



HAL
open science

Anomaly Prevision in Radio Access Networks Using Functional Data Analysis

Yosra Ben Slimen, Sylvain Allio, Julien Jacques

► **To cite this version:**

Yosra Ben Slimen, Sylvain Allio, Julien Jacques. Anomaly Prevision in Radio Access Networks Using Functional Data Analysis. IEEE GlobeCom 2017, Dec 2017, Singapour, Singapore. hal-01613475

HAL Id: hal-01613475

<https://inria.hal.science/hal-01613475v1>

Submitted on 9 Oct 2017

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.

Anomaly Prevision in Radio Access Networks Using Functional Data Analysis

Yosra BEN SLIMEN

Orange Labs, Belfort, France
Université Lyon 2, ERIC, Lyon, France
Email: yosra.benslimen@orange.com

Sylvain ALLIO

Orange Labs, Belfort, France
Email: sylvain.allio@orange.com

Julien JACQUES

Université de Lyon, Université Lyon 2,
ERIC EA3083, Lyon, France
Email: julien.jacques@univ-lyon2.fr

Abstract—In order to help the network maintainers with the daily diagnosis and optimization tasks, a supervised model for mobile anomalies prevention is proposed. The objective is to detect future malfunctions of a set of cells, by only observing key performance indicators that are considered as functional data. Thus, by alerting the engineers as well as self-organizing networks, mobile operators can be saved from a certain performance degradation. The model has proven its efficiency with an application on real data that aims to detect capacity degradation, accessibility and call drops anomalies for LTE networks.

Index Terms—Anomaly prevention, Multivariate functional data prevision, Functional data analysis, LTE network, Troubleshooting, Optimization

I. INTRODUCTION

The mobile telecommunication industry has and is still undergoing interesting changes resulted by the introduction of new technologies and services. The operators and manufacturers of mobile equipments are undertaking huge efforts to adapt the cellular networks to the new technologies, while aiming to maintain the level of service of the current networks. Consequently, the operation of the radio network is becoming increasingly complex in an environment where the fault management (troubleshooting) and network optimization are still a manual process. They are accomplished by experts in diagnosis. These latter are personnel dedicated to daily analyze the main performance indicators and other information, in order to detect problems in different cells and to solve them after diagnosing the causes. Ensuring that the malfunctions in the cells are rapidly solved is important because if a cell is temporarily non-operational, probably neighbouring cells will also be affected and a degradation in performance of a cluster of cells will be resulted. However, the growing size of cellular networks, together with their increasing complexity, make it very difficult for a human to analyze such a large amount of information. For this reason, Self-Organizing Networks (SON, [1]) are being proposed in order to automate network procedures which significantly reduces the costs. The SON covers three different functions: self-configuring, self-optimization and self-healing.

In order to automate these functions, some solutions can be found in the literature. Some studies propose to define the relationship between symptoms and causes through supervised models. In [2] and [3], this relationship is expressed by dependency graphs in order to automate the troubleshooting

of mobile networks through Bayesian networks. In [4] and [5], the relationship is defined by compiling expert knowledge about symptoms and their causes, into causal graphs. For interference optimization purposes, this relationship is defined by a liner regression model such as [6] or a genetic algorithm such as [7]. The challenge with these methods is that historical records of previously solved problems are needed. However, labeled cases i.e. those associating identified faults with their symptoms are hard to get from live networks. This is due to three main reasons: (1) the absence of a normalized troubleshooting procedure; (2) the difficulty in determining the real fault that degraded the performance since it is frequent that the cause remains unknown even if the problem is solved (by a reset for example); (3) since the evolution of networks is so fast, the knowledge rapidly becomes obsolete and the expertise thus gathered from experts is generic.

To cope with this problem, solutions based on unsupervised techniques have been proposed. Several works have demonstrated the utility of using self-organizing maps. For instance, in [8], the objective is to automate the fault detection phase of the troubleshooting. In [9] and [10], the objective is to analyze multidimensional 3G network performance data in order to aid in the manual fault diagnosis. In [11], the work has been devoted to cope with uncertainty in the data by using fuzzy logic theory. A co-clustering model is proposed in [12] in order to provide a simplified representation of the observed data for an easier analysis. However, experts are still needed to interpret the obtained clusters. The interpretation is not an easy task especially when the number of clusters are high or when data into one cluster are not homogeneous enough to be easily understandable. Moreover, these algorithms are able to detect and to correct the problem after the damage is done and the network performance has been already degraded.

With the presented work, the objective is to anticipate the anomalies since it is more interesting to predict the malfunctions of mobile networks when they didn't happen yet. With this anomaly prevision, diagnosis experts could have the chance to correct the problem before it occurs and thus to save the quality of services of mobile operators. The data used by the proposed model are Key Performance Indicators (KPIs) which are measurements responsible for evaluating the network's performance. They are defined by mathematical formulas derived from different counters and computed peri-

odically from the network with different temporal granularities (weekly, daily, hourly or less). For instance, Figure 1 illustrates a sample of $p = 30$ KPIs for $n = 20$ daily observations. From

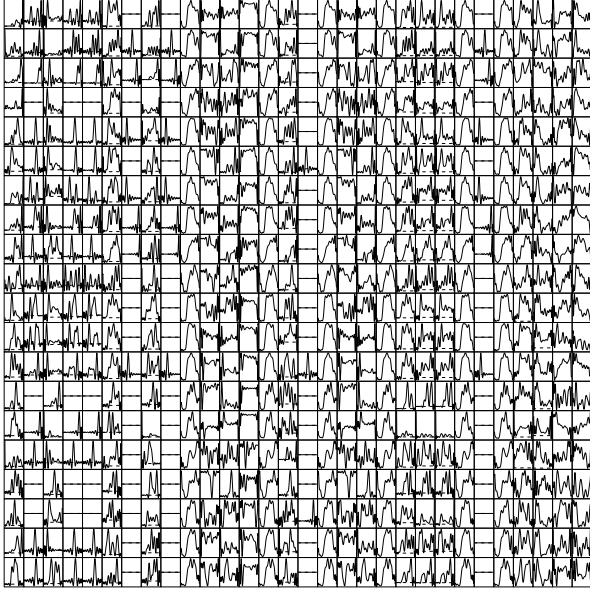


Fig. 1. An example of a functional data set composed of 20 observations (rows) and 30 KPIs (columns)

a statistical perspective, these KPIs are considered as functional data [13] which have become a commonly encountered type of data. With the advance of modern technology, more and more data are being recorded continuously during a time interval (or intermittently at numerous discrete time points). They become frequent, not only in the telecommunication field, but in numerous other domains like medicine, economics and chemometrics (see [13] for an overview). Functional data is the observation (sample path) of a stochastic process $X = \{X(t), t \in T\}$, where T can be for instance a time interval, or any other continuous subset.

The novelty of our work is that it allows to anticipate the anomalies in radio access networks which may be helpful for mobile operators to maintain a top quality of services. Besides, it takes into consideration the functional aspect of the KPIs and, up to our knowledge, the proposed model is the first prevision model for multivariate functional data.

The paper is organized as follows. Section II introduces the malfunction prevision model for multivariate functional data i.e. for the curves of multiple KPIs per observation. The behavior of the model is studied on real data extracted within an internal tool of Orange France in Section III. Three use cases are considered: problems related to call drops, capacity degradation and accessibility degradation in LTE networks.

II. ANOMALY PREVISION MODEL

The proposed method offers a supervised technique that aims to prevent anomalies in mobile networks. It is assumed that a labeled dataset is provided. The dataset is related to a specific use case (the target anomaly to detect) with

the corresponding KPIs. It contains data related to normal behaviour of the network as well as data related to previous failures. The model is window-based: the size of the window, the step of the window as well as the prediction horizon are parameters to be set. Each observation corresponds to the set of the different KPIs curves, for one cell and for one window. The labeling is assured by observing if an anomaly will occur in the prediction horizon.

Statistically speaking, the data \mathbf{X} under study are a sample of n observations. Each observation X_i is described by a set of p curves and a label. The curves are the functional features that correspond to the daily evolution of p KPIs. The statistical model underlying data, represented by multivariate curves, is a stochastic process with continuous time:

$$\mathbf{X} = \{(\mathbf{X}_i(t), label_i)\}_{t \in [0, T], 1 \leq i \leq n}$$

$$\text{with } \mathbf{X}_i(t) = (X_{i1}(t), \dots, X_{ip}(t))' \in \mathbb{R}^p, \quad p \geq 1$$

and $label_i=1$ if there is a problem in the prediction horizon given $\mathbf{X}_i(t)$, 0 otherwise.

The approach is composed of three steps. Since the collected KPIs have discrete values, the first step of the proposed model is to retrieve the functional nature of KPIs. By considering that the numbers of the observed KPIs and cells could be huge and that the observation duration could be long, a dimensionality reduction seems inevitable. Therefore, a Functional Principal Components Analysis (FPCA, [13]) should be applied. Once the labeled dataset is resumed in terms of principal components, a classification algorithm can be applied. This classification allows to predict future malfunctions given the observed KPIs for a specific cell and for a fixed window.

A. From discrete data to functional data

The main difficulty when dealing with functional data consists in the fact that these latter belong to an infinite-dimensional space, whereas in practice, KPIs are observed at discrete time points and with some noise. Thus, in order to reflect the functional nature of the KPIs, a smoothing may be considered. Smoothing methods consider that the true curve belongs to a finite-dimensional space spanned by some basis of functions such as trigonometric functions, B-splines or wavelets (see [13] for a detailed study). Smoothing assumes that each observed curve x_{ij} ($1 \leq i \leq n$, $1 \leq j \leq p$) can be expressed as a linear combination of basis functions $\{\phi_{j\ell}\}_{\ell=1, \dots, M_j}$:

$$x_{ij}(t) = \sum_{\ell=1}^{M_j} a_{ij\ell} \phi_{j\ell}(t), \quad t \in [0, T], \quad (1)$$

where $\{a_{ij\ell}\}_{\ell=1, \dots, M_j}$ are the basis expansion coefficients. These coefficients can be estimated by least square smoothing for instance [13]. In this work, due to the nature of the KPIs under study, the same B-spline basis $\{\phi_{\ell}\}_{\ell=1, \dots, M}$ is used for all the functional features. The choice of the basis as well as the number of basis functions strongly depends to the nature of data. Hence, they can be set empirically.

B. Principal components analysis for functional data

From the set of functional data, it is interesting to have an optimal representation of curves into a functional space of reduced dimension. The main tool to answer this request is the principal components analysis for functional data (FPCA, [13]). It consists in computing the principal components C^h and principal factors f^h of the Karhunen-Loeve expansion:

$$X(t) = \mu(t) + \sum_{h \geq 1} C^h f^h(t), \quad t \in [0, T]. \quad (2)$$

When curves are assumed to be decomposed into a finite basis of function (1), FPCA consists in a usual PCA of the basis expansion coefficients using a metric. This latter is defined by the inner products between the basis functions. In theory, the number of principal components are infinite. However, in practice, the curves are observed at discrete time points and they are approximated on a finite basis of functions. For this reason, the maximum number of components one can compute is equal to the number M of basis functions used for approximation.

In this work, in order to project all the data onto the same FPCA space, functional principal components analysis is applied to the whole data set of curves \mathbf{x} . Moreover, in order to reduce the dimensionality of the problem, only the first $m \leq M$ principal components are considered. This latter is fixed empirically so that the principal components express a given part of the total variance. As a result, each curve is defined by a vector of its principal components of size m .

C. Classification and anomaly prevention

Once the labeled dataset is transformed in terms of principal components, a training data with numeric values and binary classes is ready for a training phase. The training is assured by any suitable classification algorithm that deals with numeric variables and binary classes. Examples of the algorithms that can be used are neural networks, support vector machines and decision trees [14]. The classification allows to create a trained model that is able to discern a pattern describing the normal behaviour of the network and a pattern describing suspicious behaviours.

This trained model will be used in order to predict if there will be a malfunction in the future for new observations. Given the smoothing basis $\{\phi_{j\ell}\}_{\ell=1, \dots, M}$, the FPCA basis and the learned classification model, prediction is as follows: (1) for each new observation, a smoothing is first applied for every KPI using $\{\phi_{j\ell}\}_{\ell=1, \dots, M}$; (2) the observation in terms of functional features is then projected on the same FPCA basis used in the training phase; (3) given the obtained principal components, the learned model will then predict if there will be any anomaly in the prediction horizon.

III. EXPERIMENTAL STUDY

The aim of this section is to apply the anomaly prevision model on Long Term Evolution (LTE) cells to evaluate its efficiency through real KPI measurements. In order to present where our application is located in terms of LTE system,

Figure 2 illustrates the architecture of the Evolved Packet System (EPS) bearer. The EPS is composed of the radio access part E-UTRAN (Evolved Universal Terrestrial Radio Access Network), the core part EPC (Evolved Packet Core) and the Packet Data Network (PDN). Our experiment focuses on anomaly prevention concerning the radio access part of LTE networks. In this part, KPI measurements are performed through the Enhanced Radio Access Bearer (E-RAB) service. The role of the E-RAB is to transport packets of EPS bearer between the user equipment (UE) and the EPC and it is generated from a combination of radio bearer and S1 bearer. LTE KPIs are mainly classified into five classes [15], [16]:

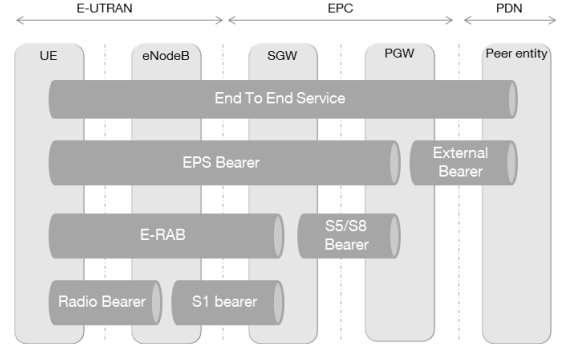


Fig. 2. EPS bearer architecture

- **Accessibility:** it regroups measurements that allow operators to gather information related to the mobile services' accessibility for the subscriber.
- **Retainability:** it measures how many times a service was interrupted or dropped during use. Thus, it prevents the subscriber from using the service and the operator cannot charge for it.
- **Mobility:** it measures how many times a service was interrupted or dropped during a subscriber's handover or mobility from one cell to another.
- **Integrity:** it measures the high or low quality of a service while the subscriber is using it (latency and throughput).
- **Availability:** it measures a service's availability for the subscriber.

Three of the five classes of LTE KPIs are addressed in this paper. The first use case aims to detect retainability anomalies by analyzing call drop problems. The second use case aims to detect accessibility anomalies by analyzing radio and S1 bearers setup success rate. The third use case aims to detect integrity anomalies by analyzing capacity degradation through delay and throughput. Two families of experimentation are distinguished. The first family, denoted by *Family1*, considers that only one KPI i.e. one uni-variate functional feature is capable of predicting the anomaly. The second family, denoted by *Family2*, considers that determining an anomaly needs the analysis of a set of KPIs i.e. multivariate functional features. In the following, a description of the data and the experimentation over *Family1* and *Family2* are presented.

A. Network description

The analysis of the proposed model has been conducted in a real LTE network in an urban area with a population of nearly one million. It corresponds to Lyon, a big city in France. Figure 3 illustrates the geographical area used for the data extraction. For the *Family1*, the data are generated within



Fig. 3. The geographical area used for the data extraction

191 cells with different parameters and they are situated at different locations and thus, they reflect different environment conditions. Table I summarizes the main parameters of the LTE network. As for *Family2*, 704 cells are used. The reason behind the choice of a bigger study zone is that the labels of normal and problematic observations are unbalanced which may mislead the model in its training phase. The dataset is first balanced by ignoring some observations. Therefore, the 191 cells used for *Family1* provides insufficient data for *Family2* and a bigger number of cells is needed. The period

TABLE I
PARAMETERS OF THE REAL LTE NETWORK

Network layout	Urban environment		
Technology	Alcatel LTE FDD		
System frequency band	800 MHz	1800 MHz	2600 MHz
Max transmit power	46 dBm	44.8 dBm	46 dBm
System bandwidth	10 MHz		
Number of cells for <i>Family1</i>	82 (43%)	9 (5%)	100 (52%)
Number of cells for <i>Family2</i>	296 (42%)	38 (5%)	370 (53%)
Days under observation	14		

of observation covers two weeks from December 5, 2016 to December 18, 2016. This period contains some ordinary workdays, weekends, the beginning of holidays and a special event of festival of lights that has been held in Lyon from December 8 to December 10. Therefore, normal behavior of the network can be found in this period as well as problematic observations. These latter correspond to anomalies related to capacity degradation, call drops and accessibility problems due to the huge number of users in the cells.

For all the following experiments, the size of the window is fixed to one day. The step of the window is equal to the prediction horizon so that no replications nor holes are possible. The KPIs are extracted with a granularity of 15 minutes (therefore, each daily KPI contains 96 values). 80% of the labeled dataset is used for the training phase and the remaining 20% is used for the test phase.

B. Experiments of the anomaly prevision model in case of uni-variate functional feature

The objective of this experiment is to test the efficiency of the model at predicting future problems by using only one KPI. The test is held over ten KPIs. Three among them belong to Retainability class and more specifically they correspond to dropped calls indicators. Two KPIs belong to Accessibility class and they correspond to radio and S1 bearers setup success rate. The last five KPIs belong to Integrity class and they are indicators of throughput, delay and traffic volume (see Table II).

The data contains missing values. They are easily treated by applying a smoothing for each curve, since it allows to gain the functional behavior of the daily KPIs which is an advantage when dealing with functional data. The used smoothing basis is B-splines where the number of basis functions is empirically set to $M = 20$. A FPCA for univariate data is then applied. The number of principal components m is chosen so that at least 80% of the information is covered. The horizon prediction varies from 3 hours to one day. The classification is performed with a neural network having 2 layers and 100 iterations. The evaluation is held in terms of the following performance indicators [17]:

- Confusion matrix (M): it is a matrix that contains information about actual and predicted classification. Two groups are considered: "group0" for normal behaviour in the prediction horizon and "group1" for a future malfunction in the network. A confusion matrix M is such that $M_{k,z}$ is equal to the number of observations known to be in group k but predicted to be in group z .
- F-measure (F): it is the weighted harmonic mean of the precision and the recall of the test. Recall calculates the proportion of malfunctions predicted by the model among the real malfunctions that should have been detected. Precision calculates to which point we can trust the model if a malfunction is predicted. The F-measure is high when both precision and recall are high.
- Correct-classification rate (C): it calculates the number of correct predictions among all the predicted observations.

The results are presented in Table II. We notice that for all the tests, the correct classification rate is promising, varying from 70% to 97%. Moreover, the false alarms generated by the model are moderate as shown by the F-measure that varies from 66% to 98%. With the confusion matrices, we can verify that the model does not have a tendency to predict one class over another. Consequently, the model has proven its efficiency in anomaly prevention related to retainability, accessibility and

TABLE II
EXPERIMENTAL RESULTS FOR PERFORMANCE DEGRADATION PREVENTION USING UNI-VARIATE FUNCTIONAL FEATURE

Problem	KPI	24 hours			12 hours			6 hours			3 hours			
		M	F	C	M	F	C	M	F	C	M	F	C	
Retainability	DROP_CALL_RATE	0 1			0 1			0 1			0 1			
		0 44 3	0.83	0.84	0 186 50	0.78	0.79	0 705 278	0.77	0.75	0 1161 350	0.77	0.77	
		1 13 18			1 46 169			1 257 899			1 311 1094			
Retainability	UECTXDROP_RATE	0 1			0 1			0 1			0 1			
		0 57 10	0.8	0.81	0 227 43	0.85	0.85	0 708 124	0.83	0.84	0 1966 374	0.82	0.84	
		1 16 52			1 36 221			1 138 641			1 324 1613			
Retainability	ERAB_DROP_RLF_RATE	0 1			0 1			0 1			0 1			
		0 84 35	0.74	0.72	0 373 143	0.78	0.76	0 576 195	0.74	0.75	0 693 215	0.79	0.79	
		1 36 101			1 116 438			1 180 545			1 141 677			
Accessibility	S1_CNX_ESTAB_SR	0 1			0 1			0 1			0 1			
		0 69 39	0.66	0.73	0 104 56	0.7	0.75	0 178 22	0.90	0.92	0 286 32	0.93	0.94	
		1 1 38			1 2 66			1 7 137			1 5 264			
Accessibility	LTE_CSSR	0 1			0 1			0 1			0 1			
		0 56 20	0.76	0.79	0 86 36	0.77	0.8	0 167 18	0.93	0.94	0 287 21	0.94	0.95	
		1 6 41			1 1 63			1 0 123			1 10 248			
Integrity	TAUXCONGDL_CAUSEPDCCH	0 1			0 1			0 1			0 1			
		0 96 16	0.92	0.92	0 179 11	0.96	0.96	0 354 15	0.97	0.97	0 626 18	0.98	0.97	
		1 0 97			1 4 192			1 8 343			1 15 648			
	Integrity	AVGUSERMAC_UL_THROUGHPUT	0 1			0 1			0 1			0 1		
			0 34 6	0.8	0.79	0 177 48	0.82	0.84	0 535 132	0.83	0.83	0 1318 239	0.86	0.85
		1 13 37			1 12 141			1 71 490			1 222 1396			
Integrity	UE_RRC_CONNECTED_AVG_PERCELL	0 1			0 1			0 1			0 1			
		0 142 30	0.85	0.86	0 363 55	0.88	0.89	0 994 95	0.9	0.91	0 1701 151	0.92	0.92	
		1 15 131			1 31 327			1 105 945			1 138 1672			
Integrity	LTE_RADIO_DL_DELAY	0 1			0 1			0 1			0 1			
		0 45 10	0.7	0.77	0 195 83	0.88	0.89	0 728 309	0.72	0.77	0 2147 585	0.77	0.81	
		1 2 24			1 9 105			1 62 489			1 203 1298			
Integrity	LTE_DL_TRAFFIC_VOLUME	0 1			0 1			0 1			0 1			
		0 63 8	0.81	0.85	0 196 22	0.87	0.88	0 662 144	0.84	0.84	0 1901 460	0.81	0.83	
		1 10 38			1 30 173			1 102 628			1 269 1603			

integrity degradations when only one uni-variate functional feature is used for the prediction.

C. Experiments of the anomaly prevision model in case of multivariate functional features

In this experiment, the objective is to test the efficiency of the model at predicting future problems related to the three use cases by using a set of KPIs. The data is extracted from 704 cells. For each curve, a smoothing with B-spline basis is used with $M = 20$. A FPCA for multivariate data is then applied for the dataset of smoothed curves. The horizon prediction varies from 3 hours to one week and the evaluation is held by the same performance metrics that are used for the uni-variate case. Since the neural network, in this experiment, takes a long time before it converges, a decision tree is used instead. The resulted confusion matrices are presented in Table III. Figure 4 presents the correct classification rates and the F-measure results of the prevention model applied to the three use cases when multivariate KPIs are used for the prediction. We notice that the algorithm has the potential to prevent from future problems since the correct classification rate is greater than 70% even for a long prediction horizon, equal to one week. The model does not have a tendency to predict a specific class nor to generate false alarms as proven by the confusion matrices and by the F-measures. The performance of the model decreases when the prediction horizon increases which is expected. The data does not allow to have a prediction horizon bigger than one week since the extraction covers two weeks. The proposed model supposes that the network configuration is

fixed. However, network behaviour is complex and it exhibits behavioural dynamics on multiple seasonal timescales (special events, fault conditions, and the non-stationarity over time as traffic volumes generally grows...), so for the proposed model to maintain accuracy, different models related to different geographical areas should be trained separately. The training should be programmed periodically with updated training data.

IV. CONCLUSION

The paper presents an anomaly prevision model that aims to detect future anomalies in mobile networks by observing key performance indicators. The KPIs are considered as functional data. Through a smoothing step, a functional principal components analysis and a classification phase, the model is able to prevent from a performance degradation. It is useful for engineers as well as SONs for troubleshooting and optimization purposes. The model is window-based and it has proven its efficiency through a real data application on LTE networks. The objective of this application is to detect future malfunctions related to degradation in cells capacity, accessibility and call drops. One advantage of the proposed approach is that it allows to take into consideration the temporal dynamic of the KPIs evolutions which explains the good prediction results obtained in the experimentation. Another advantage is that it is capable of dealing with missing data. Moreover, up to our knowledge, it is the first prevision model for multivariate functional data since no reference can be found related to this topic.

TABLE III
CONFUSION MATRICES FOR PERFORMANCE DEGRADATION PREVENTION USING MULTIVARIATE FUNCTIONAL FEATURES

Problem	KPIs	3 hours		6 hours		12 hours		24 hours		48 hours		72 hours		1 week	
Retainability	UECTXDROP_RATE	0	1	0	1	0	1	0	1	0	1	0	1	0	1
	ERAB_DROP_RLF_RATE	0 1248 275	0 1084 201	0 873 139	0 590 102	0 420 116	0 209 46	0 61 20							
		1 279 1291	1 228 1077	1 183 886	1 154 624	1 72 378	1 76 326	1 15 69							
Accessibility	S1_CNX_ESTAB_SR	0	1	0	1	0	1	0	1	0	1	0	1	0	1
	RRC_cnx_estab_SR	0 1085 76	0 639 106	0 466 58	0 291 58	0 204 48	0 125 44	0 94 51							
		1 60 1046	1 67 587	1 56 442	1 69 300	1 31 211	1 48 163	1 13 73							
Integrity	TAUXCONGDL_CAUSEPDCCH	0	1	0	1	0	1	0	1	0	1	0	1	0	1
	UE_RRC_CONNECTED_AVG_PERCELL	0 367 18	0 291 28	0 255 35	0 189 32	0 130 23	0 75 15	0 50 15							
	AVGUSERMAC_UL_THROUGHPUT	1 60 387	1 36 316	1 32 230	1 14 163	1 9 98	1 23 79	1 5 53							

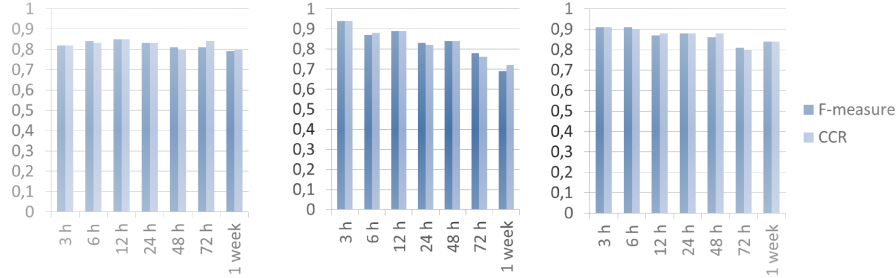


Fig. 4. Experimental results in terms of F-measure and correct classification rate (CCR) for performance degradation prevention regarding Retainability (left), Accessibility (middle) and Integrity (right) anomalies using multivariate functional features

As future work, an application of the model on other use cases could be interesting. Moreover, the data extracted within two weeks was not enough to determine the maximum size of the prediction horizon that the model could reach. For this reason, an application on the different use cases with large extractions should be considered. Another interesting perspective is to consider mixed data. Actually, besides the KPIs, engineers deal with other source of information such as alarms and the network parameters. Hence, an anomaly prevision model for mixed data will be proposed which gives more flexibility to the model. Finally, in order to automatically generate up-to-date models, a dynamic version will be considered in which adding new information will not lead to re-learn the model but to sum-up old trainings and to generate a new model in an incremental way.

REFERENCES

- [1] 3GPP, "Telecommunication management; Self-Organizing Networks (SON); Concepts and requirements," 3rd Generation Partnership Project (3GPP), TS 32.500, Jul. 2008. [Online]. Available: <http://www.3gpp.org/ftp/Specs/html-info/32500.htm>
- [2] R. Barco, V. Wille, and L. Dez, "System for automated diagnosis in cellular networks based on performance indicators." *European Transactions on Telecommunications*, vol. 16, no. 5, pp. 399–409, 2005. [Online]. Available: <http://dblp.uni-trier.de/db/journals/ett/ett16.html#BarcoWD05>
- [3] R. Khanafer, B. Solana, J. Triola, R. Barco, L. Moltsen, Z. Altman, and P. Lázaro, "Automated diagnosis for UMTS networks using bayesian network approach," *IEEE Trans. Vehicular Technology*, vol. 57, no. 4, pp. 2451–2461, 2008. [Online]. Available: <http://dx.doi.org/10.1109/TVT.2007.912610>
- [4] J. Lu, C. Dousson, and F. Krief, "A Self-diagnosis Algorithm based on Causal graphs," in *ICAS 2011*, Venice/Mestre, Italy, May 2011, pp. 1–6. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-01011398>
- [5] J. Lu, C. Dousson, B. Radier, and F. Krief, *Towards an Autonomic Network Architecture for Self-healing in Telecommunications Networks*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 110–113. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-13986-4_16
- [6] M. I. Tiwana, "Automated RRM optimization of LTE networks using statistical learning," Theses, Institut National des Télécommunications, Nov. 2010. [Online]. Available: <https://tel.archives-ouvertes.fr/tel-00589617>
- [7] S. C. Ghosh, B. P. Sinha, and N. Das, "Channel assignment using genetic algorithm based on geometric symmetry," *IEEE Transactions on Vehicular Technology*, vol. 52, no. 4, pp. 860–875, July 2003.
- [8] G. A. Barreto, J. C. Mota, L. G. Souza, R. A. Frota, L. Aguayo, J. S. Yamamoto, and P. E. Macedo, "A new approach to fault detection and diagnosis in cellular systems using competitive learning," in *Proceedings of the VII Brazilian Symposium on Neural Networks (SBRN04)*, 2004.
- [9] J. Laiho, K. Raivio, P. Lehtimäki, K. Hatonen, and O. Simula, "Advanced analysis methods for 3g cellular networks," *IEEE Transactions on Wireless Communications*, vol. 4, no. 3, pp. 930–942, May 2005.
- [10] M. Kylväjä, K. Hatonen, P. Kumpulainen, J. Laiho, P. Lehtimäki, K. Raivio, and P. Vehviläinen, "Trial report on self-organizing map based analysis tool for radio networks [gsm applications]," in *2004 IEEE 59th Vehicular Technology Conference. VTC 2004-Spring (IEEE Cat. No.04CH37514)*, vol. 4, May 2004, pp. 2365–2369 Vol.4.
- [11] A. Gómez-Andrades, P. M. Luengo, E. J. Khatib, I. de la Bandera, I. Serrano, and R. Barco, "Methodology for the design and evaluation of self-healing LTE networks," *IEEE Trans. Vehicular Technology*, vol. 65, no. 8, pp. 6468–6486, 2016. [Online]. Available: <http://dx.doi.org/10.1109/TVT.2015.2477945>
- [12] Y. B. Slimen, S. Allio, and J. Jacques, "Model-Based Co-clustering for Functional Data," Dec. 2016, working paper or preprint. [Online]. Available: <https://hal.inria.fr/hal-01422756>
- [13] J. O. Ramsay and B. W. Silverman, *Functional data analysis*, 2nd ed., ser. Springer Series in Statistics. New York: Springer, 2005.
- [14] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, ser. Springer Series in Statistics. New York, NY, USA: Springer New York Inc., 2001.
- [15] K. Ralf and G. Karsten, *LTE Signaling: Troubleshooting and Optimization*. Chichester, West Sussex, U.K.: Wiley, 2010.
- [16] 3GPP, "Telecommunication management; Key Performance Indicators (KPI) for Evolved Universal Terrestrial Radio Access Network (E-UTRAN): Requirements," 3rd Generation Partnership Project (3GPP), TS 32.451, Aug. 2008. [Online]. Available: <http://www.3gpp.org/ftp/Specs/html-info/32451.htm>
- [17] A. Gunawardana and G. Shani, "A survey of accuracy evaluation metrics of recommendation tasks," *J. Mach. Learn. Res.*, vol. 10, pp. 2935–2962, Dec. 2009. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1577069.1755883>