Health monitoring: a machine learning approach for anomaly detection in multi-sensor networks

Hajee, Bram; Wisse, Kees; Mohajerin Esfahani, P.

DOI
10.34641/clima.2022.262

Publication date
2022

Document Version
Final published version

Published in
CLIMA 2022 - 14th REHVA HVAC World Congress

Health monitoring: a machine learning approach for anomaly detection in multi-sensor networks

Bram Hajee [a], *Kees* Wisse [a], *Peyman* Mohajerin Esfahani [b]

[a] DWA, The Netherlands

[b] Delft Center for Systems and Control, Delft University of Technology

Abstract. Multi-sensor networks are becoming more and more popular in order to assess the post-occupancy performance of smart buildings, since they enable continuous monitoring with a high spatial resolution of the occupancy, thermal comfort and indoor air quality. An urgent, but poorly attended topic in this field is the automated detection of sensor anomalies. For example, CO_2-sensors can perform auto-calibration, during which the data is not reliable. Without identifying the poor reliability of this data, any analysis based on it may be misleading. Automated detection and diagnosis of multi-sensor anomalies is a challenging task due to the complex characteristics of each data point, the variety of data points and the sheer number of data points. As a result, rule-based algorithms require an extensive expert-based set of rules, which makes them sensitive to threshold values and case specific exceptions. Machine learning algorithms can overcome these issues, but they require datasets with labelled sensor anomalies to do diagnosis. Acquiring such labelled datasets is labour intensive and therefore expensive. In this paper we show the potential of a transition from an unsupervised to a supervised machine learning approach. The unsupervised algorithm is used to detect anomalies and to identify anomaly classes of interest. This enables for labelling such classes efficiently in order to train classifiers for multiple classes of anomalies. The unsupervised and supervised algorithms are employed in parallel during the transition, allowing for the simultaneous detection of unknown anomaly classes and diagnosis of known anomaly classes. The improved performance of the combined classifier compared to unsupervised detection is shown by the precision-recall curve. Though the presented approach is rather generic, it does have some limitations. Because a window-based approach is used, only time windows can be detected as being anomalous, not the exact time. Also, we focus on the detection of sudden anomalies and the approach does not allow for detecting stationary or trend anomalies.

Keywords. Multi-sensor networks, anomaly detection and diagnosis, machine learning, HVAC
DOI: https://doi.org/10.34641/clima.2022.262

1. Introduction

Health monitoring in buildings is, apart from Covid-19, an important topic as people spend a lot of hours in the indoor environment. Evaluation standards are commonly based on samples that are taken from a selection of rooms in a limited time (see for example [1]). Multi-sensor networks enable the possibility to assess thermal comfort (temperatures), indoor air quality (CO_2, TVOC, relative humidity), light and sound with a high spatial resolution using continuous monitoring. Together with occupancy measurements, it's becoming a standard within so-called smart buildings.

Continuous monitoring with a high spatial resolution also opens new perspectives for continuous commissioning instead of periodic commissioning or initial commissioning after the construction phase [2].

An urgent, but poorly attended topic in this field is the accuracy of the sensors and potential sensor anomalies. Applying sensors at such a large scale will lead to a preference for low-cost solutions with potential poor performance related to accuracy. The large scale applications also challenge the maintenance of the sensors.

Automated fault detection and diagnosis (FDD) for sensor networks is therefore inevitable. This is enhanced by the fact that sensors play also an important role in FDD on the HVAC system level. The consideration of sensor faults together with a component fault can increase the difficulty of FDD exponentially [3]. This can be the case especially for temperature, CO_2 and humidity measurements as they can be involved in control loops of room supply of heating and cooling by variable air volume terminal units, chilled beams, ceiling panels etc. Specific anomaly detection focussed on sensors will strongly improve the FDD on the component and system level.

1.1 Anomalies in multi-sensor networks

Anomalies in sensors occur due to different causes. Examples are inaccurate measurement position of the sensor, connection losses to the data acquisition, blocking of sensors by objects and calibration issues. CO_2 sensor calibrations degenerate over time and some CO_2 sensors perform auto-calibration procedures with a possible erroneous outcome.

To deal with the vast amounts of data, pre-processing is often performed in the local sensor unit, which makes the performance also sensible to software updates. New versions of the pre-processing software may for example not be compatible with the older versions of the hardware.

In order to reduce the data transfer from local to central data processing systems may also use the change-of-value (CoV) principle. The local sensor unit only provides a new sample to the central system when the measured value is changed related to the previous measured value. On the level of the central system, it is difficult to distinguish between the regular missing values due to the CoV principle and temporal connection losses of the multi-sensor unit. Fig. 1 shows connection losses in temperature measurements. As can be seen, the connections losses are hard to identify by looking at the single sensor in isolation. Alternatively, they are easily detected when looking at all sensors combined.

Anomalies in temporal data are often categorized as either point, contextual or collective anomalies [4, 5, 6]. A point anomaly is a single instance and a collective anomaly is a collection of instances that is anomalous as a whole (see Fig. 2). Furthermore, we can distinguish sudden, stationary and trend anomalies. Continuous blocking of a sensor for example will give a stationary error. Slowly degenerating CO_2 sensors will give a trend error, while a CO_2 sensor that performs auto calibration will give a sudden error. In this paper, we focus on sudden errors.

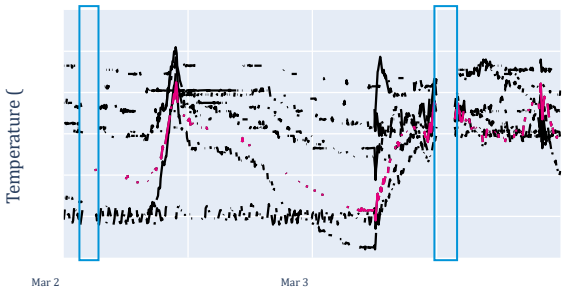

Fig. 1 - Analysis of a single sensor (pink) in relation to its neighbouring sensors (black) for the detection of connection losses (blue).

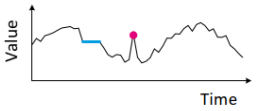

Fig. 2 - Point (pink) and collective anomalies (blue).

1.2 Related work

Fault detection and diagnosis for building systems and HVAC applications show a long history of techniques which already has been applied. Kim and Katipamula provide an overview of the different methods with their applications, from quantitative as well as qualitative model-based diagnostic methods and process history-based methods [7].

Process history-based and qualitative model-based methods are the most popular methods. Quantitative model-based methods desire an explicit mathematical expression of the system as well as accurate input parameters. For multi-sensors, this should include thermal models, air quality models, light models and sound models. From the qualitative model-based models the rule-based algorithms show the most extensive track record. However, they require an extensive expert-based set of rules, which makes them sensitive to threshold values and case-specific exceptions.

An overview of artificial intelligence-based FDD methods is given by Zhao, Li, Zhang et al. [3]. They divide the data-driven detection methods into classification based methods and unsupervised detection methods. From the unsupervised methods, principle component analysis is the most popular method, while from the classification based methods multi-class classification has received the most attention.

Related work from other disciplines than HVAC can be found in the field of wireless sensor networks, see for example Chen, Li and Huang [8].

1.3 State of the art and its limitations

One common problem for the application of artificial intelligence-based FDD is the lack of labelled datasets. Zhao, Li, Zhang et al. [3] indicate it as one of the main shortcomings of data-driven methods. The availability of normal data, separated from faulty data is often a prerequisite for applying the available methods which are described in the review [3]. However, in practical applications manufacturers of multi-sensors don't provide 'normal' datasets.

This problem is not limited to HVAC applications. Also, Chen, Li and Huang mention this problem for wireless sensor network datasets in general and use anomaly insertion methods to test their detection methods [8].

Acquiring labelled datasets is labour intensive and therefore expensive. Efficient labelling is a possible way out. Efficient labelling can be performed by starting with unsupervised learning followed by supervised learning for the most common faults or faults with a high impact.

In this paper, we show the potential of a transition from an unsupervised to a supervised machine learning approach. The main contribution of the paper is the development of a systematic framework for this transition and a proof of concept in the application domain of multi-sensors using a real-world dataset.

The unsupervised algorithm is used to detect anomalies and to identify anomaly classes of interest. This enables for labelling such classes efficiently in order to train classifiers for multiple classes of anomalies. The unsupervised and supervised algorithms are employed in parallel during the transition, allowing for the simultaneous detection of unknown anomaly classes and diagnosis of known anomaly classes.

It will be shown how parallel operation of supervised and unsupervised provides a way of efficient labelling compared to conventional querying and labelling of datasets. Efficient labelling, however, can still lead to extensive datasets which have to be queried and labelled. Therefore we also show how limited querying and labelling performs when unsupervised and supervised learning are used in parallel.

Results are shown for a real-world dataset from a case study: a Dutch office building utilized with multi-sensors.

2. Research methods

2.1 Case study

The case study comprises a Dutch office building in the Netherlands, which is utilized with BRT-35 ceiling multi-sensors [9]. The following parameters are measured: occupancy of the room (passive infrared detection), infrared temperature, CO_2 concentration, relative humidity, sound pressure level and light intensity. The data is processed by the local sensor unit, while the data is stored in the database of the building management system.

The storage uses the change-of-value (CoV) principle, i.e. a value is only stored when it has changed within a certain resolution. The results of this paper are based on the averaged values on a time grid of 3 minutes. In this case study, the results are presented for the infrared temperature sensors. Due to CoV principle, connection loss is an important anomaly that is challenging to detect. This is one of the labelled anomaly classes and it is used in this paper to demonstrate the transition from unsupervised detection to supervised classification.

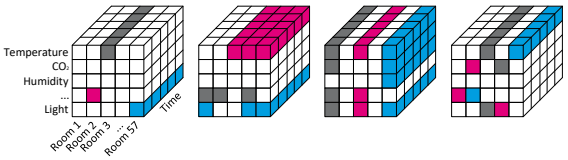

Fig. 3 - The four categories for analyzing the data in a multi-sensor network.

As illustrated by Fig. 3, the data analysis can be categorized into four categories. In the first category, each sensor at each location is analyzed in isolation on a purely temporal basis. In the second category, spatial relations are used as well, e.g. to compare a CO_2 sensor to neighbouring CO_2 sensors. In the third category, temporal and sensorial relations are used, e.g. to compare a CO_2 sensor to a temperature sensor at the same location. In the fourth category, temporal, spatial and sensorial relations are used, e.g. to compare a CO_2 sensor at a certain location to a temperature sensor at another location. For this paper, we mainly focus on the first category, but we also show the potential benefit of adding the second category. Starting with the temporal basis allows also the application of the concept for other HVAC-applications like flow, electricity consumption etc.

The dataset consists of data gathered from 57 multi-sensors from February 20, 2020 until July 10th, 2020 (139 days). Labelling is performed per day (see section 3.1), so we have 7923 day records available. With the manual labelling of the temperature sensors, the following numbers of labels apply: 5434 unlabelled, 2489 anomalies. Of these anomalies, 1653 are labelled as connection loss and 836 as other anomalies.

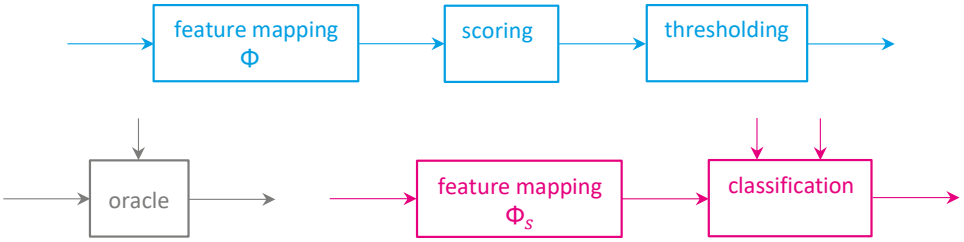

Fig. 4 - Framework for unsupervised querying (blue), manual labelling (grey) and supervised classification (pink).

Furthermore, the labelling was done by an expert. For each window of a day, visually deviating patterns in the entire dataset of 57 sensors were labelled as being anomalous. Note that the deviating pattern was judged in the temporal context of the sensor itself. This process was performed for temperature all temperature sensors.

2.2 Proposed framework

In this section, the proposed framework for learning and labelling is further discussed. We aim to detect and diagnose anomalies coming from the sensor network such that the results can be visualized in a dashboard. As illustrated in Fig. 4 together with Tab.1, the framework can be divided into three sections; unsupervised querying, manual labelling and supervised classification. Likewise, the framework can be summarized with one equation for each section:

$$l_u = \tau_\theta\left(h_u(\Phi_u(d))\right) \tag{2.1}$$

$$l_m = oracle(d, l_u) \tag{2.2}$$

$$l_s = h_s(\Phi_s(d), l_m, s) \tag{2.3}$$

Tab. 1 - Signal and function descriptions.

Notation	Name	Domain
d	Raw data	R^n
x_u, x_s	Unsupervised, supervised instances in feature space	R^o, R^p
l_u, l_m, l_s	Unsupervised, manual, supervised labels	$\{0,1\}, N, \{0,2,3,\dots\}$
s	Anomaly score	R_+
Φ_u, Φ_s	Unsupervised, supervised feature mapping	$R^n \to R^o$, $R^n \to R^p$
h_u	Scoring function	$R^o \to R_+$
h_s	Classifier	$R^p \times R_+ \times N \to \{0,2,3,\dots\}$
τ_θ	Thresholding function	$R_+ \to \{0,1\}$
oracle	Oracle	$R^n \times \{0,1\} \to N$

For the unsupervised querying, the raw data d is first mapped to x_u in the feature space using the feature mapping function $\Phi_u : R^n \to R^o$, with $o \ll n$ for dimensionality reduction. A scoring function h_u is then applied to x_u to generate an anomaly score s. Lastly, a threshold function τ_θ is applied to s for a given threshold θ to generate labels l_u, where $l_u = 0$ implies normal data and $l_u = 1$ implies anomalous data.

The manual labelling is performed by a human expert, who provides manual labels l_m for the instances where $l_u = 1$. Data selected by the query algorithm is presented to the human expert for a single sensor and for the corresponding day within the context of all other days of that sensor. Visually deviating patterns are labelled as being anomalous within their corresponding class ($l_m \in \{1,2,\dots\}$) and are labelled normal otherwise ($l_m = 0$).

For the supervised classification, the raw data d is first mapped to x_s in the feature space using the feature mapping function Φ_s. A classifier h_s is then applied to x_s using the anomaly scores s and manual labels l_m to generate supervised labels l_s. Where $l_s = 0$ implies normal data and other values imply a specific class of anomalies.

In practice, l_u is used for unsupervised anomaly detection, while l_s is used for supervised anomaly detection and diagnosis. The framework, therefore, allows us to transition from detection towards detection and diagnosis, by manually labelling instances where $l_u = 1$ and using those manual labels l_m for supervised classification.

Assuming that l_u is a good indication of anomalousness, the transition is more efficient compared to traditional querying. Additionally, the unsupervised nature of the querying allows us to query novelties. The latter is important as we need to label novelties to construct new supervised classification blocks for previously unknown anomalies. Note that once l_s is available (i.e. h_s is trained using l_m), it can be used in parallel with l_u to generate combined labels $l_c \in N$, see equation 2.4.

$$l_c = max(l_u, l_s) \tag{2.4}$$

2.3 Performance metrics

The performance of the proposed framework is evaluated in two ways. Firstly, the labelling efficiency is evaluated by comparing the fraction of anomalies queried with l_u-based querying to those queried with traditional querying. This is tested using the manually obtained labels for the entire dataset. As explained in section 2.1, these labels should not be confused with the labels obtained through l_u-based querying.

Because l_u depends on the chosen threshold θ, the performance should be measures as a function of θ. This is commonly done by ranking all instances by their anomaly score s and subsequently applying the varying threshold θ to convert the scores to labels l_u [4, 10].

Secondly, the detection performance is evaluated by comparing the combined labels l_c to the unsupervised labels l_u. But this can not be done directly, because $l_c \in N$ can be used for both detection and diagnosis while l_u can only be used for detection. Thus, compare the detection performance of the two by transforming l_c to combined detection labels $l_{cd} \in \{0,1\}$, see equation 2.5.

$$l_{cd} = \begin{cases} 0, & l_c = 0 \\ 1, & l_c \geq 1 \end{cases} \quad (2.5)$$

The labels are used to compute whether an instance is a true positive (TP), false positive (FP), true negative (TN) or false negative (FN) for each threshold. Subsequently, the precision ($PR = TP/TP + FP$) and recall ($RE = TP/TP + FN$) are computed to create a precision-recall curve. The area under the curve (AUC) is computed and used as a performance measure for detection. However, a downside to using the AUC is that it does not put emphasis on the precision. Precision is an important performance measure when addressing an important drawback of many rule-based systems: too many false positives, resulting in the loss of attention of users of the anomaly detection system. Therefore, the recall for a precision of 95% ($RE_{PR=95\%}$) and for a precision of 98% ($RE_{PR=98\%}$) are used as well.

3. Framework application

The main contribution of this paper is the proposed framework, which can be exercised with various functions for Φ_u, h_u, Φ_s and h_s. These functions can be adapted to the application at hand and the functions used here are described in this chapter. Furthermore, this chapter starts with a description of the window-based approach that is applied to the framework.

3.1 Window-based approach

Anomaly detection methods are either prediction-based or window-based [4, 5]. The window-based approach is applied in this paper. The main consideration for designing the window is that the raw data is periodic in nature, i.e. seasonal, weekly and daily. Daily windows are used because the number of seasonal and weekly cycles in the data is quite low and it allows for the detection of sudden anomalies, rather than trend or stationary anomalies.

Consequently, the data d is partitioned into non-overlapping, equally sized data windows d_{W_i} such that each day is represented by one window $W_i \subseteq t \in \{1, \dots, T\}$ with $1 \leq i \leq w$. This is done through the feature mapping functions Φ_u and Φ_s. These functions respectively generate a single o- and p-dimensional point in the feature space for each window.

3.2 Unsupervised querying

The unsupervised nature of the querying allows us to query novelties so that we can obtain manual labels for new anomaly classes and train a classifier on those labels. However, the unsupervised nature also makes automated feature learning and extraction challenging [4]. Therefore, Φ_u is constructed using expert domain knowledge only, i.e. without knowledge of the querying performance. Φ_u consists of a set of seven features $\varphi_{u,1:7}$ and thereby reduces the number of dimensions to seven. Tab. 2 gives mathematical descriptions of $\varphi_{u,1:7}$.

Tab. 2 - Feature mapping function Φ_u.

Feature	Description
$\varphi_{u,1:2}(W_i)$	$\min_{t \in W_i} d(t), \max_{t \in W_i} d(t)$
$\varphi_{u,3}(W_i)$	$\sum_{t \in W_i} d(t)/\lvert W_i \rvert$
$\varphi_{u,4}(W_i)$	$\min_{t \in W_i} d(t) - d(t-1)$
$\varphi_{u,5}(W_i)$	$\max_{t \in W_i} d(t) - d(t-1)$
$\varphi_{u,6}(W_i)$	$\lvert \{d(t) : t \in W_i\} \rvert$
$\varphi_{u,7}(W_i)$	$\sum_{t \in W_i} \left\lVert d(t) - \frac{1}{w} \sum_{j=1}^{w} d\left(t + j\frac{T}{w}\right) \right\rVert$

The textual descriptions of $\varphi_{u,1:7}$ are respectively the minimum, the maximum, the average, the minimum first order difference, the maximum first order difference, the number of unique values and the Euclidean distance to the average window of a window W_i.

As we are interested in the detection at the individual sensor level, Φ_u is applied to each sensor individually. However, the scoring function h_u and thresholding function τ_θ are applied to all sensors combined so that the detection at the sensor level is done with the knowledge of the behavior of the other sensors. Note that all sensors are treated equally here and no information (e.g. spatial information) is used to group them.

The unsupervised scoring function h_u: $R^o \rightarrow R_+$ first scales the feature space using median-MAD scaling, as it is robust to anomalies and it centers the data around the origin [11, 12]. The median-MAD scaling function is given by equation 3.1, where x is a vector containing a feature's values in the feature space and x' is the scaled version of that vector.

$$x' = \frac{x - median(x)}{median(|x - median(x)|)} \quad (3.1)$$

An unsupervised learning algorithm is then applied to the scaled feature space in order to generate scores s. Unsupervised learning algorithms for anomaly detection can broadly be categorized as statistical, proximity-based and information-theoretic [4, 5, 6]. Two of the most established algorithms are the k^{th} nearest neighbour (k^{th}-NN)

and the local outlier factor (LOF) algorithms [4, 10, 13]. k^{th}-NN is better suited for finding global anomalies, while LOF is better suited for finding local anomalies [10]. Since we are mainly interested in global anomalies, we apply k^{th}-NN.

Please note that the unsupervised k^{th}-NN algorithm should not be confused with the supervised k-NN he k^{th}-NN algorithm finds the distance to an instances k^{th} nearest neighbour and we use that distance as anomaly score s. For k, we use half the size of the dataset [14].

3.3 Supervised classification

For the supervised classification, two sets of the feature mapping function Φ_s and the classifier h_s are constructed to demonstrate the transition from detection towards detection and diagnosis. Both sets aim to detect and diagnose detection losses.

Classification set 1: $\Phi_{s,1}$ computes the longest period during which the values do not change for each sensor individually. The classifier $h_{s,1}$ is then applied to all sensors combined.

Classification set 2: $\Phi_{s,2}$ computes the longest period during which the values of all sensors do not change. The classifier $h_{s,2}$ is then applied to the resulting feature space.

Two sets are constructed to demonstrate the strength of utilizing the sensor network for anomaly detection and diagnosis on a single sensor level for connection losses. Note that both $h_{s,1}$ and $h_{s,2}$ scale the feature space through median-MAD scaling, see equation 3.1.

Though the proposed framework allows us to efficiently obtain positive manual labels, only a small subset of the data is generally labelled as such. Additionally, very few or no negative labels are obtained. Given these constraints and the need to distinguish classes of anomalies from the normal data, we are limited to a specific case of binary classification called positive unlabeled (PU) learning [15]. In PU-learning, a learner only has access to positive and unlabeled instances.

In order to obtain a binary classification problem, both positive and negative labels are required to train the classifier. As the oracle provides mostly positive labels, there is a lack of negative labels in the training set. But since normal data is abundant, we can leverage the anomaly score s to select instances that are likely normal, i.e. instances with a low anomaly score. Those instances are then assumed to be normal and used to enrich the training set, see Fig. 5. This is done such that the normal of positive labels equals the number of negative labels.

For the resulting classification problem, we use a k-NN classifier, which is optimized with respect to the leave-one-out classification error. As described by equation 2.4, the output l_s of the classifier overrules the output l_u of the unsupervised detection.

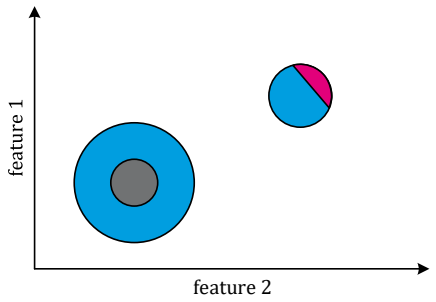

Fig. 5 - Selecting likely normal instances (grey) from unlabeled data (blue) to enrich the training set, which would otherwise consist of only manually labelled anomalous instances (pink).

4. Results

4.1 Labelling efficiency

As described in section 2.3, the labelling efficiency using l_u-based querying is compared to conventional querying. With conventional querying, a human annotator sequentially inspects the data day by day. The strategies are compared by comparing the percentage of anomalies labelled as a function of the percentage of instances queried for both query strategies, see Fig. 6.

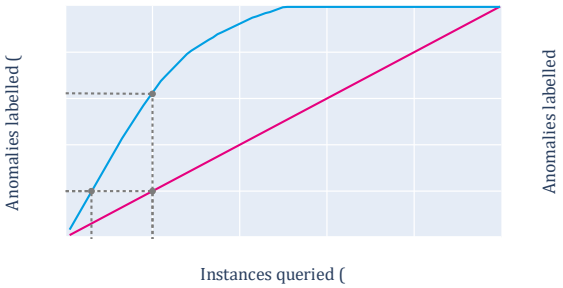

Fig. 6 - Labelling efficiency for anomalies using l_u-based (blue) and conventional querying (pink).

When querying 20% of all instances using conventional querying, 20% of all anomalies will be labelled. But when querying 20% of all instances using l_u-based querying, 62% of all anomalies will be labelled. Alternatively, to label 20% of all anomalies using l_u-based querying, only 6% of all instances need to be queried.

Because supervised labels l_s are generated for connection losses, as described in section 2.3, we also evaluate the labelling efficiency for this specific class of anomalies, see Fig. 7.

When querying 20% of all instances using conventional querying, 20% of all connection losses will be labelled. But when querying 20% of all instances using l_u-based querying, 69% of all connection losses will be labelled. Alternatively, to label 20% of all connection losses using l_u-based querying, only 7% of all instances need to be queried.

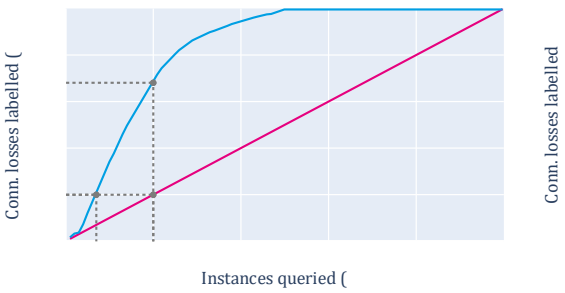

Fig. 7 - Labelling efficiency for connection losses using l_u-based (blue) and conventional querying (pink).

The results show the improved labelling efficiency for labelling anomalies and a specific anomaly class (i.e. connection losses) using l_u-based querying compared to conventional querying. The improved efficiency allows the oracle to label more anomalies for a given number of queries. Or alternatively, it allows the oracle to reduce the number of queries to label a given number of anomalies and thereby reduce the resources required for human annotation.

4.2 Detection performance

As described in section 2.3, the combined performance is evaluated by comparing the combined detection labels l_{cd} to the unsupervised labels l_u with precision-recall curves. Naturally, this performance is a function of the number of labelling instances. But though the labelling is done efficiently, the process could still be labour intensive if a large number of labels is required for classification.

In this section, we show the combined performance when only 1% of all instances are queried, i.e. those with the highest anomaly scores s. This provides us with 27 labelled connection losses. Additionally, the 27 instances with the lowest s are selected to complete the training set, while assuming they are normal.

For classification set 1, the feature mapping function $\Phi_{s,1}$ is applied to the raw data and classifier $h_{s,1}$ is trained on the training set. $h_{s,1}$ is then applied to the remaining instances, i.e. the test set, to generate supervised labels l_s. Fig. 8 shows the resulting performance when equation 2.4 and equation 2.5 are applied.

For classification set 2, we use the knowledge that connection losses often occur for all sensors simultaneously. So if we have a label for 1 sensor for a day, we assume all sensors have a connection loss and use that for training the classifier. Again, Fig. 8 shows the resulting performance when equation 2.4 and equation 2.5 are applied.

The performance can now be measured by the AUC of the precision-recall curves, see Tab. 3. Though as stated in section 2.3, this metric does not put emphasis on the precision, which is important for preventing false positives. Therefore, we also take into account $RE_{PR=95\%}$ and $RE_{PR=98\%}$.

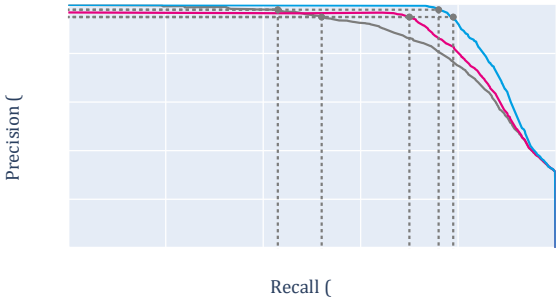

Fig. 8 - The precision-recall curve for unsupervised detection (grey), combined detection using classification set 1 (pink) and classification set 2 (blue).

Tab. 3 - AUC, $RE_{PR=95\%}$ and $RE_{PR=98\%}$ for the different configurations.

Configuration	AUC	$RE_{PR=95\%}$	$RE_{PR=98\%}$
Unsupervised detection	87%	52%	43%
Combined detection using class. set 1	88%	70%	-
Combined detection using class. set 2	90%	78%	76%

Clearly, the combined detection using classification set 2 performs the best, as it scores the highest on all metrics. Whether the unsupervised detection is preferred over the combined detection using classification set 1 is depends on the application though. The latter does have a higher AUC and $RE_{PR=95\%}$, but its main drawback is that it never achieves a precision higher than 97% (and therefore does not have a $RE_{PR=98\%}$). This is due to the false positives generated by the supervised classification, which overrule the output of the unsupervised detection.

5. Discussion and future direction

The proposed framework is rather generic in nature, but it does have two limitations resulting from the framework application. Firstly, the window-based approach results in the detection of anomalous windows. But if a short anomalous event occurs that renders the window anomalous, the detection is not able to single out the event itself within the window. Secondly, we focus on the detection of sudden anomalies and the approach does not allow for detecting stationary or trend anomalies.

Furthermore, we have shown that an improvement in labelling efficiency can be achieved by querying instances with a high anomaly score. Though a drawback of this strategy is that it does not allow us to target edge cases for querying. Future research could focus on utilizing active learning methods to query edge cases for partially labelled anomaly classes.

When instances are manually labelled by the oracle and they are used to train a classifier, we have seen that the false positives provided by the classifier limit the precision of the whole framework. This is a direct result of the supervised labels overruling the unsupervised labels through equation 2.4. It is therefore advised to use only supervised classifiers with high precision in combination with unsupervised detection. Furthermore, a specific feature mapping function and binary classifier need to be designed for each anomaly class. To reduce the required work associated with classifying multiple anomaly classes, it is worth investigating the usage of a multi-class classifier and automated feature extraction.

Note that the unsupervised/supervised approach not only improves anomaly detection but also provides diagnosis for labelled classes.

6. Conclusions

A common problem for the application of machine learning-based FDD is the lack of labelled data. In this paper, we have shown that the proposed framework allows us to efficiently obtain labels and use those labels to transition from unsupervised detection towards supervised diagnosis. This was shown using a real-world dataset of multi-sensors in an office building.

Additionally, we have shown that the (mostly positive) labels can be used in combination with the anomaly scores to train a classifier using PU-learning. This results in an improved precision-recall curve, and additionally allows us to generate supervised labels for diagnosis. This has been shown for connection losses of the multi-sensor, an urgent topic when multi-sensors operate according to the CoV principle with its frequently missing values in the time series.

7. Acknowledgements

The authors acknowledge the financial support of the Dutch RVO TKI Urban Energy through the project Brains for Buildings.

8. References

[1] Neumann C., Yoshida H, Choinière et al. (ed.) Commissioning tools for existing and low energy buildings, 2010, IEA Annex 47 report 2.

[2] The Well Performance Verification Guidebook Q3 2021. WELL Building Institute 20.

[3] Zhao Y., Li T., Zhang C., Artificial intelligence-based fault detection and diagnosis methods for building energy systems: Advantages, challenges and the future. Renewable and Sustainable Energy Reviews. 2019; 109: 85-101.

[4] Aggarwal C. C., Outlier Analysis. Springer; 2015.

[5] Chandola V., Banerjee A., Kumar V., Anomaly detection: A survey, ACM computing surveys (CSUR). 2009; 41(3): 1–58.

[6] Zhang Y., Meratnia N., Havinga, P., Outlier detection techniques for wireless sensor networks: A survey. IEEE communications surveys & tutorials. 2010; 12(2): 159–170.

[7] Kim W., Katipamula S. A review of fault detection and diagnostics methods for building systems. Science and Technology for the Built Environment. 2018; 24:1, 3-21.

[8] Chen L., Li G., Huang G., A hypergrid based adaptive learning method for detecting data faults in wireless sensor networks. Information Sciences. 2021; 553: 49-65.

[9] BR Controls, BRT-35 Plafond multi-sensor, V20200124. [Download 2020 Nov 2] Available from www.brcontrols.com.

[10] Goldstein M., Uchida S., A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. PloS one. 2016; 11(4), p. e0152173.

[11] Kandanaarachchi S., Muñoz M. A., Hyndman R. J., Smith-Miles K. On normalization and algorithm selection for unsupervised outlier detection. Data Mining and Knowledge Discovery, 2020; 34 (2): 309–354.

[12] Rousseeuw P.J., Hubert M., Anomaly detection by robust statistics. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2018; 8(2), p. 1236.

[13] Breunig M.M., Kriegel H.-P., Ng R. T., Sander J., Lof: identifying density-based local outliers. In: Proceedings of the 2000 ACM SIGMOD international conference on Management of data, 2000, p. 93–104.

[14] Hajee B., Anomaly detection in multi-sensor networks: an active learning approach. Master Thesis Delft University of Technology. 2020; p.24.

[15] Bekker J., Davis J., Learning from positive and unlabeled data: a survey. Machine learning. 2020; 109:719-760.

Data Statement

The datasets generated during and/or analysed during the current study are not available due to commercial restrictions.