IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 16 May 2024, accepted 5 August 2024, date of publication 15 August 2024, date of current version 27 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3444704

APPLIED RESEARCH

Temporal and Multivariate Similarity Clustering of 5G Performance Data

JAKUB MAZGUŁA[1,2], DARIUSZ KRÓL[1], (Senior Member, IEEE), AND IRENEUSZ JABŁOŃSKI[3,4]

[1]Faculty of Information and Communication Technology, Wrocław University of Science and Technology, 50-370 Wrocław, Poland
[2]Nokia Solutions and Networks, 54-130 Wrocław, Poland
[3]Faculty of Physics, Brandenburg University of Technology, 03046 Cottbus, Germany
[4]Fraunhofer Institute for Photonic Microsystems, Branch Integrated Silicon Systems, 03046 Cottbus, Germany

Corresponding author: Jakub Mazguła (jakub.mazgula@nokia.com)

This work was supported by the Polish Ministry of Education and Science with funds allocated from ''Implementation Doctorate'' Program.

ABSTRACT The performance of 5G mobile network cells is highly impacted by their evolving configuration and temporal environmental conditions, such as the number of connected devices or resource utilization. Evaluation of the performance of such a system is a complex task that requires the simultaneous analysis of multiple indicators and inspires the research community to work on zero-touch network service management. In this paper, we present a novel time series clustering method - Temporal and Multivariate Similarity Clustering (TMSC) - that incorporates Dynamic Time Warping with Limited Warping Length and Spectral Clustering, allowing for radio cell grouping based on realization of multiple Key Performance Indicators. We evaluated TMSC against state-of-the-art algorithms at a practical task of identifying cell configuration differences by clustering their performance metrics with a limited set of observations. The proposed algorithm outperformed other methods regarding the Normalized Mutual Information score achieved for more than 95% of the cases studied. We also display the potential for method generalization by evaluating it at the hand gesture recognition task, which yields satisfactory results.

INDEX TERMS 5G network performance, radio cell management, spectral clustering, dynamic time warping, multivariate time series, temporal similarity.

I. INTRODUCTION

The 5G cellular networks are complex systems with hundreds of features controlled by thousands of configuration parameters. Temporal radio cell performance depends on its hardware, software configuration, and the environment in which it operates. The state of the environment (e.g. traffic load, channel conditions, number of connected users) and resulting performance (e.g. throughput, block error rate, spectral efficiency) vary significantly over time. Automated management of such a complex system to achieve an expected performance profile is challenging and inspires the research community to work on zero-touch network service management [1], [2], [3], [4].

Significant heterogeneity in configuration-performance profiles throughout the operator network makes network management troublesome. One proposed solution to simplify this problem is dimensionality reduction by cell clustering.

Clustering in the context of cellular networks usually refers to grouping radio cells in the geographic domain for interference coordination or mitigation [5], cell load balancing [6], user mobility management [7] or energy efficiency [8].

In network performance management, clustering is a relatively new problem. Various equipment vendors and mobile network operators may approach this differently. In this context, three management scenarios can be distinguished based on their level of advancement:

1) Basic network management: no methodology defined for cell clustering, network control realized in a mass manner (e.g. unified management actions)
2) Expert-centric network management: relatively subjective cell clustering realized using expert-defined heuristics based on preselected analytics and metrics

The associate editor coordinating the review of this manuscript and approving it for publication was Nurul I. Sarkar.

(e.g. by grouping cells based on bandwidth, duplex, antenna setup, frame configuration, etc.)
3) Zero-touch networks: ongoing research toward self-evaluating and self-managing networks, with objective cell grouping

To the authors' best knowledge, the problem of cell clustering concerning temporal environmental and performance characteristics in 5G cellular networks has not been solved yet. Although time series clustering has been widely studied, its application to performance analysis in cellular networks has been limited, as presented in the related work section II-B.

In this paper, we propose a new algorithm, Temporal and Multivariate Similarity Clustering (TMSC), which simultaneously explores patterns in a set of time-variant cell characteristics that facilitates multivariate cell clustering and outperforms state-of-the-art clustering algorithms.

The structure of the article is as follows. In Section II, we formulate the problem (also concerning the state of the art), identify the patterns observed in the 5G Key Performance Indicators (KPI), and summarize the related work. Section III introduces the TMSC algorithm in detail, followed by its evaluation on a data set of KPIs from live mobile operator networks in Section IV. In Section V, we discuss the generalization of the TMSC method to other domains, which we exhibit with an experiment on a hand gesture data set, and describe the algorithm's limitations. Finally, we conclude the article and indicate further steps in Section VI.

II. PROBLEM FORMULATION

Radio cell performance assessment is critical to successful 5G network deployments. Although measuring instantaneous performance in a single domain is relatively simple and can be done by, e.g. carrying out drive tests [9] or download speed trials, evaluating performance in a longer time horizon is more complicated. Cell performance is highly dependent on environmental conditions [10]. For example, user throughput decreases as user density increases [11]. The conditions, such as the number of connected users, the volume of the data they download/upload, or the speed at which they are moving, vary significantly in time, and hence, to properly evaluate cell performance, the knowledge of its operating environment is necessary. Moreover, monitoring multiple performance metrics over an extended period is generally required to understand the radio cell capacity achieved [12].

It should be noted that data collection, transmission, storage, and analysis generate additional costs for the network operator and / or the equipment vendor. Therefore, the amount of data required to produce satisfactory results is an essential criterion for potential solutions to the problem. This translates into the number of observed cells and KPIs and the duration of the observation period.

A specific case of performance evaluation is when a certain network feature is enabled in a subset of cells. Due to varying environmental conditions, a direct comparison between cells with feature enabled and those with feature disabled might not be feasible. This refers to the following research question: ***is it possible to identify differences in 5G radio cell configuration by cell clustering using temporal environmental and performance metrics?***

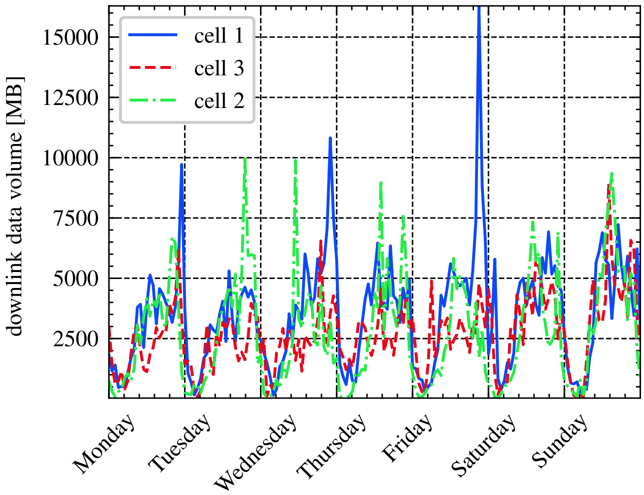

FIGURE 1. Example variable realization for downlink data volume measured in three different cells with the same configuration and similar location.

A. DATA CHARACTERISTICS

To evaluate network performance, telecommunications experts choose a set of relevant indicators from available Performance Management (PM) data, which are often cross-correlated and, therefore, should be analyzed jointly as multivariate time series [12]. Mobile network KPIs are time series collected with different granularity, spanning from (rare) 5- and 15-minute intervals to (most popular) hourly or daily aggregations. Temporal granularity affects signal characteristics, such as volatility, with 5-minute and daily intervals displaying the highest and lowest standard deviation, respectively. We focus on the KPIs collected every hour to balance data availability with its carried information content. The hourly measured signals tend to be noisy, although they show evident daily and weekly trends. These variations are strongly related to human activity (making phone calls, watching videos, uploading short clips) and mobility (driving to and from work, walking around the city, picking up children from school) during the day and at night, as well as on weekdays versus weekends [13], [14]. An exemplary time series, showing clear daily trends, for a selected KPI is presented in Fig. 1. There are also seasonal changes and anomalies (e.g., sports events or partial network failures, changing traffic locally); however, a time frame of one week is a good representation of the behavior of a cell, as it captures the majority of the conditions under which the given radio cell operates typically.

To cluster cells based on their KPIs, one must find similarities (or dissimilarities) among observations for different objects, usually by defining a distance measure. In this context, it is vital to understand the physical interpretation

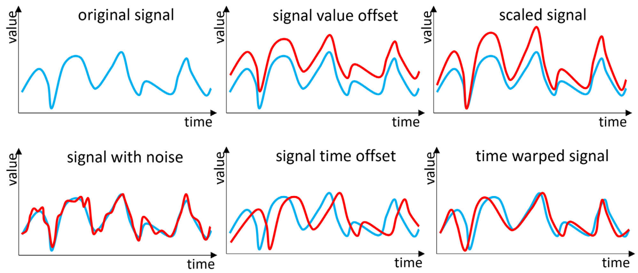

FIGURE 2. Different types of signal patterns. Top row, from left: original, value offset and scaled signals. Bottom row, from left: noisy, time offset and time-warped signals.

of these time series, as similarity in the context of mobile network metrics is different from the classical shaped-based approach to comparing time series [15]. Instead of only comparing the shapes of the signals, we also want to consider the value range of the variable and time distortions. The distance measure for telecommunication KPIs should be less sensitive to noise (the signals tend to be noisy by nature of the measurement observation and the system itself working in a dynamic environment), time offset (time shifts might come from physical locations of the cells, how the users move during the day, or even time zones), and time warping (for example, peak hours might be longer in one location than other, with cells functioning almost identically otherwise). On the other hand, the measure should be more sensitive to the range of values of the signals and their shape because the maximum absolute and relative values of KPIs indicate performance in the edge cases. Fig. 2 presents different signal patterns observed experimentally in KPIs measured in 5G networks. In this example, the distance measure is expected to yield higher values (meaning less similarity) for scaled and value offset signals and lower values (meaning more similarity) for noisy, time offset, and time-warped signals compared to the original signal.

These patterns, as identified and explained above, can also be found in systems of other nature, e.g., measurements recorded by accelerometers used for gesture recognition [16], [17] or melt curves used for fungal identification [18]. This brings us to the conclusion that solving the problem for exemplary data (here, 5G performance metrics) can contribute to the generalized problem of data clustering for complex systems operating in fluctuating and noisy conditions or affected by measurement imperfections.

B. RELATED WORK

Cellular data clustering is not widely discussed in the literature, while reliable and efficient data modeling remains challenging for the research community and telecommunication engineers [10]. In [19], Chidean et al. used *Kullback-Leibler divergence* and *k-Means* clustering to identify urban areas based on aggregated network usage data. However, the proposed approach is univariate, as network usage indicators are summed up into a single variable for analysis, which is only possible for indicators with the same unit (in this case, count of incoming/outgoing calls, received/sent text messages, and established Internet connections). Mahdy et al. performed univariate time series clustering to expand the training set for deep recurrent neural networks that forecast traffic load. The best results were reported with the use of *k-Means for time series data* (`TimeSeriesKMeans`) clustering algorithm with *Soft Dynamic Time Warping* (`softDTW`) distance metric [20]. In [21], Wang et al. developed a compelling base station (BTS) performance evaluation method, which utilizes the *FastDTW* algorithm, auto-encoder, and *k-Means* clustering. The authors clustered the base stations into three performance categories based on a time series of four selected KPIs. The precision of the method was evaluated by comparing the distributions of the KPIs in each group with the mean values according to the KPI standards provided by the mobile network operator from which the data were collected. Li et al. presented a different approach and performed a cell performance analysis by extracting the variable deviation at each time step, representing it as a vector of distribution statistics to which *k-Means* clustering was applied. For the multivariate case, representative vectors for each variable were concatenated. This method captures the temporal dynamics of a variable, but loses information about its value range [22]. Lu et al. performed cell clustering by comparing the statistical and temporal characteristics of the variables separately and applying the *k-Medoids* algorithm to the weighted distance matrix. The results were evaluated using *silhouette score* and cluster averages [23]. A more advanced description was reported by Wang et al., who used *Self Organizing Maps* (SOM) and *k-Means* algorithms to extract the behavior patterns of cells of the Long Term Evolution (LTE) mobile network and group them. They used data sets collected from a live LTE European network [24]. Finally, Kajó et al. created a clustering method *DANCE* based on *Decorrelating Adversarial Networks* and used it to recognize different groups of mobile network users in a simulated scenario [25]. Note that the mobile user data was collected from a simulation with subsecond granularity at the user equipment (UE) level. Therefore, they do not exhibit the same characteristics described in Section II-A.

III. TEMPORAL AND MULTIVARIATE SIMILARITY CLUSTERING ALGORITHM

Having defined the signal patterns in the measured indicators in section II-A, and motivated by the physical context of the complex system under study, we propose a new algorithm called Temporal and Multivariate Similarity Clustering, or TMSC. The algorithm finds similarities between multivariate time series, allowing for a controlled level of time distortion and/or warping, and groups observations into a predetermined number of clusters. The algorithm is designed to work with variables of different units and magnitudes. Although not necessary for the calculation, for the best results, it is recommended that the clustered time series are aligned in time and are not fragmented by missing values,

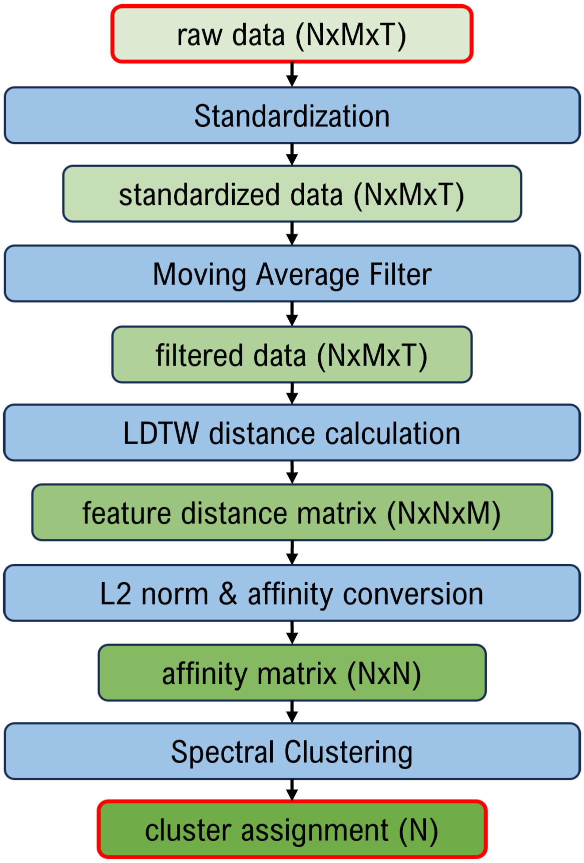

FIGURE 3. The flow chart of Temporal Multivariate Similarity Clustering algorithm. Blue blocks represent functional blocks, and green blocks represent data format steps, with matrix dimensions in brackets.

with a more detailed explanation provided in Section III-B. Fig. 3 graphically represents the flow of the procedure, and the following sections describe the functional blocks of the procedure in detail.

A. DATA PREPROCESSING

Data preprocessing aims to equalize the impact of all variables on distance calculation and facilitate trend matching in data. This can be achieved by KPI standardization, which allows the abstraction of variable units while keeping value range differences between observations of the same variable, and the moving average calculation reduces noise in the variables, smooths out short-term fluctuations, and highlights longer-term trends, as described below.

Given N observations, described each by M variables of length T, the input data set is a matrix $\mathbf{X} = (\mathbf{x}_{ijk})$ where $0 \leq i \leq N$, $0 \leq j \leq M$ and $0 \leq k \leq T$.

In the first step, each variable is standardized among all observations to have a mean of 0 and a standard deviation of 1:

For given j and $\mathbf{x}_j = (\mathbf{x}_{ijk})$, where $0 \leq i \leq N$ and $0 \leq k \leq T$, $\mathbf{x}'_j$ is calculated

$$\mathbf{x}'_j = \frac{\mathbf{x}_j - mean(\mathbf{x}_j)}{std(\mathbf{x}_j)} \tag{1}$$

and $\mathbf{X}' = (\mathbf{x}'_j) = (\mathbf{x}'_{ijk})$.

Next, a simple moving average with a window of length w is applied to standardized data:

$$\mathbf{x}''_{ijk} = \frac{1}{w} \sum_{t=k-w+1}^{k} \mathbf{x}'_{ijt} \tag{2}$$

and $\mathbf{X}'' = (\mathbf{x}''_j) = (\mathbf{x}''_{ijk})$.

B. LDTW DISTANCE CALCULATION

The state-of-the-art method for dealing with time-warped time series is *Dynamic Time Warping* (DTW). It is a distance measure for time series, first introduced in [26] and commonly used in the fields of industrial engineering [27], bioinformatics [28], transportation [29], gesture recognition [16] and flight maneuver recognition [30].

Dynamic Time Warping between two time series $\mathbf{x} = (x_0, \ldots, x_{n-1})$ and $\mathbf{y} = (y_0, \ldots, y_{m-1})$ is formulated as the following optimization problem:

$$DTW(\mathbf{x}, \mathbf{y}) = \min_{\pi} \sqrt{\sum_{(i,j)\in\pi} (x_i - y_j)^2}, \tag{3}$$

where $\pi = [\pi_0, \ldots, \pi_{K-1}]$ is a path of length $0 < K <= n + m$ that satisfies the following properties:

- it is a list of index pairs $\pi_k = (i_k, j_k)$ with $0 \leq i_k < n$ and $0 \leq j_k < m$
- $\pi_0 = (0, 0)$ and $\pi_{K-1} = (n-1, m-1)$
- the following inequalities hold for all $k > 0$:
  - $i_{k-1} \leq i_k \leq i_{k-1} + 1$
  - $j_{k-1} \leq j_k \leq j_{k-1} + 1$

Although *DTW* is widely used in time series comparison tasks, it suffers from the pathological alignment problem, which leads to an abnormally large number of links between two sequences [31]. Therefore, Zhang et al. introduced *Dynamic Time Warping Under Limited Warping Path Length* (LDTW), which limits the total number of links during the optimization process of *DTW* (by setting an upper bound on K) and therefore effectively oppresses the pathological alignment [31].

In terms of data characteristics explained in Section II-A, *LDTW* allows a controlled level of time warping (including time offset) while at the same time remaining sensitive to value offset or scaling of a signal. We used *LDTW* implementation from the `tslearn` [32] library, with our improvements, such as memory preallocation and matrix algebra replacing loops, which improved calculation speed (at the expense of higher memory usage), as presented in Fig. 4.

LDTW is applied pairwise to all observations on each variable separately, resulting in a feature distance matrix $\mathbf{D} = (\mathbf{d}_{nmj})$ where $0 \leq n, m \leq N$ and $0 \leq j \leq M$.

C. DIMENSIONALITY REDUCTION

Since the distance matrix is calculated for each feature (variable) separately, each pair of observations has a feature distance vector of length M. Clustering algorithms that work

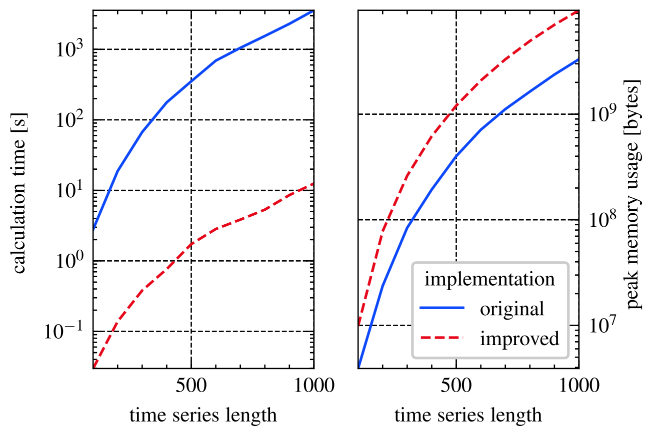

FIGURE 4. *LDTW* calculation time and peak memory usage vs. time series length for original (solid blue line) and improved (dashed red line) implementations on Lenovo P14s notebook (Intel Core i7 vPRO 10th gen, 32GB RAM).

on a precomputed distance or affinity matrix, such as *Spectral Clustering*, described in Section III-D, require a scalar distance for each pair of observations. Hence, after evaluating multiple vector metrics, the *Euclidean* norm (L2) was chosen and applied to each vector of distances of characteristics, resulting in a scalar distance matrix $\mathbf{D}' = (d'_{nm})$, of size NxM, where

$$d'_{nm} = \sqrt{\sum_{j=0}^{M} (d_{nmj})^2}. \tag{4}$$

D. CLUSTERING

Taking into account the characteristics of 5G KPIs, as systematized in Fig. 2, and the need to distinguish between identified patterns, empirical experiments were carried out, which led to the selection of *Spectral Clustering* [33] as the best algorithm for the given task among other state-of-the-art algorithms.

Spectral clustering uses the eigenvalues of the affinity matrix of the data to reduce dimensionality before clustering in fewer dimensions. This work uses the Python implementation from `scikit-learn` library [34]. It worked efficiently with similarity matrix $\mathbf{S}$ created from distance matrix calculated in Section III-C by the following transformation

$$\mathbf{S} = \max(\mathbf{D}') - \mathbf{D}'. \tag{5}$$

IV. RESULTS

Due to the complexity of cellular networks, their temporal performance dependence on the environment, and the need to obtain the approval of the mobile network operator for data collection, it isn't easy to get a real-world ground truth set for algorithm evaluation.

To evaluate TMSC against existing methods (as listed below), we collected KPIs from a selected set of radio cells with significantly different configurations, which, according to telecommunication engineers, should result in performance deviations under similar environmental conditions [35]. Then a task of identifying differences in configuration, based on cells' chosen Key Performance Indicators measurements, was set for TMSC and the state-of-the-art clustering algorithms: *k-Means* [36], *DBSCAN* [37], *Spectral Clustering* [33], *Gaussian Mixture Models (GMM)* [38] as well as the *DANCE* method proposed in [25].

TABLE 1. Configuration for selected classes. Specific bands are not provided due to privacy concerns.

class	frequency band	bandwidth [MHz]	duplex
1	band 1	5	FDD
2	band 2	100	TDD
3	band 3	10	FDD

TABLE 2. Eight chosen Key Performance Indicators with units. 5G SA is short for 5G Stand Alone.

ID	KPI	unit
1	average downlink end-user throughput	Mbit/s
2	average downlink end-user throughput (5G SA only)	Mbit/s
3	maximum downlink end-user throughput	Mbit/s
4	maximum downlink end-user throughput (5G SA only)	Mbit/s
5	data volume transmitted in downlink	MB
6	average cell throughput	Mbit/s
7	downlink physical resource block utilization	%
8	average number of connected users	#

TABLE 3. Results of two-sample *Kolmogorov-Smirnov* test for goodness of fit applied to configuration classes and Key Performance Indicators.

	K-S statistic			*p*-value		
classes	1&2	1&3	2&3	1&2	1&3	2&3
KPI ID						
1	0.86	0.45	0.87	0.00	0.00	0.00
2	0.95	0.62	0.97	0.00	0.00	0.00
3	0.89	0.30	0.88	0.00	0.00	0.00
4	0.95	0.43	0.95	0.00	0.00	0.00
5	0.69	0.37	0.78	0.00	0.00	0.00
6	0.98	0.39	0.98	0.00	0.00	0.00
7	0.12	0.24	0.33	0.00	0.00	0.00
8	0.64	0.29	0.70	0.00	0.00	0.00

A. DATA SET

We nominated three classes of radio cells that differ in the following configuration aspects: operating frequency band, bandwidth, and duplexing method, as presented in Table 1. For the chosen classes, we collected eight network KPIs listed with their units in Table 2, aggregated hourly for one week (a total of 168 samples for each KPI and cell) from live networks of large European operators. The data obtained was then analyzed for possible missing values, and only cells with less than 5% missing values were used for the study. The averages for each cell replaced the missing values. There were 9209 cells left, with 3851, 1658, and 3700 cells for bands 1, 2, and 3, respectively.

To test whether there are significant differences between the selected classes, we performed a two-sample *Kolmogorov-Smirnov* test pairwise to the three configurations on each KPI separately [39]. The results, in the form of *Kolmogorov-Smirnov statistic* and respective *p*-value are presented in Table 3. All comparisons yielded *p*-value practically equal to 0, therefore the null hypothesis that the samples from two compared classes come from identical distributions must be rejected. However, significant differences in the univariate distributions of the KPIs indicated by the *Kolmogorov-Smirnov* test do not indicate performance differences between cells from these classes, as this comparison does not include the environmental conditions under which the cells operated.

B. EVALUATION METHOD

Instead of clustering the entire data set at once, we evaluated TMSC with smaller batches of cells to verify how it performs with limited data. As explained in Section II, an algorithm that does not require a large data set is desired. To this end, we randomly drew 100 cells from each class, resulting in 300 observations. The sample set was then clustered using the proposed method and reference algorithms. Since TMSC uses *Spectral Clustering*, we provide results for *Spectral Clustering* with *Euclidean* distance to verify the importance of the selected distance measure. Similarly, *k-Means*, *GMM*, and *DBSCAN* also used the *Euclidean* distance. We experimented with the parameters of each method and selected those that achieved the best possible results.

Standard cluster analysis approaches (e.g., *silhouette score*, *Calinski-Harabasz index*) depend on the distance metrics adopted and, therefore, cannot be a reference for the direct comparison of the clustering results between methods that use different distance metrics [40]. Since we had the ground truth labels but were solving a clustering problem (not classification), the selected metric was the *Normalized Mutual Information* (NMI) score [41], which is widely used as a similarity measure to evaluate the performance of clustering and classification algorithms, because it is independent of the distance metrics used and the absolute values of the labels.

With $\mathbf{X}$ being a set of predicted labels and $\mathbf{Y}$ a set of ground-truth labels Normalized Mutual Information of $\mathbf{X}$ and $\mathbf{Y}$ is defined as

$$\mathbf{NMI}(X;Y) = \frac{2\mathbf{I}(X;Y)}{\mathbf{H}(X)+\mathbf{H}(Y)}, \tag{6}$$

where $\mathbf{I}$ is Mutual Information of X and Y given by:

$$\mathbf{I}(\mathbf{X};\mathbf{Y}) = \sum_{\mathbf{Y}}\sum_{\mathbf{X}} P_{(\mathbf{X},\mathbf{Y})} \log\left(\frac{P_{(\mathbf{X},\mathbf{Y})}}{P_X P_Y}\right), \tag{7}$$

and $\mathbf{H}$ is entropy given by:

$$\mathbf{H}(\mathbf{X}) = -\sum_{\mathbf{X}} P_X \log P_X. \tag{8}$$

NMI takes values from 0 (no mutual information) to 1 (perfect match).

The random batch selection procedure was repeated 100 times to capture the empirical cumulative distribution functions of NMI for each of the algorithms. The results curves are presented in Fig. 5.

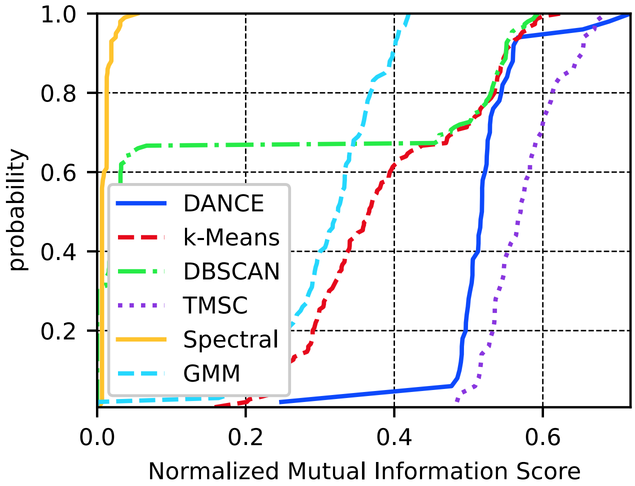

FIGURE 5. Empirical Cumulative Distribution Function of Normalized Mutual Information Score for evaluated clustering algorithms. ***TMSC*** is depicted with purple dotted line, ***DANCE*** - dark blue solid, ***k-Means*** - red dashed, ***DBSCAN*** - green dashed-dotted, ***Spectral Clustering*** - solid yellow and ***Gaussian Mixture Models*** - dashed light blue.

Spectral clustering with *Euclidean* distance metric achieved the worst results, with the NMI in the range of 0 to 0.06, which was expected and highlighted the importance of distance measure selection, compared to the results of our proposed algorithm. Surprisingly, *DBSCAN* was second worst, producing volatile results, failing for the bottom 30% of batches, and achieving results for the top 35% of sample sets comparable to those of *k-Means*, which produced a wide range of NMI from 0.16 to 0.6 and beat the *Gaussian Mixture Models* that delivered maximum NMI value of 0.42. TMSC, providing consistent NMI from 0.48 to 0.69, outperformed all other methods for 95% batches, except for 5% of cases where the second best algorithm *DANCE* was better. It should be noted that *DANCE* employs artificial neural networks for the clustering task; therefore, it requires large amounts of data to learn patterns, and the batch size in this task could have been too small to reach its full potential (a disadvantage concerning operating on relatively short sequences of measurement data, as expected by practical applications in telecommunications and other fields, as commented above in the paper).

V. GENERALIZATION OF TMSC APPLICATION

Because of the difficulty in determining a ground truth set for 5G performance KPIs measured in live networks and demonstrating the applicability of the TMSC algorithm to time series data collected from various systems, a clustering exercise was performed on a data set from other application domains with similar signal characteristics.

We decided to focus on classification problems, where having proper labels allows for reliable method comparison. Classification results are typically evaluated with one or more metrics: accuracy, precision, recall, or F1 score [42]. However, in some cases, the authors of the publication also report the confusion matrix, which allows the computation of Normalized Mutual Information as presented in Section V-A.

A. CONFUSION MATRIX TO NORMALIZED MUTUAL INFORMATION

Confusion matrix of $\mathbf{X}$ and $\mathbf{Y}$ can be interpreted as a table of conditional probabilities $P_{(\mathbf{X}|\mathbf{Y})}$ [43], [44].

Following the Normalized Mutual Information definition in Eq. 6, probabilities $P_{(\mathbf{X},\mathbf{Y})}$, P_X, and P_Y are required for NMI calculation.

The marginal probability P_Y is known from the design of the experiment. The joint probability $P_{(\mathbf{X},\mathbf{Y})}$ is given by:

$$P_{(\mathbf{X},\mathbf{Y})} = P_{(\mathbf{X}|\mathbf{Y})} \cdot P_{\mathbf{Y}} \tag{9}$$

and marginal probability $P_{\mathbf{X}}$ is given by:

$$P_X = \sum_{\mathbf{Y}} P_{(\mathbf{X},\mathbf{Y})} \tag{10}$$

which allows the calculation of Normalized Mutual Information.

B. GESTURE RECOGNITION

The data set selected for evaluation is a gesture recognition data set introduced in [16].[1] The data consists of a time series of X-, Y- and Z-axis acceleration measurements from a hand-held device, collected for eight participants who replicated eight preselected hand gestures multiple times over seven days. The data set available for download is incomplete compared to what was reported in [16]. It contains 55 (15 in the training set and 40 in the testing set) measurements for each gesture, and the information about the participant and the day of the experiment is not present in the data. The authors indicated significant differences in the measurements collected from different participants and even from the same participant on other days and adjusted their methodology accordingly. The authors tested their proposed algorithm using two approaches: evaluating accuracy for each participant and each day separately and evaluating accuracy for each participant regardless of the measurement date. They achieved NMI (calculated from the confusion matrix as in Section V-A) of 0.949 and **0.838** for the former and latter approaches, respectively. The second approach disregards the date information, and without participant information, the task becomes even more complex. Therefore, we compared our clustering results to that result.

We clustered the entire training set, the test set, and the joint data set and obtained NMI of **0.827**, **0.79**, and **0.816** respectively, which is close to the value reported for the participant-specific evaluation. The result is highly satisfactory, as the TMSC was not explicitly designed for this task.

[1]The data set is available at https://www.timeseriesclassification.com/description.php?Dataset=UWaveGestureLibrary (Accessed January 1, 2024)

C. LIMITATIONS

Although the proposed algorithm has been shown to achieve good results, it has limitations. First, it is designed for signals with specific characteristics (potential offset or warping in the time domain, noisy measurements, signal scaling), as described in Section II-A. Second, it requires calculating the pairwise distance using *LDTW* for each variable. Therefore, the computational complexity of the proposed method is $\mathcal{O}(N^2MT^3)$ for the number of samples N, the number of variables (KPIs) M, and the length of the time series T. On the other hand, the second-best reported method, *DANCE*, requires dedicated resources (preferably a powerful Graphics Processing Unit (GPU)) and a significant amount of training time. Specific steps for TMSC could be explored to avoid these difficulties, such as signal quantization, computation parallelization, or the use of hash functions. Lastly, the number of clusters should be known or found using internal clustering evaluation methods.

VI. CONCLUSION

We have defined and systematized a set of patterns observed in experimental data, particularly in performance data measured in 5G cellular networks. We proposed a clustering algorithm, Temporal and Multivariate Similarity Clustering (TMSC), designed to cluster 5G mobile network performance data. The algorithm combines existing techniques, such as Dynamic Time Warping and Spectral Clustering, into a novel chain, allowing efficient clustering of 5G network performance data. We evaluated it against state-of-the-art algorithms in identifying radio cell configuration differences based on time series realization of eight Key Performance Indicators. With a limited sample size, essential for practical use in network performance management, TMSC outperformed all other methods in at least 95% of cases. The result obtained is particularly important when considering learning methods (e.g., those utilizing artificial neural networks, as in [25]), as it allows operation on short data sequences, which are unavailable in alternative methods (especially significant in industrial applications, e.g., in 5G networks and/or processes with significant dynamics).

The algorithm can help the mobile network operator or equipment vendor in cell clustering for various management tasks, e.g., identifying best/worst performing cells, generating virtual clusters for large-scale feature activation, enabling feature gain estimation concerning environmental conditions, etc.

To analyze the potential for generalization of TMSC, we experimented with and evaluated the algorithm in the hand gesture recognition task, producing an outcome comparable to a method specifically designed for that purpose.

We identified the limitations of the procedure and indicated promising mitigation steps. Another potential research area is to explore the possibility of incorporating regularization methods, which have yielded promising results in [25].

REFERENCES

[1] A. Rizwan, M. Jaber, F. Filali, A. Imran, and A. Abu-Dayya, "A zero-touch network service management approach using AI-enabled CDR analysis," *IEEE Access*, vol. 9, pp. 157699–157714, 2021.

[2] C. Benzaid and T. Taleb, "AI-driven zero touch network and service management in 5G and beyond: Challenges and research directions," *IEEE Netw.*, vol. 34, no. 2, pp. 186–194, Mar. 2020.

[3] A. Oi, R. Sato, Y. Suto, K. Sakata, M. Nakajima, and T. Furukawa, "A study on automation of network maintenance in telecom carriers for zero-touch operations," in *Proc. 21st Asia–Pacific Netw. Oper. Manage. Symp. (APNOMS)*, Sep. 2020, pp. 1–6.

[4] W. B. Abbas, Q. Z. Ahmed, F. A. Khan, N. S. Mian, P. I. Lazaridis, and P. Sureephong, "Designing future wireless networks (FWN)s with net zero (NZ) and zero touch (ZT) perspective," *IEEE Access*, vol. 11, pp. 83301–83321, 2023.

[5] F. Adachi, R. Takahashi, and H. Matsuo, "Enhanced interference coordination and radio resource management for 5G advanced ultra-dense RAN," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring)*, May 2020, pp. 1–5.

[6] S. Bassoy, M. Jaber, M. A. Imran, and P. Xiao, "Load aware self-organising user-centric dynamic CoMP clustering for 5G networks," *IEEE Access*, vol. 4, pp. 2895–2906, 2016.

[7] M. Joud, M. García-Lozano, and S. Ruiz, "User specific cell clustering to improve mobility robustness in 5G ultra-dense cellular networks," in *Proc. 14th Annu. Conf. Wireless On-Demand Netw. Syst. Services (WONS)*, Feb. 2018, pp. 45–50.

[8] W. Dghais, M. Souilem, H. R. Chi, A. Radwan, and A. M. Taha, "Dynamic clustering for power effective small cell deployment in HetNet 5G networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2020, pp. 1–5.

[9] M. Sousa, A. Alves, P. Vieira, M. P. Queluz, and A. Rodrigues, "Analysis and optimization of 5G coverage predictions using a beamforming antenna model and real drive test measurements," *IEEE Access*, vol. 9, pp. 101787–101808, 2021.

[10] A. Mahimkar, Z. Ge, X. Liu, Y. Shaqalle, Y. Xiang, J. Yates, S. Pathak, and R. Reichel, "Aurora: Conformity-based configuration recommendation to improve LTE/5G service," in *Proc. 22nd ACM Internet Meas. Conf.*, New York, NY, USA, Oct. 2022, pp. 83–97.

[11] E. J. Oughton, Z. Frias, S. van der Gaast, and R. van der Berg, "Assessing the capacity, coverage and cost of 5G infrastructure strategies: Analysis of The Netherlands," *Telematics Informat.*, vol. 37, pp. 50–69, Apr. 2019.

[12] M. Sirbu and G. Harman, "Performance verification for 5G NR deployments," Ericsson AB, Stockholm, Sweden, Tech. Rep. GFMC-19:000148, Mar. 2019. [Online]. Available: https://web.archive.org/web/20230726202136/https://www.ericsson.com/4ac658/assets/local/reports-papers/white-papers/wp_5g_nrperformanceverification.pdf

[13] J. Moysen, F. Ahmed, M. García-Lozano, and J. Niemelä, "Big data-driven automated anomaly detection and performance forecasting in mobile networks," in *Proc. IEEE Globecom Workshops (GC Wkshps*, Dec. 2020, pp. 1–5.

[14] J. Yin, W. Li, X. Wang, X. Ye, and Y. Ouyang, "4G 5G cell-level multi-indicator forecasting based on dense-MLP," 2022, *arXiv:2209.05989*.

[15] S. Aghabozorgi, A. S. Shirkhorshidi, and T. Y. Wah, "Time-series clustering—A decade review," *Inf. Syst.*, vol. 53, pp. 16–38, Nov. 2015.

[16] J. Liu, L. Zhong, J. Wickramasuriya, and V. Vasudevan, "UWave: Accelerometer-based personalized gesture recognition and its applications," *Pervas. Mobile Comput.*, vol. 5, no. 6, pp. 657–675, Dec. 2009.

[17] F. M. Caputo, P. Prebianca, A. Carcangiu, L. D. Spano, and A. Giachetti, "Comparing 3D trajectories for simple mid-air gesture recognition," *Comput. Graph.*, vol. 73, pp. 17–25, Jun. 2018.

[18] S. Lu, G. Mirchevska, S. S. Phatak, D. Li, J. Luka, R. A. Calderone, and W. A. Fonzi, "Dynamic time warping assessment of high-resolution melt curves provides a robust metric for fungal identification," *PLoS ONE*, vol. 12, no. 3, Mar. 2017, Art. no. e0173320.

[19] M. I. Chidean, L. I. Jiménez Gil, J. Carmona-Murillo, and D. Cortés-Polo, "Information theory based clustering of cellular network usage data for the identification of representative urban areas," *Digit. Commun. Netw.*, pp. 1–13, Jul. 2023, doi: 10.1016/j.dcan.2023.07.002.

[20] B. Mahdy, H. Abbas, H. Hassanein, A. Noureldin, and H. Abou-zeid, "A clustering-driven approach to predict the traffic load of mobile networks for the analysis of base stations deployment," *J. Sens. Actuator Netw.*, vol. 9, no. 4, p. 53, Nov. 2020.

[21] X. Wang, Y. Jin, and Y. Yu, "A mobile network performance evaluation method based on multivariate time series clustering with auto-encoder," in *Proc. 2nd Int. Conf. Telecommun. Commun. Eng. (ICTCE)*, New York, NY, USA, 2018, pp. 33–37.

[22] S. Li, G. Francini, and E. Magli, "Temporal dynamics clustering for analyzing cell behavior in mobile networks," *Comput. Netw.*, vol. 223, Mar. 2023, Art. no. 109578.

[23] S. Lu, B. Qian, L.-G. Zhao, and Q. Sun, "Mobile networks classification based on time-series clustering," in *Proc. IEEE 5th Int. Conf. Electron. Commun. Eng. (ICECE)*, Dec. 2022, pp. 65–71.

[24] S. Wang and R. Ferrús, "Extracting cell patterns from high-dimensional radio network performance datasets using self-organizing maps and K-means clustering," *IEEE Access*, vol. 9, pp. 42045–42058, 2021.

[25] M. Kajó, J. Schnellbach, S. S. Mwanje, and G. Carle, "Clustering mobile network data with decorrelating adversarial nets," in *Proc. NOMS IEEE/IFIP Netw. Oper. Manage. Symp.*, Apr. 2022, pp. 1–9.

[26] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-26, no. 1, pp. 43–49, Feb. 1978.

[27] G. Gins, P. Van den Kerkhof, and J. F. M. Van Impe, "Hybrid derivative dynamic time warping for online industrial batch-end quality estimation," *Ind. Eng. Chem. Res.*, vol. 51, no. 17, pp. 6071–6084, 2012.

[28] H. Skutkova, M. Vitek, P. Babula, R. Kizek, and I. Provaznik, "Classification of genomic signals using dynamic time warping," *BMC Bioinf.*, vol. 14, pp. 1–7, Aug. 2013.

[29] Y. Xi, D. Huang, Y. Yuan, Z. Liu, K. Anish, and N. Zheng, "Improved dynamic time warping algorithm for bus route trajectory curve fitting," *J. Transp. Eng., A, Syst.*, vol. 147, no. 8, Aug. 2021, Art. no. 04021044.

[30] W. Tian, H. Zhang, H. Li, and Y. Xiong, "Flight maneuver intelligent recognition based on deep variational autoencoder network," *EURASIP J. Adv. Signal Process.*, vol. 2022, no. 1, p. 21, Dec. 2022.

[31] Z. Zhang, R. Tavenard, A. Bailly, X. Tang, P. Tang, and T. Corpetti, "Dynamic time warping under limited warping path length," *Inf. Sci.*, vol. 393, pp. 91–107, Jul. 2017.

[32] R. Tavenard, J. Faouzi, G. Vandewiele, F. Divo, G. Androz, C. Holtz, M. Payne, R. Yurchak, M. Rußwurm, K. Kolar, and E. Woods, "Tslearn, a machine learning toolkit for time series data," *J. Mach. Learn. Res.*, vol. 21, no. 118, pp. 1–6, 2020.

[33] U. V. Luxburg, "A tutorial on spectral clustering," *Statist. Comput.*, vol. 17, no. 4, pp. 395–416, Dec. 2007.

[34] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Nov. 2011.

[35] M. Panek, A. Pomykała, I. Jabłoński, and M. Woźniak, "5G/5G+ network management employing AI-based continuous deployment," *Appl. Soft Comput.*, vol. 134, Feb. 2023, Art. no. 109984.

[36] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. TIT-28, no. 2, pp. 129–137, Mar. 1982.

[37] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," *Knowl. Discovery Data Mining*, vol. 96, no. 34, pp. 226–231, 1996.

[38] J. D. Banfield and A. E. Raftery, "Model-based Gaussian and non-Gaussian clustering," *Biometrics*, vol. 49, no. 3, p. 803, Sep. 1993.

[39] J. L. Hodges, "The significance probability of the Smirnov two-sample test," *Arkiv För Matematik*, vol. 3, no. 5, pp. 469–486, Jan. 1958.

[40] M. M. Gösgens, A. Tikhonov, and L. Prokhorenkova, "Systematic analysis of cluster similarity indices: How to validate validation measures," in *Proc. Int. Conf. Mach. Learn. (PMLR)*, 2021, pp. 3799–3808.

[41] C. Studholme, D. L. G. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Pattern Recognit.*, vol. 32, no. 1, pp. 71–86, Jan. 1999.

[42] M. Hossin and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluations," *Int. J. Data Mining Knowl. Manage. Process.*, vol. 5, no. 2, pp. 1–11, Mar. 2015.

[43] R. Q. Quiroga and S. Panzeri, "Extracting information from neuronal populations: Information theory and decoding approaches," *Nature Rev. Neurosci.*, vol. 10, no. 3, pp. 173–185, Mar. 2009.

[44] I. Kononenko and I. Bratko, "Information-based evaluation criterion for classifier's performance," *Mach. Learn.*, vol. 6, pp. 67–80, Jan. 1991.

JAKUB MAZGUŁA received the B.Sc. degree in mathematics and the M.Sc. degree in mathematics for industry and commerce from Wrocław University of Science and Technology, Poland, in 2013 and 2015, respectively. Since 2015, he has been held various research positions with Nokia and Nokia Bell Labs in Poland and USA. Currently, he leads a Data Science Team with Nokia, focused on creating advanced analytics of mobile network performance. His research interest includes the modeling and optimization of complex systems.

DARIUSZ KRÓL (Senior Member, IEEE) is currently a Professor with Wrocław University of Science and Technology; the Vice-Dean of General Affairs with the Faculty of Information and Communication Technology; and the Head of the Knowledge Engineering Group, Department of Applied Informatics. He has held positions as a Marie Curie Senior Research Fellow with Bournemouth University and a DAAD Research Fellow with Friedrich-Alexander University. He has 30 years of academic experience and about 20 years in R&I projects (as a CEO and the Head of research and development). He has published over 150 peer-reviewed papers, co-authored three books, and co-edited four. He was an invited lecturer for many events throughout the U.K., Europe, and Asia, in the last ten years. His research and teaching interests are oriented toward knowledge engineering applications, data quality, computational intelligence, and cognitive technologies for adaptive and sustainable manufacturing systems. He is a Project Evaluator with the European Commission. His prestigious contribution in the field was rewarded by IBM and granted by the European Commission, the Polish National Center for Science, the Polish Ministry of Science and Higher Education, and DAAD. He has a great deal of hands-on experience with EU-funded projects and programs. Moreover, he has been elected to the Polish Accreditation Committee in recognition of their professional standing.

IRENEUSZ JABŁOŃSKI received the M.Sc., Ph.D., and D.Sc. degrees in electronics from Wrocław University of Science and Technology, Poland, in 1999, 2004, and 2015, respectively. He currently leads the "AI-Based Systems" Group, Fraunhofer IPMS, Germany, and the Brandenburg University of Technology Cottbus-Senftenberg, Germany. His research interests include complex systems measurements, sensing and communication technologies, edge devices, AI/ML techniques in systems modeling and data processing, multisensor data fusion, and cyber-physical systems.