Online Nonparametric Anomaly Detection based on Geometric Entropy Minimization

Yasin Yılmaz
Department of Electrical Engineering
University of South Florida
Tampa, FL 33620
Email: yasiny@usf.edu

Abstract—We consider the online and nonparametric detection of abrupt and persistent anomalies, such as a change in the regular system dynamics at a time instance due to an anomalous event (e.g., a failure, a malicious activity). Combining the simplicity of the nonparametric Geometric Entropy Minimization (GEM) method with the timely detection capability of the Cumulative Sum (CUSUM) algorithm we propose a computationally efficient online anomaly detection method that is applicable to high-dimensional datasets, and at the same time achieve a near-optimum average detection delay performance for a given false alarm constraint. We provide new insights to both GEM and CUSUM, including new asymptotic analysis for GEM, which enables soft decisions for outlier detection, and a novel interpretation of CUSUM in terms of the discrepancy theory, which helps us generalize it to the nonparametric GEM statistic. We numerically show, using both simulated and real datasets, that the proposed nonparametric algorithm attains a close performance to the clairvoyant parametric CUSUM test.

Keywords—anomaly detection, change detection, k-nearest-neighbor graph, discrepancy theory, entropy minimization

I. INTRODUCTION

Anomaly detection is an important problem with various applications, such as cybersecurity, quality control, fraud detection, fault detection, and health care [1]. It deals with identifying patterns that deviate from a nominal behavior. Although not exactly the same, it is sometimes considered equivalent to outlier detection, which deals with how well a given data sample fits to the nominal behavior [2, 3]. Sequential detection methods, on the other hand, considers also the temporal information in the sequence of data samples regarding a possible anomaly. In particular, sequential change detection looks for abrupt and persistent changes [4]. Moreover, sequential methods provide timely detection capabilities.

Parametric methods assume probabilistic models for both nominal or anomalous states. In practice, it is difficult to know the anomalous distribution, hence model mismatch limits the applicability of parametric approaches only to specific types of anomaly whose probability distribution is well aligned with the assumed model. To that end, statistical outlier detection deals only with significant deviations from the nominal distribution. However, in high-dimensional problems, it is even difficult to know the nominal distribution. Hence, recently, effective nonparametric methods for anomaly detection based on minimum entropy sets are proposed [2, 4]. Nevertheless, these methods are solely based on outlier detection, and thus cannot use the temporal information regarding anomaly in the data samples.

In this paper, for timely and accurately detecting abrupt and persistent anomalies, we develop a nonparametric method that is simple enough to work with high-dimensional datasets. Providing key insights to both outlier detection and change detection we merge the desirable properties of nonparametric methods and online methods. Specifically, we show asymptotic links between nonparametric Geometric Entropy Minimization (GEM) and parametric outlier detection; and also provide a novel interpretation of the sequential Cumulative Sum (CUSUM) method using the discrepancy theory, which dates back to Hermann Weyl’s work in 1916 [5].

In Section II, we formulate the problem and provide relevant background information. Then, in Section III, we present our proposed anomaly detection method, and in Section IV, evaluate its performance through numerical experiments. Finally, Section V concludes the paper.

II. PROBLEM FORMULATION AND BACKGROUND

We monitor a system online by getting sequential observations $X_t = \{X_1, X_2, \ldots, X_t\}$ of $d$-dimensional independent vectors $X_t$ in time. In this paper, we model an anomaly as persistent outliers in the observations, and aim to accurately detect such anomalies in a timely fashion using a practical nonparametric approach applicable to high-dimensional datasets (e.g., big data problems).

A. Change Detection

Consider an abrupt and persistent change in the probability distribution of the observations at an unknown point in time. Let $f_0$ denote the nominal probability distribution of $X_t$ before any change, and $f$ denote the actual distribution of $X_t$ We formulate the problem as a composite binary hypothesis test as

$$H_0 : f = f_0, \forall t$$
$$H_1 : f = f_0, t < \tau, \text{ and } f \neq f_0, t \geq \tau,$$  \hspace{1cm} (1)

where $\tau$ denotes the unknown change time. As seen in the formulation, in the change detection problem, $H_0$ (i.e., nominal distribution) is assumed true at the beginning, and the objective is to statistically detect the potential switching of the true distribution to an anomalous distribution. Hence, the change detection problem is fundamentally different than the standard binary hypothesis testing problem, in which

$$H_0 : f = f_0, H_1 : f \neq f_0,$$  \hspace{1cm} (2)
and the objective is to find the true hypothesis (H₀ or H₁) from the beginning.

Standard hypothesis testing (either fixed-sample-size or sequential) for anomaly detection (see (2)) outputs a decision: nominal (φ = H₀) or anomalous (φ = H₁), whereas in change detection (see (1)) testing continues until the output decision is H₁, i.e., until an anomaly is detected [4]. Accordingly, the objective in the former is to maximize the detection probability \( P(\phi = H₁ | H₀) \) while satisfying a false alarm (i.e., false positive) constraint \( P(\phi = H₁ | H₀) ≤ α \), whereas the objective in the latter is to minimize the average detection delay while again satisfying a false alarm constraint. Hence, the term quickest detection is also used [4].

The minimax performance criterion [7]

\[
\inf_T \sup_\tau \sup_{X_\tau} \mathbb{E}_T[(T - \tau)^+ | X_\tau]
\]

subject to \( E_\infty[T] ≥ β \),

is commonly used to formulate the change detection problem. In [3], “ess sup” denotes essential supremum, a concept in measure theory that is in practice equivalent to supremum; \( (\cdot)^+ = \max\{\cdot, 0\} \); \( E_\tau \) is the expectation with respect to the change time; and accordingly \( E_\infty \) is the expectation when there is no change at all. The minimax performance criterion in (3) minimizes the worst-case average detection delay subject to a false alarm constraint, represented by a lower bound \( β \) on the expected alarm time when there is never a change.

It is known that the widely used cumulative sum (CUSUM) algorithm [8], is optimum with respect to (3) when both the nominal distribution \( f_0 \) and the anomalous distribution \( f_1 \), to which \( f \) switches at time \( τ \), are completely known [4]. The CUSUM procedure is given by

\[
T_c = \min\{t : \max_{1 ≤ j ≤ t} S^c_j ≥ h_c\},
\]

\[
S^c_j = \sum_{i=1}^j \log \frac{f_1(X_i)}{f_0(X_i)},
\]

where \( T_c \) is the stopping time, \( S^c_j \) is the running log-likelihood ratio from time \( j \) to time \( t \), and \( h_c \) is a threshold selected to satisfy the false alarm constraint in (3) with equality, i.e., \( E_\infty[T] = β \). Operationally, CUSUM stops the first time the evidence against \( H₀ \) is sufficiently large, achieving quickest detection among its competitors satisfying the same false alarm constraint. However, CUSUM is a parametric method which requires the knowledge of \( f_0 \) and \( f_1 \) up to the parameters, limiting its use in high-dimensional problems where \( d ≥ 2 \). In generalized CUSUM, which estimates the parameters of \( f_0 \) and \( f_1 \) using maximum likelihood estimation, only asymptotic optimality is achievable [4]. To tackle high-dimensional problems we resort to nonparametric methods, such as GEM, which is discussed next.

B. Outlier Detection via GEM

Parametric change detection methods, in particular CUSUM, enable timely detection of certain anomaly types in which the anomalous distribution is known, as well as the nominal (i.e., baseline) distribution (e.g., change in the mean or variance of a Gaussian distribution). Outlier detection deals with the general problem of detecting unknown anomaly types (see (1) and (2)) by considering only the likelihood under the nominal distribution. For instance, a data point is declared as an outlier if it lies outside the most compact set of data points under the nominal distribution, called the minimum volume set.

The minimum volume set of level \( α \) is given by

\[
Ω_α = \arg \min_{\mathcal{A}} \int_{\mathcal{A}} dx \text{ subject to } \int_{\mathcal{A}} f_0(x) dx ≥ 1 - α,
\]

where \( x \) is a data point, i.e., a realization of the random variable \( X_t \), \( \mathcal{A} \) is the acceptance region for \( H₀ \) in which a data point is deemed nominal, and \( α \) is the significance level, i.e., constraint on the false alarm probability. In [5], \( Ω_α \) minimizes the Lebesgue measure (i.e., volume) in \( \mathbb{R}^d \) among the subsets of data points satisfying the same false alarm constraint \( α \) to minimize the interference with anomalous data points, i.e., to minimize \( P(\phi = H_0 | H_1) \) and thus to maximize the detection probability \( P(\phi = H_1 | H_0) \). Indeed, the detector based on the minimum volume set has a strong optimality property: it is the uniformly most powerful test when the nominal distribution \( f_0 \) is Lebesgue continuous and has no flat spots over its support set, and the actual distribution \( f \) is a linear mixture of \( f_0 \) and the uniform distribution [2]. It is also known that the minimum volume set \( Ω_α \) coincides with the minimum entropy set which minimizes the Rényi entropy while satisfying the same false alarm constraint [3].

In high-dimensional datasets \( (d ≥ 2) \), even if \( f_0 \) is known, it is very computationally expensive (if not impossible) to determine \( Ω_α \). Hence, in the literature, there are various methods for learning minimum volume sets [9]. One of them, called Geometric Entropy Minimization (GEM), is shown to be very effective with high-dimensional datasets while asymptotically achieving the performance of minimum volume set [2]. Specifically, from a training set \( \mathcal{X}^N \) of \( N \) data points distributed according to \( f_0 \), it first forms a k-nearest-neighbor (kNN) Euclidean graph \( G = (\mathcal{X}^N, E) \) with \( K \) vertices \( \mathcal{X}^N_K \in \mathcal{X}^N \) and \( kK \) edges

\[
E = \{e_{i(l)} : i = 1, \ldots, K; l = 1, \ldots, k\},
\]

where the edge length \( |e_{i(l)}| \) is the Euclidean distance between the \( i \)th data point and its \( l \)th nearest neighbor in the graph, and the vertices \( \mathcal{X}^N_K \) are chosen by minimizing the total weighted edge length

\[
L_k(\mathcal{X}^N_K) = \sum_{i=1}^K \sum_{l=1}^k |e_{i(l)}|^γ,
\]

over all possible \( K \)-point subsets of \( \mathcal{X}^N \), where \( γ > 0 \) is the weight. Then, with a new data point \( X_t \), it recomputes the kNN graph described above over the extended set \( \mathcal{X}^N \cup \{X_t\} \), resulting in the updated vertices \( \mathcal{X}^{N+1}_K \in \mathcal{X}^{N+1} \cup \{X_t\} \). If \( X_t \in \mathcal{X}^{N+1}_K \), the new data point \( X_t \) is classified as nominal; otherwise anomalous.

In [2], using the asymptotic theory of Euclidean graphs \( \mathcal{X}^{N+1}_K \) is shown to converge to the minimum volume set (and accordingly the minimum entropy set) \( Ω_α \) as

\[
\lim_{K, N → ∞} K/N → 1 - α.
\]

Since the original GEM algorithm, which, for each \( X_t \), computes \( \mathcal{X}^{N+1}_K \) over all possible \( K \)-point subsets of \( \mathcal{X}^N \cup \{X_t\} \), has exponential computational complexity, a simpler variant...
based on bipartite kNN graph (BP-GEM) is proposed in [3]. BP-GEM significantly decreases the complexity of GEM from \(O(dK^2(N)^k)\) to \(O(dN^{(8+3d)/(4+2d)})\) while maintaining the theoretical guarantees of GEM [3]. Specifically, BP-GEM randomly partitions the training data set \(X^N\) into two sets \(X^{N_1}\) and \(X^{N_2}\), where \(N_1 + N_2 = N\), and finds the vertices \(X_{K}^N\) in \(X^{N_1}\) by minimizing

\[
L_k(X_{K}^N, X^{N_2}) = \sum_{i=1}^{K} \sum_{l=k-s+1}^{k} |e_{i(l)}|^\gamma, \quad (7)
\]

over all possible \(K\)-point subsets of \(X^{N_1}\), where \(|e_{i(l)}|\) is the Euclidean distance from point \(i\) in \(X_{K}^N\) to its \(l\)th nearest neighbor in \(X^{N_2}\). \(1 \leq s \leq k\) is a fixed number introduced for convenience, and \(0 < \gamma < d\) is the weight. This initial graph with vertices \(X_{K}^N\) is computed only once at the beginning. Then, with the inclusion of each new data point \(X_t\), we do not need to re-determine the graph vertices \(X_{K}^N\) over \(X^{N_1} \cup \{X_t\}\) every time, since for each \(i\) in \(X_{K}^N\), the nearest neighbors \(\{s(l)\}\) are selected from the separate set \(X^{N_2}\), and thus the total edge length \(\sum_{i=k-s+1}^{k} |e_{i(l)}|^\gamma\) does not change, as opposed to the original GEM. Instead, we only need to compute the total edge length \(\sum_{i=k-s+1}^{k} |e_{X_t(l)}|^\gamma\) for the new point \(X_t\), and choose the \(K\) points with smallest total edge lengths from \(X_{K}^N \cup \{X_t\}\) as the new graph vertices \(X_{K}^{N+1}\), with the decision function

\[
\phi = \begin{cases} 
0 & \text{if } D_t \leq 0 \\
1 & \text{if } D_t > 0,
\end{cases} \quad (9)
\]

\[
D_t = \sum_{l=k-s+1}^{k} |e_{X_t(l)}|^\gamma - \sum_{l=k-s+1}^{k} |e_{X_t(l)}|^\gamma, \quad (10)
\]

\[
\lim_{N_1, N_2 \to \infty} D_t \text{ monotonically } \sim \lim_{N_1, N_2 \to \infty} \log \frac{f_0(x_\alpha)}{f_0(x_t)}, \quad (11)
\]

and

\[
\lim_{N_1, N_2 \to \infty} D_t = \lim_{N_1, N_2 \to \infty} \log \frac{f_0(x_\alpha)}{f_0(x_t)}, \quad (12)
\]

where \(x_\alpha\) is a boundary point of \(\Omega_\alpha\) (see [5]), \(N_1\) and \(N_2\) are the size of two partitions in the training set, and the test statistic \(D_t\) is the difference between the total edge lengths of the new point \(X_t\) and \(X_K\), the \(K\)th point in \(X_{K}^N\), which has the largest total edge length in \(X_{K}^N\).

**Proof:** The original decision rule of BP-GEM (given as choose \(H_0\) if \(X_t \in X_{K}^N+1\) and choose \(H_1\) otherwise) can be restated in terms of a test statistic. To see this, note that \(X_t \in X_{K}^N+1\) when the new point replaces the \(K\)th point in the previous best set of vertices \(X_{K}^N\), i.e., its total edge length is smaller than that of the \(K\)th point, as shown in [9]. The asymptotic properties in [11] and [12] follow from the asymptotic optimality of BP-GEM. It is known [3] that the \(H_0\)-region of [9] converges to the minimum volume set \(\Omega_\alpha\), whose decision rule can be stated as \(H_0\) if \(f_0(X_t) \geq f_0(x_\alpha)\), i.e., \(\log \frac{f_0(x_\alpha)}{f_0(X_t)} \leq 0\), and \(H_1\) otherwise, hence the sign property in [12]. Assume that, as \(N_2 \to \infty\), also \(k \to \infty\) such that the total edge length \(L_k(X_t)\) of a point \(X_t\) remains a constant. In that case, \(L_k(X_{t_1}) < L_k(X_{t_2})\) for all \(X_{t_1}\) and \(X_{t_2}\) such that \(f_0(X_{t_1}) > f_0(X_{t_2})\). Since \(D_t = L_k(X_t) - L_k(X_K)\), we have the monotonicity property stated in [11]. Note also that \(X_K \to x_\alpha\) as \(N_1, K \to \infty\) such that \(K/N_1 = 1 - \alpha\).

Proposition 1 shows the structural resemblance of \(D_t\) to the log-likelihood ratio between the boundary point \(x_\alpha\) and \(X_t\). To see the physical relationship consider the case where \(f_0\) is from the exponential family, i.e., \(f_0 = e^{-\delta(X, \theta)}\) where \(\theta\) is the parameter vector and \(\delta(X, \theta)\) is a distance term causing the exponential decay in the probability density function. In this case, \(\log \frac{f_0(x_\alpha)}{f_0(X_t)} = \delta(X_t, \theta) - \delta(x_\alpha, \theta)\) is a distance metric that is similar to \(D_t\) as shown by (13). They also asymptotically share a very similar structure (see Proposition 1).

This theoretical similarity between \(D_t\) and the log-likelihood ratio \(\log \frac{f_0(x_\alpha)}{f_0(x_X)}\) motivates us to use the nonparametric BP-GEM approach in online anomaly detection, in a similar fashion the parametric CUSUM algorithm uses log-likelihood ratio (see [4]).

**B. Online Nonparametric Anomaly Detection**

Instead of classifying each point by hard-thresholding \(D_t\) as in (9) and treating each outlier as an anomaly, we model an anomaly as persistent outliers in the observations and treat \(D_t\) as a positive/negative evidence for anomaly. For timely and accurate detection, we propose to accumulate such anomaly evidence in time, i.e., use also the history, as opposed to

**A. Analysis of BP-GEM for Outlier Detection**

The proposed detector is motivated by the theoretical foundations explained below.

**Proposition 1** (Test statistic of BP-GEM). The BP-GEM algorithm proposed in [3] actually treats the hypothesis testing problem

\[
H_0 : X_t \in \Omega_\alpha, \quad H_1 : X_t \not\in \Omega_\alpha, \quad (8)
\]
sequence/vector.

where our function measures the maximum increase in a norm in \( [10] \). The function in \( [10] \) measures the maximum discrepancy theory \([5]\), and defined similarly to the discrepancy norm in \((10)\). The function in \((10)\) measures the maximum bidirectional change (i.e., increase or decrease) in a vector, whereas our function measures the maximum increase in a sequence/vector.

Before presenting the proposed ODIT algorithm, we recall the CUSUM formulation that recursively updates its test statistic \([4]\).

\[
\bar{\Delta} = \max \{ \Delta_t \geq h_c \}, \quad \bar{S}_t = \max \left\{ \bar{S}_{t-1} + \log \frac{f_1(X_t)}{f_0(X_t)}, 0 \right\}, \quad S_0 = 0.
\]

In computing the discrepancy function, since we are only interested in the increase from the minimum value in the past, as shown in Fig. 1 in the proposed algorithm we clip the minimum value always at zero as in \((15)\), which yields the following stopping rule and the easy-to-compute recursive update rule for the ODIT statistic

\[
T_c = \min \{ t : \Delta_t \geq h_c \}, \quad \bar{S}_t = \max \left\{ \bar{S}_{t-1} + \log \frac{f_1(X_t)}{f_0(X_t)}, 0 \right\}, \quad S_0 = 0.
\]

The similarity between \( \Delta_t \) and \( \Delta_t \) is not necessarily equal to the current time \( t \). Actually, \( \max_{1 \leq j \leq t} S_j^t \) is the accumulated log-likelihood ratio since the time the minimum value of \( S_t \) took place until time \( t \), i.e., \( S_t - S_{t_{\min}} \). However, note that the original procedure, given by \([4]\), stops the first time \( \max_{1 \leq j \leq t} S_j^t \) exceeds a threshold \( h_c \), which always occurs at the maximum point after the minimum occurs. That is, at the stopping time \( \max_{1 \leq j \leq t} S_j^t = g(\ell_i) \), i.e., the stopping times obtained through \([4]\) and \((14)\) are identical.

The utility of Theorem 1 is that it expresses the CUSUM procedure in terms of a general discrepancy metric which is applicable to any number sequence. Specifically, according to Theorem 1 CUSUM stops when the discrepancy, measured by \( g(\ell_i) \), of the observations with respect to \( f_0 \) is large enough. On the other hand, the original CUSUM procedure is completely based on the log-likelihood ratio, and thus not readily applicable to the ODIT statistic \( \Delta_t \), which is similar, but not the same as the running log-likelihood ratio \( S_t \).

Our discrepancy function \( g(\cdot) \) is motivated by the discrepancy theory \([5]\) and defined similarly to the discrepancy norm in \([10]\). The function in \([10]\) measures the maximum bidirectional change (i.e., increase or decrease) in a vector, whereas our function measures the maximum increase in a sequence/vector.

In this section, we provide numerical results to compare the proposed nonparametric ODIT detector with the parametric clairvoyant CUSUM detector and the generalized CUSUM detector, which estimates the model parameters. Following the simulations in \([2]\) and \([8]\), we first test the scenario in which the nominal distribution \( f_0 \) is a 2-dimensional Gaussian with zero mean and diagonal covariance with standard deviation \( \sigma = \ldots \)
the parameters from some training data using the maximum probability, and compared it to that of G-CUSUM which exactly knows $f_0$ and $f_1$ exactly, and the generalized CUSUM detector, which exactly knows $f_0$, but estimates the uniform distribution upper bound as 0.9 instead of the true value 1 in $f_1$.

0.1; and the anomalous distribution $f_1$ is a mixture of $f_0$ and 2-dimensional uniform distribution on $[0, 1]^2$: $f = 0.8 f_0 + 0.2 U$. The training set consists of $N = 10000$ points with partitions $N_1 = 1000$ and $N_2 = 9000$. The test set contains 500 points with anomaly time at 100. Thus, after $t = 100$, we see an anomalous point on average every five time instance. We set $\alpha = 0.05$, $K = \alpha N_1$, $k = 1$, and $s = 1$. As seen in Fig. 2 the proposed nonparametric detector well approximates the parametric optimum CUSUM detector, which knows $f_0$ and $f_1$ exactly. It achieves near-optimum performance while being computationally simple and free of assumptions on the nominal and anomalous probability distributions. Furthermore, the proposed detector significantly outperforms the generalized CUSUM (G-CUSUM), which is used in practice as it estimates the unknown parameters using the maximum likelihood approach. Note the significant performance gap even with a small error in estimating the anomalous distribution.

We also test our proposed algorithm on the “Heterogeneity Human Activity Recognition Dataset” [11] obtained from the UCI Machine Learning Repository [12]. This dataset contains accelerometer and gyroscope data from smartphones and smartwatches about 6 activities: biking, sitting, standing, walking, stair up, and stair down. We used the smartwatch accelerometer data, which contains around 3.5 million data points with 5 numeric features. Focusing on the activity transitions we tested the online detection performance of our algorithm in terms of average detection delay vs. false alarm probability, and compared it to that of G-CUSUM which fits multivariate Gaussian models to $f_0$ and $f_1$ by estimating the parameters from some training data using the maximum likelihood approach.

V. Conclusion

We developed a computationally efficient nonparametric method for timely detection of abrupt and persistent anomalies. The proposed algorithm combines the GEM outlier detection approach with the CUSUM sequential change method. New insights into GEM and CUSUM were provided to effectively combine them. The introduced relation between CUSUM and the discrepancy theory enables extensions of the CUSUM procedure to new applications, as we showed for nonparametric anomaly detection. Similarly, the relation of GEM to parametric outlier detection enabled us to use the simple nonparametric outlier detection approach for timely detection of statistical changes. Numerical results on both simulated and real datasets justified the effectiveness of the proposed algorithm.

REFERENCES

[1] V. Chandola, A. Banerjee, and V. Kumar, “Anomaly detection : A survey”, ACM Comput. Surv., vol. 41, no. September, pp. 1–58, 2009.
[2] A. O. Hero III, “Geometric entropy minimization (GEM) for anomaly detection and localization”, In Proc. Advances in Neural Information Processing Systems (NIPS), pp. 585–592, 2006.
[3] K. Sriphanich and A. O. Hero III, “Efficient anomaly detection using bipartite k-NN graphs”, In Proc. Advances in Neural Information Processing Systems (NIPS), pp. 478–486, 2011.
[4] M. Basseville and I. V. Nikiforov, Detection of abrupt changes : theory and application, Englewood Cliffs, NJ: Prentice Hall, 1993.
[5] H. Weyl, “¨Uber die Gleichverteilung von Zahlen mod. Eins”, Math. Ann., vol. 77, pp. 313–352, 1916.
[6] H. V. Poor, An Introduction to Signal Detection and Estimation, New York, NY: Springer New York, 1994.
[7] G. Lorden, “Procedures for reacting to a change in distribution”, Ann. Math. Statist., vol. 42, no. 6, pp. 1897–1908, 1971.
[8] E.S. Page, “Continuous inspection schemes”, Biometrika, vol.41, pp.100–115, 1954.
[9] C. Scott and R. Nowak, “Learning minimum volume sets”, Journal of Machine Learning Research, vol. 7, pp. 665–704, April 2006.
[10] B. A. Moser and T. Natschlager, “On stability of distance measures for event sequences induced by level-crossing sampling”, IEEE Trans. Signal Process., vol. 62, no. 8, pp. 1987–1999, 2014.
[11] A. Stisen, et al., “Smart Devices are Different: Assessing and Mitigating Heterogeneity for Activity Recognition”, In Proc. 13th ACM Conference on Embedded Networked Sensor Systems (SenSys 2015), Seoul, Korea, 2015.
[12] M. Lichman, UCI Machine Learning Repository, http://archive.ics.uci.edu/ml Irvine, CA: University of California, School of Information and Computer Science, 2013.