left(m\mu_\mu, m\sigma_\mu^2\right)$. The values of μ_μ and σ_μ are chosen such that $m\mu_\mu - k\sqrt{m}\sigma_\mu = y_{min}$, $m\mu_\mu + k\sqrt{m}\sigma_\mu = y_{max}$ for a preselected value of k. Large k and small σ_μ^2 yield more model regularization. The posterior distribution is, then, approximated via a Markov Chain Monte Carlo (MCMC) sampling (Kapelner and Bleich, 2016). Further details about BART can be found in Chipman et al. (2012).

4.1. BART-based Critical Components Identification

BART-based feature selection approaches have been proposed in Hill (2011); Bleich et al. (2014); Linero (2018). In this work, we adopt the approach of Bleich et al. (2014), which measures the importance of a feature X_j as the 'feature inclusion proportion', $\text{FIP}(X_j)$, i.e., the ratio between the number of times each feature is split, divided by the total number of feature splittings, in the model.

Bleich et al. (2014) propose a feature selection scheme based on a 'null' permutation distribution of the feature inclusion proportion $\text{FIP}(X_j)$ obtained by:

(i) permuting the output y^i to break its relationship with the features;
(ii) rebuilding the 'null' BART model using the permuted output and unpermuted features to obtain the null inclusion proportions of all features;
(iii) repeating steps (i) and (ii) several times using different random permutations of the data to estimate the distribution of null inclusion proportion of each feature.

The identification of the subset of critical features $\mathbf{X}^\star$ (thus, the subset of critical components $\mathbf{C}^\star$) is performed using the strategy of Bleich et al. (2014), according to which a feature X_j is selected if its average inclusion proportion $(\frac{1}{r}\sum_{i=1}^{r}\text{FIP}^i(X_j))$ over r replicates of BART, exceeds the $1-\alpha$ quantile of its own null distribution.

The open-source R packages bartMachine (Kapelner and Bleich, 2018) is used to perform the BART-based 'local' feature selection strategy of Bleich et al. (2014) (Section 4.1), which will be referred to as 'BART' in the latter sections.

5. Case Study

We consider a CTI formed by $p = 50$ components, in which each component can be in five states, $D \in \{1, 2, 3, 4, 5\}$ corresponding to healthy, partially degraded, degraded, very degraded and failure, respectively. The components perform transitions among the states at random times. Figure 1 shows the possible stochastic state transitions corresponding to: degradation (from $D = 1$ to $D = 2$, from $D = 2$ to $D = 3$ and from $D = 3$ to $D = 4$), partial restoration (from $D = 4$ to $D = 3$, from $D = 3$ to $D = 2$ and from $D = 2$ to $D = 1$), failure (from $D = 4$ to $D = 5$) and complete repair (from $D = 5$ to $D = 1$). Table 1 reports the time-invariant transition rates, $\lambda_j^{D' \to D''}$ of component j from the state D' to D'', $D' \neq D''$. Each component C_j is monitored by a signal X_j directly measuring its state D.

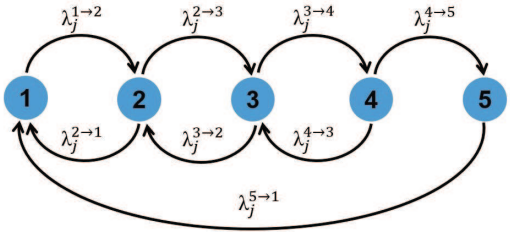

Fig. 1. State transitions of a CTI component

We assume that the CTI can fail due to two cascading failures:

(i) Component C_{11} performs a transition from state 4 to state 5, which can cause an ordered sequence of events leading to the transitions of components C_{12}, C_{13}, C_{14}, C_{15} and C_{16} into state 5 and the consequent failure of the CTI. The probability of failure propagation between any two components in the sequence is set to 0.95 and the time necessary for the malfunction propagation follows a uniform distribution in the interval $[1, 20]$ minutes;
(ii) Component C_{21} performs a transition from state 4 to state 5, which can cause an ordered sequence of events leading to the transitions of components C_{22}, C_{23}, C_{24}, C_{25} and C_{26} into the state 5 and the consequent failure of the CTI. The probability of failure propagation between any two components in the sequence is set to 0.95 and the time necessary for the malfunction propagation follows a uniform distribution in the interval $[1, 30]$ minutes.

The CTI critical components are those involved in the two cascading failures, i.e. C_{11} C_{12}, C_{13}, C_{14}, C_{15} ,C_{16}, and C_{21}, C_{22}, C_{23},

Table 1. Transition rates in hours^{-1}.

Component C_j	Transition rates			
$j = 1, 2, 3, 6 \ldots, 10,$ $11, 12, 17, 21, 22, 35, 36$	$\lambda_j^{1\to2} = 0.5$ $\lambda_j^{2\to1} = 0.5$	$\lambda_j^{2\to3} = 0.02$ $\lambda_j^{3\to2} = 0.01$	$\lambda_j^{3\to4} = 0.5$ $\lambda_j^{4\to3} = 0.4$	$\lambda_j^{4\to5} = 0.01$ $\lambda_j^{5\to1} = 0.2$
$j = 4, 5, 13, 14, 18, 19, 20, 23,$ $24, 27, \ldots, 34, 38, 39$	$\lambda_j^{1\to2} = 0.3$ $\lambda_j^{2\to1} = 0.3$	$\lambda_j^{2\to3} = 0.005$ $\lambda_j^{3\to2} = 0.01$	$\lambda_j^{3\to4} = 0.4$ $\lambda_j^{4\to3} = 0.4$	$\lambda_j^{4\to5} = 0.01$ $\lambda_j^{5\to1} = 0.2$
$j = 15, 16, 25, 26, 37,$ $40, \ldots, 50$	$\lambda_j^{1\to2} = 0.4$ $\lambda_j^{2\to1} = 0.4$	$\lambda_j^{2\to3} = 0.005$ $\lambda_j^{3\to2} = 0.01$	$\lambda_j^{3\to4} = 0.4$ $\lambda_j^{4\to3} = 0.4$	$\lambda_j^{4\to5} = 0.01$ $\lambda_j^{5\to1} = 0.2$

C_{24}, C_{25} , C_{26}.

The CTI behavior is simulated for 720 days and the signals $X_j, j = 1 \ldots 50$ assessing the j-th component degradation state are collected every 2 hours, as well as the corresponding CTI safe(0)/failure(1) state Y. Therefore, a dataset $\mathcal{D} = \{(\mathbf{x}^i, y^i)\}_{i=1}^n$ formed by $n = 8642$ patterns is obtained. The simulated dataset is unbalanced, being the fraction of positive patterns ($y^i = 1$) over the total number n of simulated patterns equal to 5.3%.

5.1. Critical Components Identification

We consider the following three performance metrics to quantify the ability of identifying the critical components of the BART-based (Section 4.1) method:

$$\text{precision}^c = \frac{TP^c}{TP^c + FP^c}, \tag{2}$$

$$\text{recall}^c = \frac{TP^c}{TP^c + FN^c}, \tag{3}$$

$$F_1^c = \frac{2 \cdot \text{precision}^c \cdot \text{recall}^c}{\text{precision}^c + \text{recall}^c}, \tag{4}$$

where TP^c denotes the number of components correctly identified as critical; FP^c the number of components incorrectly identified as critical; FN^c the number of components incorrectly identified as non-critical. Therefore, the precision^c indicates the fraction of components correctly identified as critical over all the selected components; the recall^c indicates the fraction of components correctly identified as critical over all the critical components; the F_1^c score is a summary score that balances the previous two. The values of precision^c, recall^c and F_1^c fall within the range of $[0, 1]$: the larger the value, the more satisfactory the performance.

The application of the proposed BART feature selection strategy requires to properly set the parameter α (Section 4.1). This is done by adopting a trial-and-error approach in which the BART feature selection is repeated using various values of α and the performances of the BART classifiers built using the selected features are evaluated on a validation set. We have considered the F_1 metric to evaluate the classification performance and used 50% of the patterns of $\mathcal{D}$ for feature selection, 40% for the classifier training, the remaining 10% for estimating their classification performance. The most satisfactory classification performance ($F_1^p = 1$) has been obtained with $\alpha = 0.15$.

Using this setting, the proposed method identifies 10 among the 12 critical components (recall$^c = 0.833$, precision$^c = 1$, $F_1^c = 0.909$). The CTI critical components C_{12} and C_{22}, which are at the beginning of the two cascading failures, are not selected. This is due to the fact that they are characterized by a larger number of failures (80 and 51 respectively), which do not lead to the system failures, than the other critical components (23.4 on average).

5.2. Robustness of the Critical Components Identification Method

The robustness of the proposed method has been verified when the number of CTI components (p) is increased from 50 to 200 by adding

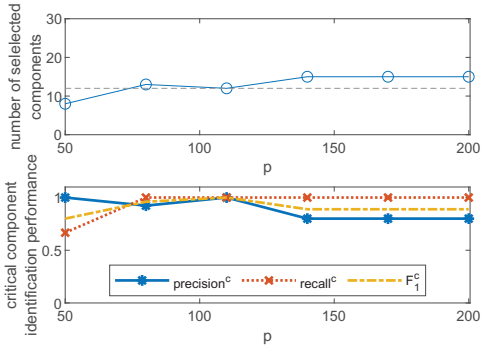

Fig. 2. Critical components identification performances, when the number of CTI components p increases from 50 to 250.

noncritical components. Figure 2 shows the obtained performance in terms of critical components identification. Notice that when p increases, all critical components are identified (recall^c = 1), although few non-critical components are also selected (precision^c < 1). Furthermore, when the total number of components exceeds 140, the performance becomes stable with the identification of a set of 15 components, which includes 3 noncritical components.

6. Conclusion

This work proposes a data-driven method for the identification of CTI critical components in those cases in which the system functional logic structure is unknown. The method is based on the application of a feature selection technique based on BART to operational signal data. Its application to a synthetic case study has shown its capability of identifying most of the CTI critical components. Its main limitations are the difficulty of identifying critical components at the beginning of cascading failures and the tendency of identifying few non-critical components when the overall number of CTI components increases. Nevertheless, the obtained results encourage the use of data-driven methods for investigating the risk and reliability of CTIs, whose components are normally monitored.

References

Bleich, J., A. Kapelner, E. I. George, and S. T. Jensen (2014). Variable selection for bart: An application to gene regulation. Annals of Applied Statistics 8(3), 1750–1781.

Blum, A. L. and P. Langley (1997). Selection of relevant features and examples in machine learning. Artificial intelligence 97(1-2), 245–271.

Boardman, J. and B. Sauser (2006). System of Systems - the meaning of of. In 2006 IEEE/SMC International Conference on System of Systems Engineering, Number April, Los Angeles, CA, USA, pp. 118–123. IEEE.

Bolón-Canedo, V. and A. Alonso-Betanzos (2019). Ensembles for feature selection: A review and future trends. Information Fusion 52, 1–12.

Chandrashekar, G. and F. Sahin (2014). A survey on feature selection methods. Computers and Electrical Engineering 40(1), 16–28.

Chipman, H. A., E. I. George, and R. E. McCulloch (2012). BART: Bayesian additive regression trees. Annals of Applied Statistics 6(1), 266–298.

Eusgeld, I., C. Nan, and S. Dietz (2011). System-of-systems approach for interdependent critical infrastructures. Reliability Engineering & System Safety 96(6), 679–686.

Genuer, R., J. M. Poggi, and C. Tuleau-Malot (2010). Variable selection using random forests. Pattern Recognition Letters 31(14), 2225–2236.

Genuer, R., J.-M. Poggi, and C. Tuleau-Malot (2015). VSURF: Variable selection using random forests. The R Journal 7(December), 19–33.

Guyon, I., S. Gunn, A. Ben-Hur, and G. Dror (2005). Result Analysis of the NIPS 2003 Feature Selection Challenge. In L. K. Saul, Y. Weiss, and L. Bottou (Eds.), Advances in Neural Information Processing Systems 17, pp. 545–552. MIT Press.

Hill, J. L. (2011). Bayesian nonparametric modeling for causal inference. Journal of Computational and Graphical Statistics.

Kapelner, A. and J. Bleich (2016). bartMachine: Machine Learning with Bayesian Additive Regression Trees. Journal of Statistical Software 70(4).

Kapelner, M. A. and J. Bleich (2018). bartMachine: Bayesian Additive Regression Trees. https://cran.r-project.org/package=bartMachine.

Keating, C. B., J. J. Padilla, and K. Adams (2008). System of systems engineering requirements: challenges and guidelines. Engineering Management Journal 20(4), 24–31.

Linero, A. R. (2018). Bayesian Regression Trees for High-Dimensional Prediction and Variable Selection. Journal of the American Statistical Association 113(522), 626–636.

Salcedo-Sanz, S., L. Cornejo-Bueno, L. Prieto, D. Paredes, and R. García-Herrera (2018). Feature selection in machine learning prediction systems for renewable energy applications. Renewable and Sustainable Energy Reviews 90, 728–741.

Stetco, A., F. Dinmohammadi, X. Zhao, V. Robu, D. Flynn, M. Barnes, J. Keane, and G. Nenadic (2019). Machine learning methods for wind turbine condition monitoring: A review. Renewable Energy 133, 620–635.

Tibshirani, R. (1996). Regression Selection and Shrinkage via the Lasso. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 58(1), 267–288.

Zio, E. (2016). Challenges in the vulnerability and risk analysis of critical infrastructures. Reliability Engineering and System Safety 152, 137–150.