This article has been accepted for publication in IEEE Communications Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LCOMM.2024.3411768

A Scoring Framework for the Unsupervised Identification of the Relevance of Metrics in Cell Degradation Clusters

Javier Villegas, Sergio Fortes, *Senior Member, IEEE*, Juan Cantizani-Estepa, Javier Rasines Suarez,
Raúl Martín Cuerdo, and Raquel Barco

***Abstract*—The complexity of cellular networks makes them difficult to manage and prone to failures. This has led to the application of artificial intelligence mechanisms for failure management tasks. Nevertheless, the availability of labeled data on failure cases is limited, making unsupervised techniques the most relevant to apply. However, these techniques require network experts to analyze the results through a costly process to link them to specific cases. Intending to ease the labeling of cases, the present work proposes a framework to determine the relevant metrics for any clustering. Furthermore, the framework is evaluated using data from real cellular networks.**

***Index Terms*—Cellular networks, Failure management, Labeling, Feature ranking, Jensen-Shannon Divergence**

I. Introduction

Cellular networks are becoming increasingly complex with the coexistence of different radio technologies and will grow further with the adoption of future technologies. The increasing number of technologies, network elements, and user equipment make cellular network management a growingly difficult task for operators, which might result in performance degradation due to undetected problems. Moreover, newly deployed technologies are expected to present their own possible failures [1]. In order to make network operation efficient and sustainable, the paradigm of Zero-Touch network and Service Management (ZSM) [2] is proposed by the European Telecommunications Standards Institute (ETSI). This paradigm aims to completely automate the management of networks and the services on top, reducing Operation Expenditure (OpEx) and Capital Expenditure (CaPex).

Here, the identification of the root cause of a problem can be faced with different Machine Learning (ML) approaches. Some of the most common are analyzing network logs [3] and the use of supervised classifiers with additional information, such as users' positioning [4], [5]. However, classifiers for network failures have the shortcoming of needing vast amounts of labeled data which is most commonly unavailable. For this reason, failure detection techniques using unsupervised learning are the preferred approach. Here, clustering methods can group samples by types of network status based on the statistical relationships found in the metrics (e.g., counters, Key Performance Indicators (KPIs), alarms) [6], [7], [8].

Nevertheless, clustering algorithms' output needs to be manually labeled by network experts to be useful for failure troubleshooting. To solve this, algorithms have been developed to automatically label clusters [9], [10] in the field of computer science. However, these do not consider the nature of metrics, complicating their application in cellular networks.

Principal Component Analysis (PCA) [11] can be used to determine the most important variables in the data. However, this needs selecting the number of components, which can vary with the data and special treatment for time series [12]. Furthermore, time series relationships can be explained using Dynamic Factor Analysis (DFA)[13], although this also requires modeling based on the data.

Therefore, with the objective of easing the task of characterizing cell degradations, this work proposes a framework to establish the relevance of the metrics on a cluster basis. It aims to provide the relevance of metrics across multiple iterations of unsupervised failure detection on cellular networks from an initial set of parameters. Thus, providing a higher level of automation compared to previous approaches. Furthermore, this work also proposes an automatic labeling scheme using the metrics' relevance based on generic, predefined rules. Then, the results are evaluated against the labels defined by network experts for a set of real cellular network data.

In this way, the remainder of this manuscript is organized as follows. In Section II the proposed framework to automatically characterize clusters is presented. Then, in Section III the results obtained using clusters obtained from the data of a live network is presented. Lastly, the conclusions of this work are presented in Section IV.

II. Framework

The proposed framework for the identification of the relevant metrics is depicted in Fig. 1. This work parts from a

This work has been partially funded by contract Ref.- 8.06/5.59.5705-3 IDEA, “Desarrollo de casos de uso para el diseño, optimización y dimensionado de redes móviles - Líneas B1 y D1”, in the framework of incentives by Agencia IDEA, by Ministerio de Asuntos Económicos y Transformación Digital and European Union - NextGenerationEU within the framework “Recuperación, Transformación y Resiliencia under the project MAORI, Ministerio de Universidades (grant FPU21/04472) and Universidad de Málaga through the “II Plan Propio de Investigación, Transferencia y Divulgación Científica”. (*Corresponding author: Sergio Fortes*).

Javier Villegas, Sergio Fortes, Juan Cantizani Estepa and Raquel Barco are with the Instituto Universitario de Telecomunicación (TELMA). Universidad de Málaga. CEI Andalucía TECH E.T.S.I. Telecomunicación (e-mail: {jvc, sfr, jce, rbm}@ic.uma.es).

Javier Rasines Suarez is with Ericsson - GAIA Sweden (e-mail: javier.rasines@ericsson.com).

Raúl Martín Cuerdo is with Ericsson - NDO SW R&D (e-mail: raul.r.martin@ericsson.com).

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

This article has been accepted for publication in IEEE Communications Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LCOMM.2024.3411768

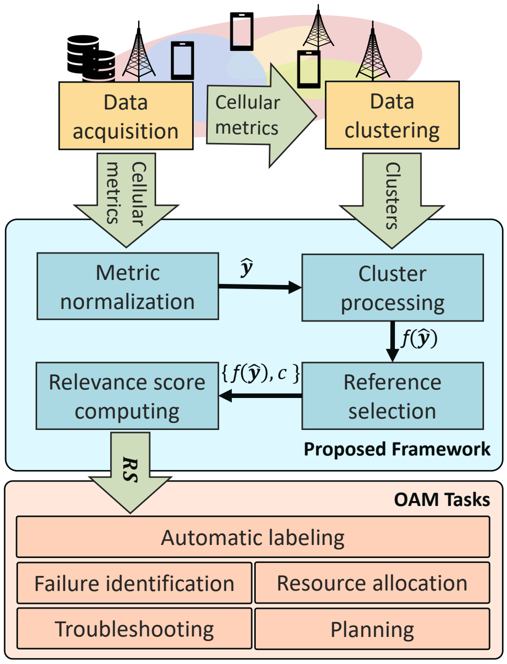

Fig. 1. Proposed framework architecture.

scenario where the cellular network metrics are acquired and clustered based on cell status by external systems. From this point, "Metric normalization" normalizes the acquired metrics. Then, "Cluster processing" associates metrics with their corresponding clusters and computes the probability distribution function for each of the metrics in each of the clusters. In this way, the metrics are in the desired format regardless of the transformations applied to achieve the clusters. Next, "Reference selection" establishes the cluster with the lowest degradation as cluster with normal status. Finally, "Relevance score computing" estimates the relevance for each metric.

A. Metric normalization and cluster processing

The first step of the proposed framework is the metric normalization of the cellular data. Accordingly, the metrics have been normalized using a process of min-max scaling, a process that is usually present in most data pipelines. In this particular case, the present work relies on a normalization process previously implemented and refined by a team of network experts. However, it is worth noting that the complementary of the normalized metrics is used for those considered degraded at low values. This is defined as

$$\hat{y}_i := 1 - \hat{y}_i. \tag{1}$$

In this way, the closer a metric is to 1 the more degraded the metric is, simplifying the interpretation of any metric's values.

Furthermore, the input data of the framework are daily samples of several time series of metrics derived from performance counters gathered hourly at cell level. In this manner, each sample contains 24 hours of values of several metrics obtained from a single cell during a day. In order to overcome the difficulties introduced by time series, the data of each cluster is used to compute its probability distribution for each of the metrics. Then, the Jensen-Shannon Divergence (JSD)[14] is used to generate a relevance ranking for each of the clusters.

B. Reference selection

This block is responsible for establishing a cluster as a reference for the calculation of the relevance metric. In order to estimate the importance of a metric, it is necessary to use a second distribution to compare against, therefore, it is required to establish a cluster as a reference. In this manner, clusters are compared pair-wise against the reference, allowing the system to score the clusters regardless the total number of these. Here, it is assumed that a cluster with samples without degradation exists and can be taken advantage of to highlight the degradation of other clusters. This will help to fulfill the framework's objective of easing the labeling of clusters by network experts.

Therefore, the cluster identified as the one with a lower level of degradation is selected as a reference to compute the relevance of the metrics. This cluster can be chosen by taking advantage of the normalization of the metrics, where high values represent a higher level of degradation. Hence, the level of degradation of each cluster can be estimated as the sum of the average values of each metric. Therefore, the selected cluster as reference is expressed as

$$c = \arg\min_{j}(\sum_{i} \mathrm{mean}(\widehat{\mathbf{y}}_{i,j})), \tag{2}$$

where $\widehat{\mathbf{y}}_{i,j}$ is the normalized samples of the metric i of cluster j and c is the cluster selected as the reference. Here, the sum of the means for each metric is the accumulated mean degradation.

C. Relevance Score Computing

This block performs the calculations to obtain a metric to measure the relevance of each of the metrics in each cluster. For this reason, the use of a new figure of merit named RS (Relevance Score) is proposed and defined as

$$\mathrm{rs}_{i,j} = \mathrm{jsd}_{i,j}\,\mathrm{sgn}(\widehat{\mathbf{y}}_{i,c}^{75th} - \widehat{\mathbf{y}}_{i,j}^{75th})\sqrt{|\widehat{\mathbf{y}}_{i,c}^{75th} - \widehat{\mathbf{y}}_{i,j}^{75th}|}, \tag{3}$$

where $\mathrm{rs}_{i,j}$ is the RS for the metric i in the cluster j, $\widehat{\mathbf{y}}_{i,j}^{75th}$ is its 75th percentile value, sgn is the sign operator and jsd is the Jensen-Shannon Distance [14]. Here, the 75th percentile aims to extract a representative value of the metric during cell usage. This is chosen based on the rationale that cells are usually loaded for 10-12 hours (i.e., daytime) and are expected to be highly loaded during the busy hours, hence, values of interest are expected to be between the 50th percentile and the 90th percentile. Moreover, the jsd is defined as

$$\mathrm{jsd}_{j,i} = \sqrt{\mathrm{JSD}(f(\widehat{\mathbf{y}}_{i,j})||f(\widehat{\mathbf{y}}_{i,c}))}, \tag{4}$$

where JSD is the Jensen-Shannon Divergence obtained from comparing the probability distribution of the metric i from cluster j, $f(\widehat{\mathbf{y}}_{i,j})$ and the one for the metric i from cluster c, $f(\widehat{\mathbf{y}}_{i,c})$. Furthermore, the JSD can be defined as

$$\mathrm{JSD}(P||Q) = \frac{1}{2}(\mathrm{D}(P||M) + \mathrm{D}(Q||M)), \tag{5}$$

where P and Q are probability distributions whose divergence will be computed, and M is a probability distribution defined following

$$M = \frac{1}{2}(P + Q), \tag{6}$$

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

This article has been accepted for publication in IEEE Communications Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LCOMM.2024.3411768

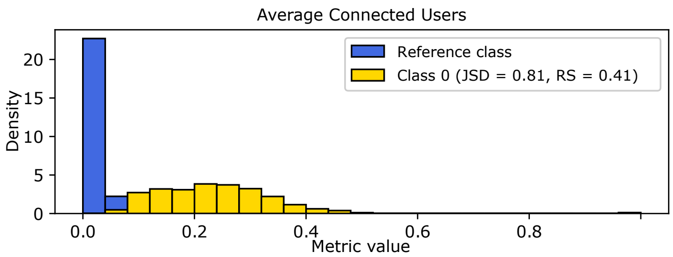

Fig. 2. Comparison between the JSD and the RS.

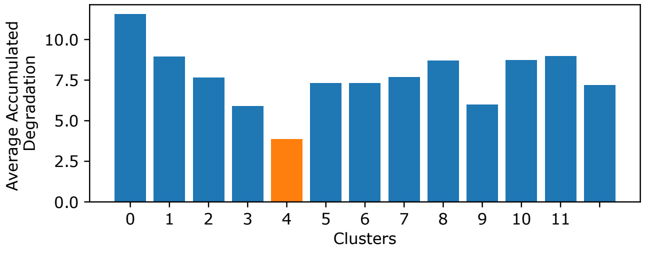

Fig. 3. Accumulated mean degradation for each cluster.

where D is the Kullback-Leibler Divergence [15], which is defined as

$$\mathrm{D}(P||Q) = \sum_{x\in X} P(x)\log(\frac{P(x)}{Q(x)}). \tag{7}$$

The adopted metric presents two significant properties. The first one is that its absolute value is symmetric regardless of the order the probability distributions are compared. This symmetry is one of the advantages of the Jensen-Shannon Divergence over the Kullback-Leibler Divergence. The second property is that since the Jensen-Shannon Divergence and the metric are normalized, the resulting value range of the proposed metric is bounded between 0 and 1. These properties simplify the process of comparing the different scores obtained across all metrics and clusters.

In order to provide a vision of the objective of the RS, Fig. 2 presents an example where the probability density function of a reference metric is shown to be highly different from the metric of the class whose relevance is being measured. Still, in both cases, the level of degradation is not high (i.e., metric value < 0.5). In this case, the legend shows that the JSD value is high (0.81), but, in contrast, the RS value is considerably lower (0.41). In the former, the difference between both distributions is still taken into account, but, since the level of degradation is not much higher than in the reference class, this metric does not receive a very high relevance (i.e., RS > 0.7), what indicates the advantages of the defined RS over JSD.

The output of the framework is the score for each metric in each cluster. These scores can be used to determine the most relevant metrics, helping to analyze the underlying issues contained in each cluster. In this manner, these scores could be used for tasks such as feature selection or the labeling of failures. However, the present work focuses on the latter, demonstrating in the evaluation the labeling capabilities of applying predefined rules.

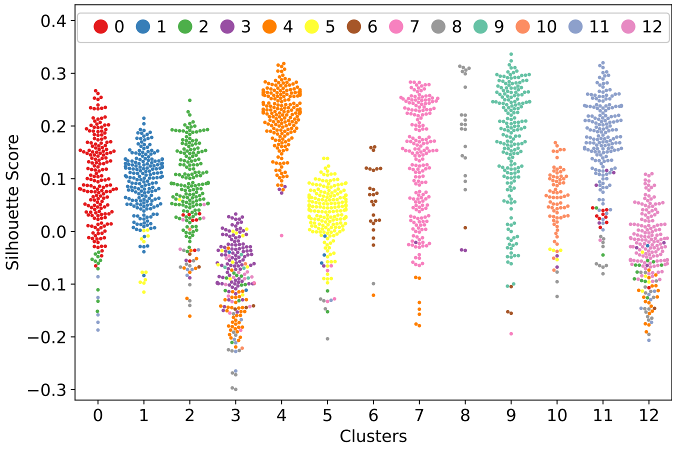

Fig. 4. Silhouette score per sample and closest cluster representation (Maximum of 250 samples per cluster).

III. Evaluation

The evaluation of the proposed framework has been carried out using a dataset from a live cellular network with more than 25000 daily samples available, each of them composed of 33 metrics with 24 hourly values each, amounting to a total of more than 600000 hourly samples. The daily samples are gathered at cell level from cells all over a country over a period of two weeks. These samples have been assigned into 13 different categories using a deep learning classifier based on a Long Short-Term Memory (LSTM) network [16] external to this work. This model was trained using a training dataset obtained using a Deep Embedded Clustering (DEC) model [17] and manual refinement by network experts. In this way, the categories from the classifier can be used as ground truth to evaluate the results, given that specific labels for failure/normal status were established for each of them by network troubleshooting experts.

In order to provide a deeper level of insight into the data used for the evaluation, Fig. 4 shows the distributions of silhouette score values for the samples in the different clusters, with each of the samples being colored according to the cluster they are closest to. The distance to a cluster is defined as the Euclidean distance from to sample to the centroid of the cluster. In this manner, it is possible to observe if a sample corresponds to the cluster was assigned to or if, otherwise, it might have been misclassified.

Furthermore, Fig. 4 shows that, although silhouette values are not very high, samples have a positive value of silhouette and are in the closest cluster even in cases where the silhouette value is negative. These low values in silhouette might be caused by the complexity introduced by using time series and by grouping samples by general type of issues rather than by very specific issues, which would result in a high number of clusters with fewer samples. Lastly, it can be seen that cluster 3 has many samples that are closer to other clusters and cluster 12 also has some. The former is caused by the

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

This article has been accepted for publication in IEEE Communications Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LCOMM.2024.3411768

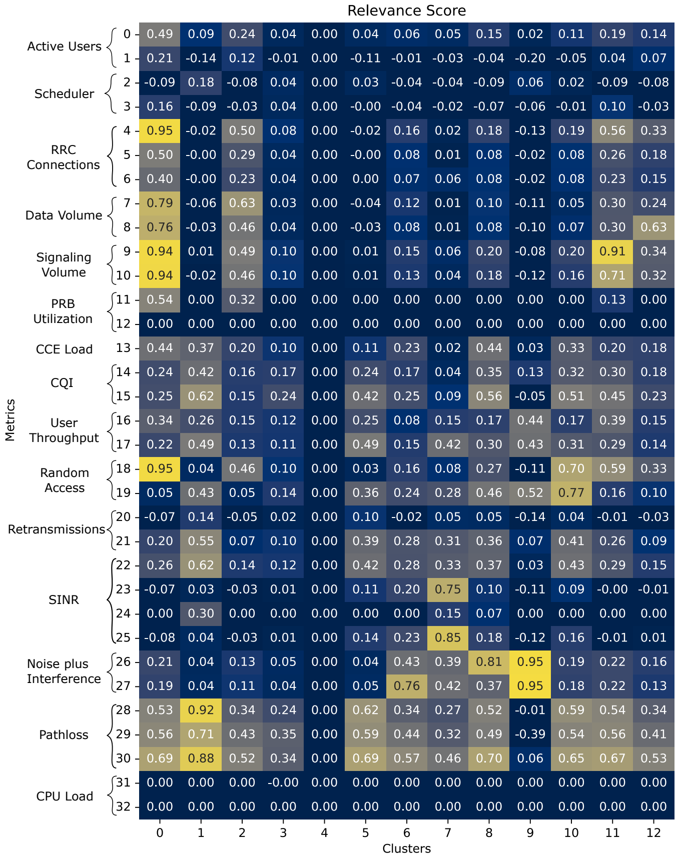

Fig. 5. Obtained Relevance Scores.

definition of the cluster, which is being an umbrella cluster that gathers samples that are too degraded to be deemed normal but are not degraded enough to be considered an issue. The latter's misclassification rate is due to being a cluster without an outstanding degradation pattern, as will be seen later.

In the first step of the evaluation process, the reference cluster for the dataset has been determined according to (2). Consequently, the accumulated mean degradation of each cluster can be seen in Fig. 3, highlighting the minimum value, which in this case corresponds to cluster 4 and, therefore, is chosen as the reference cluster.

Next, the proposed relevance score computing process is applied, calculating the RS for each metric for all clusters compared to cluster 4, shown in Fig. 5, where the x-axis indicates the cluster number, and the y-axis corresponds to the metrics. Moreover, although the value of the RS already quantifies the degradation, it might not provide a clear vision of what it means for the proposed use case. Therefore, to solve this, four additional ranges of values have been defined to categorize the results. These ranges are $[-1, 0]$, $[0, 0.4]$, $[0.4, 0.7]$, and $[0.7, 1]$, which correspond to less degraded than usual, normal, impacted, (i.e., partially degraded), and degraded metric, respectively.

Then, applying these ranges to the RS values obtained several conclusions can be drawn. Firstly, observing rows presented in Fig. 5, it can be seen that metrics 12, 31, and 32 are filled with zeroes, which means that those metrics have a value that is the same throughout the dataset and can be removed from it as they do not provide relevant information. In addition to this, every score within metrics 1, 2, 3, 20, and 24 are in the second RS range ($[0, 0.4]$), which means that those metrics' values are normal and, therefore, they could be taken out of the dataset for the detection of these issues without affecting the performance. Moreover, if a row has many values in the third ($[0.4, 0.7]$) or fourth ($[0.7, 1]$) RS ranges, it might be indicative that the metric is degraded globally across the network.

Secondly, focusing on the columns of the matrix, it can be seen that all the values of cluster 4 are 0 because it is the reference cluster. Furthermore, it is observed that in cluster 3 the values are low as well because this cluster corresponds to a network status where metrics' values are slightly higher than in the optimal status, but still not considered a degradation in performance. For instance, clusters 0, 2, 11, and 12 are related to traffic (metrics 4 to 13). In this case, cluster 0 has degraded metrics related to signaling and data usage in both uplink and downlink (metrics 7 to 10). In contrast, cluster 2 only has a degradation in the data usage in the downlink (metric 7), cluster 11 in the signaling (metrics 9 and 10), and cluster 12 in the data usage in the uplink (metric 8).

Thirdly, the defined value ranges could be used to provide a simple description of the degradations found in a cluster. For example, cluster 9 has degradation in metrics related to the noise and interference in the uplink (metrics 26 and 27) and slight degradation in user throughput (metrics 16 and 17) and random access success rate (metric 19). In this manner, "strong degradation in metrics 26 and 27, and slight degradation in metrics 16, 17, and 19" could be used as a preliminary label for the cluster.

Finally, Table I summarizes all the impacted and degraded metrics along with the label defined by experts. It can be seen in the table that most of the automatically selected metrics could be expected according to the expert label. Moreover, it is worth noting that cluster 3 does not present any particular metric as it was defined as a cluster to gather samples that cannot be considered normal but either correspond to a type of failure. Hence, the proposed RS shows to be a relevant measure for the importance of the different metrics for each of the clusters obtained using clustering techniques. This allows the identification of the degradation presented in each cluster, as well as the key differences among similar clusters, being a key complement to provide value to unsupervised clustering results.

IV. Conclusions & Outlook

This work has proposed a framework and associated mechanisms to automatically establish the relevance of metrics in different clusters obtained using clustering algorithms, via a proposed importance score derived from the Jensen-Shannon Divergence. The provided score is able to estimate metric importance considering the divergence between the distribution of its values as well as its overall degradation compared to a reference cluster. Furthermore, the cluster used as reference is obtained automatically based on the degradation of the metrics so, the cluster chosen as reference is the one considered to be the normal status of the network.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

This article has been accepted for publication in IEEE Communications Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LCOMM.2024.3411768

TABLE I
COMPARISON OF EXPERT LABELS AGAINST AUTOMATICALLY SELECTED IMPACTED AND DEGRADED METRICS.

No.	Expert label	Automatic label	
		Impacted	**Degraded**
0	Overloaded cell	DL Active Users, RRC connected users, DL PRB Utilization, CCE Load and Pathloss.	RRC Connection attempts, DL/UL Data Volume, DL/UL Signaling Volume and Random Access Attempts
1	Coverage problems	CQI, UL Throughput, Random Access Success, UL Retransmissions and PUSCH SINR	Pathloss
2	Downlink overload	RCC Connection Attempts, DL† /UL Data Volume, DL/UL Signaling Volume, Random Access Attempts and Pathloss	-
3	No issue	-	-
4	Normal	-	-
5	Overshooting	CQI, UL Throughput, PUCCH SINR and Pathloss†	-
6	Noisy PUCCH	Noise plus Interference in PUSCH and Pathloss	Noise plus Interference in PUCCH
7	PUCCH Degradation	UL Throughput, Noise plus Interference in PUCCH and Pathloss	PUCCH SINR
8	Noisy PUSCH	CCE Load, CQI, Random Access Success and Pathloss	Noise plus Interference in PUSCH and Pathloss
9	Noisy PUSCH and PUCCH	DL/UL Throughput and Random Access Success	Noise plus Interference in PUSCH/PUCCH
10	Degraded RACH	CQI, Retransmissions and Pathloss	Random Access Attempts and Success
11	Signaling Overload	RCC Connection Attempts, CQI, Random Access Attempts and Pathloss	DL/UL Signaling Volume
12	Uplink overload	Pathloss, UL Data Volume†	-

† Most degraded metric, although not in the degraded range.

The results obtained have proven that the proposed framework can properly provide information on the relevancy of metrics in each cluster. This information is useful to facilitate to network experts the process of labeling clusters obtained using unsupervised mechanisms. Once clusters are labeled, the information can be used in the detection and troubleshooting of problems in live networks. Furthermore, the score provided is useful to establish which metrics do not add information for the detection of issues. In this manner, avoiding those metrics can improve the performance of the detection method and save resources in a production environment.

Future work will focus on developing a system capable of using the information provided by the RS to automatically generate labels that can be understood by network experts.

REFERENCES

[1] A. Tarrias, S. Fortes, and R. Barco, "Failure Management in 5G RAN: Challenges and Open Research Lines," *IEEE Network*, pp. 1–7, 2023.

[2] M. Liyanage, Q.-V. Pham, K. Dev, S. Bhattacharya, P. K. R. Maddikunta, T. R. Gadekallu, and G. Yenduri, "A survey on Zero touch network and Service Management (ZSM) for 5G and beyond networks," *Journal of Network and Computer Applications*, vol. 203, p. 103362, 2022.

[3] M. Mdini, G. Simon, A. Blanc, and J. Lecoeuvre, "Introducing an Unsupervised Automated Solution for Root Cause Diagnosis in Mobile Networks," *IEEE Transactions on Network and Service Management*, vol. 17, no. 1, pp. 547–561, 2020.

[4] S. Fortes, C. Baena, J. Villegas, E. Baena, M. Z. Asghar, and R. Barco, "Location-Awareness for Failure Management in Cellular Networks: An Integrated Approach," *Sensors*, vol. 21, no. 4, 2021.

[5] H. Dinaki, S. Shirmohammadi, E. Janulewicz, and D. Cote, "Deep Learning-Based Fault Localization in Video Networks Using Only Client-Side QoE," *IEEE Transactions on Artificial Intelligence*, vol. 1, pp. 130–138, 10 2020.

[6] A. Gómez-Andrades, P. Muñoz, I. Serrano, and R. Barco, "Automatic Root Cause Analysis for LTE Networks Based on Unsupervised Techniques," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 4, pp. 2369–2386, 2016.

[7] J. Santos, P. Leroux, T. Wauters, B. Volckaert, and F. De Turck, "Anomaly detection for Smart City applications over 5G low power wide area networks," in *NOMS 2018 - 2018 IEEE/IFIP Network Operations and Management Symposium*, 2018, pp. 1–9.

[8] A. Margaris, I. Filippas, and K. Tsagkaris, "Hybrid Network-Spatial Clustering for Optimizing 5G Mobile Networks," *Applied Sciences*, vol. 12, no. 3, 2022.

[9] L. A. Lopes, V. P. Machado, and R. d. A. L. Rabêlo, "Automatic cluster labeling through Artificial Neural Networks," in *2014 International Joint Conference on Neural Networks (IJCNN)*, 2014, pp. 762–769.

[10] L. E. S. Silva, V. P. Machado, S. S. Araujo, B. V. A. de Lima, and R. d. M. S. Veras, "Using Regression Error Analysis and Feature Selection to Automatic Cluster Labeling," in *Progress in Artificial Intelligence*. Cham: Springer International Publishing, 2021, pp. 376–388.

[11] I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," *Philosophical Transactions of the Royal Society of London Series A*, vol. 374, no. 2065, p. 20150202, Apr. 2016.

[12] J. D. Hamilton and J. Xi, "Principal Component Analysis for Nonstationary Series," National Bureau of Economic Research, Working Paper 32068, January 2024.

[13] P. C. Molenaar, J. G. De Gooijer, and B. Schmitz, "Dynamic factor analysis of nonstationary multivariate time series," *Psychometrika*, vol. 57, pp. 333–349, 1992.

[14] B. Fuglede and F. Topsoe, "Jensen-Shannon divergence and Hilbert space embedding," in *International Symposium onInformation Theory, 2004. ISIT 2004. Proceedings.*, 2004, pp. 31–.

[15] J. M. Joyce, *Kullback-Leibler Divergence*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 720–722.

[16] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 11 1997.

[17] J. Xie, R. Girshick, and A. Farhadi, "Unsupervised deep embedding for clustering analysis," 2016.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/