Multivariable Anomaly Detection Framework for Multi-sensor Network

From rule-based to data-driven

Jingru Feng

Master of Science Thesis

Delft University of Technology

Delft Center for Systems and Control

Multivariable Anomaly Detection Framework for Multi-sensor Network

From rule-based to data-driven

Master of Science Thesis

For the degree of Master of Science in Systems and Control at Delft University of Technology

Jingru Feng

December 8, 2022

Faculty of Mechanical, Maritime and Materials Engineering (3mE) · Delft University of Technology

The work in this thesis was supported by Adviesbureau DWA. Their cooperation is hereby gratefully acknowledged.

Abstract

With the demand for more information on the building's indoor climate, a massive amount of multi-sensors are mounted in buildings. Sensor anomalies in smart buildings lead to higher energy consumption and a less comfortable indoor climate. In the state of practice, rule-based approaches were proposed to detect and diagnose sensor anomalies in the building sensor network. However, as the number of sensors is growing and the types of sensors are becoming more diverse, rule-based approaches become more and more limiting and expensive.

This project proposed a data-driven detection and diagnosis framework that transfer the knowledge from one sensor to the other sensors in the network for detecting and diagnosing sensor anomalies in the smart building. In the proposed framework, all the data from one room that contains targeted anomalies is chosen to be the training set. Principal component analysis (PCA) is used to extract the features, and support vector machine (SVM) is used to model the classifier. This framework is tested specifically on all the CO2 sensors in a smart building. Functions of detecting a single anomaly, detecting mixed anomaly and diagnosis anomaly are evaluated and compared with the state of practice. Moreover, the influence of anomaly rate on the performance of the proposed method is investigated. In order to test the sensitivity to the training dataset, a final experiment was performed where the room that provide the training data was changed to a different room.

Table of Contents

List of Figures

List of Tables

Acknowledgements

When I was young, I always wanted to contribute to sustainability. I try my best to perfect garbage classification, water and energy saving and discover things that can help protect the environment. As I grow up, I begin pursuing an engineering degree. And I realise that the knowledge I gain during my study can be used to help achieve sustainability in a more effective way and the outcome can be more influential. This project is technically my first project in that I use technology to help step forward to net-zero emission. I would like to thank all my supervisors for giving me this chance to explore sustainability and apply my knowledge to a real-world project.

I would like to thank my supervisor P. Mohajerin Esfahani for his assistance during the writing of this thesis. He provided guidance not only in academic research that guide me through this thesis but also in a lifelong beneficial lesson on how to be a good engineer. I would like to thank my supervisor B. Hajee and K. Wisse from Adviesbureau DWA. I have learned a lot from working with them. I obtained valuable knowledge in how to build up an algorithm, evaluate the algorithm and motivate myself. More importantly, I learned much about Dutch working culture. Last but not least, I would like to thank my family and friends, who have been supporting me during the whole year. I feel very lucky to have them in my life.

I believe graduating is not the end, but a new beginning of my study.

Delft, University of Technology — Jingru Feng
December 8, 2022

“There is only one heroism in the world: to see the world as it is, and to love it.”
— *Romain Rolland*

Chapter 1

Introduction

1-1 Research background

As the increase of the population, energy consumption is rising. The increment in energy consumption is bringing many impacts on the world. This increment is leading to global warming and global energy shortage. Global warming and global energy shortage are very much of concern. Two of the consequences of global warming and the energy shortage are climate change and energy price increases. These consequences have raised people's awareness of the massive energy consumption. In 2022, the whole world is in the middle of the energy transition because the goal is to reach net-zero operation by 2050 [1].

Among the energy consumption, buildings are responsible for 36% of the total amount [2]. Most of the energy is spent on maintaining the indoor climate. Therefore, the building and construction sector is playing important roles in this global transition. To cope with the massive energy consumption, many companies, such as Adviesbureau DWA (DWA), are researching and developing smarter and more sustainable solutions. These solutions such as digital twins, data-driven climate control and energy monitoring platform, dedicate to giving human being a better indoor climate but meanwhile minimise the usage of energy. They rely heavily on the analysis and usage of big data. To satisfy the increasing demand for big data, more and more sensors are mounted in the building. These sensors assist the building management system (BMS) in monitoring and controlling the building's indoor environment.

However, the sensors sometimes would give measurements that do not comfort the expected behaviours. They are abnormal events or patterns that deviate from normal or expected events. These data are called anomalies.

The cause of the sensor anomalies can be two scenarios. It could be the sensors themselves are broken or have errors. For example, sensors need periodic calibration. If one sensor is not correctly calibrated, the output data will have an offset. It could also be that unexpected events happened in the environment that make the sensors output abnormal values. For example, the heating, ventilation and air conditioning (HVAC) system could be suffering from malfunctioning or inefficient operations. These types of events could lead sensors to output an extremely high or low value over a period.

The sensor is an important part of the building climate control system. As shown in the figure 1-1, BMS determines setpoints for indoor conditions. The setpoints can be the temperature in the room, CO2 density in the room and so on. The controllers receive the setpoints and then will control the actuators (HVAC) to influence the system. The system in this control loop is the office. The occupancy of the human will bring disturbance to the indoor environment. Sensors are the feedback part of the control loop. Sensors can not only "tell" the system what the current conditions are, but their historical measurement can be used for analysis and for BMS to determine new setpoints.

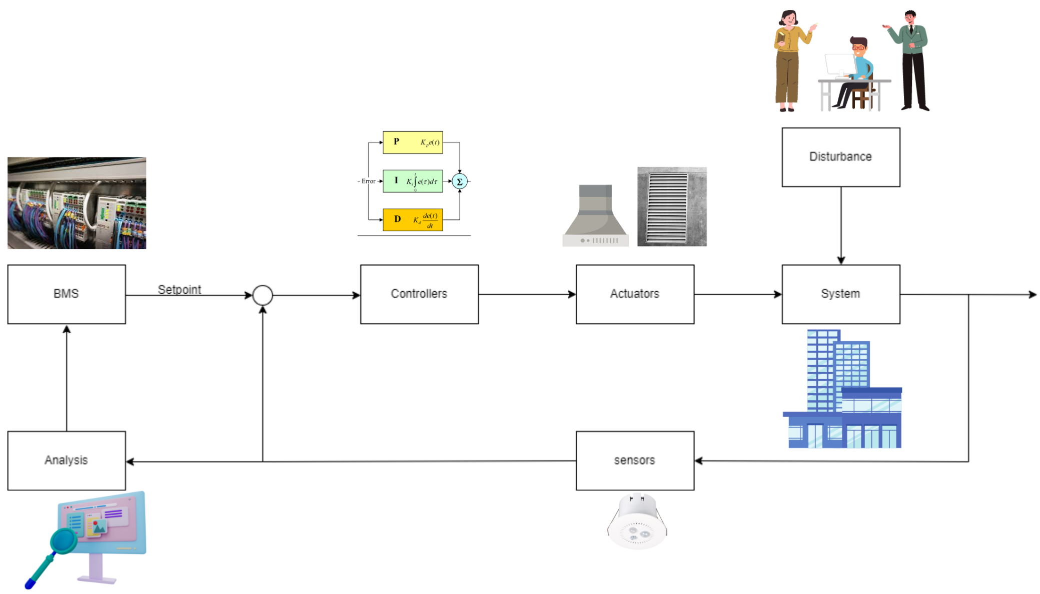

Figure 1-1: The control loop in the BMS

Sensor anomalies have great impacts on the performance of the climate control system and sensor anomalies can affect the performance evaluation of the company. The sensor provides the feedback for the control system, if sensors cannot provide decent feedback, the indoor environment will not end up where the setpoints are. Sensor anomalies can lead to higher energy consumption because the system has to work extra to compensate for the effect caused by sensor anomalies. Moreover, for companies like DWA, nowadays their performance contract is data-driven. This means their performance will be evaluated using the data from the sensors. Companies could face unnecessary punishment during the performance evaluation if the sensors are outputting abnormal data. Therefore, the reliability of the sensors is of paramount importance, and automated anomaly detection and diagnosis in sensor networks are essential.

1-2 State-of-practice solutions and current challenges

The current state-of-practice methods to detect and diagnose in the practical are rule-based approaches. Rule-based approaches are drawing up a set of rules according to the expert's experience with the data. The data points that fall out of the rules are considered anomalies. For example, in the figure 1-2, one expert can set a rule for CO2 sensors, that the value of the CO2 sensor must be under 800 ppm [1]. If the value exceeds the range, the value is considered abnormal.

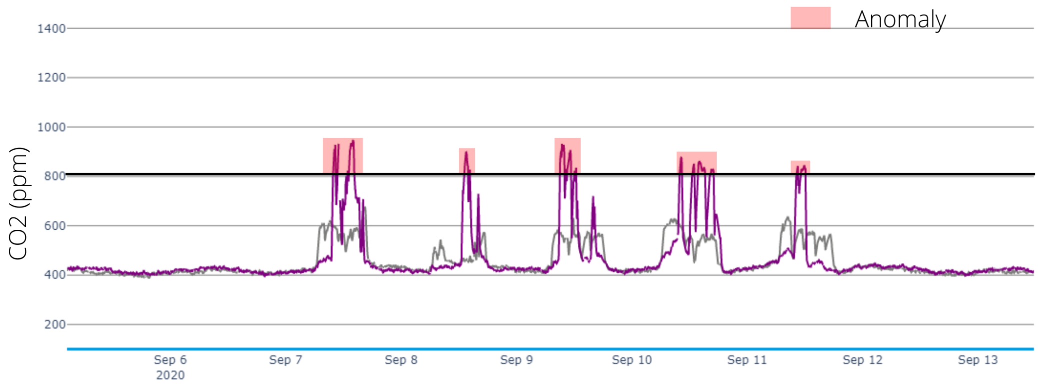

Figure 1-2: An example of rule-based approaches

However, as the scale of the sensor networks is growing, variables in smart building systems are usually multi-dimensional[3]. The rule-based approaches have limitations in the real data problem. To detect and diagnose anomalies among a big amount of sensors, the rules are becoming very complex. In a smart building, there will be sensors mounted in different parts of the building. The data series have a high dimensionality. The location, size of the room, and direction that the rooms are facing and such factors can influence the normal range of the sensors of the rooms. Setting rules for such a big sensor network is labour-intensive and requires empirical knowledge. It is desired for the companies in the industry to develop an easier solution that can be done by a less experienced person, and the solution is robust in a dynamic environment.

Moreover, the rule-based approach is giving out false positive detection. False positives cause inconvenience for the employee. Having a false positive means that the system gives an alert that raise the responsible employees' attention but there is nothing that needs to be fixed. This type of alert is cost unnecessary time and energy of the people. So this type of mistake needs to be eliminated as much as possible.

1-3 Proposed solutions

To break the limitation of the state of practice, data-driven methods are proposed. The main objective of this project is to build and test an anomaly detection and diagnosis framework

[1]Since this project only focuses on CO2 sensors, for all the figures that show signals from one or more CO2 sensors, the y-axis titles are CO2 density (unit:ppm). The y-axis title is only shown in the figure 1-2

that aims to increase the reliability of the building sensor networks and efficiency of the indoor climate control. This project tackles the challenges that happen in the current state-of-practice approaches.

Data-driven methods, compare to rule-based methods, are easily implementable. They can learn human's opinions by composing a hypothesis function that maps the original data or the features of the data to a score or a label space. For the challenge of solving the high false positive rate, a proper classifier can be selected to achieve it. Because parameters in some classifiers can be tuned to lower the false positive rate.

Based on the proposal, the corresponding research questions are:

- In practice, most of the datasets provided by companies are unlabelled. The insufficient labelled dataset is a challenge in the data-driven method. To tackle this challenge, an approach to selecting the initial dataset is developed. Among the high dimensional data, an approach to efficiently build a good quality initial training set for the machine learning algorithm is found.
- To cope with the high dimensionality of the data, a proper feature extraction or selection approach is developed. Therefore, the second challenge of the research is about finding a set of good features which allow us to represent the high-dimensional data using a low dimension features.
- A complete framework is developed based on the previous two research questions. This framework takes historical data and the new monitoring data as input, then give the label to the new monitoring data on a daily based as output. Moreover, the framework can be updated based on iterations.
- At the end, an evaluation of the proposed method is performed. The evaluation compares the proposed method to the state-of-practice rule-based approach. Some criteria for the assessment are discussed.

1-4 Thesis outline

This paper is split into 6 chapters. Chapter 1 is a general introduction to the research background, current challenge and the proposed approach. Chapter 2 introduces the indoor climate control system and sensors network this project closely works with and it further introduces the sensor anomalies this project focuses on. Chapter 3 provides the details of preprocessing the dataset, selected classification model, and other preliminary. Chapter 4 elucidates the proposed anomaly detection and diagnosis framework. In chapter 5, the evaluation criteria are presented and the result of the proposed method is compared with the state-of-practice rule-based methods. Chapter 6 concludes the work in the report and provides highlights for future directions.

Chapter 2

System Description

In this chapter, the system that this project is working closely with is introduced. Firstly, there is a description of the sensors and sensor networks that this project is going to work on. Secondly, the climate control system is introduced. In addition, this chapter will introduce the type of anomaly in the sensor network that will be detected.

This project, together with Adviesbureau DWA (DWA), developed a data-driven framework for anomaly detection in smart buildings. As this project aims at developing a methodology for a real-world project. A real-time dataset will be needed. DWA is a consultancy company that provides solutions for making buildings more and more sustainable. DWA as a good company in the industry cares about this topic. Thus, they provide the dataset and utility for this project.

2-1 Climate control system

The air-water systems control the air quality and temperature in the building. Take the DWA Gouda office as an example. As shown in figure 2-1, there are air distribution pipes which transfer fresh air from the outdoor environment to the indoor. The fresh air, assumed $26°C$, goes through the heating and the cooling coil where its temperature is modified based on the demand. For example, in the summertime, the temperature of the fresh air will be modified to $18°C$. The air will be distributed to the rooms which are occupied by people and demand cold air. There are water pipes that distribute hot or cold water into the radiator in every room which will add heat or coldness to the rooms.

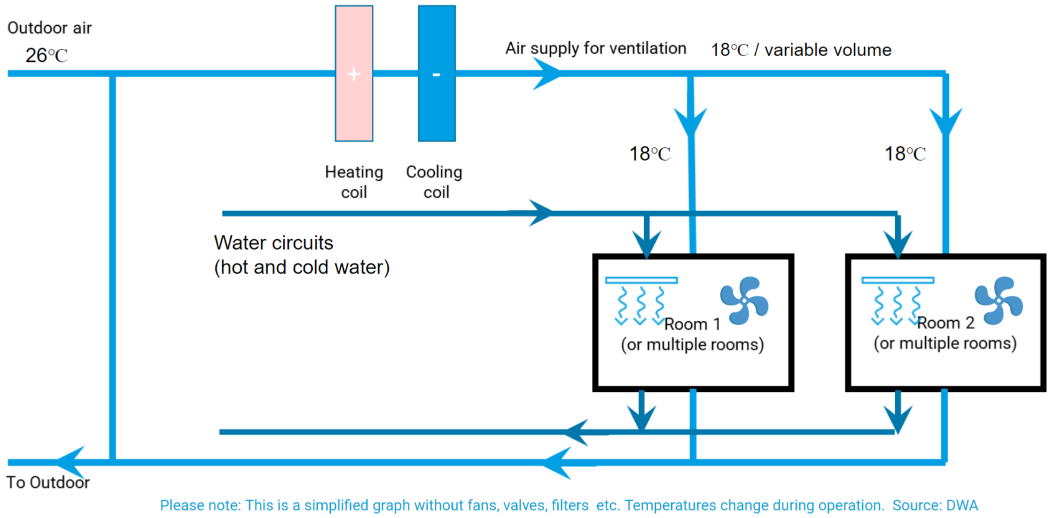

Figure 2-1: The climate control system in DWA

2-2 Multi-sensors and the sensor networks

Multi-sensors combine multiple individual sensors in one box. Multi-sensor in buildings can measure more than one relevant parameter in a location. In this project, the BRT-35 multi-sensors from BRControls© are used (figure 2-2). Each multi-sensor measures 6 parameters at one location: CO2, temperature, light, sound, (relative) humidity and occupancy. In this project, the sensors' measurements are read every 3 minutes. So, the signals are time series with a sample time of 3 minutes. The time series from the sensors are stored using the principle of "Change of Value". The principle of Change of Value means the sensor will check if the new measurement has changed within a small range. If not, the value will not be stored. The aim of this data-storing principle is to save storage space. If the sensor only records the change of the value instead of recording every data sample, the required storage of the data will be less.

Figure 2-2: BRT-35 sensor

The sensor networks consist of 57 independent multi-sensors. These sensors are distributed in the 2 floors owned by DWA Gouda office as shown in figure 2-3. From the figure 2-3, it can be read that in the company, there are conference rooms (grey), stairwell (green), hallway and open space (white), open working space (blue and not surrounded by the black line), closed working space (blue and surrounded by the black line) and single-person concentrating rooms (yellow). Every red dot represents a multi-sensor.

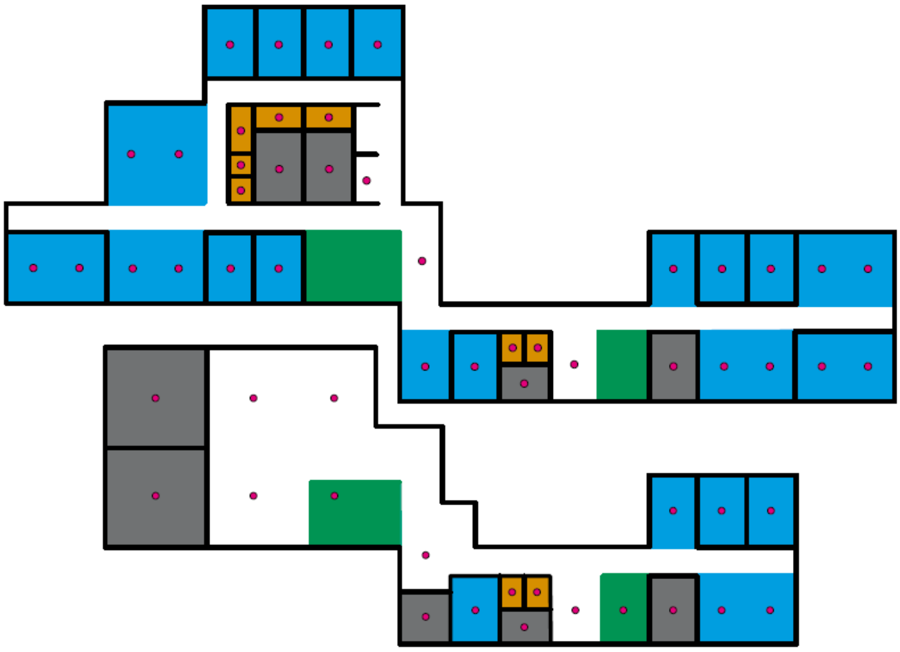

Figure 2-3: Floor map of DWA Gouda Office

2-3 Anomaly types and formulation

The definition of the anomaly in this context is the unusual patterns in the time-series data. As stated in the previous chapter, these unusual patterns can be caused by the error in the sensors themselves or by the occurrence of unusual events in the environment. In this project, we consider the abnormal behaviour of the sensors resulting from both scenarios.

The anomalies that occur in this sensor network can be represented by the following categories. Instant anomaly, constant anomaly, graduate drifting anomaly, bias anomaly and miss anomaly[4]. Explanation and examples of each type of anomaly are shown below:

- Instant anomaly: instant anomaly is referred to the sudden change between two consecutive readings from the same sensor. This type of anomaly normally can be easily recognised by visual inspection. As shown in the figure 2-4, one of the typical instant anomalies is the lower peak. This type of anomaly has a slighter impact on the building management system compared to the other types of anomalies since it is corrected by itself by the next sample time. Therefore, the instant anomaly will be removed during the data preprocessing.

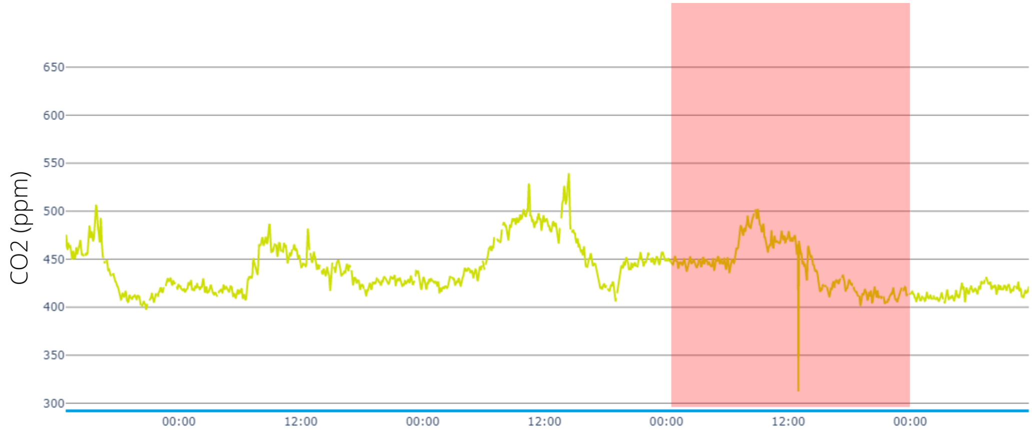

Figure 2-4: An example of the instant anomaly

- Constant anomaly: the measurements from the sensor that is deviating from the expected pattern and lasting for a period. Figure 2-5 shows an instance of constant anomaly. Due to the malfunction of the ventilation in the building on the day, the CO2 value climbs to a high point and only recovers until the ventilation is back to work.

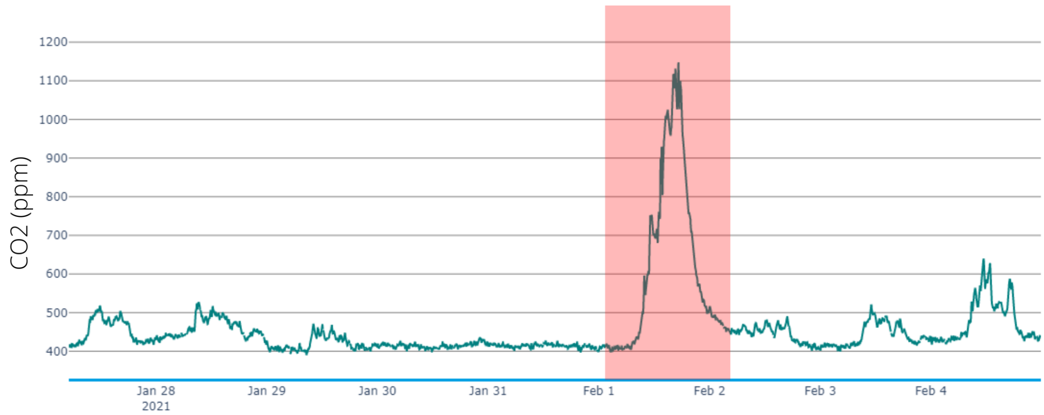

Figure 2-5: An example of the constant anomaly

- Gradual drifting anomaly: the observation of the sensor is gradually drifting. In a longer period, the gradual drifting will accumulate and result in a large discrepancy between the real values and the sensor measurements. Like what is shown in figure 2-6, in the time frame, there was no activity going on in the office, the observation of the sensor was supposed to be around a constant value. However, due to the degeneration of the sensor, the value is slowly dropping. It happens commonly in the sensors, which is why the sensors are equipped with an auto-calibration function.

- Bias anomaly: a constant offset in the measurements for a period. As shown in figure 2-7, the observation all of a sudden has been lifted to a different level. This type of anomaly has the biggest impact on the performance of indoor climate control.

- Miss anomaly: the measurement is missing for a time period. This could be caused by

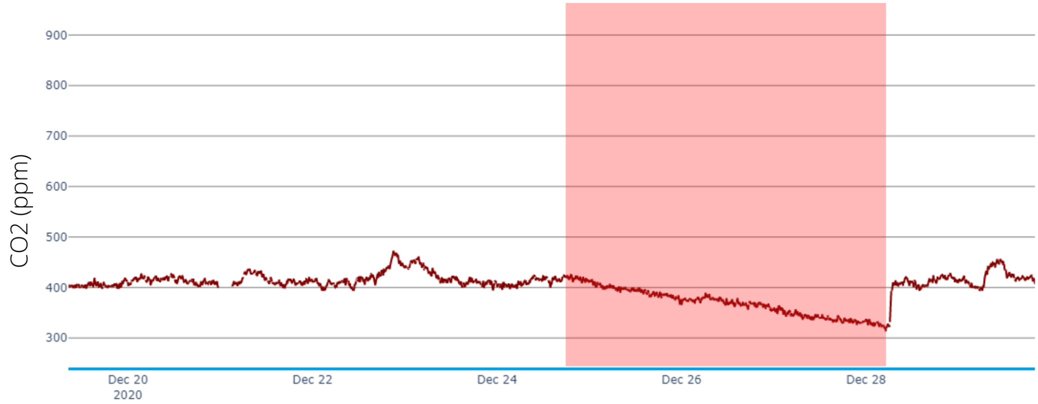

Figure 2-6: An example of the gradual drifting anomaly

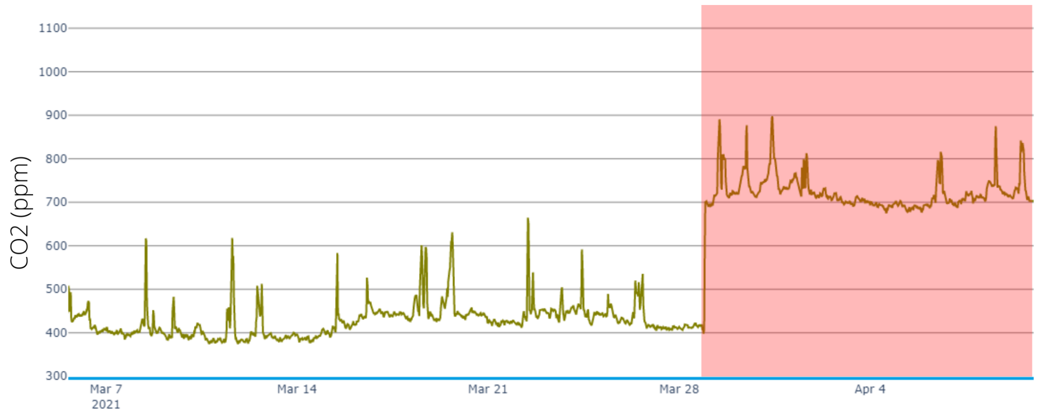

Figure 2-7: An example of the bias anomaly

connection loss, incorrect mounting of the sensors or maintenance. Figure 2-8 shows an example of the miss anomaly.

2-4 Dataset and data dimensions

The available dataset at DWA consists of the data from August 1, 2020 to December 9, 2021. The dataset includes data from every sensor in the office.

The data measured from the multi-sensors is in the form of time series. Each multi-sensor contains 6 individual sensors. So, the signals from one sensor are 6-dimensional time series. In total, there are 57 sensors distributed in the office. This complex sensor network produces high-dimensional data. However, sometimes only part of the sensors and their data will be selected for the analysing purpose based on the need. The input can be categorised into four types based on the usage of the data. As shown in figure 2-9, the first type only takes temporal information of individual sensors as input. The second type will focus on the same type of sensors but in multiple rooms. The third type will focus on the multiple sensors in a single room. And the fourth type takes different sensors located in different rooms.

This project will focus on the second type of input. The proposed framework will be applicable

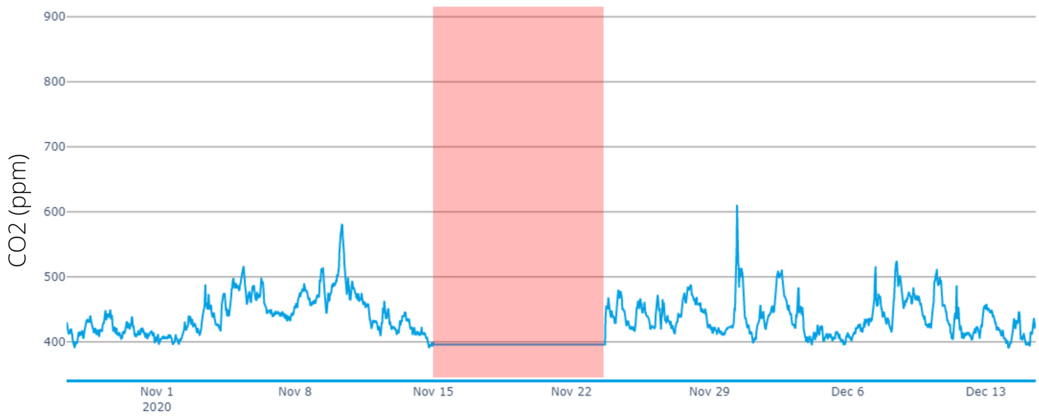

Figure 2-8: An example of miss anomaly

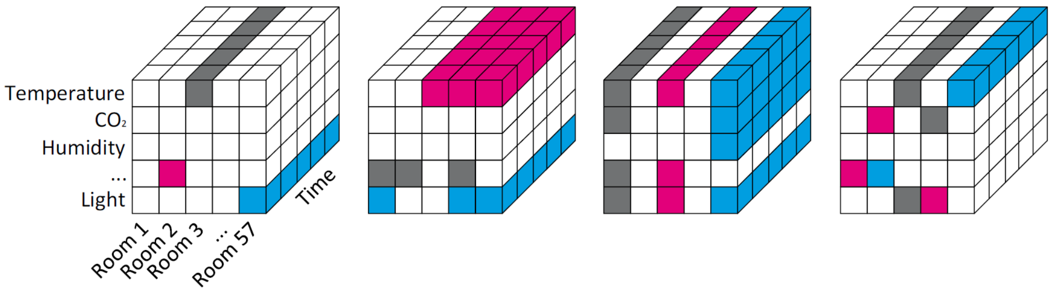

Figure 2-9: The four categories for analysing the data in a multi-sensor network

to detecting and diagnosing the bias anomaly and gradual drifting anomaly of the CO2 sensors in every room.

Chapter 3

Preliminary

This chapter provides detailed information on the following content. First of all, the procedure of data preprocessing which includes data recovery, down-sampling and outliers removal is presented. Secondly, this chapter explained the method of forming the initial training set. Then, the concept of feature extraction based on principal component analysis (PCA) is elucidated. Last but not least, the classification algorithm support vector machine (SVM) is introduced.

3-1 Data pre-processing

The data in the dataset are the measurements from the sensors. The real-world data is incomplete, contains noise and redundant. Before the data can be used in the training and testing of the algorithm, they need to be pre-processed. To begin with, the missing value in the dataset needs to be recovered. Besides the missing value, there are also outliers in the dataset. A step of outlier removal is necessary. In addition, the original data is sampled every 3 minutes. The number of data points for each day is high. This would cost much time for computation in the later steps. And this massive data point would bring complexity. Therefore, down-sampling is used to produce an approximate sequence and shrink the data.

3-1-1 Data recovery

As mentioned in chapter 2, the sensor network adopts the Change of Value principle to store the measurement. The values that are not changing compared to the previous reading are stored as empty values in the system. As a result, the data read from the sensors is incomplete.

The data need to be reconstructed due to the missing value. Based on the storing principle, zero-order hold interpolation is adopted to reconstruct the data. The zero-order holder holds the value of the previous sample for one or a few sample intervals until new a value is read. The principle of the zero-order holder is corresponding with the change of value storing principal.

As shown in the equation 3-1, assume a current time point $x(n)$ is missing, the interpolated value for $x(n)$ is inherited from the previous value.

$$x(n) = x(n-1), \quad if\ x(n) = \text{"nan"} \tag{3-1}$$

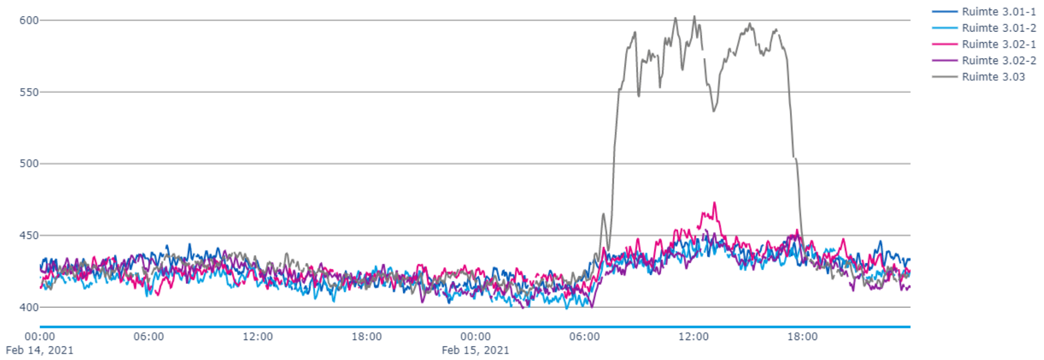

Figure 3-1: Data before zero-order holder

As shown in figure 3-1, the signal is not continuous. The broken parts are where the empty values locate. In figure 3-2, the empty values are interpolated with the value of their previous samples.

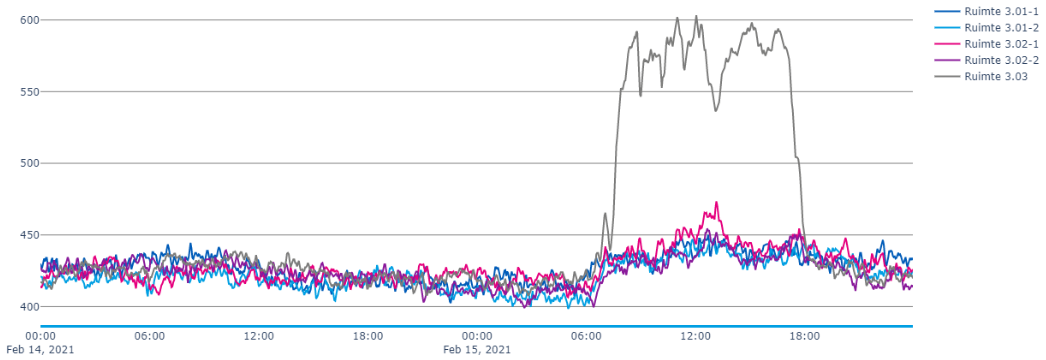

Figure 3-2: Data after zero-order holder

3-1-2 Outliers removal

In the dataset, there are instant anomalies. These anomalies are in the form of upper peaks or lower peaks. In the dataset that this project is working with, there are only lower peaks. As mentioned in chapter 2, the instant anomaly has the least slight impact on the control system. But instant anomaly has an influence on detecting and diagnosing other anomalies. Therefore, detecting and removing the instant anomaly is part of the data preprocessing.

The lower peaks are detected using the rule presented in table 3-1. Assume a current time point $x(n)$ and a threshold H, the condition of judging if $x(n)$ is a lower peak or not is shown in table 3-1

Table 3-1: Rule for detecting a lower peak

Condition	Detection result
$x(n-1) - x(n) >= H$	$x(n)$ is a lower peaks
$x(n-1) - x(n) < H$	$x(n)$ is a normal data point

In this project, $H = 40$. After removing the lower peaks that are being detected at time point n, $x(n)$ becomes an empty value. Therefore, $x(n)$ is interpolated using the average between the previous value $x(n-1)$ and the next value $x(n+1)$.

$$x(n) = \frac{x(n-1) + x(n+1)}{2} \tag{3-2}$$

Figure 3-3 shows one part of the data that is before outliers removing and figure 3-4 shows the same part of data that is after outliers removing.

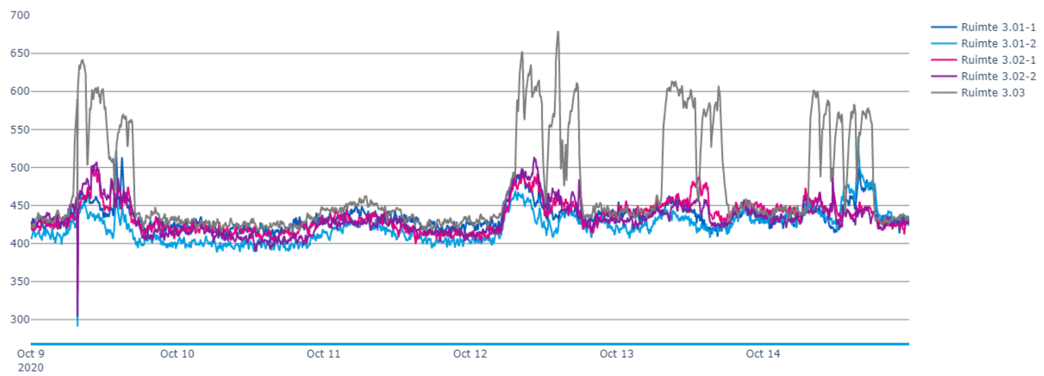

Figure 3-3: Data before outliers removing

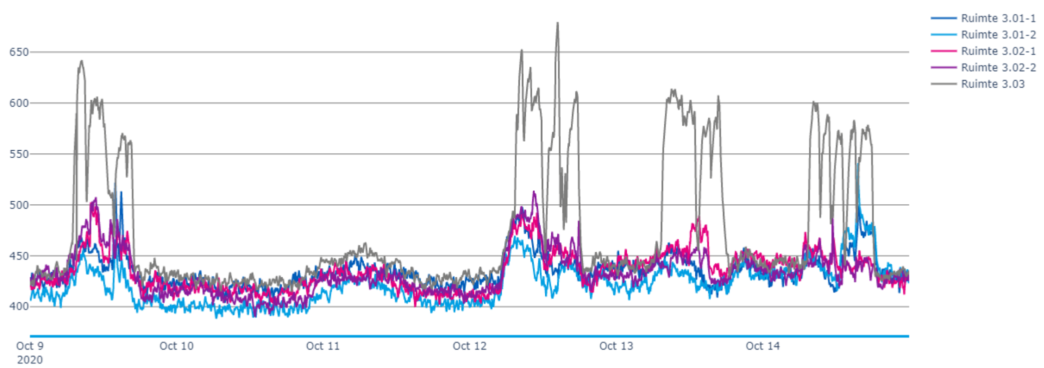

Figure 3-4: Data after outliers removing

3-1-3 Down-sampling

As mentioned in chapter 2, the original sample time of the data is 3 minutes. With the original sample time, each day would contain 480 data points. Having high-dimensional data cost time in the later steps. As part of reducing the dimensions of the data, down-sampling is an efficient way. Besides, with a short sample time, there would be noise in the signals. The measurement noise can be smoothed by down-sampling.

For this project, a new sample time of 30 minutes is chosen. With the new sample time, there are only 48 data points in each data sample. Figure 3-5 shows one part of the data that is before down-sampling and figure 3-6 shows the same part of data that is after down-sampling.

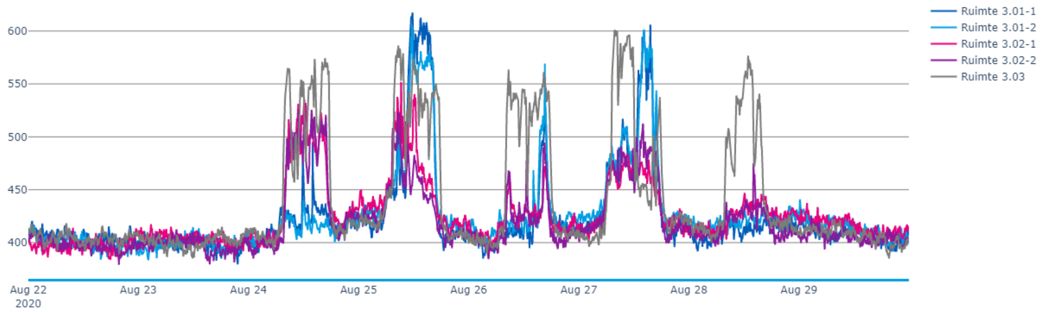

Figure 3-5: Data before down-sampling

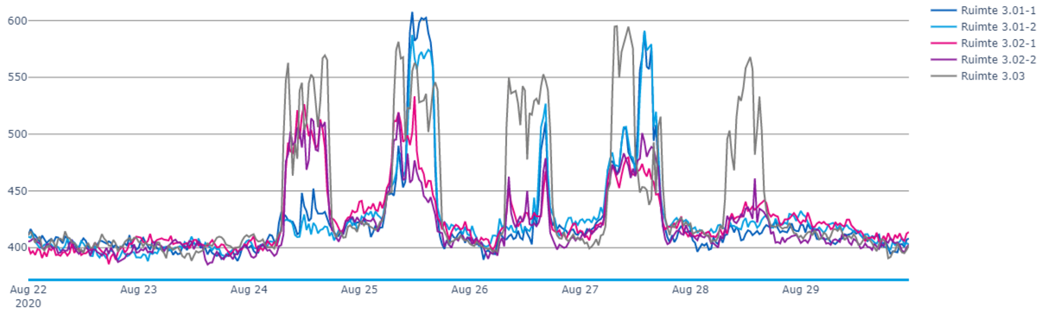

Figure 3-6: Data after down-sampling

3-2 Preparing dataset

In machine learning, each data sample has one or more labels which indicate its properties. In this project, the goal is to detect and diagnose the anomaly in the sensor signals. It is to detect and diagnose on a daily base. One data instance consists of the data from one day. In the labelling of the dataset, each instance obtains one or more labels. Assume the label for a data instant from j^{th} room on i^{th} day is denoted by l_{ij}. It is known that $i \in 1, 2, ..., 494$ and $j \in 1, 2, ..., 57$. The label of a data instant $l_{ij} \in \{0, 1, 2, ..., m\}$ where 0 is the label for normal data and m is the number of the anomaly classes in the framework. It is because this project only focuses on detecting and diagnosing two types of anomaly, $m = 2$ and $l_{ij} \in \{0, 1, 2\}$.

In machine learning, the labels of the dataset can be used for training the classifier and evaluation. However, in the practice, the signal is normally unlabelled, because labelling is heavily manual work. Creating labels for such a big dataset is unrealistic. It is more demand to pick out representative data among the big dataset. Having an efficient labelling approach and selecting a representative dataset is important.

In a building, it is natural to have more than one room. In a multi-zone space like an office, it is common that more than one sensor is mounted. Due to the airflow and diffusion, the indoor climate in different rooms has influences on each other[5].

Based on the relation, one room will be picked out and its data is labelled. This room will be referred to as the "training room". Room 3.34 contains both of the anomalies that this project wants to detect. Therefore, this room is picked as a training room. The classifier model will be built on the knowledge of one room and will be applied to other rooms. This method is referred to as transfer learning[6].

The labelling procedure is shown in the appendix A.

3-3 Feature extraction based on PCA

Feature extraction is to extract useful information from the dataset. The features of the dataset should be able to represent the information of the original dataset as much as possible. In the state of practice, the expert-defined features were part of the rule-based approaches. Utilising handcrafted features whose quality relies on the judgements of human experts. Existing methods take advantage of handcrafted features, resulting in overlooking certain cases or being inadequate. Automatic feature extraction method including PCA and its variants[7] is commonly used[8]. These methods do not require expert knowledge and try to make use of as much information as possible.

The principal component analysis[9] is one of the most common methods The principal component analysis is an unsupervised and widely used feature extraction method. The objective of this method is to project the high dimensional data to linear subspaces with lower dimensions meanwhile keep as much informativeness as possible.

As shown in figure 3-7, the two-dimensional data can have two principal components in total. The projection to the first PCA dimension has the largest variance. The PCA's second dimension is orthogonal to the first one.

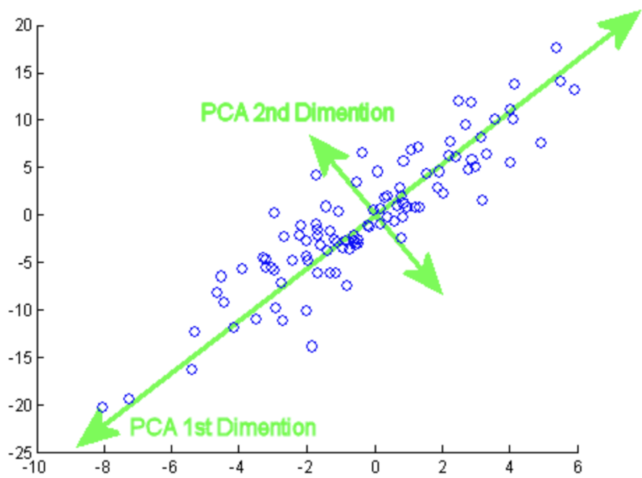

Figure 3-7: Principal component analysis visual illustration[10]

The first component is the subspace that has the lowest distance to each data point and can maximise the variance of the dataset.

3-3-1 Principal components

Consider a dataset X with the dimension of $N \times p$ where $N = 494$ is the number of days. Each day has p data samples. The dataset is

$$X = [x_1 \ x_2 \ x_3 \ \dots \ x_N]^T$$

where

$$x_i = [x_i(1) \ x_i(2) \ \dots \ x_i(p)], \quad i = 1, 2, \dots, N$$

.

The principal components are a set of vectors with p dimensions where p is equal to the dimensions of the data. The i^{th} vector is called the i^{th} principal component. The i^{th} component is orthogonal to $(i-1)^{th}$ component and while it fits in with the dataset best.

The dataset need to be standardised.

$$Y = \frac{X - u}{s} \tag{3-3}$$

where u is the mean value of the whole dataset, s is the standard deviation of the dataset.

Let Σ denote the covariance matrix of Y, they have the following relation:

$$\Sigma = Y^T Y \tag{3-4}$$

The eigenvalue decomposition result in

$$\Sigma = Q \Lambda Q^{-1} \tag{3-5}$$

where Λ is a diagonal matrix filled with the eigenvalues of Σ and Q is eigenvector matrix with each column is eigenvector corresponding each eigenvalue.

As Y is a matrix with dimension of $N \times p$, it has p principal components in total. These principal components can be calculated by multiplying the original dataset Y with eigenvector matrix Q. Let P denote the principal component matrix,

$$P = YQ \tag{3-6}$$

3-3-2 The optimal number of features

It is obvious that the principal component matrix P has the same dimension as X. PCA aims at making use of the first few components which summarise most of the original information and discard the rest of the components. The more components are preserved, the more information is kept.

The number of the feature is determined by Kaiser's rule[11]. Kaiser's rule look at the explained variance, which is the sorted eigenvalues from large to small. Larger eigenvalues means more information is loaded onto the corresponding component. It is because the eigenvalue of 1 means the component contains equivalent information of a single variable. Kaiser's rule looks at how much information each eigenvalue contains and suggests dropping the components which have eigenvalues that are less than 1.

Moreover, let ρ_i denote explained variance ratio of the eigenvalue λ_i, ρ_i can be expressed by equation:

$$\rho_i = \frac{\lambda_i}{\sum_{i=1}^{p} \lambda_i} \tag{3-7}$$

By applying eigenvalue decomposition on the training room dataset, figure 3-8 shows the explained variance and accumulation of explained variance ratio of each component of the training room:

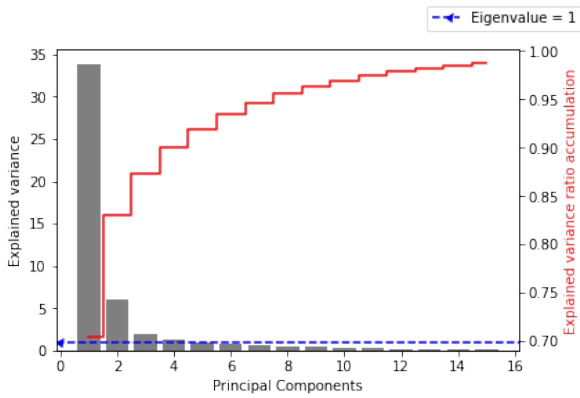

Figure 3-8: Percentage of Variance (Information) for each principal component

In conclusion, by keeping the primary four components, around 90% of the original information can be preserved.

3-4 Support vector machine

Support vector machine[9] is a linear classification method. In a dataset represented by a p-dimensional feature map, the SVM is to build up one or several hyperplanes with dimensions of $p-1$ to separate the classes. The hyperplane between two different classes has the maximum margin between the hyperplane and the two classes. A visual illustration is shown in figure 3-9. For the dataset with two classes and two-dimensional features. The optimal hyperplane that separates the two classes is a line. The closest point from both classes to the line has the possible largest distance.

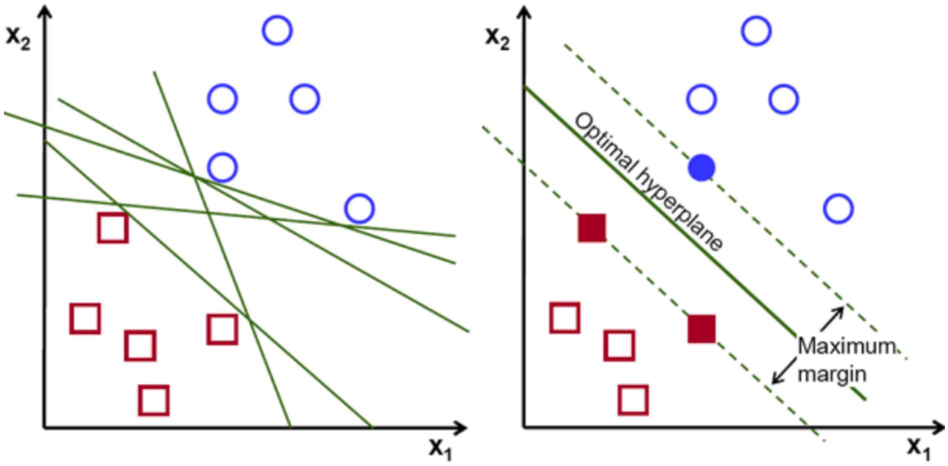

Figure 3-9: Support Vector Machine Visual Illustration[12]

The SVM classification boundary can be represented by:

$$f(x) = \mathbf{w}^T x + b \tag{3-8}$$

where x is the feature vector, $\mathbf{w}$ is the weight vector and b is the offset from origin.

And the SVM solves the following optimisation problem:

$$\begin{aligned} &\min_{\mathbf{w},b,} \frac{1}{2}\mathbf{w}^T\mathbf{w} \\ s.t. \quad & y_m(\mathbf{w}^T x_m + b) \geq 1\,,\; m = 1, \ldots, n \end{aligned} \tag{3-9}$$

where y is the label, x is the training data and n is the total number of the training sample.

However, in most cases, the classes are not linearly separable. An L2 penalty[13] with regularisation parameter C and slack variable γ. In addition, SVM allows projecting the data $x \in \mathcal{X}$ from the training set into $p+1$ dimensions using kernel trick $\phi(x)$.

$$f(x) = \langle \mathbf{w}, \phi(x) \rangle + b \tag{3-10}$$

The kernel function $K(x_p, x_q)$ is the inner product of the two feature vector's kernel trick projection:

$$K(x_p, x_q) = \phi(x_p)^T \phi(x_q)$$

In this project, the Gaussian kernel is chosen.

$$K(x_p, x_q) = \exp -\frac{\|x_p - x_q\|^2}{\sigma} \tag{3-11}$$

where default σ is the number of features.

SVM solves the following primal optimisation problem that separates the data and yields the maximum margin [14]:

$$\begin{aligned} \min_{\mathbf{w},b,\gamma} \ & \frac{1}{2}\mathbf{w}^T\mathbf{w} + C_m\sum_{m=1}^{n}\gamma_m \\ s.t. \quad & y_m(\langle \mathbf{w}, \phi(x_m)\rangle + b) \geq 1 - \gamma_m \\ & \gamma \geq 0\,,\, m = 1,\ldots,n \end{aligned} \tag{3-12}$$

The dual problem to the primal problem is:

$$\begin{gathered} \min_{\alpha} \frac{1}{2}\alpha^T Q\alpha - e^T\alpha \\ s.t. \quad \mathbf{y^T}\alpha = 0 \\ 0 \leq \alpha_m \leq C_m\,,\, m = 1,\ldots,n \end{gathered} \tag{3-13}$$

It is very common that in anomaly detection, the number of positive samples is much less compared to the negative samples. This is called an unbalanced dataset. The influence of it is that the classification boundary will tend to miss out on many positive examples. One way to avoid that is to weight the regularisation parameter C by class[15]. Let C_m denote the regularisation parameter for class m and

$$C_m = Ck_m$$

where k_m is the weight inversely proportional to the frequency of class m.

$$k_m = \frac{n}{mn_m} \tag{3-14}$$

where n_m is the total number of samples for class m

The solution of the optimisation is a decision function[15]:

$$\mathrm{sgn}(\mathbf{w}^T\phi(x) + b) = \mathrm{sgn}\left(\sum_{m=1}^{n} y_m\alpha_m K(\mathbf{x}_m, \mathbf{x}) + b\right) \tag{3-15}$$

Chapter 4

Proposed Framework

In this chapter, an overview of the proposed framework is presented. This framework takes one-day data from all the CO2 sensors as inputs. The inputs are preprocessed. The preprocessing includes data recovery, outlier removal and down-sampling. After the preprocessing, the same feature extraction model is applied to the data. Then, the classification model makes a prediction on the extracted features. The outputs of this framework are labels of the CO2 sensor data or anomaly scores. The label is the prediction of which class the data belongs to. The anomaly score is a score that describes how abnormal the signal is. The choice between the labels and anomaly scores is based on the requirement.

4-1 The logic of the framework

The aim of the framework is to give labels to the monitoring data from CO2 sensors in multi-zone space on a daily base. In this project, there are 57 CO2 sensors mounted in different rooms in the office, The logic of the framework is shown in figure 4-1:

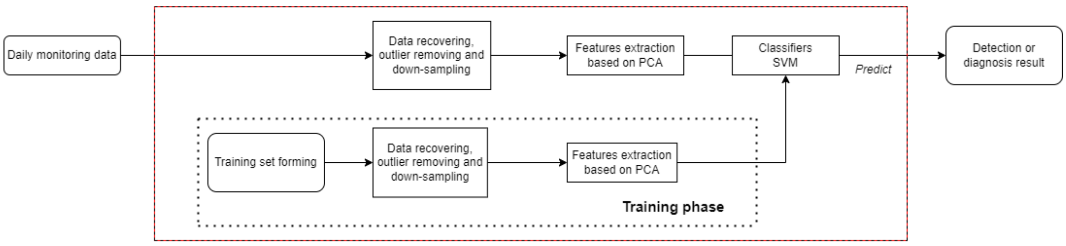

Figure 4-1: Framework

As shown in the figure 4-1, there is a training phase where historical data forms a training set,

classifiers are trained by historical data. The training phase starts with the forming of the dataset. The data that is selected to be the training data will be preprocessed. After data preprocessing, the feature extraction based on principal component analysis (PCA) is applied. The dimension of the data will be reduced to the chosen number of components. Then the features are used to train classifiers.

The outputs of the framework are labels or anomaly scores. The choice can be made depending on requirements. The labels are results from classification models, and the anomaly scores are results from regression models. As the classification method and regression method are convertible, these two results are also convertible.

In this project, the classifiers are trained for application in four scenarios. Their functions are respectively detecting bias anomaly, detecting gradual drifting anomaly, detecting mixed bias and gradual drifting anomaly and diagnosing the two anomalies. The performance of the four scenarios is evaluated in the chapter 5.

Chapter 5

Result

In this chapter, the performance of the proposed framework is evaluated. For the framework, there are four classifiers which are used in four different scenarios: detecting bias anomaly, detecting gradual drifting anomaly, detecting mixed bias and gradual drifting anomaly and diagnosing the two anomalies. The framework with the four different classifiers is evaluated separately. The goal this project aims to achieve is to tackle the challenges of breaking the limitations of rule-based approaches. The proposed method is evaluated individually and then compared to the rule-based approach. Moreover, to find out the influence of the anomaly rate on the framework's performance, a relevant experiment is conducted. Last but not least, to validate the training room selection, and to find out how sensitive the performance is to the different training rooms, the sensitivity test is done. Data from all the rooms take turns to be the training room.

Evaluation criteria are selected based on the goal of this project.

5-1 Evaluation criteria

To evaluate the performance of the detection and diagnosis, the following criteria are adopted. For evaluating the performance of detection, accuracy and precision are adopted. For evaluating the performance of the diagnosis, the accuracy and confusion matrix are adopted.

5-1-1 Accuracy

The accuracy is defined as the rate of correctly predicted samples. It can be calculated by first summing the number of true positive (TP) and the number of true negative (TN) and then divided by the total number of predictions. As shown in the equation 5-1

$$accuracy = \frac{TP + TN}{Total} \tag{5-1}$$

5-1-2 Precision

Precision stands for how many correctly detected positive samples are among all the positive detection. As one of the challenges in this project is to lower the false positive (FP) rate, this criteria is helpful to view the difference in true positive detection between the two approaches.

$$precision = \frac{TP}{TP + FP} \tag{5-2}$$

5-1-3 Confusion matrix

The confusion matrix is a table with one dimension of the actual label and another dimension of the predicted label. The confusion matrix allows visualisation of the performance of a multi-classification algorithm. Thus, the confusion matrix is suitable for evaluation in the multi-class classification.

5-2 The result of detecting the bias anomaly

The bias anomaly classifier is the classifier that gives the result of if the incoming signal contains a bias anomaly or not.

5-2-1 Bias anomaly classifier's stand-alone performance

Below show the accuracy and precision score in one plot.

Figure 5-1: Accuracy and precision of detecting the bias anomaly, with the proposed classifier

Most of the rooms have above 90% of accuracy. Except for some rooms. If look into the data of these rooms, the bias anomalies there are not very obvious. Figure 5-2 shows that four of the rooms that have low accuracy also have unapparent bias anomalies. Note that according to appendix A, all the data after March 28th contains bias anomalies.

As for precision, most of the rooms have it above 90%. And there are a few rooms where accuracy drops but precision is still high, and there are a few rooms with low precision but high accuracy.

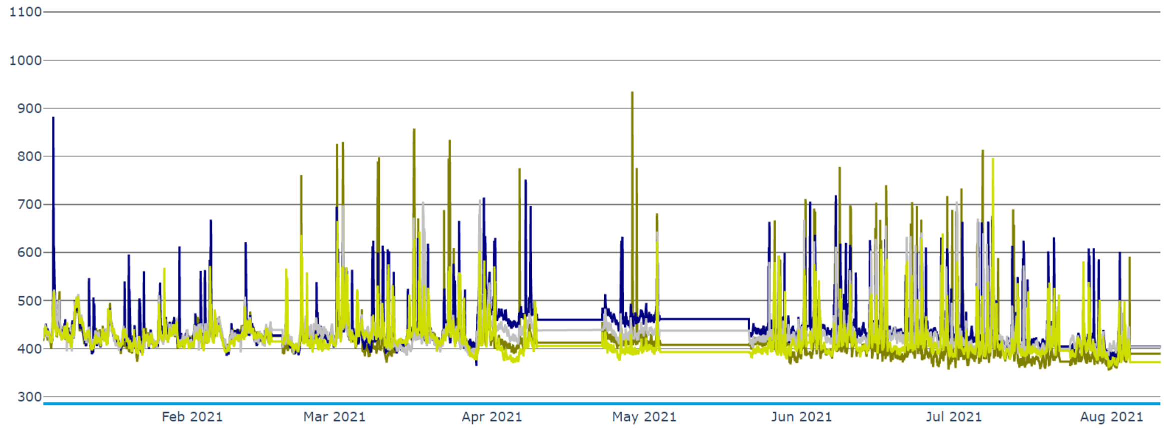

Figure 5-2: Four rooms that do not have obvious bias anomalies

5-2-2 The bias anomaly classifier compares to the rule-based approach

The rule used to detect bias anomaly is: the weekend night data of the CO2 sensors are less than 493 ppm. It is because the rule can only work on weekend data, all the weekend data are extracted from the original dataset. This way, the anomaly rate will not change. The extracted data are used to compare the performance of the proposed classifier and the rule-based approach.

Below is the accuracy of the proposed classifier and rule-based approach.

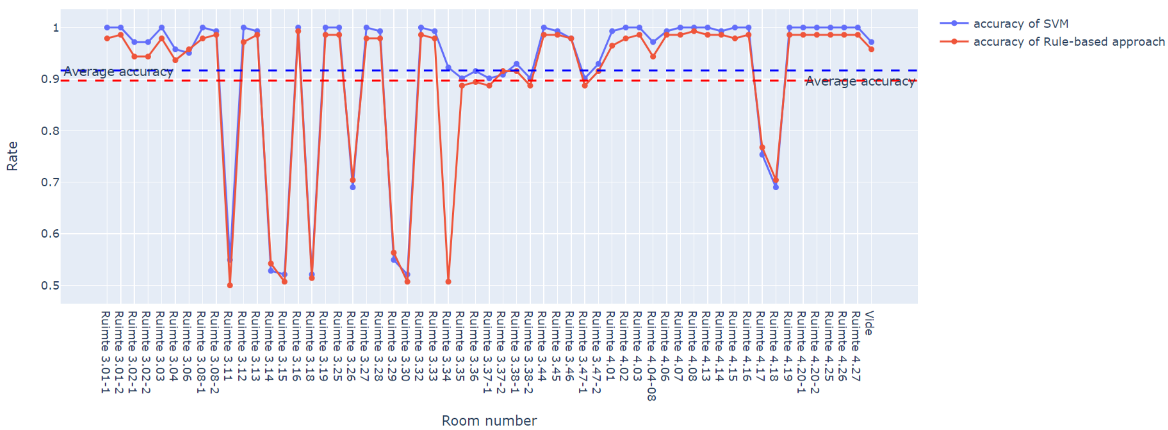

Figure 5-3: Accuracy of detecting the bias anomaly, with the proposed classifier and rule-based approach

The accuracy of the proposed classifier is aligned well with the rule-based approach and performs generally better. Room 3.34 has a negative bias; SVM can detect it, but the rule-based approach cannot. And in the previous subsection, the SVM performed badly in some rooms that did have not distinguishable bias anomalies. For the rule-based approach, these anomalies are also hard to be detected.

Below is the precision of the proposed classifier and rule-based approach:

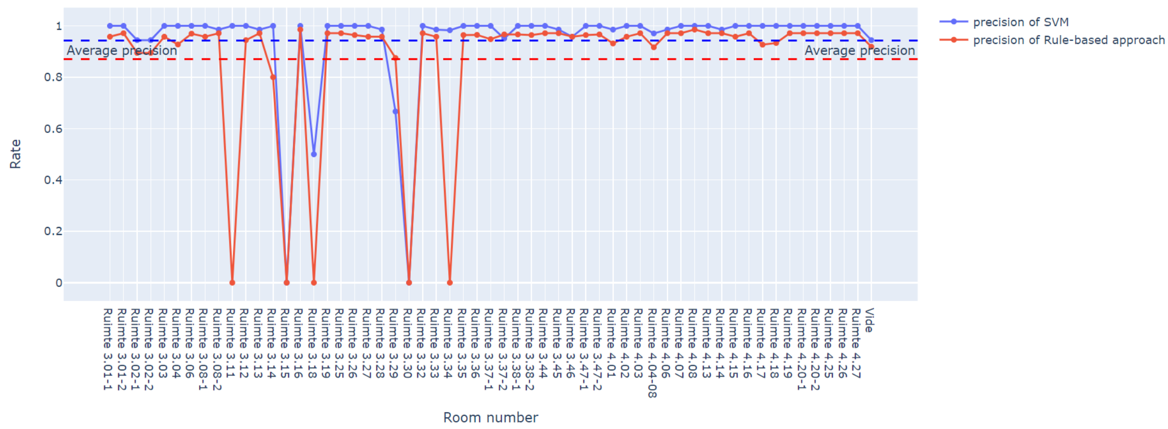

Figure 5-4: Precision of detecting the bias anomaly, with the proposed classifier and rule-based approach

SVM generally outperforms the rule-based approach. Worth noticing that there are many 100% precision for the SVM. There are a few points of the rule-based approach that has 0% precision which means that they have classified 0 true positive samples. These rooms have either an unapparent bias or a bias that is lower than normal data. With an unapparent bias, the SVM is also not able to detect well. As for the bias that is lower than normal data, SVM is performing decently but rule-based is not doing well.

5-3 The result of detecting the gradual drifting anomaly

In this section, the gradual drifting anomaly classifier is evaluated. The output of this classifier is a label that indicates if the incoming signal contains a gradual drifting anomaly or not.

Below shows the accuracy and precision of the classifier in one plot:

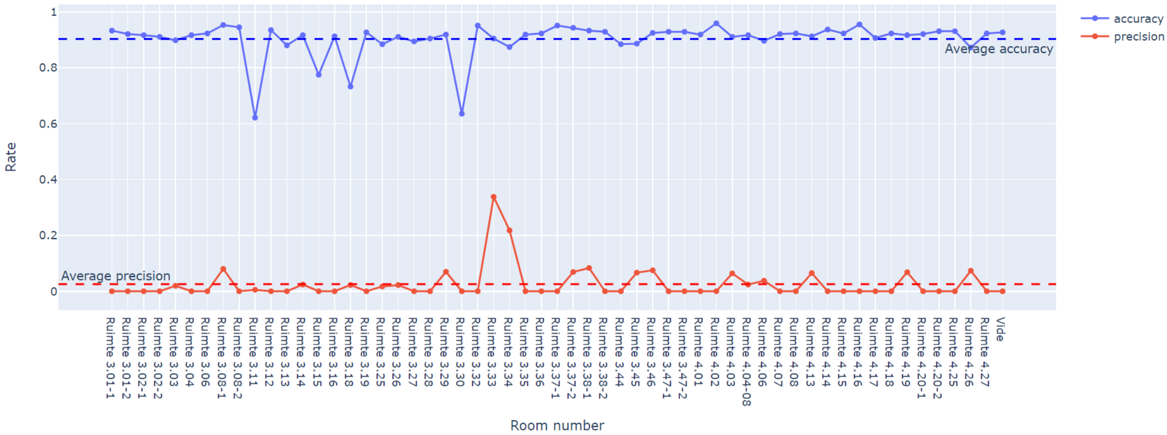

Figure 5-5: Accuracy and precision of detecting the gradual drifting anomaly, with the proposed classifier

As most of the accuracy is above 90%, except for some 4 rooms that are obviously lower than others. These four rooms during the bias anomaly period, have a lightly negative bias, which might have confused the classifier. As the precision is very low. In this scenario, there are two possible reasons. One of the reasons is that some room doesn't contain this anomaly. There are 37 rooms that do not have gradual drifting anomalies. For these rooms, since there is no positive example at all, the precision scores are 0.

For the rest of the rooms, the possible reason for the low precision is that the classifier has detected zero positive samples and possibly a high amount of false positive samples. To further analyse the low precision for these rooms, the true positive (TP), true negative (TN), false positive (FP) and false negative (FN) rates are shown below.

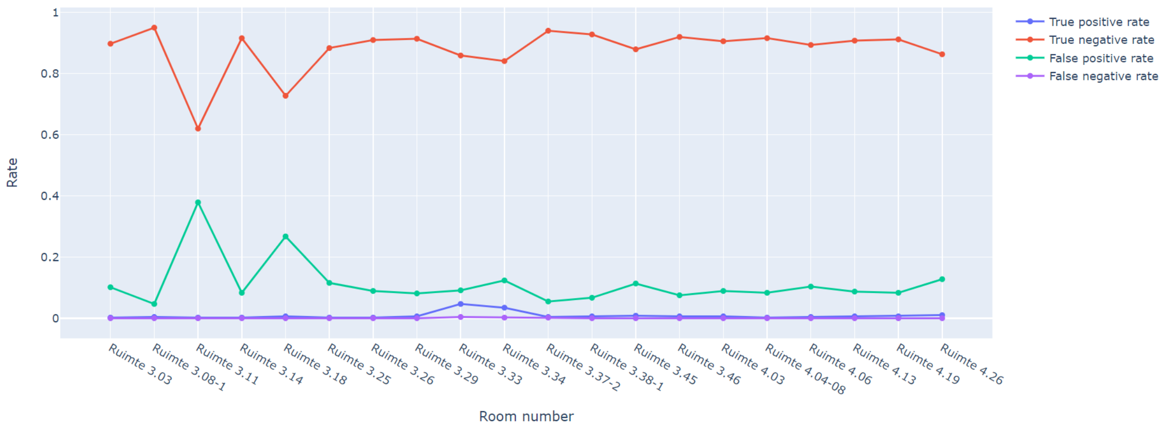

Figure 5-6: TP, TN, FP and FN of the proposed classifier detecting the gradual drifting anomaly in selected rooms

It is known that $TP + TN + FP + FN = 1$. The graph provides a better insight into the low precision rate problem. As shown in the graph, it can be concluded that there are many correctly detected negative samples. This is a good sign because this means that the false positive rate is low. These samples contribute most to the high accuracy. There are four lower peaks in the true negative line and correspondingly four upper peaks in the false positive line. This means that for these four rooms, many negative samples are detected as positive. It is explained in the previous chapter that these four rooms have bias anomalies that lower than normal data. This appearance of the anomaly influence the performance of the gradual drifting anomaly classifier. Also, these four rooms have a low accuracy because of that.

Besides, the low true positive rate and low false negative rate are apparent. This indicates that some positive examples are not detected. The reason is that the features of the gradual drifting anomaly are not distinguishable. The classifier tends to classify the data that has lower values as a gradual drifting anomaly. If looking at the feature map, the gradual drifting anomaly is not separable.

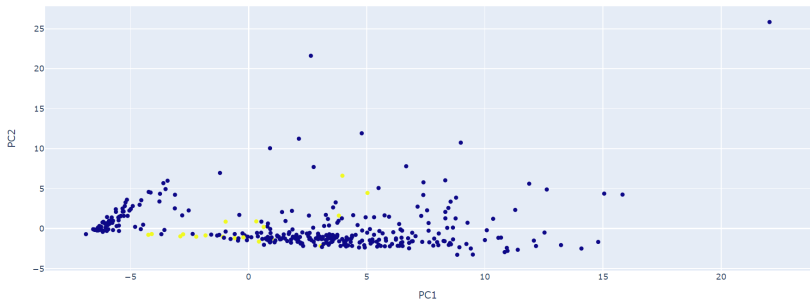

Figure 5-7: 2D feature map of normal data (blue dots) and gradual drifting anomaly data (yellow dots)

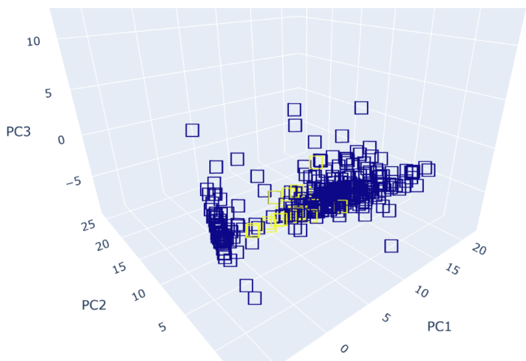

Figure 5-8: 3D feature map of normal data (blue boxes) and gradual drifting anomaly data (yellow boxes)

5-3-1 Gradual drifting anomaly classifier compares to the rule-based approach

The rule used to detect gradual drifting anomaly is that the CO_2 data cannot be lower than 390ppm. Below is the accuracy of the proposed classifier and the rule-based approach:

Figure 5-9: Accuracy of detecting the gradual drifting anomaly, with the proposed classifier and rule-based approach

In terms of the weekend data, the rule-based approach is generally better than the proposed classifier. The average accuracy of the rule-based approach is around 5% more than the proposed framework. For the proposed classifier, there were four rooms having a relatively low accuracy. This was analysed in the previous subsection. It can be seen that the rule-based approach also has trouble performing well in detecting the gradual drifting anomaly in the four rooms. It is also worth noting that for room 3.34, the rule-based approach performs badly when the proposed classifier performs decently. This is because room 3.34 had a bias anomaly that has a negative offset. The negative offset has caused the value of the CO2 density to be lower than expected. As a result, if a room has a negative offset in the signal, it would be likely to be detected as a gradual drifting anomaly.

Below is the precision of the proposed classifier and the rule-based approach.

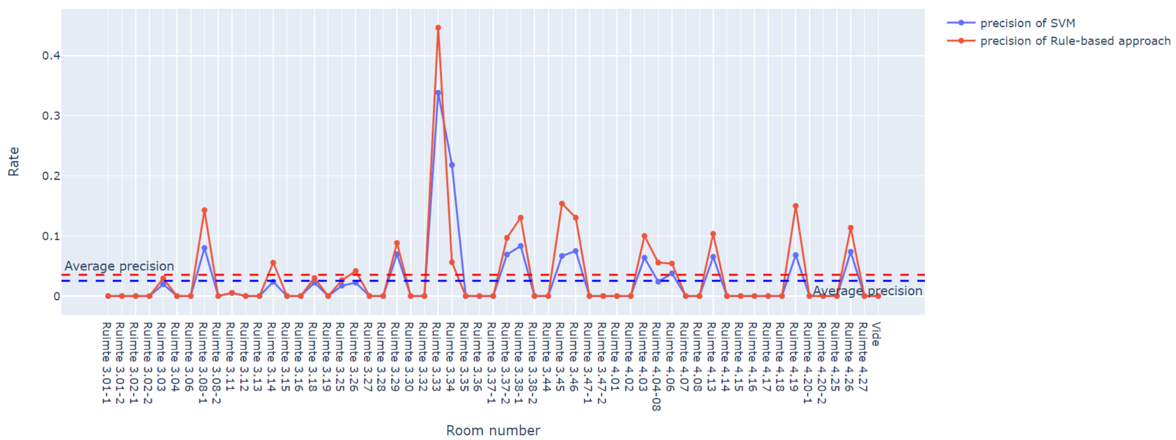

Figure 5-10: Precision of detecting the gradual drifting anomaly, with the proposed classifier and rule-based approach

The precision is generally low. The rule-based approach performs slightly better than the proposed classifier. Among all the rooms, room 3.33 has the highest precision. It is because for room 3.33, the gradual drifting anomaly leads to lower CO2 values and it becomes more obvious than in other rooms.

5-4 The result of detecting the mix of two anomalies

This classifier aim at detecting if the incoming signal contains the bias anomaly or gradual drifting anomaly. This classifier is referred as detector.

5-4-1 Detecting mixed anomalies stand-alone performance

Below shows the accuracy and precision of the classifier in one plot:

Figure 5-11: Accuracy and precision of detecting the mix of both anomaly, with the proposed classifier

The objective of the detector is to separate the normal data and non-normal data. In the dataset, both of the anomalies are included but their types are not stated. The detector can tell if the coming signals are from one of the two anomalies or not, but it cannot tell which type of anomaly they belong to.

As the graph shows, the data from most of the rooms can be correctly classified. The average accuracy is around 90.9%. Most of the rooms have more than 90% of accuracy. There are a few rooms that obtain low accuracy. It is noticeable that these rooms are the same as the rooms that obtain low accuracy in classifying bias anomalies. The trend for accuracy is similar to the trend for accuracy in detecting bias anomalies. The main reason for these rooms having low accuracy is that these rooms have non-distinguishable bias anomalies. It is because the bias anomaly takes up 40% of the dataset. If the detector is not able to detect bias anomalies very well, the overall performance would not be good.

The average precision score is around 94.8%. The data from most of the rooms have precision scores above 90%. Rooms 3.15 and 3.18 have low precision of 60% and 50% respectively. Rooms 3.14 and 3.30 have a precision of 81.7% and 78.7% respectively.

5-4-2 The detector compares to the rule-based approach

The rule that acts as a detector is that on the weekend night the difference from the mean value is no more than 10%. As the rule can only work on weekend data, the weekend data are extracted to test the performance of the proposed classifier and the rule-based approach. But for training the detector, the data from both weekdays and weekends are used.

Below is the accuracy of the proposed classifier and the rule-based approach.

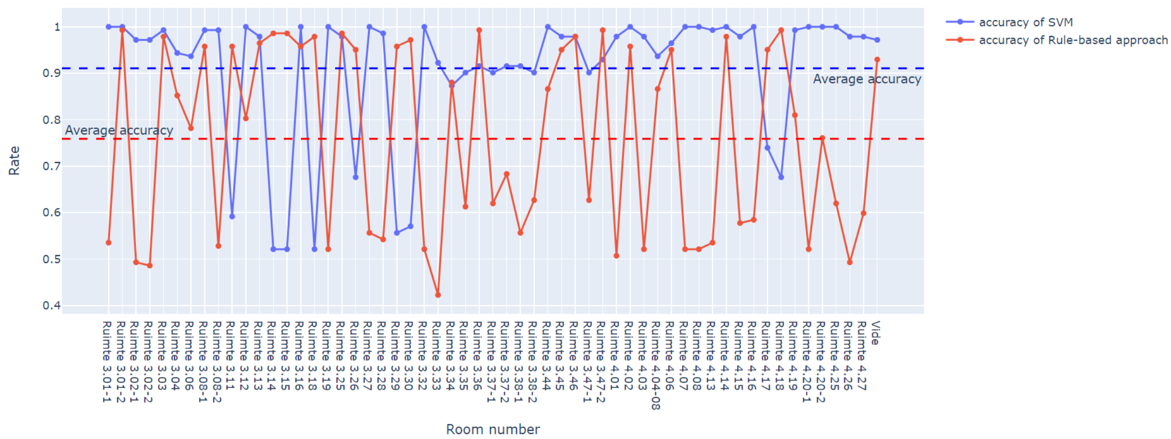

Figure 5-12: Accuracy of detecting the mixed anomalies, with the proposed classifier and rule-based approach

Most of the time, the rule-based does not perform as well as SVM. The rule-based approach's performance is unstable as it fluctuates violently. However, for those rooms that have low accuracy with the proposed detector, the rule-based approach performs very well on them. The rule-based approach relies on the average, but during the period where bias anomaly happened, because most of the rooms have high offset, the average of all the signals is already high. Based on the rules, some rooms that have a bias anomaly that is close to the average values would not be detected. But the rooms that have very low offset, even though the proposed detector cannot detect them, since their CO2 value is too far from the average and can be detected.

Below is the precision of the proposed classifier and the rule-based approach.

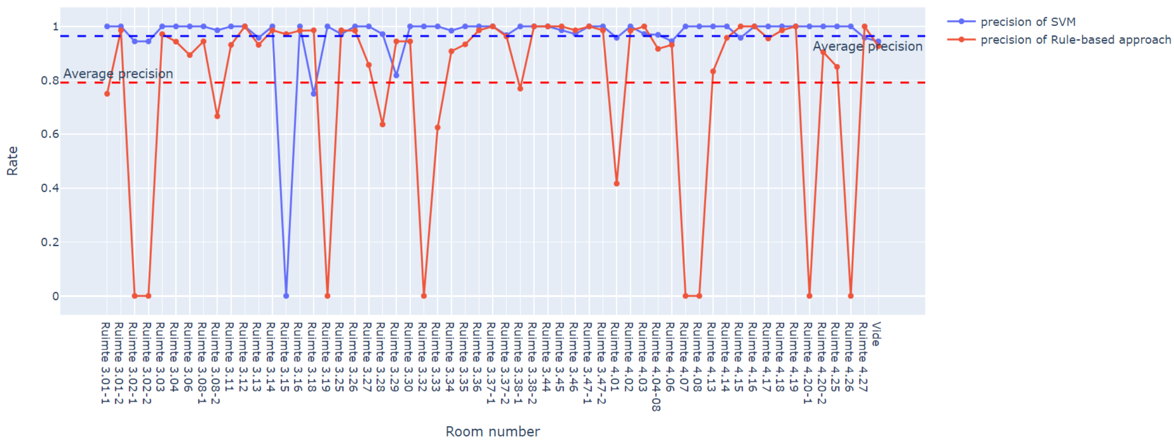

Figure 5-13: Precision of detecting the mixed anomalies, with the proposed classifier and rule-based approach

In the SVM classifier, there is one room that has 0 precision. This is because there is no positive example that is detected as true. By comparing the figure 5-12 for the rule-based classifier, the room that has high accuracy would also result in a high precision score.

There are 8 rooms that have precision scores of 0. By looking at the original data, one of the common features of these rooms is that they have a very high offset.

5-5 The result of diagnosing the two anomalies

In the diagnosing framework, the classifier gives the result of if the incoming signal is one of the two anomalies or not, and it tells which anomaly it belongs to.

There is no rule designed to diagnose these two anomalies. Thus, only the results of evaluating the proposed classifier will be shown.

Below is the overall accuracy of the diagnosis which is calculated by summing up all the correctly classified examples and dividing by the total number of examples.

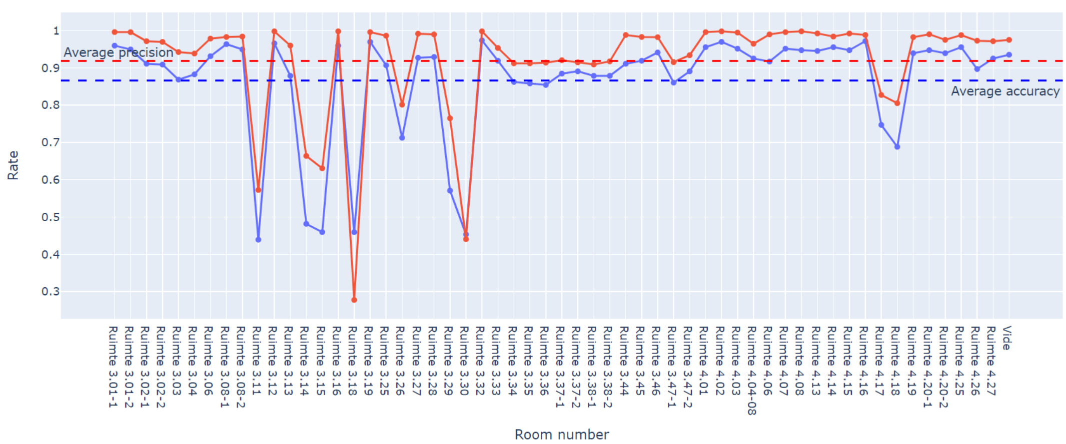

Figure 5-14: Accuracy and precision of diagnosing the two anomalies, with the proposed classifier

The accuracy and precision score have similar trends. There are rooms 3.11 3.14 3.15 3.18 3.29 3.30 4.17 and 4.18 have relatively lower accuracy and precision. Other rooms have accuracy above average and precision scores above average. The components in the confusion matrix are used to further analyse the performance of classifying different classes.

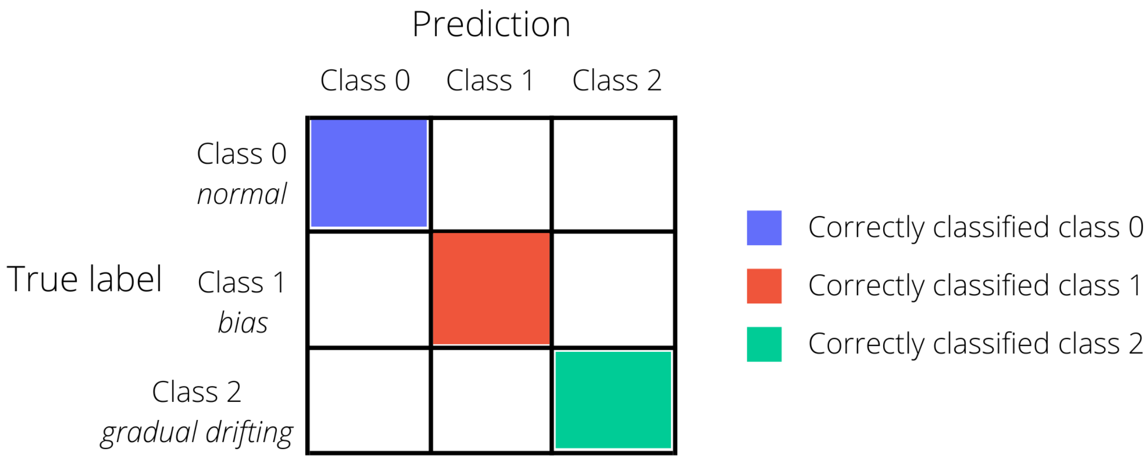

Figure 5-15: Confusion matrix with correct detection

As known, the accuracy of a classifier is calculated by the sum of the correctly classified samples and divided by the total number of examples. If illustrated using a confusion matrix as shown in the figure 5-15, the accuracy is calculated by summing the number in the coloured block and being divided by the total number of samples.

But accuracy does not show how many of the examples from each class are correctly detected. Therefore, to show how many examples from each class that is being detected correctly, a more specific matrix is used.

A confusion matrix is shown in the figure 5-15, the three coloured blocks respectively stand for how much normal data is detected as normal, how many bias anomalies are detected as bias anomalies and how many gradual drifting anomalies are detected as gradual drifting anomalies. The values in the coloured blocks divided by the total number of samples show the contribution each class gives to the accuracy.

Below is the contribution each class gives to the overall accuracy:

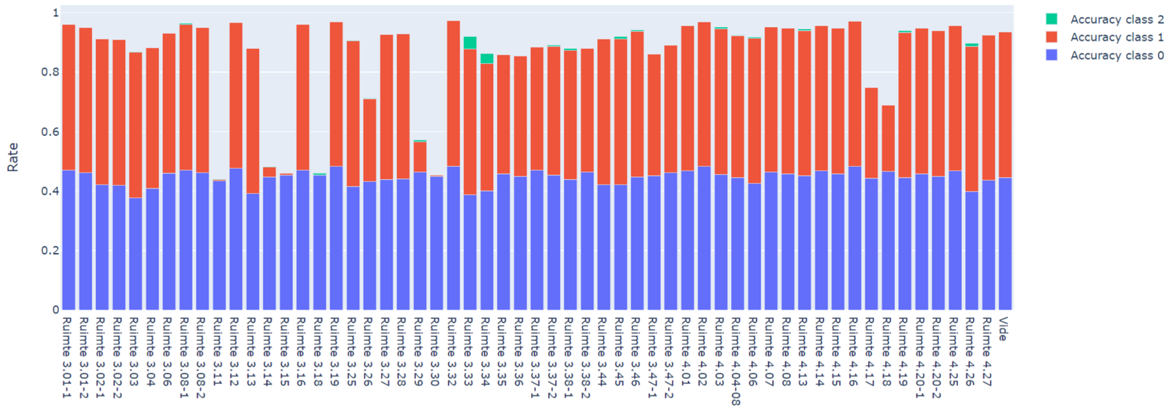

Figure 5-16: Individual class's contribution to the overall accuracy

It can be read from the graph, that every room has similar accuracy in class 0. This means that for all the rooms, the diagnosing classifier is able to detect normal data correctly. For the 6 rooms that have low accuracy, it can be told by the figure 5-16 that the low accuracy is caused by the classifier not being able to detect bias anomaly (class 1). When the accuracy goes down, it is because of classifier cannot detect class 1 but not because of class 0.

However, the figure 5-16 is not able to show how many examples are detected wrongly. It is also not able to show the performance in diagnosing class 2 very well. To also show the proportion of wrongly classified samples, three more components from the confusion matrix are added.

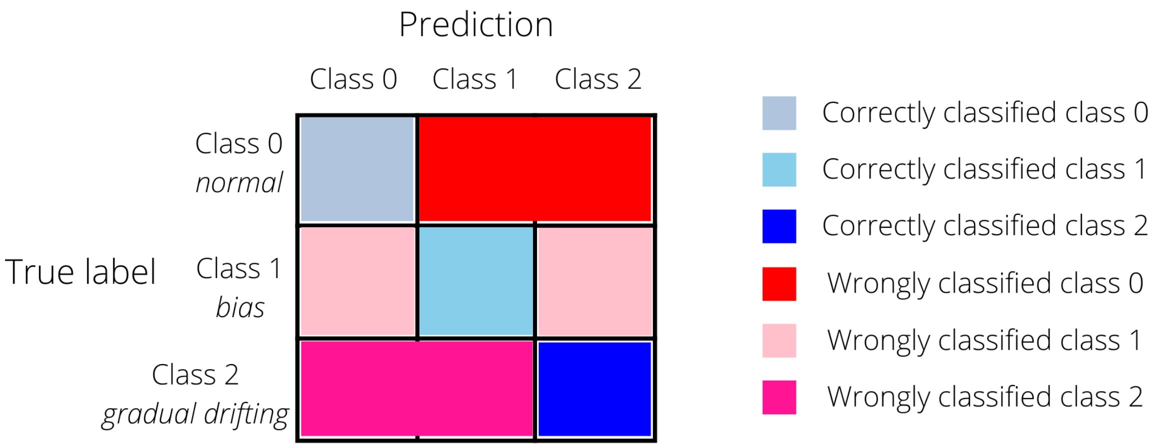

Figure 5-17: Confusion matrix with correct and wrong detection

As shown in the figure 5-17, the correctly and wrongly classified classes are represented by different colour blocks. The six indices sum up to 1.

Below is the contribution each class gives to the overall accuracy and the contribution each class gives to the overall rate of misclassification:

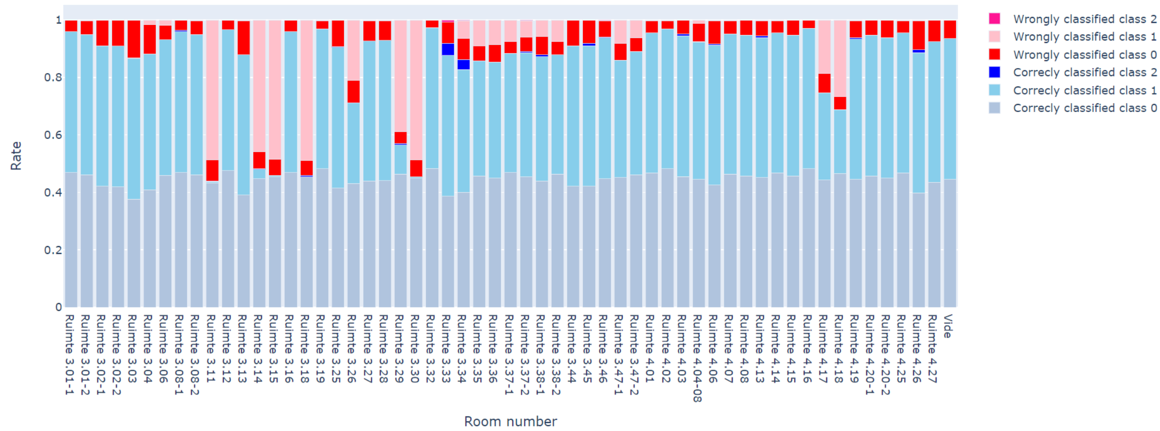

Figure 5-18: Accuracy and misclassification of each class, with the proposed classifier

The diagnosis classifier has stable performance in detecting most of the class 0 correctly. Figure 5-18 proves that for rooms that have low accuracy, wrongly classified class 1 is the main reason for the accuracy drop. For the well-performed room whose accuracy is above 80%, the drops are because of wrongly classifier class 0.

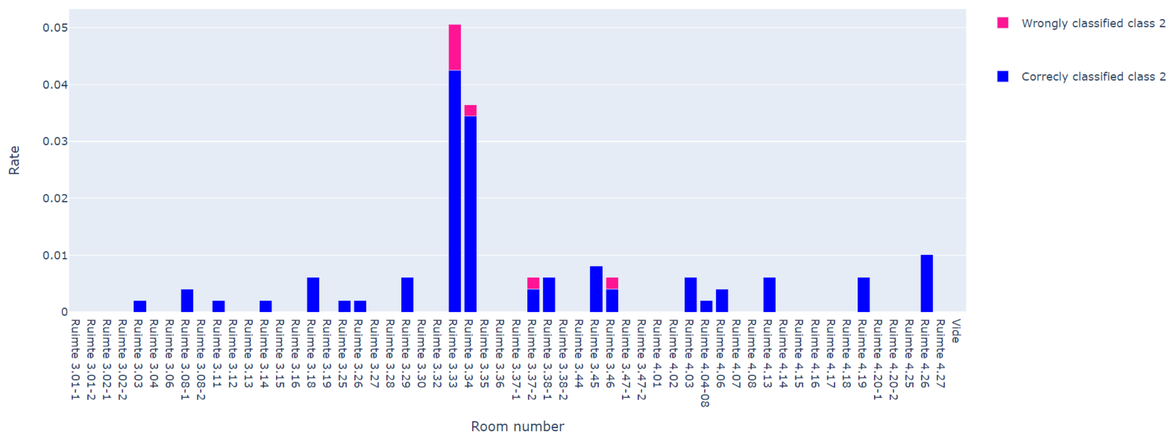

Figure 5-19: The proportion of correctly and wrongly classified class 2 examples

Figure 5-19 only shows the correctly classified gradual drifting anomaly (class 2). As aforementioned, there are only 20 rooms that have gradual drifting anomalies. The diagnosing classifier is able to predict most of the class 2 samples correctly. For most of the room, there is no wrongly predicted class 2 sample. There are only 4 rooms that have data from class 2 which is incorrectly predicted. For them, the wrong prediction of class 2 takes less than 20%.

5-6 The influence of the anomaly rate on performance

Sometimes, the number of anomalies is very low, because the occurrence of anomalies is rare. This section is designed to find out how sensitive the proposed classifier reacts to the ratio of the anomaly.

The bias classifier is used in this evaluation because there is a sufficient amount of bias anomaly, the anomaly rate can vary from a small amount to the amount of the normal data. Room 3.34 was chosen to be the training room and room 3.33 was chosen to be the test room.

Based on the appendix A, there are 494 days of data out of which 242 samples are positive. To begin with, 20 positive samples are randomly drawn from the positive sample set. In every iteration, the number of positive examples is increased by 10, until it is 240. Same as the other evaluation, accuracy and precision are chosen to evaluate the performance.

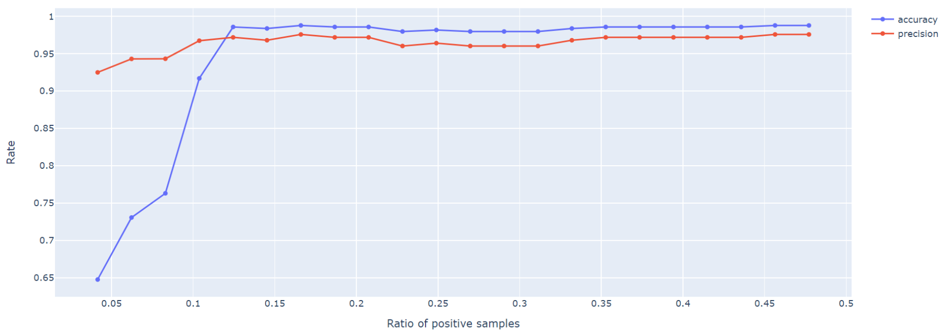

Figure 5-20: Accuracy and precision of detecting bias anomaly, with the changing anomaly rate

The graph shows that the accuracy was only 65% present when there were only around 4% of the anomaly in the dataset. However, the accuracy rises quickly to 98.5% when the anomaly rate only increases to 12%. The accuracy remains around 98% when the anomaly rate keeps increasing. A similar trend applies to precision. The precision is 92.5% when the anomaly rate is only around 4%. But it slightly increases to 97.1% when the anomaly rate is around 4%. Then it hovers around 97.1% no matter how much the anomaly rate ascent.

5-7 Validating room selection

This is to find out the impact of training classifiers using different training rooms. In the previous section, the classifier is trained using the data of room 3.34. In this section, all the rooms take turns to be the training rooms. The classifier trained by the training room will be tested on the data from all rooms. For one training room, there are 57 test rooms. As a result, there are 57 accuracy and precision scores. So, in this section, the x-axis is the training room number, and the y-axis is the average accuracy and precision of all test rooms.

5-7-1 Validating room selection for bias classifier

Figure 5-21: Average accuracy and precision of bias classifier, based on the varying training room

There are a few low points in the graph. The low points are the low accuracy and precision scores from 5 rooms. The training rooms that do not perform well are those with non-distinguishable bias anomalies. The precision score and the accuracy have similar trends. Despite the room that has low accuracy and precision score, other rooms have an accuracy of around 87% and a precision of 89.1%. This confirms transfer learning approach works. With classifiers trained by datasets from different rooms, the performance is still similar. The only requirement is that the training room needs to have distinguishable bias anomalies.

5-7-2 Validating room selection for gradual drifting classifier

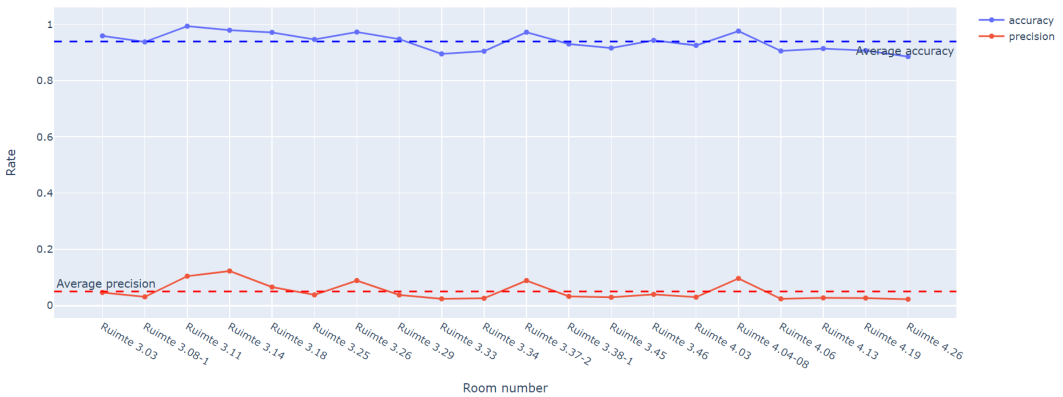

Figure 5-22: Average accuracy and precision of gradual drifting classifier, based on the varying training room

For the gradual drifting classifier, the performance of the classifier is not sensitive to the change in the training room. For accuracy, they are around 93.9%. For precision, they are around 4.9%.

5-7-3 Validating room selection for detector

Figure 5-23: Average accuracy and precision of detector, based on the varying training room

The trend of the detector is similar to the bias classifier. There are four training rooms having relatively low accuracy and precision. As known, these four rooms have non-distinguishable bias anomalies. They perform badly in being a training room for the detector. Except for the four rooms, the other rooms have an accuracy of around 86.4% and a precision score of around 88.7%.

5-7-4 Validating room selection for diagnosis classifier

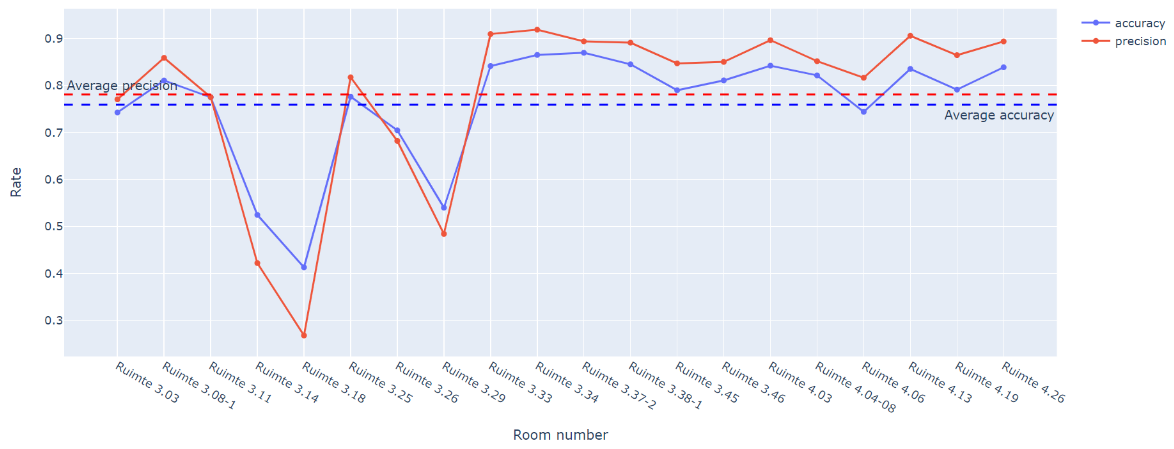

Figure 5-24: Average accuracy and precision of diagnosis classifier, based on the varying training room

Rooms 3.14, 3.18 and 3.29 have been mentioned in subsection 5-2-1, and they do have not distinguishable bias anomalies. Because of that, in subsection 5-2-1, they have bad performance in being a test room. This is the reason that they also perform badly when they are the training rooms. Despite the three rooms, other rooms perform stably. They have an accuracy of around 83% and a precision score of around 89%.

5-8 Conclusion

The framework is evaluated. Accuracy, precision and confusion matrix are the metrics to measure the performance. The bias anomaly classifier outperforms the rule-based approach. Both accuracy and precision are above 90%. There are rooms that have a not obvious anomaly. Their data is very similar to the normal data. For these rooms, the classifier and the rule-based method are not able to detect the bias anomaly. There are rooms that have bias anomalies with a negative offset. The proposed method is able to detect them while the rule-based method is not. For the gradual drifting classifier, the accuracies of most of the rooms are above 90%. However, the precision is surprisingly low, with an average of only 2.38%. One reason for low precision is that some rooms do not have a gradual drifting anomaly, the precision scores are 0 for those rooms. Other rooms have a high true negative rate but a low true positive rate. By looking at the feature maps of both 2D and 3D, it can be told that the PCA is not very efficient in creating separable features for this type of anomaly. For detecting this anomaly, the rule-based approach outperforms the proposed method. The rule-based approach results in a higher average accuracy and slightly higher average precision score. But the rule-based approach takes bias anomaly with negative offset as a gradual drifting anomaly while the proposed method does not. As for the detector, the average accuracy is 91.1% and the average precision is 94.7%. It is because there are 47% of the data samples are bias anomalies, and the performance of the detector is relatable to the ability to detect bias anomalies. The rooms that include non-distinguishable bias anomalies have low accuracy and precision. The rule-based approach performs unstably. The accuracy and precision are floating violently. So the rule-based approach is not reliable for detecting a mix of more than one anomaly. Finally, the diagnosis classifier has an average accuracy of 86.5% and an average precision score of 92%. By looking into the confusion matrix, it can be concluded that most of the normal samples and gradual drifting anomalies can be classified correctly. For the rooms that have non-distinguishable bias anomalies, the bias anomalies are hard to be detected.

Experiment on how the anomaly rate influence performance was conducted. The bias anomaly classifier is chosen to be the classifier in this experiment. It turns out that when the anomaly rate is 4% or higher, the data can train a decent classifier.

To find out the influence of the change of training room, validation of the room selection is done. In conclusion, only the rooms that have unnoticeable bias anomaly that is bad at being the training room. For example, for the rooms that have non-distinguishable bias anomaly, the bias classifiers, detector and diagnosis classifiers trained by them always have lower accuracy and precision scores. For the rest of the rooms, the change of the training room does not have much effect on the performance, because the accuracy and precision only change slightly when the training room changes.

Chapter 6

Conclusions and future work

6-1 Conclusions

This project aims to develop and test an anomaly detection and diagnosis framework for the sensor network in smart buildings. In a multi-zone space like an office, a big amount of sensors are mounted to monitor each subspace. This increases the demand for the sensors.

This project concludes the challenges of anomaly detection in a complex multi-sensor network faced by the state of the practice rule-based approach. To break the limitation, a data-driven approach based on machine learning is proposed.

The framework contains a training stage. In the training stage, a classifier is trained to give labels to the daily data of a CO2 sensor. The raw data is noisy and high-dimension. In order to smooth the signal and reduce the dimension, the data is preprocessed. An automatic feature extraction method, principal component analysis (PCA), is used for feature extraction. support vector machine (SVM) is the chosen classification method.

As in the multi-space, the airflow and temperature diffusion, the data collected by the sensors in the same space have relations. One room is chosen to be labelled and be the training dataset. The data from the rest of the rooms are used for testing. This approach is referred to as transfer learning. Based on the requirement, four classifiers are trained. They are specially used for detecting bias anomalies, detecting gradual drifting anomalies, detecting mixed of the two anomalies and diagnosing the two anomalies.

Concluding from the result, there are limitations to the proposed approach. First of all, the framework is not able to detect anomalies that are not obvious. For example, the non-distinguishable bias anomaly and some parts of the gradual drifting anomaly look like normal data. These anomalies are not able to be detected. Secondly, there are some daily data that contains both anomalies. The current diagnosis classifier can only give labels for one type of anomaly. Last but not least, there are new and unusual patterns appearing in the data sample. The features of these data patterns do not belong to the normal data, but also do not belong to the existing two anomaly types. These patterns are called novelty. The proposed method can only classify daily data as normal or one of the two anomalies while novelty clearly does not belong to either of these classes.

6-2 Future work

Based on the limitation of the proposed approach, the following points of potential work are suggested.

- Consider the relation between rooms. In this project, transfer learning is used. But the relation between rooms is considered ideally correlated. In other words, the data from different rooms locates in the same range in the feature domain. However, it is known that the neighbour sensors tend to have stronger relations and the distant sensors may have weaker relations. A relation matrix for the rooms can be developed to represent the relation between rooms. This will help transfer learning. This way the insignificant bias and gradual drifting anomalies can be detected.
- Consider detecting a combination of two anomalies. In the bias period, there are some gradual drifting and miss anomaly mixed with it. If the classifier can tell if daily data contains both anomalies, it would be more useful.
- Add novelty detection to the framework. A novelty does not belong to a normal class or any of the two anomaly classes. The proposed method can only give labels of normal data, bias anomaly and gradual drifting anomaly. By adding a novelty detector to the framework, the anomaly that has not occurred before can be correctly classified as a novelty instead of being classified into the existing classes.

Appendix A

Labelling Methods

The goal of the thesis is to create a method that minimises the amount of label that needs to be done. The raw data is unlabelled. To evaluate the performance of the method, a labelled dataset is still required. In this appendix, the process of creating labels for the dataset is shown. The anomalies are labelled based on visual inspection.

A-1 Labelling the bias anomaly

A known sensor network error has led to the offset in every CO2 sensor. Some sensors have high offsets, some sensors have low offsets and some even have negative offsets that cause the CO2 values to be lower than their normal values. This anomaly happens in a known period. As shown in the figure A-1, from March 28th, 2021 till December 10th 2021, there is an offset in every CO2 sensor.

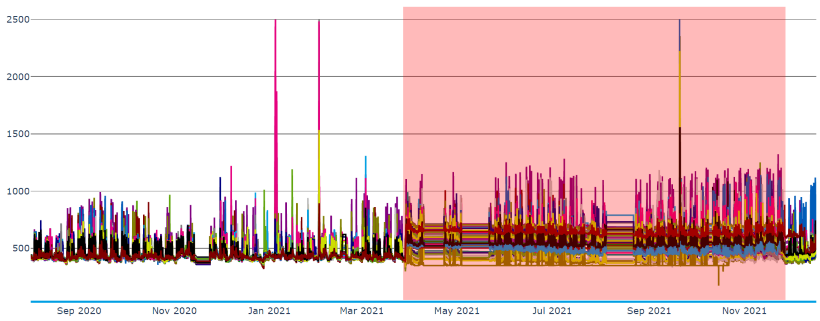

Figure A-1: Bias anomaly period

To create a label for the bias anomaly, the day that is in the range of the period, all 57 rooms are marked with "1". Otherwise, the data is marked as "0". In this case, the "1" and "0" are categorical. The label "1" stands for the daily data containing bias anomaly and the label "0" means that the daily data does not contain bias anomaly.

A-2 Labelling the gradual drifting anomaly

Gradual drifting anomaly is caused by sensor degeneration. All of the sensors had a fixed period of calibration. The calibration for every sensor happens on weekends. This means that this anomaly can only occur on the weekends. Therefore, a plot with the CO2 sensor from all the rooms is first used to mark the period that potentially contains this anomaly. One of the consequences of this anomaly is a deficient CO2 value. Keeping the features of this anomaly in mind and knowing that the density of CO2 in the outdoor environment is around 400 – 500 ppm, the labelling process begins by following the below steps.

Step 1: By looking at the data before the offset happened, we can notice samples have low values in some periods that deviate from the normal range. Based on the visual inspection of two people and the validation of an expert, 16 weekends are marked that contain these outliers.

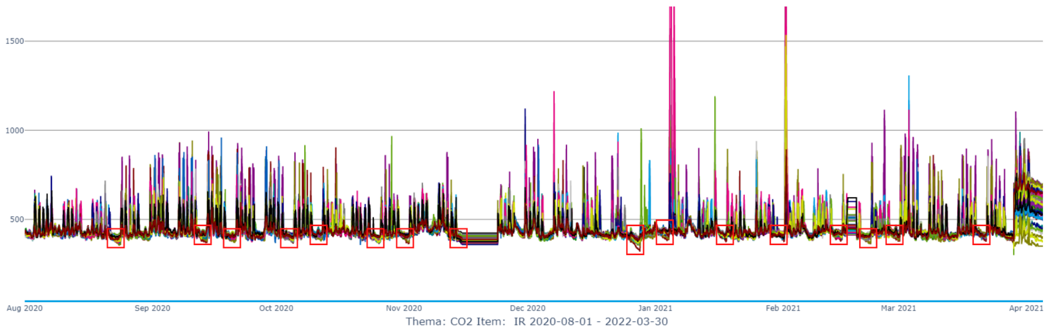

Figure A-2: Marking the weekends that have obvious outliers

Step 2: By zooming in at each marked weekend, the room that has the data that deviates from the majority, is considered a gradual drifting anomaly. As shown in the figure A-3, take the day off from December 26th to 28th as an example, there are two rooms that have obvious outliers. Thus, these two rooms on the three days are marked with the label "1". This process repeats for each weekend marked with a red box in step 1. In the end, the data instances that are not marked with "1" is the normal data. They are marked with "0".

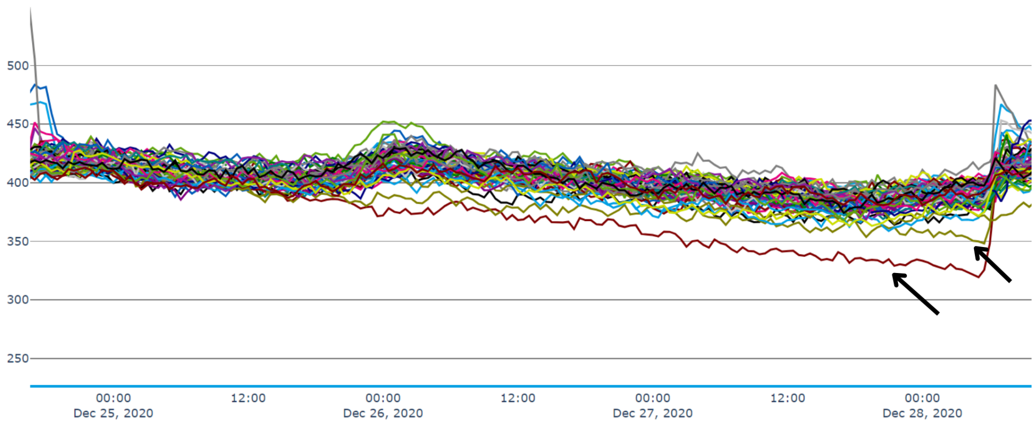

Figure A-3: An example of marking gradual drifting anomaly

A-3 Labels for detection dataset

The dataset for detection is the combination of the dataset of bias anomaly and gradual drifting anomaly. In the detection dataset, one data sample is marked with "1" if it is marked "1" in either the bias anomaly dataset or the gradual drifting anomaly dataset. The rest of the data is normal data, thus is marked with "0".

A-4 Labels for diagnosis dataset

The dataset for diagnosis is similar to the dataset for detection. Except for that, "1" is marked for the data that has a bias anomaly and "2" is marked for the data that has a gradual drifting anomaly.

Bibliography

[1] Deloitte, *Energy Transition Trends Report 2022*. Accessed 2022-7-22. [Online]. Available: https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/energy-resources/deloitte-uk-global-energy-transition-trends-2022.pdf.

[2] *2021 global status report for buildings and construction*, Oct. 2021. [Online]. Available: https://globalabc.org/resources/publications/2021-global-status-report-buildings-and-construction.

[3] S. Wang and F. Xiao, "Ahu sensor fault diagnosis using principal component analysis method," *Energy and Buildings*, vol. 36, no. 2, pp. 147–160, 2004.

[4] F. van Wyk, Y. Wang, A. Khojandi, and N. Masoud, "Real-time sensor anomaly detection and identification in automated vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 1264–1276, 2020. DOI: 10.1109/TITS.2019.2906038.

[5] L. Wang, Y. Wang, F. Wang, and H. Wang, "Quantitative assessment of indoor co2 concentration of a comprehensive office building," in *E3S Web of Conferences*, EDP Sciences, vol. 356, 2022, p. 05 043.

[6] F. Zhuang, Z. Qi, K. Duan, *et al.*, "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2020.

[7] J. Lu, Z. Lai, H. Wang, Y. Chen, J. Zhou, and L. Shen, "Generalized embedding regression: A framework for supervised feature extraction," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 1, pp. 185–199, 2022. DOI: 10.1109/TNNLS.2020.3027602.

[8] M. S. Mirnaghi and F. Haghighat, "Fault detection and diagnosis of large-scale hvac systems in buildings using data-driven methods: A comprehensive review," *Energy and Buildings*, vol. 229, p. 110 492, 2020.

[9] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[10] P. Majumder, *Principal component analysis and its implementions*, Analytics Vidhya, May 2022. [Online]. Available: https://www.analyticsvidhya.com/blog/2022/04/principal-component-analysis-its-implementions/.

[11] H. F. Kaiser, "The application of electronic computers to factor analysis," *Educational and psychological measurement*, vol. 20, no. 1, pp. 141–151, 1960.

[12] R. Lakshman Naika, R. Dinesh, and S. Prabhanjan, "Handwritten electric circuit diagram recognition: An approach based on finite state machine," *Int J Mach Learn Comput*, vol. 9, pp. 374–380, 2019.

[13] A. Y. Ng, "Feature selection, l 1 vs. l 2 regularization, and rotational invariance," in *Proceedings of the twenty-first international conference on Machine learning*, 2004, p. 78.

[14] J. Kremer, K. Steenstrup Pedersen, and C. Igel, "Active learning with support vector machines," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 4, no. 4, pp. 313–326, 2014.

[15] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," *ACM transactions on intelligent systems and technology (TIST)*, vol. 2, no. 3, pp. 1–27, 2011.

Glossary

List of Acronyms

HVAC	heating, ventilation and air conditioning
DWA	Adviesbureau DWA
BMS	building management system
PCA	principal component analysis
SVM	support vector machine
TP	true positive
TN	true negative
FP	false positive
FN	false negative