Anomaly Detection for Multi-Region Radio Noise Measurement Data

Youngkeun Yoon
Radio Resource Research Sector
Electronics and Telecommunications Research Institute
Daejeon, Republic of Korea
ykyoon@etri.re.kr

Heonjin Hong
Radio Resource Research Sector
Electronics and Telecommunications Research Institute
Daejeon, Republic of Korea
hjhong@etri.re.kr

Ho-Kyung Son
Radio Resource Research Sector
Electronics and Telecommunications Research Institute
Daejeon, Republic of Korea
hgson@etri.re.kr

***Abstract*— This paper investigates anomaly detection in multi-region radio noise power measurements through two–dimensional Principal Component Analysis preprocessing, followed by two widely recognized machine learning–based unsupervised outlier detection methods: Density–Based Spatial Clustering of Applications with Noise and Isolation Forest. Measurement data from six regions were synthesized to simulate realistic electromagnetic noise conditions, including sporadic and non-specific anomalous received power events. The preprocessing step reduced the high-dimensional feature space to two principal components while preserving more than 97% of the total variance, thereby enabling efficient visualization and robust feature extraction for subsequent learning algorithms. For radio noise monitoring and interference management across various urban areas in South Korea, this study conducted a comparison and simulation of the performance of well–known unsupervised machine learning algorithms. Although the analysis was limited in terms of the measurement data and the applied unsupervised learning methods, the results indicated that the DBSCAN method was more effective than the Isolation Forest in detecting and capturing anomalies, suggesting its potential applicability in future related research.**

Keywords*— *Radio, Noise, DBSCAN, Isolation Forest, PCA

I. Introduction

Radio noise measurement plays a crucial role in wireless communication system design, electromagnetic com–patibility testing, and interference management. Field measurements, however, are often affected by variability arising from environmental changes, transient interference events, and hardware malfunctions. Detecting such anomalies is essential for ensuring reliable spectrum monitoring and maintaining network stability. Outlier detection can be formulated as an unsupervised learning problem when labeled anomaly data are scarce or unavailable [1]. Among the numerous methods in this domain, Density-Based Spatial Clustering of Applications with Noise(DBSCAN) [2] and Isolation Forest(IF) [3] have emerged as prominent techniques. DBSCAN is an algorithm that identifies anomalies as points located in low–density regions of the feature space. It requires minimal prior assumptions about the data distribution and can detect clusters of arbitrary shapes. IF, on the other hand, is an ensemble–based method that isolates anomalies through recursive partitioning, leveraging the fact that anomalies are easier to isolate than normal points. Both methods are non–parametric and can be effectively applied after dimensionality reduction, such as PCA [1], to improve computational efficiency and interpretability. In this study, we employ a preprocessing step to transform multi–region radio noise measurement data into a 2D space, thereby enhancing the interpretability and effectiveness of machine learning–based detection algorithms. By comparing DBSCAN and IF in the reduced feature space, we discuss their applicability to real-time radio noise monitoring systems.

II. Measurement and Analysis

We analyzed a dataset of radio noise measurements comprising six regional columns and 418 rows. All features were standardized using the z–score method, involving per-feature mean removal and variance normalization.

The measurement campaign, which was directly established and implemented by Electronics and Telecommunications Research Institute(ETRI), was designed to capture representative radio noise power levels in six major metropolitan areas selected to encompass diverse electromagnetic environments, with a particular focus on dense urban conditions. For each urban area, a predefined scenario was developed that included designated measurement routes within the city.

The measurement system was mounted on a vehicle and operated while traversing the routes, continuously recording received noise power over the specified frequency range and logging the data to a monitoring computer. The measurement setup comprised an omnidirectional broadband antenna to ensure uniform azimuthal coverage, a low-loss RF cable, a band-pass filter, a high gain low-noise amplifier, and a high-resolution spectrum analyzer serving. The spectrum analyzer was interfaced with a monitoring and control computer equipped with custom software for automated data acquisition, and real-time visualization. Field measurements were conducted across six distinct major urban areas, following the predefined scenario. At each urban site, measurements were performed continuously for several hours to a full day depending on local scheduling constraints. Frequency sweeps were configured with a predefined resolution bandwidth, and the sweep time was optimized to balance temporal and spectral granularity.

Given the high dimensionality of the standardized measurement matrix $\boldsymbol{X} \in \mathrm{R}^{n\times d}$, where n=418 denotes the number of PSD measurement instances and d=6 represents the distinct region areas, and considering the potential correlation among these regional features, PCA was applied as a preprocessing step to remove inter–feature correlations. The sample covariance matrix was computed as:

$$\boldsymbol{\Sigma} = \frac{1}{n-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}, \tag{1}$$

The eigen–decomposition,

$$\boldsymbol{\Sigma W} = \boldsymbol{W\Lambda}, \tag{2}$$

yielded the principal axes $\boldsymbol{W}$ (eigenvectors) and their corresponding explained variances $\boldsymbol{\Lambda}$ (eigenvalues). The first two eigenvectors, ω_1 and ω_2, were selected, and the dataset was projected into 2D as:

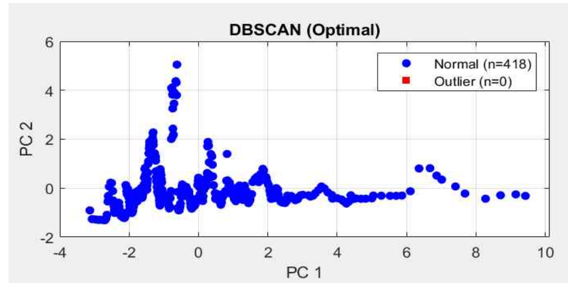
$$\boldsymbol{Z} = \boldsymbol{X}[\boldsymbol{\omega_1}, \boldsymbol{\omega_2}]. \tag{3}$$

This transformation preserved the dominant variance structure of the original data while mapping it into a 2D space, enabling direct visualization and enhancing the performance of subsequent anomaly detection algorithms such as DBSCAN and IF. DBSCAN identifies clusters as high–density regions separated by low-density boundaries, labeling points outside any cluster as noise(outliers). The algorithm operates based on two key factors: the neighborhood radius(ε) and the minimum count of data points needed to identify a core point (*minPts*). IF operates by recursively partitioning the data, isolating points that require fewer splits. In this study, the standard IF scoring mechanism was replaced with a Mahalanobis distance-based criterion to enhance robustness in PCA space. Specifically, anomalies were flagged by thresholding the squared Mahalanobis distance at the 95th percentile in the baseline configuration and at the 97th percentile in the optimized. This adjustment allowed for a clearer separation of extreme deviations from the cluster centroid while preserving sensitivity to sparse anomalies.

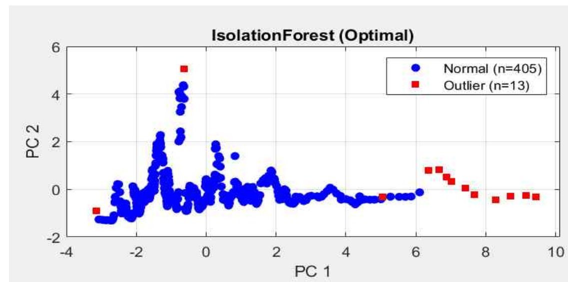
PCA was performed via singular value decomposition (SVD) on the centered matrix to obtain scores and explained-variance ratios. The first two principal components retained approximately 97.34% of the total variance(PC1: 80.63%, PC2: 16.72%), indicating that 2D embedding preserves the salient structure of the meas–urement space with minimal information loss and is suitable for 2D anomaly visualization/analysis. This high cum–ulative explained variance ratio confirms that the 2D projection was sufficient to capture the essential structure of the radio noise measurement dataset without significant loss of information, thereby facilitating efficient visualization and robust feature extraction for anomaly detection. Due to the absence of ground-truth anomaly labels, evaluation was qualitative. Performance was assessed using PCA–space scatter plots, anomaly counts, and spatial distribution patterns. In the finalized implementation, DBSCAN was run with $minPts = 3$ and $\varepsilon = 0.5$ on the full dataset, while for the high–PC1 subset ε was adaptively set to the 75th percentile of pairwise distances; the tuned ("Optimal") configuration used $\varepsilon = 0.8$ with $minPts = 3$. For IF, anomalies were flagged by thresholding the squared Mahalanobis distance at the 95th percentile in the baseline run and at the 97th percentile in the optimized run.

III. Results and Discussion

Figure 1 presents the PCA projections of the detection results for DBSCAN and IF under their respective optimal configurations. In the DBSCAN(Optimal) setting(ε=0.8, *minPts*=3), no points(n=0 out of 418) were labeled as outliers, indicating that all observations were assigned to clusters within the PCA space. The resulting distribution is compact, with high local density and no significant deviations detected. This outcome suggests that, at ε=0.8, local densities are sufficiently uniform to incorporate mild deviations into existing clusters, rendering DBSCAN highly conservative in this regime. In contrast, the IF(Optimal) configuration, employing a Mahalanobis distance–based criterion with a 97th percentile threshold, identified approximately 3.1% of the points(n=13 out of 418) as anomalies. These anomalies are scattered across the PCA space but exhibit two notable concentrations: one near the vertical spike around PC1 $\approx$ 0 and another along the extended spread near PC1$\approx$8. The Mahalanobis approach, being global and shape-agnostic, remains sensitive to points lying far from the mean in the decorrelated PCA space, even under tightened thresholds. DBSCAN can be employed to avoid over-flagging in dense, stable regimes, while the Mahalanobis-based IF can capture globally rare configurations that may be overlooked by density-based methods.



(a) DBSCAN



(b) Isolation Forest

Fig. 1. PCA Plot Results of DBSCAN and Isolation Forest(IF) Showing Normal value and Outlier Detection

IV. Conclusions

Using radio noise power measurement data collected from urban in Korea, we applied unsupervised learning methods to determine the presence of anomalies. Despite the limited scope of the dataset and conditions, DBSCAN outperformed Isolation Forest in detecting and capturing anomalies, providing more stable results. The analysis results suggest that as the magnitude of anomalies increases, the corresponding unsupervised machine learning method may exhibit greater error.

Acknowledgment

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.RS-2023-00217885, Development of integrated interference analysis technology for improving frequency utilization efficiency)

References

[1] T. Baranwal, A. Das, S. Varada, S. Das, and M. R. Haider, "Machine learning-based anomaly detection of correlated sensor data: An integrated principal component analysis–autoencoder approach," *arXiv preprint*, arXiv:2505.24044v1 [eess.SP], 2025. Available: https://arxiv.org/abs/2505.24044

[2] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. 2nd Int. Conf. Knowledge Discovery and Data Mining (KDD)*, Portland, OR, USA, Aug. 1996, pp. 226–231.

[3] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *Proc. 2008 8th IEEE Int. Conf. Data Mining (ICDM)*, Pisa, Italy, Dec. 2008, pp. 413–422.