Pattern-Based Contextual Anomaly Detection in HVAC Systems

Mohsin Munir*†, Steffen Erkel‡, Andreas Dengel*†, Sheraz Ahmed†
*Technische Universität Kaiserslautern, Germany
†German Research Center for Artificial Intelligence (DFKI) GmbH, Kaiserslautern, Germany
{firstname.lastname}@dfki.de
‡Bosch Thermotechnik GmbH, Germany
{firstname.lastname}@de.bosch.com

***Abstract*—This paper presents detailed anomaly detection evaluation on operational time-series data of Internet of Things (IoT) based household devices in general and Heating, Ventilation and Air Conditioning (HVAC) systems in specific. Due to the number of issues observed during evaluation of widely used distance-based, statistical-based, and cluster-based anomaly detection techniques, we also present a pattern-based approach for anomaly detection in HVAC time-series data. The usage and number of IoT based HVAC systems are enormously increasing and will have a major share in IoT based household devices in the near future. The operational and usage log of these devices contains different sensor values logged with time, containing normal data points, and long-term anomalies. The state-of-the-art methods for anomaly detection are unable to detect these long-term anomalies, which reflect the deteriorating effect of a sensor. The presented approach overcomes this problem by building a knowledge base of long/short -term patterns based on normal data points which keep growing over the time. In addition to the detected anomalies and in contrast to the existing methods, the presented method gives meaningful anomaly score for a number of HVAC systems. We evaluate the presented approach on real operational data, collected over the period of 2.5 years. Evaluation results show that the presented approach outperforms the state-of-the-art methods for anomaly detection with the area under the curve (AUC) value of 99.4%. Discord detection results of the proposed technique on another dataset from a different domain show the generic and adaptive nature of our technique.**

I. Introduction

Anomalies have always been of great interest to humans. In everyday life, we observe that abnormal things and happenings attract our attention. In computer science, anomaly detection refers to the technique of finding specific data points, which do not conform to the majority of the items in the dataset. Many definitions exist for anomaly detection in different contexts. Grubbs [1] defined it as: “An outlying observation is one that appears to deviate markedly from other members of the sample in which it occurs”. In some cases, anomaly detection is also used as leading indicator of unwanted events – also known as the early warning.

Each year, the usage trend, as well as the number of Internet of Things (IoT) based household devices (e.g., smart heaters, smart air conditioners, smart fire alarms, smart TVs, and smart security cameras), are enormously increasing. According to a study on green Information and Communications Technology (ICT) [2], it is predicted that by the year 2020, there will be 16 billion IoT based devices in the world with an average of more than 6 devices per person. Figure 1 shows the current and expected share of internet-connectable household devices.

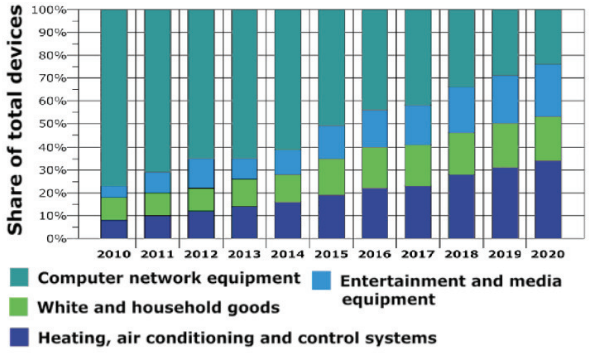


Fig. 1. Share of Internet-connectable consumer household devices [2].

It is estimated that the dominating segment in IoT based household devices by the year 2020 will be HVAC systems with a total share of almost 35% [2]. HVAC manufacturers have already enabled their devices with internet connectivity; so that they can store and analyze the operational and usage log of their devices. This connectivity with the surrounding environment and the back-end servers has opened new horizons for HVAC manufacturers and their users. The analysis of the collected data enables HVAC manufacturers and third-party companies to offer new services specifically tailored according to the customers’ need.

An important use case of analyzing the internal state of HVAC systems is to find the anomalies in the running system and finally reach the root cause of the problem. The early correct detection of such anomalies is the basis of predictive maintenance.

In the context of HVAC systems, the long-term anomalies are of great interest, where the sensor value is deviating from its normal behavior continuously and slowly. Such anomalies can be detected if data are analyzed in the context of the previous data. At this point, most of the existing distance-based and statistical-based anomaly detection techniques fail. The main reason of this failure is the co-existence of a lot of data points which have deviated from the rest of the normal data points.

There are many methods available for anomaly detection [3]–[10]. However, in most of the cases, the performance is unsatisfactory when real HVAC time-series data are presented to them. Most of the methods are unable to detect important anomalies, and/or if detected, have meaningless or misleading anomaly score. Sometimes, significant anomalies have low anomaly score as compared to less significant anomalies. Furthermore, as the existing methods generate anomaly score based on the distribution of each HVAC system, the range of anomaly score also varies from device to device. Due to this factor, it becomes difficult to select a suitable threshold which can efficiently detect anomalies in a number of same HVAC systems. The evaluation of different distance-based, statistical-based, and time-series anomaly detection algorithms shows that they are less precise to detect long-term anomalies in HVAC data set. To address these issues, we have presented an unsupervised pattern-based contextual anomaly detection technique in addition to the evaluation of existing techniques on real HVAC dataset. The presented method uses a knowledge base of only normal data points extracted from the given data. The knowledge base is updated with the passage of time where new normal points are included automatically in it and all of the anomalous points are neglected. The presented method is capable of detecting long/short -term contextual anomalies as well as point anomalies in time-series data.

Main contributions of this paper are:

- A pattern-based contextual anomaly detection approach for IoT based HVAC systems.
- Generation of meaningful anomaly score (an added advantage of the proposed method) for detected anomalies, i.e. high score to most significant anomalies and less score to less significant anomalies.
- A detailed evaluation of existing and proposed approaches on real HVAC operational dataset.

II. State-of-the-art in Anomaly detection

This section provides an overview of the commonly used methods for anomaly detection.

k**-NN** (k-nearest-neighbors) based anomaly detection is the simplest and most widely used unsupervised global anomaly detection method. This distance-based algorithm aims to find anomalous data points based on the k-nearest-neighbors distance [3]. Normally, the average distance to all the k nearest neighbors is considered as the anomaly score of a particular data point [11]. This technique is computationally expensive, highly dependent on the value of k, and may fail if normal data points do not have enough neighbors.

Breuning et al. [12] presented an unsupervised method for local density-based anomaly detection algorithm known as Local Outlier Factor (LOF). In LOF, k-nearest-neighbors set is determined for each instance by computing the distances to all the other instances. Then local density is calculated for each instance. Finally, an outlier factor of each instance is estimated by comparing its density to the estimated density of its neighbors. The basic assumption of this algorithm is that the neighbors of the data instances are distributed in a spherical manner. In some application scenarios where normal data are distributed in a linearly connected way, the spherical estimation of density becomes inappropriate [13]. Other variants of this technique are Connectivity based Outlier Factor (COF) [14] and Influenced Outlierness (INFLO) [15].

Clustering based algorithms are also used for unsupervised anomaly detection. Cluster-Based Local Outlier Factor **(CBLOF)** [4] is an anomaly detection algorithm, in which data are clustered using k-means (or any other) clustering algorithm. The anomaly score of an instance is the distance to the next large cluster. As this approach is based on the clustering algorithm, so the problem of choosing a number of clusters arises and the reproduction of same anomaly score is also an issue due to non-deterministic nature of clustering algorithms.

In addition to the clustering-based anomaly detection algorithms, Histogram-Based Outlier Score **(HBOS)** is a statistical unsupervised anomaly detection algorithm. This algorithm is computationally far less expensive as compared to the nearest-neighbor and clustering-based anomaly detection methods. HBOS works on arbitrary data by offering a standard fixed bin width histogram as well as dynamic bin width (fixed amount of items in each bin) [5].

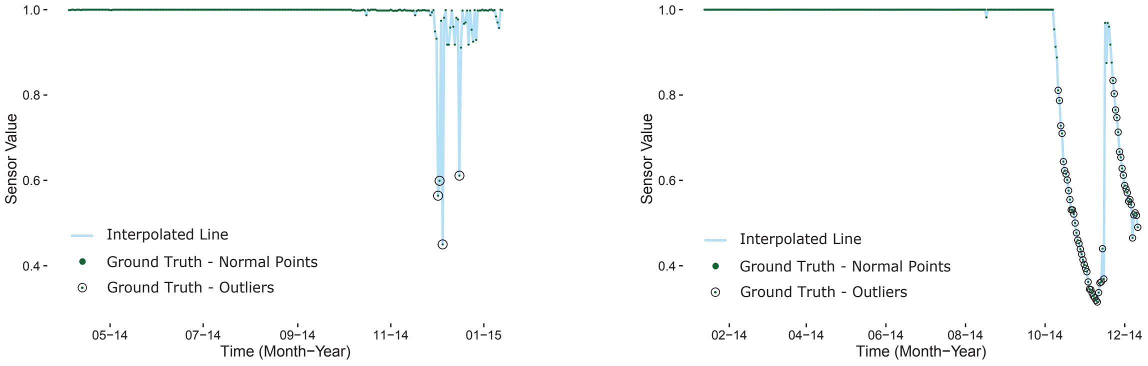
Both semi-supervised and unsupervised variants of anomaly detection algorithms exist which used one-class Support Vector Machine (SVM). An unsupervised variant of **one-class SVM** was introduced by Amer et al. [6], in which no prior training data are required. One-class SVM attempts to learn a decision boundary that achieves the maximum separation between the points and the origin. One-class SVM is sensitive to the outliers when there are no labels.

Shyu et al. [16] proposed an approach for anomaly detection based on Principle Component Analysis **(PCA)**, where a predictive model is constructed from the major and minor principle components of the normal instances. Kwitt and Hofmann [8] proposed another version of this technique.

Malhotra et al. proposed a supervised approach for anomaly detection [17] using a stacked LSTM architecture in time series data. The network is trained on non-anomalous data and used that model as a predictor. The prediction errors are used to fit a multivariate Gaussian distribution and a probability is assigned to each observation. Based on the probability and the threshold, anomalous patterns are detected. However, in deep learning based approaches, a large number of labeled training data are required, therefore, it's use is limited in real scenarios.

Twitter Inc. open sourced its anomaly detection package (**Twitter Anomaly Detection**), which is based on Seasonal Hybrid ESH (S-H-ESD) algorithm [7]. This technique is based on Generalized Extreme Studentized Deviate (ESD) test [18] to handle more than one outliers and STL (Seasonal and Trend decomposition using Loess) [19] to deal with the decomposition of time series data and seasonality trends.

Leng et al. [9] proposed a method to detect anomalous patterns in time series data. Their method has two stages. First, a time series is segmented into different patterns and then anomalous patterns are detected. Each pattern is compared to



(a)

(b)

Fig. 2. Examples of HVAC time series dataset. This dataset contains both point and contextual anomalies.

other patterns of different sizes using Dynamic Time Warping (DTW) to calculate the anomaly factor. This method highly depends on the correct time series segmentation and other thresholds. It is also not clear if this method is capable of detecting point anomalies or not. A fast variant of this anomaly detection method is proposed by Vy and Anh [20]. Another DTW based anomaly detection method for ECG data is proposed by Boulnemour et al. [10]. In this improved version of DTW, called I-DTW, time series of different lengths and periods are aligned better to each other as compared to standard DTW. For anomaly detection, a normal ECG segment and an ECG segment with the abnormality (a query segment) are given to the system. The I-DTW reconstructs normal segment onto query segment and the morphological difference between two segments less than a threshold shows anomaly. They only reported results on ECG data, so the performance of this method on HVAC dataset is not known. Also, only visual anomaly detection is shown with no information about anomaly score.

III. Datasets

A. Real HVAC dataset

This dataset contains real operational data gathered from 77 HVAC systems, which are operational at different locations in Germany. Data[1] collected from the household boilers over the span of 2.5 years are used in this paper to find operational anomalies. To analyze and control the behavior of a HVAC system, each system is equipped with a lot of sensors. These 'smart', IoT based HVAC systems transmit all of the operational data to a back-end database. A specific sensor used in boilers is analyzed in the scope of this paper, which generates a univariate time series. Each time series generated by a HVAC system is treated separately. But, the goal is to have same anomaly detection setting/threshold along all the systems. Figure 2 shows examples of normalized time series of observed sensor from two HVAC systems. In the start, when a boiler is installed, it is observed that the sensor works perfectly fine (1.0 is the normal/expected running value of the observed sensor). But, with the passage of time, the observed sensor value starts deteriorating or shows anomalous behavior on some days. These abnormal behaviors indicate that the sensor needs to be repaired before it completely breaks. Normal data points are shown in green color, whereas black circles having a dot inside them show anomalous data points. These data points serve as anomaly ground truth, which is marked by a domain expert. This dataset contains both point and long-term anomalies.

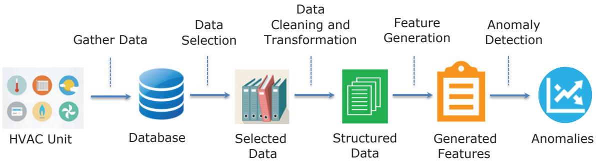


Fig. 3. A standard workflow of applying anomaly detection approach on HVAC data.

B. Real ECG dataset

An ECG time series from MIT-BIH ECG Database, available at PhysioNet [21] is also used in this paper. According to PhysioNet, the ECG readings were digitized at 360 samples per second per channel with 11-bit resolution over a 10 mV range. We used record numbered 108 from this database[2]. It is a univariate time series data with 2160 data points. This time series contains different ventricular abnormalities, which are represented as contextual anomalies in the scope of this paper.

IV. Presented Approach for Anomaly detection

This section provides a detailed insight of the presented method for anomaly detection. Figure 3 shows the complete workflow of the presented anomaly detection method which is divided into three main steps i.e. pre-processing, feature generation, and anomaly detection.

A. Pre-processing

Following are the challenges, which are faced in the pre-processing step to shape the data into a standard format. This step is further divided into two steps:

[1] Due to the data protection and privacy policy of the affiliated company, we are unable to open-source the data or provide details of the observed sensor.

[2] Direct link to database:
https://www.physionet.org/physiobank/database/mitdb/

- **Data Selection:** The purpose of this step is to select relevant data that are suitable for anomaly detection. It is possible that not all of the HVAC systems are ready for analysis, as there might be some systems, which have logged very small amount of data since their installation due to various reasons. So, we shortlisted those HVAC systems for further analysis which stayed online at least for one year. The selected systems transmit huge amount of the data, consisting hundreds of internal and external sensors. In practice, not all of the sensor data can be passed to anomaly detection algorithms. Therefore, the sensors, which are most critical to be observed are selected, based on the expert's knowledge. But, in the scope of this paper, our focus is only on one sensor.
- **Data Cleaning and Transformation:** The purpose of this step is to solve the problem of missing data. From a HVAC system to the database, the data passes through different connectivity bridges. Connectivity loss in any one of these bridges causes missing data in the database. A missing gap can be of few minute, an entire day, or even months. The systems which have missing gaps of more than consecutive 3 days are called unhealthy systems. 24 healthy systems are shortlisted for further analysis which have the maximum number of consecutive data and minimum number of missing gaps.
  The sensor data are stored in the back-end database in an unstructured format and could not be used directly for analysis purposes. So, it is transformed into a structured CSV format in which the sensor values are extracted against a unique time stamp for each HVAC system.

B. Feature Generation

The operational data logged in the database are unevenly spaced time-series data. In an unevenly spaced or an irregular time series data, observation time is not constant. In other words, the observations for each system are logged at the different time. It is, therefore, hard to find a common pattern among different systems as there is no common logging timeline among different systems. To compare different systems and the behavior of the recorded system, it is important to define the data on a common timeline. In feature generation step, we get rid of excessive and unevenly spaced time-series data and generate meaningful features out of the raw data.

In general, a feature is defined as an aggregated quantity (of some parameter) per unit. In our case, a feature is *the mean value of a selected parameter per day*. So, we take mean of the observed sensor at a particular time of day which shows the actual behavior of that sensor. This results into the data which is evenly spaced time-series and can be compared to all other systems for further analysis.

C. Anomaly detection

The presented pattern-based anomaly detection is an unsupervised technique, which builds a knowledge base of long-term patterns. The knowledge base, which is based on normal data points, keeps growing over the time. The presented approach works based on the assumption that first n instances of an operational log are normal data points and does not contain contaminated/anomalous data. This assumption is generally true, because a new device when installed for the first time, is expected to have a normal behavior. These initial n points (in our case, $n = 50$) serves as a basis of knowledge base containing normal patterns.

The size of the knowledge base (K_B) keeps growing over the time. A new point is added to the knowledge base if it is detected as a normal data point. To mark a data point as normal or anomaly, current data point (x_t), as well as its contextual information, is used. The sequence containing the context and the original value can defined as follows:

$$S_t = \bigcup_{i=t-N}^{t} x_i \tag{1}$$

Using equation 1, knowledge base (K_B) in initial state can be defined as follows:

$$K_B = \{S_N\} \tag{2}$$

This sequence/window S_t is compared with the knowledge base (K_B) by using an overlapping window of the same size as input sequence. This means that knowledge base is divided into overlapping windows of size $N + 1$. All of these overlapping windows from knowledge base are compared with the window/sequence of the current data point. The comparison between the context of the current window and all the windows in the knowledge base is done using Dynamic Time Warping (DTW). The minimum DTW distance (equation 3) finds the most similar sequence from the knowledge base.

$$dist(S_t) = min(DTW(S_t, K_i))\forall K_i \in K_B \tag{3}$$

A threshold τ is used to distinguish between a normal data point and an anomalous data point. Equation 4 defines the use of equation 3 for detection of anomalous point as well as accumulation of knowledge base.

$$f(dist(S_i)) = \begin{cases} Anomaly, & \text{if } dist(S_i) > \tau \\ S_i \bigcup K_B, & \text{otherwise} \end{cases} \tag{4}$$

Where $dist$ is a DTW distance between the current window and all the sequences in the knowledge base. All of the remaining data points which do not satisfy the first criteria in Equation 4 are marked as normal and added to the knowledge base. In this way, the knowledge base keeps on growing with time (t) incorporating long-term patterns.

V. Evaluation

This section provides a detailed analysis of the following existing state-of-the-art unsupervised anomaly detection techniques and the presented anomaly detection technique on real HVAC and ECG datasets: k-NN Global Anomaly Score [3], Clusters Based Local Outlier Factor (CBLOF) [4], Histogram Based Anomaly Score (HBOS) [5], One-class SVM [6], Twitter Anomaly Detection [7], Robust PCA (rPCA) [8].

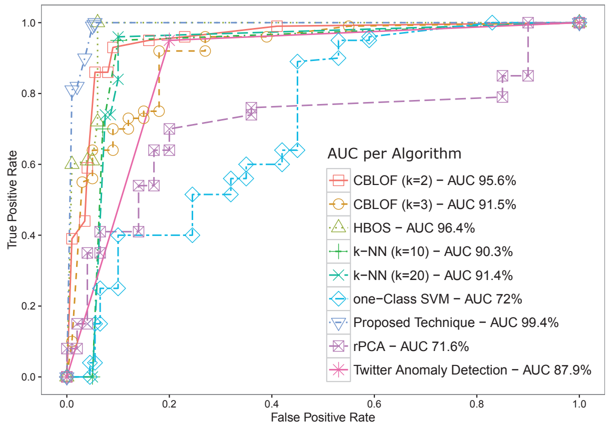


Fig. 4. ROC plot comparing the proposed technique with the state-of-the-art anomaly detection techniques.

For comparison with the existing techniques, we used the standard implementation of the above-mentioned algorithms in RapidMiner. These anomaly detection algorithms are freely available as an extension to RapidMiner. This Anomaly Detection Extension[3] contains many other unsupervised anomaly detection algorithms too. There are different hyper-parameters of the selected algorithms which need to be tuned. We did a number of experiments on different hyper-parameters and used the best parameters for the final evaluation. Same parameters are used for both datasets. A detailed evaluation is only provided for HVAC dataset (on the basis of ROC, AUC, Precision, and Recall). Whereas, for ECG dataset, visual evaluation is provided to show the issues in other anomaly detection methods and to highlight the generic nature of the proposed method.

Performance measure like receiver operating characteristic (ROC) or ROC curves and area under the curve (AUC) are used in this analysis section to show an overall performance of different algorithms and the proposed method. In ROC plot, a perfect classification would yield a point in the upper left corner, with maximum AUC. The performance of the technique is considered best whose curve is most close to the upper left corner.

Generally, all anomaly detection algorithms provide an anomaly score for a given data point. To classify a data point as normal or anomalous, a suitable threshold is required, which varies for each algorithm. It is important to find a threshold which is a true representative of the classifier.

To gain insights of each algorithm, precision and recall are also calculated. The best threshold is selected on the basis of Equal Error Rate (EER). False positive rate (FPR) and true positive rate (TPR) from ROC curve are used to find EER. Usually, the threshold where both FPR and TPR become equal is considered as the best threshold. But, in some cases, it is also possible that these two measures do not match exactly. For such cases, we calculate the difference between these two measures. The best threshold is considered at the point where the difference is minimum. Table I shows the precision and recall of all the evaluated methods on HVAC dataset. The overall precision and recall of the proposed technique are better than other anomaly detection methods. The recall of HBOS and Twitter Anomaly Detection is better than proposed technique, but a balance between precision and recall is more important than only relying on precision or recall.

TABLE I
PRECISION AND RECALL OF PROPOSED AND OTHER OBSERVED TECHNIQUES ON HVAC DATASET. THE BEST THRESHOLD IS SELECTED BY EQUAL ERROR RATE.

Algorithm	Precision	Recall
Proposed Technique	**0.91**	**0.80**
HBOS	0.47	0.90
CBLOF (K=2)	0.50	0.74
CBLOF (K=3)	0.56	0.54
k-NN (k=10)	0.34	0.54
k-NN (k=20)	0.36	0.61
one-Class SVM	0.19	0.38
rPCA	0.42	0
Twitter Anomaly Detection	0.23	0.95

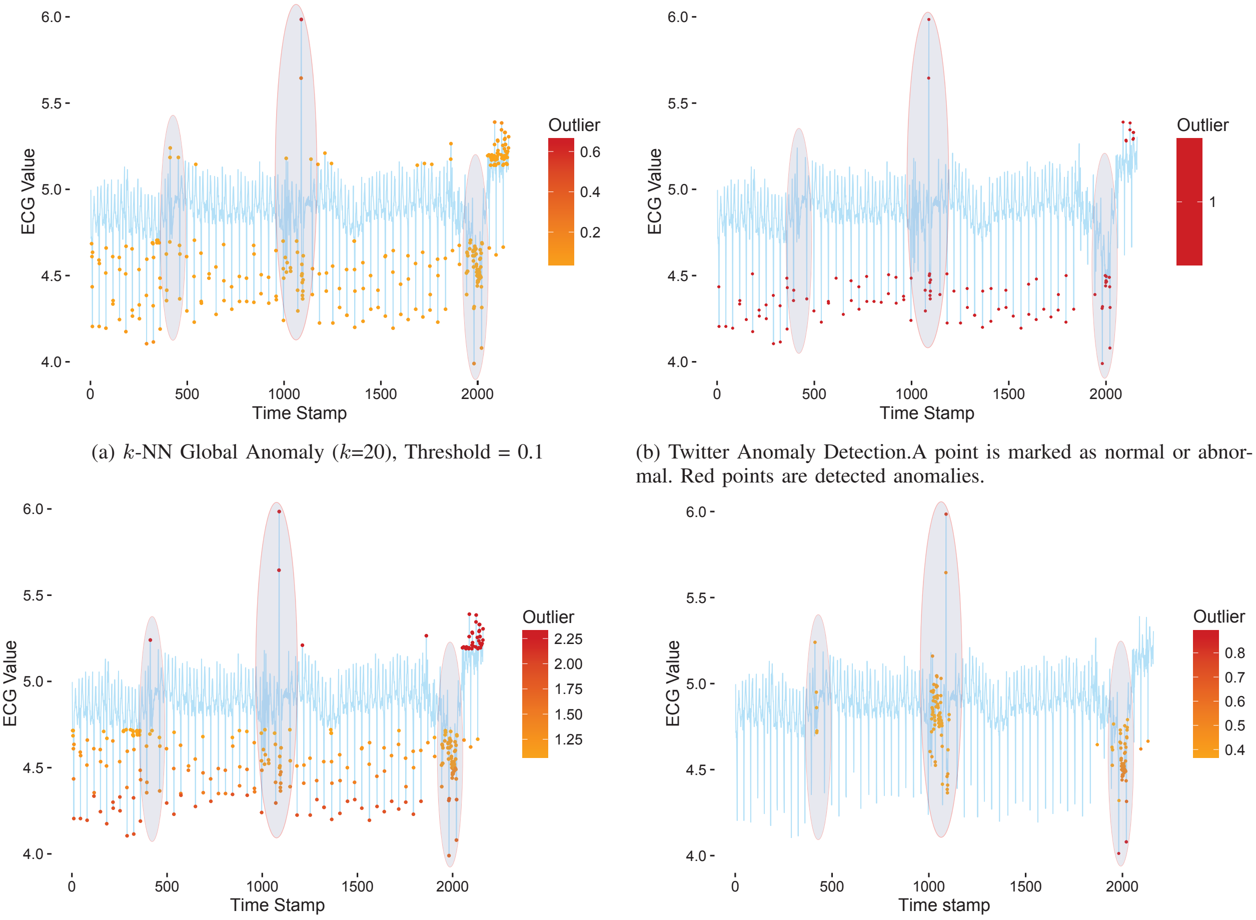
A. *Analysis*

Here, it is important to note the notation we have used for normal and anomalous data points – anomalies are considered as positive records, whereas normal data points are considered as negative records. As we are interested in anomalous data points, so they are marked as a positive hit.

The analysis of results shows that HBOS performs best on HVAC dataset from the group of the state-of-the-art techniques with maximum AUC of 96.4%. HBOS clearly performs better than distance and clustering based techniques. This is mainly of two reasons: i) The data points are spread in a way that very less number of data points lie in the same bin, which generates high anomaly score, ii) The height of histograms is normalized, so the anomaly score of different systems is more representative. It can be observed in Table I that the overall precision and recall of HBOS is better than other state-of-the-art techniques. On ECG dataset, HBOS is only able to detect point anomalies. The highlighted areas are where real anomalous points exist. Data points, which lie far from the most of the data points, are incorrectly marked as anomalies as shown in Figure 5c.

After HBOS, clustering-based technique CBLOF performs well on HVAC dataset, with AUC of 95.6%. We did experiments with the different number of clusters and report the results on two different clusters size, i.e. 2 and 3 clusters. It is observed in ROC curve, that AUC is decreased with the increase in a number of clusters for the HVAC dataset. Depending on the nature of data and number of clusters, this technique can cause a problem when a number of anomalous data points make a separate cluster because of their high co-existence. This can happen in the case of long-term anomalies, where a sensor starts to continuously give slightly deviated value. In this case, the whole cluster can be wrongly detected as normal data points. But, in the case of two clusters and existence of a relatively low number of anomalous data points,

[3] RapidMiner Anomaly Detection Extension is available at: https://marketplace.rapidminer.com/UpdateServer/faces/product_details.xhtml?productId=rmx_anomalydetection



(a) k-NN Global Anomaly (k=20), Threshold = 0.1

(b) Twitter Anomaly Detection.A point is marked as normal or abnormal. Red points are detected anomalies.

(c) HBOS, Threshold = 1.0

(d) Pattern-based Anomaly, Threshold = 0.3

Fig. 5. Color coded anomalies show anomalies detected by respective methods. Abnormal patterns exit in the highlighted area. Only the proposed method is able to detect these contextual and point anomalies.

as compared to normal data points, 2 clusters give relatively good results as compared to larger cluster number.

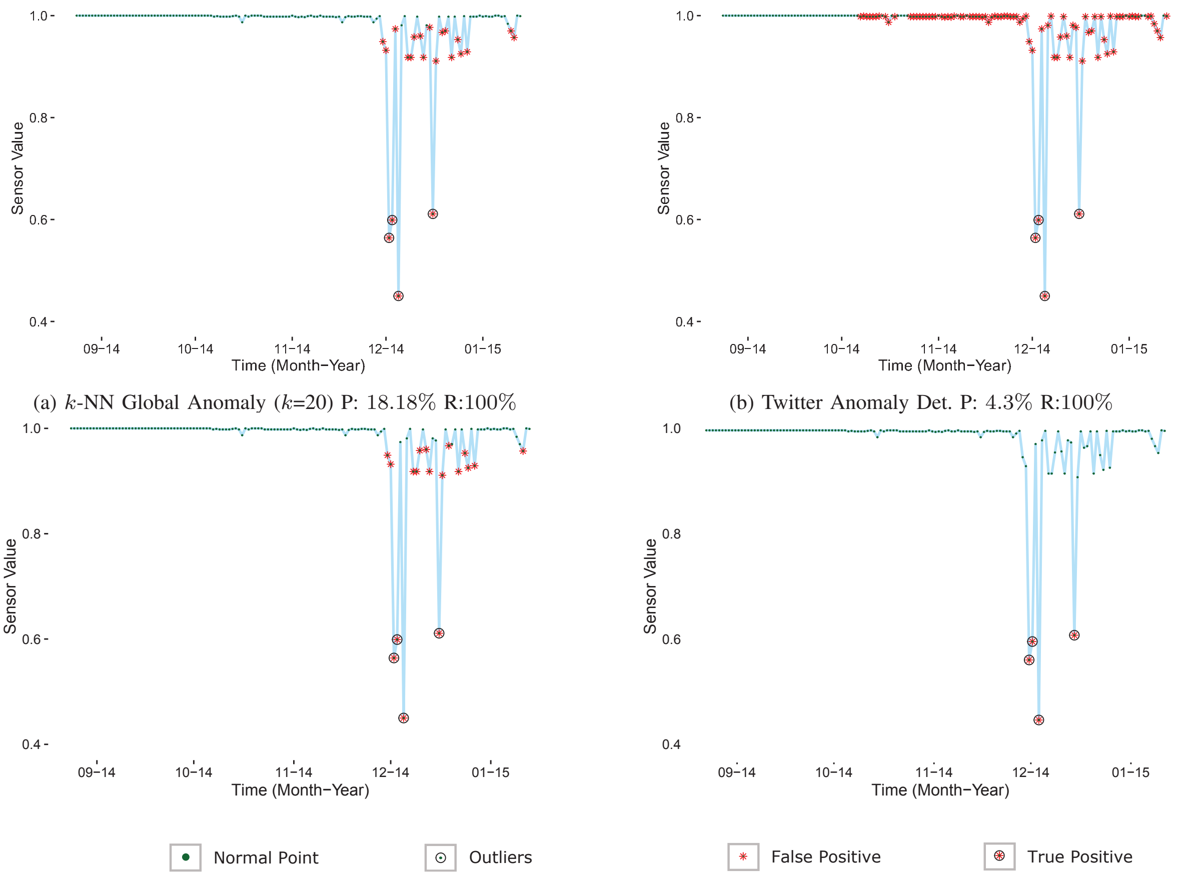
Widely used k-NN based anomaly detection technique shows AUC of 91.4% in HVAC dataset, which is less than CBLOF and HBOS. This technique is also tested on the different set of parameters and best are reported, i.e. K is set to 10 and 20. This distance-based technique is good for the detection of point anomalies as they are easily distinguishable from the rest of the data. But, in the case of time series datasets, where anomalies occur due to the change in pattern or due to the small deviations, this technique is unable to detect important anomalies. Also, in the case of long-term anomalies, this technique is unable to detect anomalies because the distance between normal and abnormal points does not change drastically. Same is observed in ECG dataset, correct anomalies are not detected in this case too (Figure 5a).

Experiments with different parameter combinations are performed using Twitter Anomaly Detection technique. In most of the cases for HVAC dataset, this technique is able to detect all of the anomalies, but with high false alarm rate. But for ECG dataset, this method is unable to detect contextual anomalies and the marked anomalies are incorrect as shown in Figure 5b. Overall results of one-class SVM and rPCA are poor on HVAC dataset.

The above-mentioned issues in different anomaly detection techniques are addressed in the proposed pattern based contextual anomaly detection technique. As shown in ROC plot, the proposed approach performs better than other anomaly detection techniques with AUC of 99.4%.

Figure 6 shows the results of some of the observed anomaly detection techniques on a HVAC system. The anomalous data points (marked with red asterisk $(*)$) detected by the mentioned technique are shown along the ground truth information. When $(*)$ is inside a black circle, it represents true positive; i.e. an anomalous data point is correctly detected by the respective algorithm. Whereas, simple asterisk shows false alarm. The purpose of this figure is to visually understand the issues of current anomaly detection techniques on HVAC dataset. Precision and recall shown in this figure are calculated for this HVAC system based on the commonly selected threshold.

Figure 6a shows the result of k-NN Global Anomaly Score technique, where all of the anomalies are detected correctly with a lot of false alarms, which degraded the overall performance and precision decreases to 18.18%. For k-NN technique, the accuracy is slightly increased with greater k on the whole dataset as shown by AUC in Figure 4. The



(a) k-NN Global Anomaly (k=20) P: 18.18% R:100%

(b) Twitter Anomaly Det. P: 4.3% R:100%

(c) HBOS P: 22.2% R:100%

(d) Pattern-based Anomaly P:100% R:100%

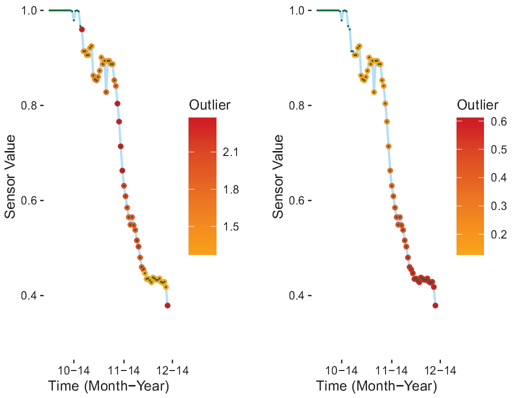
Fig. 6. Comparison of precision (P) and recall (R) of different anomaly detection techniques and the presented technique on a specific HVAC system. Here, the precision and recall are calculated for a single HVAC system. False Positive: Wrong outlier detection, True Positive: Correct outlier detection.

parameters, like the number of clusters and k in nearest neighbor techniques, cannot be generalized as they depend on the distribution of the data. Twitter Anomaly Detection technique performs worst on this system (Figure 6b). Even the small variations in data are falsely marked as anomalies by this technique. Figure 6c shows that HBOS is able to detect all anomalies correctly. However, similar to K-NN based method, it's precision is low. In comparison to existing techniques, Figure 6d shows the result of the proposed pattern-based technique, which outperforms other methods and achieves a precision and recall of 100% on this system. Also on ECG dataset, the proposed technique is clearly able to detect point and contextual anomalies (Figure 5d) which are missed by most of the other mentioned methods.

B. Problem of misleading anomaly score

In addition to the low precision and recall, another problem with most of the existing anomaly detection techniques is the misleading anomaly score, across multiple systems of the same type. Ideally, a high anomalous point should get a high anomaly score, whereas a low anomalous point should get a low score. In HVAC dataset, there are some systems in which small deviations of sensor value are observed. On such small deviations, existing algorithms give high anomaly score because of the inability of incorporating context, which makes the threshold selection process (for all systems of the same type) very difficult.

The second best technique mentioned in ROC plot for this HVAC dataset is HBOS. Figure 7a shows color-coded anomalies marked by HBOS for a HVAC system. This is an example of a long-term anomaly scenario (the normal data points in the beginning of time series are not shown here). All the anomalies are detected by HBOS in this case. But, the anomaly score is not correct when the data of the whole system is observed. The highest anomaly score is given to anomalous data point in the middle of the curve, and then anomaly score decreases towards the end of the curve (which should be increasing). It is due to the fact that when more data points occur in the anomalous region, then anomaly score decreases. In the context of HBOS algorithm, if more data points exist for a defined bin, then their possibility of being anomalous also decreases. Whereas in real data, there are cases when a lot of data points exist in the anomalous region and it is important to detect all of those anomalous points with correct anomaly score. Same anomaly score behavior is also observed in k-NN Global Anomaly Score technique for HVAC dataset. In comparison to the existing methods, the presented technique provides correct anomaly score behavior as shown in Figure 7b, in which, anomaly score increases for more



(a) HBOS

(b) Presented technique.

Fig. 7. Color-coded anomaly scores given by HBOS and the presented pattern-based technique. Anomalies toward the bottom of the curve are given relatively low anomaly score by HBOS. Whereas, correct anomaly score behavior is given by the proposed technique.

distant anomalous data points. The anomaly score generation technique makes the anomaly score true representative of the anomalies. Figure 5d also shows correct anomaly score behavior on ECG dataset. For an anomalous data point, anomaly score is calculated from the normal data points saved in the knowledge base. This gives the correct representation how far an anomalous data point is.

VI. Conclusion

This paper presents a detailed evaluation of anomaly detection techniques on real IoT based HVAC dataset and visual evaluation on ECG dataset. To get rid of different issues which occur in traditional anomaly detection techniques, like incompetency of detecting changes in pattern, and finding the correct number of clusters, a novel approach for detecting anomalies on the basis of context and history is also presented. The operational log of HVAC systems usually contains long-term anomalies. To detect these long-term anomalies, it is very important to take contextual information into account. The state-of-the-art methods for anomaly detection are unable to detect these anomalies due to the lack of contextual knowledge. The presented approach overcomes this problem by building a knowledge base of long/short -term patterns based on normal data points which keep growing over the time. In addition, it does not analyze single data point but also includes its contextual information to decide either the current point is normal or anomalous. The on-time detection of these trending long-term anomalous points can be used to give early warnings about the faulty sensor/component before the whole system breakdown. Evaluation results show that the presented method outperforms the state-of-the-art methods for anomaly detection in terms of all performance indicator measures i.e., AUC 99.4%, precision and recall of 0.91 and 0.80 respectively.

In addition to the detection of anomalies, the presented method also produces a meaningful anomaly score. This means that it gives high anomaly score to more significant anomalies and low score to less significant anomalies by considering. The anomaly score produced by the existing methods for HVAC dataset is meaningless and does not correlate with the significance of anomalies.

References

[1] F. E. Grubbs, "Procedures for detecting outlying observations in samples," *Technometrics*, vol. 11, no. 1, pp. 1–21, February 1969.

[2] P. F. Ovidiu Vermesan, *Internet of Things - Global Technological and Societal Trends From Smart Environments and Spaces to Green ICT*, ser. River Publishers Series in Communications. River Publishers, 2011.

[3] S. Ramaswamy, R. Rastogi, and K. Shim, "Efficient algorithms for mining outliers from large data sets," in *ACM SIGMOD Record*, vol. 29, no. 2. ACM, 2000, pp. 427–438.

[4] Z. He, X. Xu, and S. Deng, "Discovering cluster-based local outliers," *Pattern Recognition Letters*, vol. 24, no. 9, pp. 1641–1650, 2003.

[5] M. Goldstein and A. Dengel, "Histogram-based outlier score (hbos): A fast unsupervised anomaly detection algorithm," *KI-2012: Poster and Demo Track*, pp. 59–63, 2012.

[6] M. Amer, M. Goldstein, and S. Abdennadher, "Enhancing one-class support vector machines for unsupervised anomaly detection," in *Proceedings of the ACM SIGKDD Workshop on Outlier Detection and Description*. ACM, 2013, pp. 8–15.

[7] A. Kejariwal, "Introducing practical and robust anomaly detection in a time series," https://blog.twitter.com/2015/introducing-practical-and-robust-anomaly-detection-in-a-time-series, accessed: 2016-01-14.

[8] R. Kwitt and U. Hofmann, "Robust methods for unsupervised pca-based anomaly detection," *Proc. of IEEE/IST WorNshop on Monitoring, AttacN Detection and Mitigation*, pp. 1–3, 2006.

[9] M. Leng, X. Chen, and L. Li, "Variable length methods for detecting anomaly patterns in time series," in *Computational Intelligence and Design, 2008. ISCID'08. International Symposium on*, vol. 2. IEEE, 2008, pp. 52–56.

[10] I. Boulnemour, B. Boucheham, and S. Benloucif, "Improved dynamic time warping for abnormality detection in ecg time series," in *International Conference on Bioinformatics and Biomedical Engineering*. Springer, 2016, pp. 242–253.

[11] F. Angiulli and C. Pizzuti, "Fast outlier detection in high dimensional spaces," in *PKDD*, vol. 2. Springer, 2002, pp. 15–26.

[12] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "Lof: identifying density-based local outliers," in *ACM sigmod record*, vol. 29, no. 2. ACM, 2000, pp. 93–104.

[13] M. Goldstein, "Anomaly detection in large datasets," PhD-Thesis, University of Kaiserslautern, München, Germany, 2 2014. [Online]. Available: http://www.goldiges.de/phd

[14] J. Tang, Z. Chen, A. W.-C. Fu, and D. W. Cheung, "Enhancing effectiveness of outlier detections for low density patterns," in *Advances in Knowledge Discovery and Data Mining*. Springer, 2002, pp. 535–548.

[15] W. Jin, A. K. Tung, J. Han, and W. Wang, "Ranking outliers using symmetric neighborhood relationship," in *Advances in Knowledge Discovery and Data Mining*. Springer, 2006, pp. 577–593.

[16] M.-L. Shyu, S.-C. Chen, K. Sarinnapakorn, and L. Chang, "A novel anomaly detection scheme based on principal component classifier," DTIC Document, Tech. Rep., 2003.

[17] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal, "Long short term memory networks for anomaly detection in time series," in *European Symposium on Artificial Neural Networks*, vol. 23.

[18] B. Rosner, "Percentage Points for a Generalized ESD Many-Outlier Procedure," *Technometrics*, vol. 25, no. 2, pp. 165–172, May 1983.

[19] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning, "Stl: A seasonal-trend decomposition procedure based on loess," *Journal of Official Statistics*, vol. 6, no. 1, pp. 3–73, 1990.

[20] N. D. K. Vy and D. T. Anh, "Detecting variable length anomaly patterns in time series data," in *International Conference on Data Mining and Big Data*. Springer, 2016, pp. 279–287.

[21] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000 (June 13), circulation Electronic Pages: http://circ.ahajournals.org/content/101/23/e215.full PMID:1085218; doi: 10.1161/01.CIR.101.23.e215.