POLITECNICO MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

Executive Summary of the Thesis

Addressing Data Scarcity for Machine-Learning-based Failure Management in Microwave Networks

Laurea Magistrale in Telecommunication Engineering - Ingegneria delle Telecomunicazioni

Author: Matteo Conci

Advisor: Prof. Francesco Musumeci

Co-advisors: Nicola Di Cicco, Mëmëdhe Ibrahimi

Academic year: 2022-2023

---

1. Introduction

Failure management in communication networks is a critical issue nowadays, as a single failure in the network can lead to service disruption for thousands or millions of users at the same time. Therefore, preventing failures from occurring is a crucial task for network operators to meet Service Level Agreements (SLAs) to its customers and this requires network's data collection and analysis. Machine Learning (ML) is a possible approach used to automate and speed up the whole network management process by leveraging data retrieved monitoring the network. In our work, we take into consideration failure management in microwave networks, focusing on the failure-cause identification problem. Specifically, we use supervised machine learning models to address the classification of hardware failure in microwave networks under the condition where there is lack of abundant data from the field. The challenge of addressing these issues with a scarce number of data is generally referred to as "data scarcity". This problem harms the training and thus the performance of ML algorithms, which need a large amount of labelled data to be trained efficiently.

The main contributions of this work can be listed as follows:

- We model the hardware failure identification problem as a machine learning classification problem.
- We use different supervised learning models to identify classes (i.e., failure causes) where classification performance is poor, showing the correlation between data scarcity and poor model performance.
- We investigate different methodologies to deal with the data scarcity problem, to understand which is the most appropriate for our scope.
- We conduct an in-depth analysis on synthetic data generation as a methodology to address data scarcity, proposing a procedure to identify which classes to actually generate synthetic data on and with what percentage.

The rest of the document is organized as follows. In *Section 2* we present an overview of previous work on the use of ML for failure management in microwave networks, and on the use of synthetic data generation to deal data scarcity problem. In *Section 3* we introduce our problem and data. In *Section 4* we present the baseline supervised learning models adopted to identify classes (i.e., failure causes) where classification performance is poor. In *Section 5* we present the ML methodologies adopted to address the problem of data scarcity when performing failure-cause identification in microwave networks. In *Section 6* we conduct the in-depth analysis on synthetic data generation with SMOTE. In *Section 7* we present our final results, and in *Section 8* we report the conclusion on this work.

2. Related works

Several works have investigated the use of ML techniques to address failure management in microwave networks. In [3] the authors give an overview on supervised and semi-supervised learning approaches for automated failure-cause identification in microwave networks, showing that supervised ML enables very accurate failure identification reaching 93% classification accuracy. Then, they addressed data scarcity investigating a semi-supervised learning approach to automate labeling procedure, based on autoencoders-like Artificial Neural Networks. Other works address data scarcity by using different methodologies, such as transfer learning [4], active learning [1], data synthesis via generative adversarial network or via variational autoencoders, and data augmentation via SMOTE [2].
To the best of our knowledge, no existing work has investigated the problem of data scarcity when performing hardware failure-cause identification in microwave networks, comparing different methodologies to enhance the performance of supervised learning algorithms. Another valuable aspect in our paper is that we conduct an in-depth analysis aimed to identify on which classes and with what percentages to generate the synthetic data.

3. Problem Statement and Data

We model the failure-cause identification problem as a supervised multi-class classification problem. The data we use are provided by SIAE Microelettronica and are a collection of hardware failures states taken from a real microwave network, where devices alarms status is monitored by a Network Management System (NMS) with 15-minutes intervals. These failures events have been analysed by domain experts and labelled in 4 macro-categories and 22 micro-categories of failures. To make it suitable for the machine learning algorithms, we undertook some preprocessing actions, obtaining a resulting dataset composed of 1045 entries, i.e., 1045 different failure events, distributed in 17 different categories, a.k.a classes, of hardware failures, as shown in the following Table 1:

MACRO CATEGORY	MICRO CATEGORY	CATEGORY	DATA POINTS
0	0	**0**	13
0	1	**1**	7
0	2	**2**	21
0	3	**3**	8
0	4	**4**	264
0	5	**5**	89
1	0	**6**	87
1	1	**7**	6
1	2	**8**	130
1	3	**9**	91
2	0	**10**	29
2	1	**11**	20
2	3	**12**	15
3	0	**13**	43
3	1	**14**	39
3	2	**15**	150
3	3	**16**	33

Table 1: Categories of hardware failures in microwave network. Due to confidentiality reasons, we do not report the description of the specific failures, and instead provide a masked representation.

As shown in Table 1, the dataset is affected by data scarcity, since for several failure categories there are only few data points available.

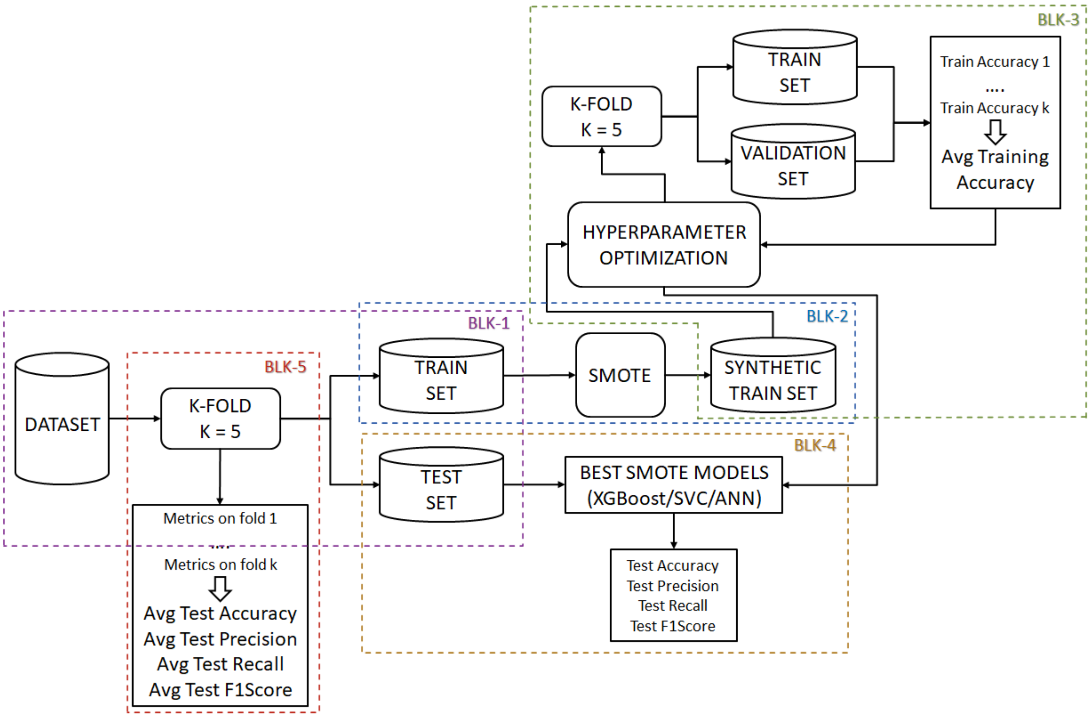

Figure 1: SMOTE pipeline with hyperparameters optimization.

4. Failure-Cause Identification with baseline Machine Learning models

In this section we present the baseline ML models used in this work, i.e., *XGBoost*, *Support Vector Classifier* (SVC) and *Artificial Neural Network* (ANN).

XGBoost is a gradient boosting algorithm that uses decision trees ensembles as base learners, i.e., an individual learner of the ensemble, to iteratively improve the predictions of the model.

SVC is a specific type of SVM (Support Vector Machine) that is optimized for classification tasks with linearly separable data. The goal of SVC is to find the optimal hyperplane that separates the data into different classes by maximizing the margin, that is the maximum distance between data points of different classes.

ANN is a machine learning model that is inspired by the structure and function of the human brain. It is composed of interconnected nodes, or neurons, that process and transmit information.

These baseline models are used to determine the baseline performance on Accuracy, Precision, Recall, and F1-score metrics, that we aim to enhance with the methodologies presented in Section 5.

5. Machine Learning methodologies to address Data Scarcity

In this section we present the methodologies used to address the data scarcity. We will present *Synthetic Minority Over-sampling Technique (SMOTE)*, *Transfer Learning (TL)*, *Auxiliary-Task Learning (ATL)*, and *Denoising Autoencoders (DAE)*.

SMOTE is a data augmentation technique that works by creating synthetic samples from the minority class, that is over-sampled by taking each minority class sample and introducing synthetic examples along the line segments joining any/all of the k minority class nearest neighbors. With this methodology our goal is to address data scarcity enriching the pre-processed dataset with synthetic data generated by SMOTE. Figure 1 illustrates the proposed pipeline for the training and testing of this methodology:

1. **BLK-1**: we split the dataset in 5 folds using *Stratified k-fold Cross-Validation* to validate the aforementioned performance metrics.
2. **BLK-2**: on the train set we apply SMOTE,

obtaining a new train set composed by real data and synthetic data.

3. **BLK-3**: we use the synthetic train set for the hyperparameters optimization; each hyperparameter set chosen is cross-validated using *Stratified k-fold Cross-Validation* to optimize the chosen performance metric that is the accuracy.
4. **BLK-4**: we test the *Best SMOTE Models*, i.e., XGBoost, SVC and ANN trained on the *Synthetic Train Set* with best set of hyperparameter, on the test set. We store the performance metrics obtained on this split.
5. **BLK-5**: at the end of the five iterations, we compute the average performance metrics for each model.

TL is a methodology based on the concept of knowledge transfer, i.e., unlike in traditional machine learning models which are trained from scratch, here the knowledge acquired by a model on one or more source tasks is transferred to a target task. The goal of applying Transfer Learning is to address data scarcity by first training an ANN model to classify macro-categories and then re-training a TL model derived from it, able to classify the 17 defined categories of hardware failures.

ATL is an extension of Multi-task Learning where the main-tasks, i.e., the ones producing the required outputs, are flanked by the auxiliary-tasks, i.e., easy to learn tasks of minor importance, and then their loss functions are combined into a final multi-task loss function that allow to benefit from both contributions. In our work the main-task is the classification on the 17 categories of hardware failure, while the auxiliary-task is the macro-category classification. We expect that this defined loss function will benefit from the auxiliary loss to enhance the classification on the 17 categories of hardware failures.

DAE is an autoencoder (AE), i.e., a self-supervised learning model trained to encode the input x into some representation c(x) so that the input can be reconstructed from that representation, that receives as input corrupted data by adding noise. DAE can address data scarcity learning not only a compressed representation of the features but also the structural relationships within them. We expect that this model will be able to learn relationships between the alarms, such as:

- if alarm “a” or set of alarms “A” are on, also alarm “b” or set of alarms “B” are on.
- if alarm “a” or set of alarms “A” are off, also alarm “b” or set of alarms “B” are off.
- if alarm “a” or set of alarms “A” are on, alarm “b” or set of alarms “B” are off.
- if alarm “a” or set of alarms “A” are off, alarm “b” or set of alarms “B” are on.

6. SMOTE Analysis

In this section we present an in-depth analysis conducted on SMOTE.

The first step of this analysis is to identify a metric that, based on the per-class F1-score, defines the criticality of the single class, i.e. how poor is the performance score on that class of hardware failures. The chosen metric is the first and second quantile score computed on the per-class F1-score given by the three *Best Baseline Models*, i.e., XGBoost, SVC and ANN trained with their respective best sets of hyperparameters. In particular, we assigned a color to each class according to its F1-score value and the quantiles, following this criteria:

- RED: if F1-score is below than or equal to first quantile (<=25%). This means that the class is considered as very critical.
- ORANGE: if F1-score is below than the second quantile (<50%). This means that the class is considered as critical.
- GREEN: if the F1-score is greater than or equal to the second quantile (=>50%). This means that the class is considered as not critical.

Once we have for each class the colors given by the three considered classifiers, we assign a global class color according to the following criteria:

- If the tree classifiers give a unique class color, that is the global color we assigned to that class.
- If the three classifiers give two different class colors, we assigned as global class color to that class the predominant one, i.e., the one given by largest number of classifiers.
- If the tree classifiers give a different class color, we assigned as global class color to that class the orange color.

The second step of this analysis is to define two

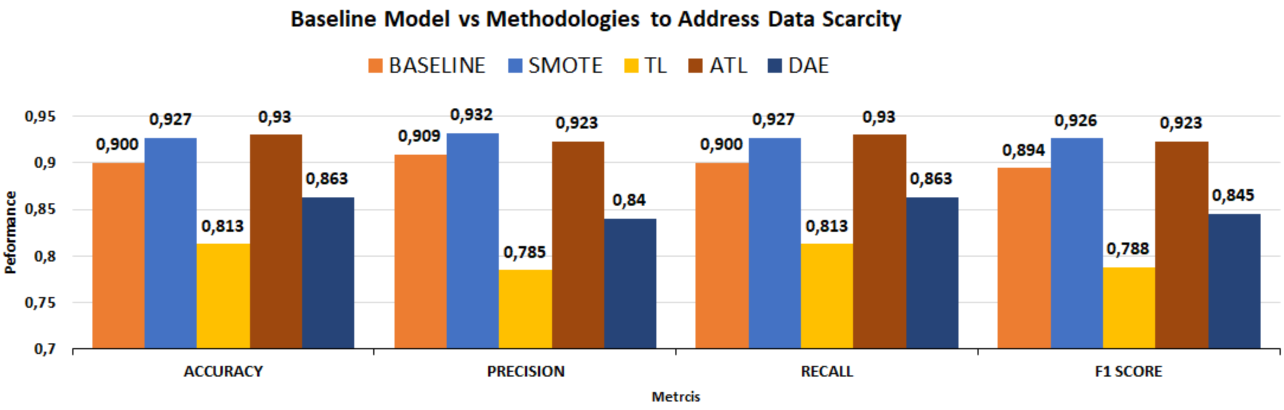

Figure 2: Performance comparison.

"sampling strategies" by which SMOTE will be applied:

- **Critical-classes**: which considers only the critical classes, i.e., the ones that have RED or ORANGE as global class color.
- **All-classes**: which considers all the classes independently from the global class color assigned.

With the purpose of identifying the amount of synthetic data to be generated on the classes, we establish four "percentages" of synthetic data generation, i.e, "12%", "35%", "50%" and "65%", where each percentage represents the number of data points, i.e., real data points plus synthetic data points, in the train set of a specific class. We compute these percentages taking as reference the class with the highest number of data points, i.e., class 4 with 264 data points.

The third step of this analysis is to understand whether it is possible to find for each class of hardware failure a combination of strategy and percentage that allows for improvements regardless of the type of classifier used. For this reason, we propose a procedure based on selecting the best strategy common to the three classifiers by taking as a starting point the best combinations found for each individual classifier.

7. Results

In this section we perform numerical evaluation of the Machine Learning (ML) methodologies adopted to address the problem of data scarcity when performing failure management in microwave networks. Figure 2 shows the results in terms of Accuracy, Precision, Recall, and F1-score we obtain using the presented methodologies. The main takeaways from the use of these methodologies are:

1. The use of TL or DAE does not improve the data scarcity problem. In fact, with TL we obtain an F1-score of 78.8% and with DAE an F1-score of 84.5%, while with the baseline classifier we obtain an F1-score of 89.4%.
2. The use of SMOTE or ATL does improve the data scarcity problem. In fact, with SMOTE we obtain an F1-score of 92.6% and with ATL an F1-score of 92.3%, while with the baseline classifier we obtain an F1-score of 89.4%.
3. Analyzing the per-class F1-score obtained with SMOTE and ATL, we observe that with SMOTE we have higher F1-score improvement on classes characterized by few points than with using ATL, as shown in the following Table 2:

CLASS	BASELINE	ATL	SMOTE
0	0.853	0.853	0.905
1	0.2	0.2	0.3
2	0.646	0.788	0.833
3	0.693	0.6	0.893
4	0.912	0.963	0.962
5	0.921	0.949	0.959
6	0.977	0.966	0.967
7	0.4	0.533	0.4
8	0.977	0.981	0.989
9	0.955	0.979	0.965
10	0.864	0.964	0.982
11	0.876	0.949	0.914
12	0.871	0.92	0.86
13	0.673	0.684	0.692
14	0.776	0.808	0.836
15	0.928	0.935	0.925
16	0.769	0.817	0.816

Table 2: Comparison between per-class F1-score values obtained using baseline model ("BASELINE"), a ML model built on Auxiliary-Task Learning ("ATL") methodology and a ML model built on SMOTE methodology.

From these results we obtain that SMOTE is the most suitable technique to deal with data scarcity, so we focus on it the in-depth analysis, summarized below.
We identify these classes as critical, i.e. class 1, 2, 3, 7, 11, 13, 14 and 16, and applying the strategies and percentages defined to SMOTE methodology, we obtain these main takeaways:

1. Regardless of the baseline classifier considered, in most cases synthetic data help to improve the global performance of the baseline classifier.
2. There are combinations of strategies and percentages that achieve better results than SMOTE without the use of the combination.
3. Generally, generating synthetics data on all the classes performs better than generating it only on critical classes. This tells us that to improve the overall performance of the classifiers, in generating the synthetic data it is not enough to consider only the critical classes but it is necessary to consider all classes, in order not to get the dataset unbalanced again towards the critical classes.
4. Generally, the highest performance is obtained when considering low percentages of synthetic data, while moving to higher percentages the performance drops.

The last result we obtain is the one regarding the proposed procedure to select the best per-class combination of strategy and percentage that allows for improvements regardless of the type of classifier adopted. From this analysis, we select the classes on which to generate the synthetic data with the respective number of synthetic data to be added. Applying this procedure we obtain a per-class F1-score improvement on 6 of 8 critical classes for XGBoost, on 2 of 8 critical classes for SVC, and on 4 of 8 critical classes for ANN. These results show us how our proposed procedure for selecting classes and percentages actually enhances the overall results, improving the F1-score of classes characterized by a low number of data points and thus affected by data scarcity.

8. Conclusion

We adopted different ML-based classification methodologies to address data scarcity in failure-cause identification in microwave networks. Overall, we obtained that SMOTE is the best performing methodology for this scope. Our in-depth analysis mainly show how generating synthetic data through SMOTE helps to improve classification on classes affected by data scarcity, as long as we understand on which classes and with what percentages to generate the synthetic data.

References

[1] Wei Guo, Xiang Zha, Kun Qian, and Tao Chen. Can active learning benefit the smart grid? a perspective on overcoming the data scarcity. In *2019 IEEE 2nd International Conference on Electronics and Communication Engineering (ICECE)*, pages 346–350, 2019.

[2] Haibo He and Edwardo A. Garcia. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9):1263–1284, 2009.

[3] Francesco Musumeci, Luca Magni, Omran Ayoub, Roberto Rubino, Massimiliano Capacchione, Gabriele Rigamonti, Michele Milano, Claudio Passera, and Massimo Tornatore. Supervised and semi-supervised learning for failure identification in microwave networks. *IEEE Transactions on Network and Service Management*, 18(2):1934–1945, 2021.

[4] Sun Zhang, Xin Hu, Zhijun Liu, Linlin Sun, Kang Han, Weidong Wang, and Fadhel M. Ghannouchi. Deep neural network behavioral modeling based on transfer learning for broadband wireless power amplifier. *IEEE Microwave and Wireless Components Letters*, 31(7):917–920, 2021.