Identify Significant Alarms from BAS Alarms Packages and Evaluate Their Impacts

Han Li
Master's Thesis
School of Architecture
Carnegie Mellon University
Pittsburgh, Pennsylvania
May 12, 2017

Advisors:
Azizan Aziz
Erica Cochran Hameen
Bertrand Lasternas

Author Note
Submitted in partial fulfillment of the requirements for the degree of Master of Science in Building Performance and Diagnostics

Copyright Declaration

I hereby declare that I am the sole author of this thesis.

I authorize Carnegie Mellon University, Pittsburgh, Pennsylvania to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I authorize Carnegie Mellon University, Pittsburgh, Pennsylvania to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Han Li

Acknowledgement

I would like to extend my heartfelt appreciation to the people who helped me to bring this study into reality.

Firstly, I would like to express my deep gratitude to Asst. Prof. Azizan Aziz, Asst. Prof. Erica Cochran Hameen, and Bertrand Lasternas, my advisors for their patient guidance, and warm support, and constructive criticisms in my thesis.

I would also like to thank Prof. Volker Hartkopf, Prof. Vivian Loftness, Prof. Khee Poh Lam, and Asst. Prof. Ömer Karaguzel for their great inspiration and guidance in my past two years' study at CMU. The things I learned become treasures for my future study and career.

I sincerely thank Asst. Prof. Pine Liu from Civil and Environmental Engineering of CMU, and Mr. Paul Sedlock from Siemens Industry Inc., for giving us valuable insights in the interviews.

I also want to thank my classmates and friends—Aakash Goliyan, Chen Zhou, Nilesh Bansal, Kevin Boyd, Siliang Lu, Xiaowen Gu, Yuqi Pan, Yuzhe Liu, Chenlu Zhang, and Zhiang Zhang for their support and encouragement in the unforgettable two years.

I would like to thank Jing Yang, for her love and support from overseas. Because of the time difference between the U.S. and China, my life is always filled with sunshine.

Last but not the least, I want to thank my parents Taiju Xiao, and Ronghai Li, for helping me to develop an ethic and allowing me to realize my potential. And of course, their unconditional love and support.

Abstract

Building systems, including Heating, Ventilation and Air-conditioning (HVAC) systems, lighting systems, and security systems are key components of modern buildings. These systems are designed to deliver an ideal built environment, ensure the quality of building service and safety. Building Automation System (BAS) is a distributed control system that helps to monitor and regulate building systems. The implementation of BAS has facilitated the building Operation and Maintenance (O&M), allowing building operators to better monitor and maintain building functions. One of the important abilities of modern BAS is raising alarms when building systems behave differently from design values. However, a BAS usually generates an overwhelming amount of alarms every day. The lack of actionable information from those alarms makes it very challenging for building operators to make corresponding O&M decisions.

The intent of this study is to find the reasons of the inefficiency of BAS alarm functions and to propose a solution which helps building operators make better O&M decisions based on the BAS alarms. This study analyzed a BAS in a university complex. First, the building's HVAC systems were investigated. Second, interviews with facility managers and BAS field engineers were conducted to identify the existing deficiencies of BAS and future user expectations. Third, a data mining framework was developed to optimize current BAS alarm management function. The goal of this framework is to help filter out trivial alarms, categorize alarms by their impact categories (e.g. equipment operations, occupant comfort, critical operations), and prioritize the alarms based on their quantitative impacts.

The data mining framework is implemented in the Gates-Hillman Center building on Carnegie Mellon University (CMU) campus as a case study. The raw BAS alarms are first categorized into occupant-related alarms, equipment-related alarms, and critical operation alarms based on their affected objects. Next, the transient energy consumption impacts and thermal comforts of the alarms are quantified by first principle calculations. The long-term impacts are then quantified with the transient impacts and alarm durations. To predict future alarm durations, a decision-tree machine learning model is built. The model could predict the occupant-related alarm durations at an approximately 80% accuracy. With both transient and long-term impacts quantified, a method which calculates the comprehensive impacts of the alarms is proposed. User could weight different impact metrics (e.g. energy consumption, thermal comfort) differently based on their preferences with the method. The framework from this study could effectively categorize and prioritize alarms from BAS, which helps building operators to make better O&M decisions.

Keywords: BAS, Alarm Package, Data Mining, Impact Quantification, O&M Decision Making

Table of Contents

List of Figures

List of Tables

Glossary of Terms

AHU	Air Handling Unit
BACnet	Building Automation and Control networking
BAS	Building Automation System
BEMS	Building Energy Management System
CMU	Carnegie Mellon University
CECI	Cumulative Energy Consumption Impact
CTCI	Cumulative Thermal Comfort Impact
Cx	Commissioning
DDC	Direct Digital Control
DM	Data Mining
ECRI	Energy Consumption Rate Impact
FDD	Fault Detection and Diagnostics
FMS	Facility Management Service
GHG	Greenhouse Gas
HVAC	Heating, Ventilation and Air Conditioning
LonWorks	Local Operating Networks
O&M	Operation and Maintenance
OPC	Open Platform Communications
PCA	Principal Component Analysis
S-Bus	Smart Bus
SVM	Support Vector Machine
TCI	Thermal Comfort Impact

VAV Variable Air Volume

1. Introduction

1.1 Overview

Commercial building accounts for 19% of total energy consumption in the United States. More than 50% of the energy consumed by commercial buildings goes toward space heating, ventilation and air-conditioning system and lighting system (ACEEE 2016). Despite such a large portion of energy consumed by commercial buildings, there is a decreasing trend of energy consumption per floor area. According to the 2012 Commercial Building Energy Consumption Summary (CBECS), there is only a 7% of energy consumption increase given the 23% increase in total floor space since 2003. This slower growing of the commercial building energy demand can be explained in part by the higher energy performance of new buildings and major renovation buildings (U.S. Energy information Administration 2016). BAS is one of the critical factors in achieving high energy efficient in those buildings. Commercial buildings implementing BAS are estimated to save an average 10% of overall energy consumption (Sustar and Goldschmidt 2007). In addition to energy savings, BAS can also help facility managers to maintain comfort levels in occupant spaces. The basic function of BAS in terms of environmental control are maintaining comfort temperature and humidity levels and providing adequate ventilation and light levels for building occupants over different seasons. Moreover, a well implemented BAS can help control HVAC and lighting system better, give facility managers necessary alarms before systems or components go wrong. This results in savings in maintenance cost and extension of equipment life.

Although BAS has the potential of optimizing energy efficient of HVAC and lighting systems, maintaining comfort levels and reducing control and maintenance efforts and costs, there are both technology and policy issues that impair its ability. For example, building control systems manufacturers add many features into their BAS solutions. They develop fancy energy dashboards, interactive tools to display equipment status on floorplans and thousands of lines of alarm generation rules. Those functionalities are designed to help facility managers to improve system operation and control. But many of the functionalities are underutilized (Munasinghe 2016). This is because for most times, facility managers only received basic training to operate the main functions of a BAS. As long as there are no complaints from occupants, facility managers just leave the systems operate by default. This could end with energy waste, poor comfort level and even malfunctions in buildings systems and equipment.

Meanwhile, a lot of efforts are made to improve the O&M of building systems. Building system Fault Detection and Diagnosis (FDD) has been the subject of intense investigation and research in recent years. Retro-commissioning and re-tuning are becoming more and more popular in industry. Researches and practices have shown that some of those measures are effective in improving building system O&M, but they also have respective limitations. Those limitations include lack of flexibility, high demand for expertise, underutilized BAS functions and uncertain costs.

Therefore, this study aims to review the status quo of BAS techniques, measurements to support O&M decisions, and investigate the possibility of using built-in alarm report function of the BAS to help O&M. At first, this study reviewed the basic concepts and technologies in commercial building BAS. It then lists the key findings in current FDD studies and retro-commissioning practices. After reviewing current O&M decision-making supports, this study presents the possibility of using built-in alarm report function of the BAS and Data Mining (DM) techniques to help facility managers make better O&M decisions.

1.2 Motivation

Buildings account for a significant portion of total energy consumption. According to U.S. Energy Information Administration, residential and commercial buildings consumed about 40% of total energy in the U.S. in 2015. BAS has become increasingly popular under such a circumstance. When properly applied, BAS enables considerable energy savings (Ahmed, Korres, Ploennigs, Elhadi, & Menzel, 2010). Modern BASs and Building Energy Management Systems (BEMS) are extremely complex—consisting of thousands of sensors, controllers and actuators. BAS keeps gathering large amount of data from the sensors in lighting, HVAC, fire protection systems. It then implements control strategies to maintain desired indoor environmental condition and save energy. Due to uncertain factors, such as weather condition, occupant behavior, operation schedules, and lack of commissioning and maintenance, building systems rarely perform as well as anticipated (Piette, Kinney, & Haves, 2001). A variety of researches have been done to detect faults in both system level and component level. But few BAS is capable of automatic fault detection and diagnostic. For instance, a BAS is used in supervising the operation of AHU in HVAC system. When certain parameters collected by sensors breach upper or lower limits, BAS raises an alarm. But it does not show the root cause of those alarms (Burton, Raftery, Kennedy, Keane, & O'Sullivan, 2013).

Besides, BAS have traditionally been the territory of control engineers and technicians writing sequences of operation into code and usually leaving them hidden from operators (Bobker, et al., 2013). Manually monitoring time series BAS data and identifying abnormal operation and system malfunction from alarms is challenging for building operators. On Carnegie Mellon University's main campus, the BAS system raises over 100,000 alarms in 4 years. Most of the alarms are unacknowledged and the facility managers can only ignore them. Mostly, they adjust the setpoints and schedules only when occupants complain. Ignoring alarms from BAS could lead to reduced thermal comfort, increased energy waste and equipment deterioration.

There are many researches in fault detection and diagnostics regarding HVAC systems and components. But given the uniqueness of the physical attributes and uncertain factors of different buildings, those complex fault detection methods rarely work in practice (Narayanaswamy, Balaji, Gupta, & Agarwal, 2014). Thus, it is worthy to investigate the causes and impact of alarms raised by BAS system.

1.3 Hypothesis

A Data Mining framework that is built on BAS alarm datasets from a university complex building, can efficiently categorize and prioritize the alarms based on potential energy consumption and thermal comfort impacts to support HVAC system operation and maintenance.

1.3.1 Sub-hypothesis

1. The framework developed in this study will streamline raw BAS alarm collection and preparation processes for data mining application.
2. The data mining methodology in this framework will accurately filter alarms with potential high impacts on energy consumption and thermal comfort level.

3. The tool implementing the framework will help facility managers to make better operation and maintenance decisions, to save energy and improve thermal comfort levels.

1.4. Deliverables

1.4.1 A Documentation

The documentation paper includes the key findings from preliminary literature review. This review identifies the deficiencies of BAS in terms of support building system Operation and Maintenance (O&M), and investigate the possibility of using BAS alarms to optimize O&M.

1.4.2 BAS Alarm Management Framework

This framework contains the steps to get actionable information from BAS alarm that can be used to support HVAC system operation.

1.4.3 Structured Survey with Results

A structured survey is used in the interview with CMU campus facility managers. The deliverables include the survey questions and responds. Survey questions can be found in 2.5.3 Interview Facility Managers.

1.4.4 Analysis Results

A brief report regarding the impacts of alarms in each impact category is included. For instance, a certain alarm has the impacts like: 1) increased energy use intensity in kWh/m^2, 2) Increased PMV by a number. The analysis also includes the ranks of alarms in those categories.

1.4.5 Recommendations

A recommendation documentation is created to provide instructions for facility managers. The instructions include a guide to choose the impact category, control sequences (e.g., turn on/off HVAC system, adjust the setpoints), or maintenance actions (e.g., check a VAV box damper).

2. Methodology

The methodology of this study consists of six parts: 1) Identify research scope, 2) Identify research hypothesis, 3) Literature review, 4) Interview with building operators and field engineers, 5) Framework development, 6) Conclusion and report. Figure 1 shows the flowchart of the methodology. This chapter explains the methodology from a high level, detailed steps and descriptions could be found in Chapter 3, Chapter 4, and Chapter 7.

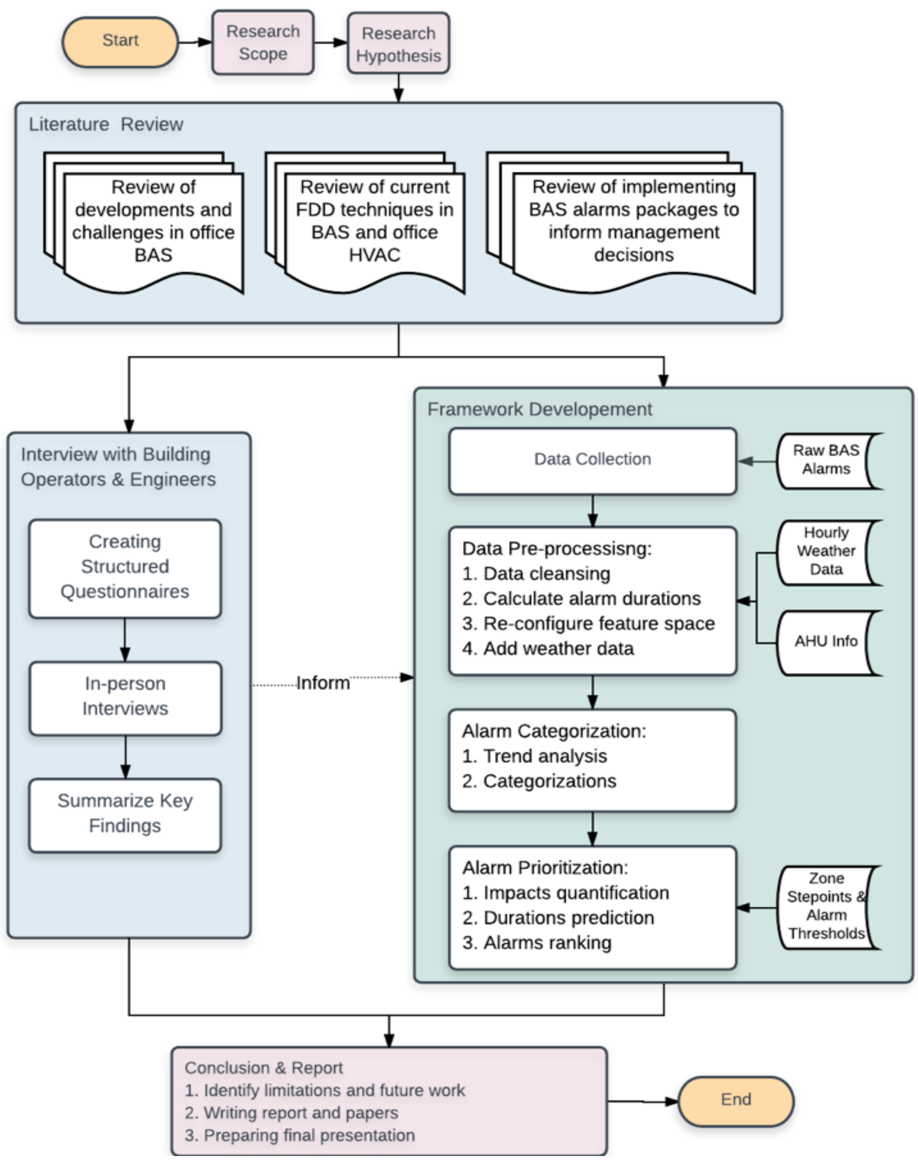


Figure 1. Flowchart of Methodology

2.1 Identify Research Scope

Office buildings are representative in commercial buildings. According to U.S. Energy Information Administration, office buildings account for 54.4% of commercial buildings in number of buildings and 51.6% of commercial buildings in floor space.

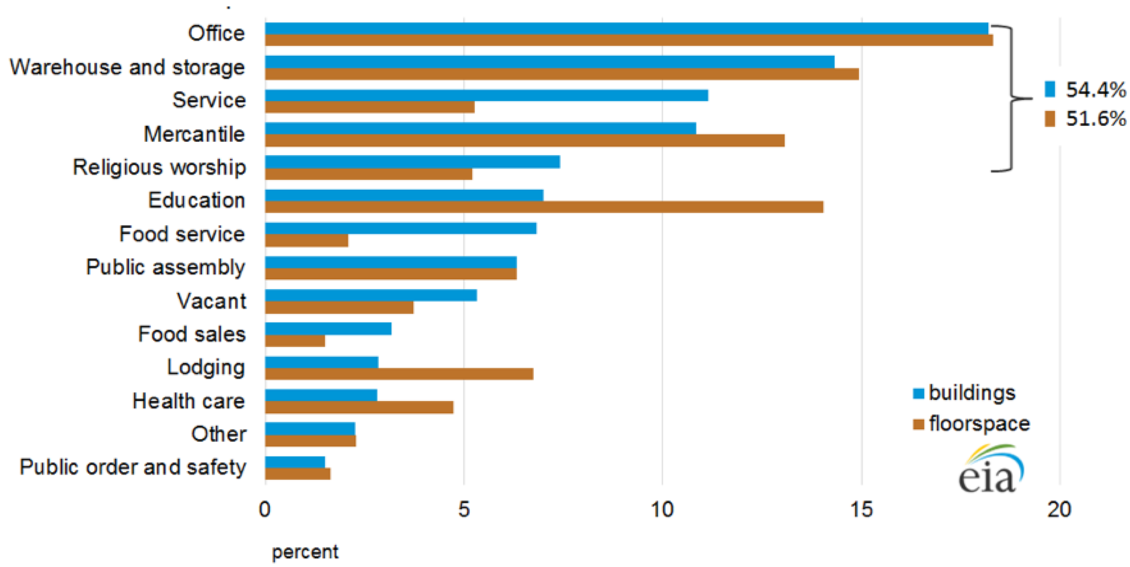


Figure 2. Percentage of Commercial Buildings by 14 Principal Activities

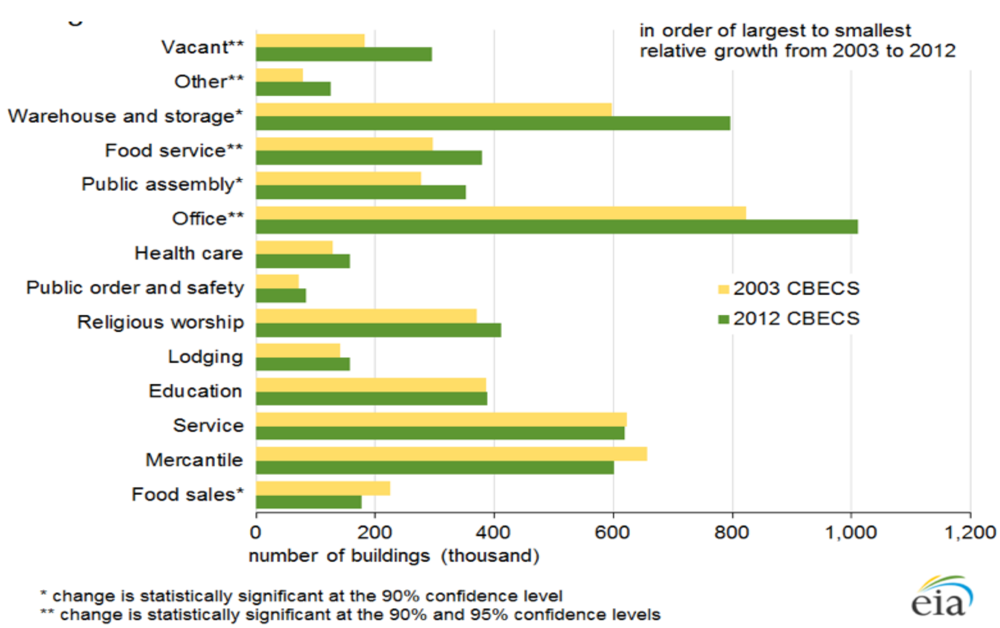


Figure 3. Number of Buildings Change from 2003 CBECS to 2012 CBECS

The system and data complexity is ideal for analysis. Another reason is that we have easy access to the BAS dashboard. The study focuses on the HVAC system of an office building on CMU campus. The building is Gates Hillman complex, which is a nine-story building composed of offices, classrooms, conference rooms, service rooms.

2.2 Research Hypothesis

The study aims to prove that BAS alarm management is good strategy of optimizing HVAC system O&M in terms of requirement of domain knowledge, cost-effectiveness, and accuracy. Thus, the author identifies the research hypothesis as discussed in Chapter 2.

2.3 Literature Review

The literature review identifies the deficiencies of BAS in terms of support building system Operation and Maintenance (O&M), and investigate the possibility of using BAS alarms to optimize O&M. The review focuses on the following contents:

- Review of basic concepts and technologies of modern BAS.
- Review of current O&M supporting techniques including FDD and building commissioning.
- Review of implementing BAS alarm packages to inform facility management decisions.
- Review of existing BAS solutions from different vendors. The focus is alarm nomenclatures, meanings and the rules of alarm generation.

The findings from the literature review can be found in Chapter 5.

2.4 Interview with Building Operators and Engineers

- An interview is conducted in this study. The goal of the interview is to understand: 1) How the current operation and maintenance is supported by BAS. 2) What the problems are with the current alarm management tool. 3) What the future needs are for the alarm management functions. The interview process includes the following parts:

- Identify interview objectives.
- Development of structured survey questions
- Summarize key findings.

- The details of the interview could be found in Chapter 6. The original survey questions and responses can be found in Appendix.

2.5 Framework Development

2.5.1 Investigation of BAS

The dashboard is a good resource of investigating BAS. In this study, the BAS dashboard by AutomatedLogic® is reviewed by the author. The main focuses are the alarm displaying function, alarm log function, and the rules behind some typical alarms.

2.5.1 Collect Raw BAS Alarms

Collect BAS alarm from existing building automation systems on Carnegie Mellon University's main campus.

2.5.2 Data Pre-processing

In this step, raw BAS alarms are cleaned and categorized firstly. Then the key features are selected and data is divided into training and testing sets.

- Clean raw data to allow tool importation.
- Manually categorize data.
- Select key alarm features such as alarm raised time, system, floor, alarm detail, acknowledged time.
- Divide original data into training and test sets.

2.5.3 Interview Facility Managers

- Contact with the FMS
- Create Structured Questionnaires
- Analyze data to identify alarm classification

2.5.4 HVAC System Information Collection

- Collect floor plan and thermal zoning information.
- Mark data point (i.e. sensor, controller, actuator) locations.

2.5.5 Data Mining Model Development

- Use K-means clustering to find typical alarm groups with similar attributes. (e.g. identify alarm patterns, durations)
- Use association rule mining to find relationship between alarms and energy consumption.
- Classify HVAC related BAS alarm based on their impact. The classifier allows user to choose prioritized impact category (i.e. Energy, Thermal comfort, CO2 level).

2.5.6 Validate Model Accuracy

- Validate model performance with test data set.
- Find evidence from literature review. (e.g. A specific type of system fault could lead to certain impact on operation cost)

2.5.7 Post Mining Application

- Rank and label alarms with different impact categories.
- Map significant alarms in system layout.
- Create a user-friendly interface for facility manager to make operation decisions.

2.6 Conclusion and Report

2.6.1 Conclusion

Draw conclusions from the data mining experiments and related literature reviews. Test the hypotheses listed in the Hypothesis section.

2.6.2 Identify Future Works

Based on the hypothesis test results, address limitations of current study and identify chances of improvements in future works.

- If the framework does not work well, analyze the reasons and address the limitations.
- If the framework works well, identify chances of improvements and boarder application in future work.

3. Review of Building System Operation and Maintenance

This chapter first reviewed: 1) basic concepts and technologies implemented in Building BAS and the benefits and challenges of using BAS, 2) current technologies that support building system O&M decision making and their pros and cons. It then investigated the potential application of using BAS alarm packages to inform building system O&M decision-making.

3.1 Benefits and Challenges of Building Automation System

3.1.1 Concepts and Technologies

BAS is a distributed control system that provides centralized monitor and control of a building's HVAC systems, lighting systems, security systems and other systems. A BAS collects data from data points—sensors, meters, and actuators in the building systems. It then analyzes the data and takes control actions or alarm operators about abnormal conditions. A typical BAS has a three-layer architecture: 1) Field layer, 2) Automation layer, 3) Management layer (Fernbach, Granzer and Kastner 2011). *Figure* 4 shows the three layers in the architecture and their corresponding roles.

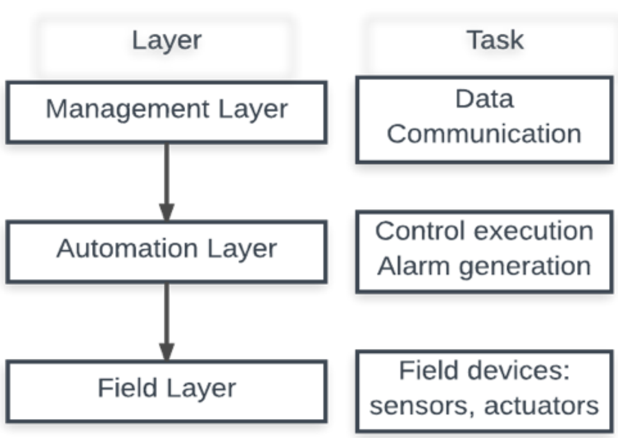


Figure 4. Three-layer Architecture of BAS

Sensors, controllers, actuators are the key components in a BAS which monitor, analyze and adjust the operation parameters of building systems. Besides the components, a BAS needs hardware and communication protocols to support the data transmission among the devices. Dashboards are interactive tools that allow facility managers to have an overview of system operations and manage alarms that highlight operational and maintenance issues. (Jadhav 2016)

Sensors. Sensors are devices in building control systems that provide monitoring functions. Most common sensors in BAS are: 1) temperature sensors, 2) humidity sensors, 3) lighting sensors, 4) occupancy sensors, 6) inductance sensors, 7) capacitance sensors, and 8) half-effect sensors. Sensors can be either wire-connected or wireless-connected.

Controllers. Controllers are devices embedded with certain rules or algorithms that take sensing data as input and convert it into the signal that actuators receive. Most controllers in BAS are in three forms: 1) scheduling and mode selection, 2) pneumatic controllers, and 3) direct digital controllers (DDC) (Jadhav 2016) In addition to traditional controllers, advanced control solutions utilizing fuzzy logic, artificial neural network, and model predictive control are becoming popular in building automation field.

Actuators. Actuators are devices that take control signal as input and take actions to regulate controllable devices such as fans, pumps, valves, lighting devices, on/off switches. Actuators are usually integrated into the controllable devices.

Communication Protocols. Communication protocols allow data communication among different parts of BAS. They consist of a set of rules to restrict the communication process. In the past, manufacturers tend to have their own protocols which limit the integration of products from other manufacturers. The lack of interoperability instigated standardization of communication protocols. In recent years, some open communication protocols and standards become popular: 1) Building Automation and Control networking protocol (BACnet), 2) Local Operating Networks (LonWorks), 3) Modbus, 4) KNX, 5) Smart Bus (S-Bus), 6) Open Platform Communications (OPC). (Jadhav 2016)

Dashboard. Dashboards are interactive tools that visualize real-time data and provide insights of system operation status for facility managers. An ideal BAS dashboard allows facility managers to monitor different systems, identify problems, and make better O&M decisions. Due to the uniqueness of buildings and systems, BAS dashboards should be customized to accurately provide system information.

3.1.2 Benefits

Researches and applications have shown great potential of Building Automation Systems in reducing building energy consumption, improving indoor environmental quality and facilitating building system operation. A well implemented of BAS could lead to benefits in different ways. Those benefits include: Environmental benefit—reduced energy demand and greenhouse gas (GHG) emission; Economic benefit—energy cost and system maintenance savings; Social benefit—productive working environment out of improved indoor environmental quality. (Simmonds & Bhattacherjee, 2015) The sensing and metering technologies of modern BAS also provide information for control decision-making. (Domingues, Carreira, Vieira, & Kastner, 2015)

3.1.3 Challenges

Despite the good features of BAS, numerous researches show that the complexity of modern building systems and the lack of commonly agreed knowledge in BAS have led to troubles in facility management. One major issue is the lack of agreement on BAS concepts and terminology. There are numerous building automation standard technologies available in the market. Examples are KNX, Local Operating Network (LonWorks), ZigBee, Building Automation and Control networking protocol (BACnet), etc. And for each of those standard technologies, the fundamental BAS concepts such as grouping, notification, scheduling and commanding are not unified. One of the undesirable results is that one vendor's solutions incompatible with other vendors'. In addition, BAS solutions are often too complex for both end-users and developers to use. And a solution is most likely being tailored for a certain working environment and lacks flexibility when facing condition changes. (Domingues, Carreira, Vieira, & Kastner, 2015) Those difficulties in BAS applications could cause malfunctions and bad maintenance, which lead to discomfort for occupants, energy waste and reduced equipment life.

3.2 Operation & Maintenance Support

Efficient building operation could maintain occupant comfort, reduce energy consumption and maintenance cost. As discussed earlier, although a BAS has the potential to optimize building operation, the complexity of systems, lack of interoperability among BAS products and technologies, and limited expertise of building operators often weaken this potential. In reality, extra efforts are needed to support operational decisions. Building Commissioning (CX) and FDD are two commonly used approaches to identify system deficiencies and operational improvement opportunities.

3.2.1 Building Commissioning

Building commissioning is a systematic process to identify and correct problems that lead to energy waste and system malfunctions in existing buildings. It can address problems started from the building's design and construction phase, or problems developed throughout the building's life. (Evan 2009)

Benefits. Based on the phases in which the Commissioning (Cx) process occurs during a building's lifecycle, there are Cx, Retro-commissioning (RCx), and Ongoing Cx. A review of the literature shows that Cx can support O&M decisions in achieving energy saving, GHG emission reduction. For example, in a meta-analysis of Cx experience from a database containing 643 commercial buildings across 26 states in the U.S., the median annual whole building energy saving was found to be 16% for existing buildings and 13% for new constructions. (Evan 2009) Besides the direct energy savings, Cx could also help to reduce GHG emission and maintain the comfort level in occupant spaces.

Limitations. Despite the benefits of building Cx, there are certain problems that limit Cx's potential in supporting O&M. First, the cost of Cx varies from building to building. Factors like HVAC system type, space function, and expected outcomes can affect the Cx cost-effectiveness significantly. In some cases, the energy use could increase after Cx. Secondly, a Cx often involves owners, designers, contractors, facility managers and Cx authority. System deficiencies may be identified and corrected during Cx or RCx processes, but facility managers gained limited training in routine operations. This usually results in energy savings not lasting for a long time. (Evan 2009) Thirdly, due to the lack of regulation on assessing the quality of Cx, the outcomes of Cx are often less-than-satisfactory. (Lord, et al. 2016)

3.2.2 Fault Detection and Diagnostics

Fault Detection and Diagnostics is widely implemented in industrial process control and automotive and aerospace engineering to pinpoint and diagnose operational problems. As the monitor and control technologies evolve in recent years, FDD has been an area of intense research in building field, especially HVAC systems. Based on the knowledge used to diagnosis the cause of system faults, FDD methods could be classified as model-based approaches and data-driven approaches. (Katipamula and Brambley 2005)

Benefits. A review of the literature shows that FDD can identify abnormal conditions in HVAC systems and support O&M. For example, Beghi, et al. (2016) developed a semi-supervised data-driven FDD method for HVAC water chillers. They implemented Principal Component Analysis (PCA) model for fault detection tasks. They assessed the model against a test dataset and found satisfactory fault detection result for two kinds of anomalies in screw-chillers. Narayanaswamy, et al. (2014) developed an unsupervised data-driven method to detect anomalies in Variable Air Volume (VAV) boxes. In their approach, zones with significantly different attributes would be grouped into different clusters firstly. Then they use those clusters to detect abnormal zone controller configurations. This method has shown good performance in terms of detecting anomalies and reducing false alarms. Li and Wen (2014) proposed a data-driven model-based FDD method to detect abnormal conditions in Air Handling Units (AHUs). Their method combined wavelet transform and PCA to avoid the impact on dynamic weather changes. The test results show their method could detect common AHU false like heating/cooling coil valve leaking and outdoor air damper stuck with no false alarm.

Limitations. Although those approaches have shown acceptable fault detection performance in their respective domain, none of them is able to cover other prevalent faults in HVAC systems. Guo, et al., (2015) presented an online sensor monitoring and fault detection technique and the key sensor sets selection approach to optimize fault detection results. They tried to address common faults in HVAC system. But the test results show a wide range of accuracy (33% to 100%) in detecting different faults. Besides their poor practicality in detecting abnormal conditions in HVAC systems from a comprehensive perspective, studies have also shown that

FDD methods have limitations in fault diagnosis. For example, the popular data-driven PCA based FDD methods do not tell the cause-effect relationship of system faults. Yu, et al., (2014) Moreover, due to the interoperability issues of BAS and requirement of strong domain knowledge, developing effective FDD methods for a specific building can be time-consuming and costly

3.2.3 Data Mining Applications

The existing BAS data sets are good resource of information which has great importance for better building system operation and energy management. Data Mining (DM) is a promising technique for unveiling hidden patterns in large scale data sets. A review of literature shows that many researchers use DM to find hidden patterns and to optimize building operation. (Ahmed, Korres, Ploennigs, Elhadi, & Menzel, 2010) presented a method which uses DM techniques to find relationships between building characteristics and energy performance. In their study, three classifiers—Naïve Bayes, Decision Tree and Supporting Vector Machine (SVM) were developed for estimation of building performance indicators (i.e. thermal condition, illuminance and demand for heating, cooling or artificial lighting). They trained and tested the three classification models. They found that Decision Tree model achieved 92% overall accuracy when using air temperature along with weather data to predict thermal condition, it also achieved 99.9% accuracy in predicting if room illuminance is enough or not. Naïve Bayes model achieved 91% overall accuracy in predicting demand for heating, cooling or artificial lighting. (Xiao & Fan, 2014) investigated the use of data mining techniques in improving building operational performance. They proposed a framework for mining BAS database. The framework consists of raw BAS data collection, data exploration, data partitioning, knowledge discovery, post-mining and application. They tested the framework on the BAS database from the tallest building in Hong Kong. The clustering and association rule mining algorithms developed in the knowledge discovery step were found effective in identification of changes in building operation strategies, identification of non-typical building operation conditions and fault detection of power consumption sensors. The same group of researchers then applied this framework in discovering temporal knowledge in BAS data set. (Fan, Xiao, Madsen, & Wang, 2015) proposed a data mining methodology for knowledge discovery in time series BAS data. Unsupervised DM techniques including clustering, association rule mining was used to detect temporal abnormal conditions and to characterize building power consumption dynamics. Their studies present generic ways to discover hidden knowledge in massive data generated by modern BAS. However, in both studies researchers could address only DM with BAS sensed data sets. And their solutions are still complex for guiding daily operations.

3.3 Alarm Management for Control Systems

Control systems for industrial processes are being continuously developed. Recent studies in the field of automation have shown that alarm management system along with sensing technologies can provide a good support for decision making and management. Urban & Landryová (2016) conducted an analysis of alarm logs in the field of marine technology to identify and analyze abnormal situations that could affect process safety. The alarm packages they used in the study are log files from a vessel control system. Their study supports the development of an engineering tool which allows operators to decide which alarms need immediate attention and which alarms could be postponed. To our knowledge, there is no similar study in building control field that addresses alarm filtering and ranking issues. Furthermore, current FDD methods provide many insights in how to detect system faults and what the faults' potential costs are. However, very few of them provides instructions for building operators to optimize O&M. Thus, it is valuable to create a BAS alarm management tool which could help facility managers prioritize alarms and provide actionable information.

3.3.1 Issues in BAS Alarms

BAS have traditionally been the territory of control engineers and technicians writing sequences of operation into codes and usually leaving them hidden from operators (Bobker, et al., 2013). Manually monitoring time series BAS data and identifying abnormal operations and system malfunction from alarms is challenging for building operators. On Carnegie Mellon University's main campus, the BAS system raises over 100,000 alarms in 4 years. Most of the alarms are unacknowledged and the facility managers have to ignore them. Mostly, they adjust the setpoints and schedules only when occupants complain. Ignoring alarms from BAS could lead to reduced thermal comfort, increased energy waste, and equipment deterioration.

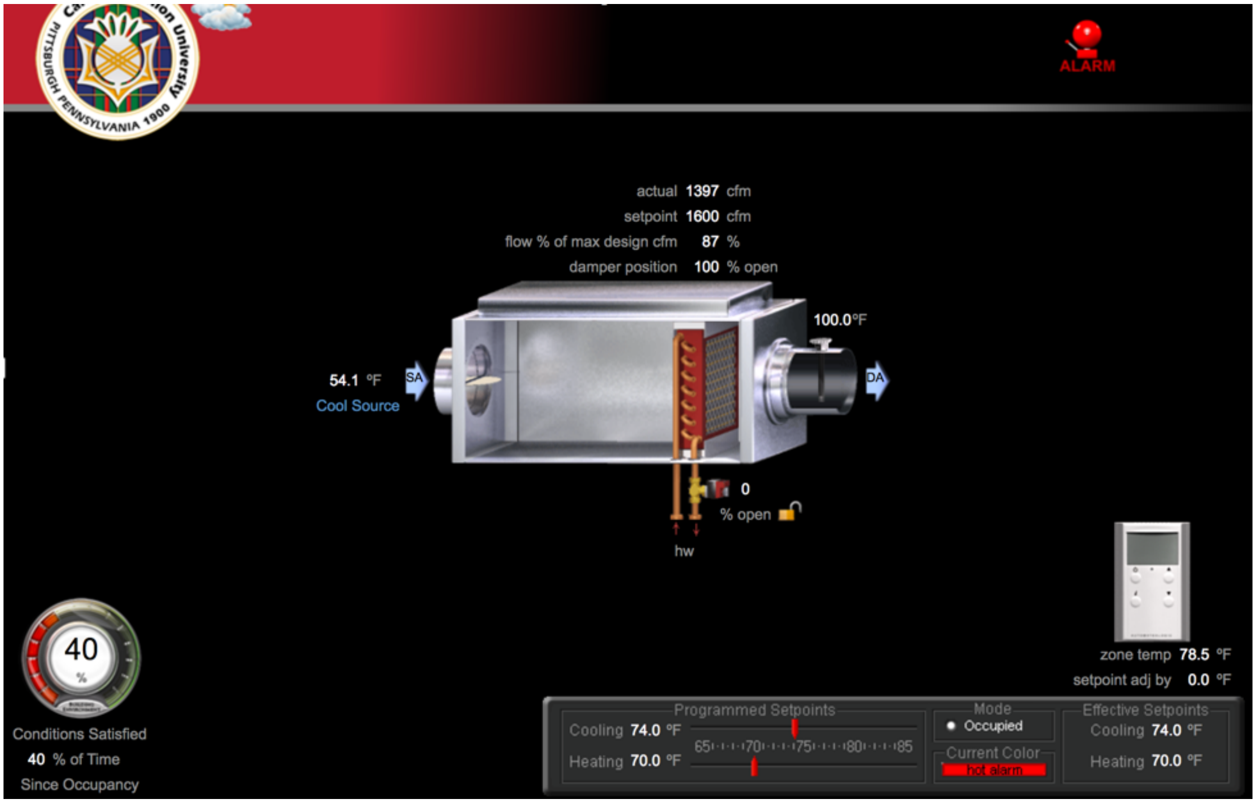


Figure 5. Examples of False Alarms from CMU BAS

Note: Wrong control logic. Keep heating the primary air when zone temperature is above the upper limit of the deadband.

Part of the reasons that the BAS generates such a huge amount of alarms is that the alarm-generation rules are wrong in certain situations. Some examples of typical false alarms we identified from CMU campus BAS are shown in Table 1:

Table 1. Examples of Unreasonable Rule Settings

Setting	Alarm	Result
Cooling setpoint unreasonably low	Zone temperature high	Unnecessary cooling demand
Heating setpoint unreasonably high	Zone temperature low	Unnecessary heating demand
Wrong cooling control logic	Zone temperature high	Cooling when heating is required
Wrong heating control logic	Zone temperature high	Heating when cooling is required
Different deadbands in same space	Zone temperature high/low	Simultaneously heating or cooling

Alarms with unreasonable rule settings shown in Table 1 have various undesirable outcomes: 1) zone thermal comfort level would be decreased, 2) Unnecessary cooling or heating demand would lead to energy waste, 3) the deluge of alarms with unreasonable rules would conceal those with critical information. It once again shows the significance of an alarm filtering and ranking tool which can help facility managers make O&M decisions.

3.4 Current BAS Alarm Management

Currently, the alarm management function of BAS does not provide building operators enough actionable information. Figure 6 shows the logic of the alarm management function. The BAS monitors the building. The rules embedded in the BAS generates alarms when the monitored parameters exceed the rule specifications. However, the rules are written by BAS designers and building operators usually don't have access to them. It is impossible for building operators to keep track of the alarms and acknowledge them manually when the BAS generates a large amount of alarms every day. Sometimes, they don't take actions until occupants complain about certain problems in their surrounding indoor environment quality.

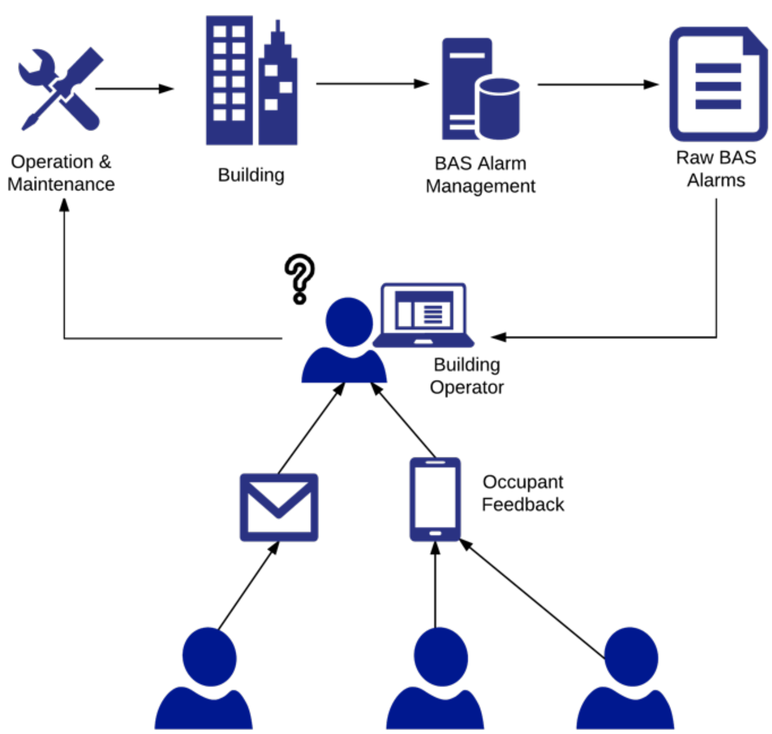


Figure 6. Current Alarm Management Function

4. Interviews with Building Operators and Engineers

The topic of this thesis is to optimize the alarm management function for facility managers so that they can make informed operation and maintenance decisions. To achieve that goal, it is necessary to understand what the pros and cons are about the current alarm management tool, and the user needs for future tool. Thus, a set of interviews are used to collect the opinions of building operators. This chapter introduces our interview target, survey development, and the key findings from the interviews.

4.1 Interview goal

The goal of this interview has two parts.

4.1.1 About alarm management

The first part is to get users' experience about the current BAS alarm management tool and their expectations for the future alarm management tool (if any). More specifically, the following aspects are considered: (1) The way that a user gets notified by the BAS when an alarm occur, (2) The information that a user receives along with the alarm notification, (3) The decision that a user make when an alarm occurs and the reason for that. (4) The pros and cons of current alarm management tool, (5) The prospective functions that can help a user make better operation and maintenance decisions and how those functions can benefit.

4.1.2 About fault detection and diagnostics

The second part is to generalize the processes and approaches of fault detection and diagnosis based on experts including facility managers and BAS engineers. The detailed aspect includes: (1) The main steps and references that facility managers use for fault diagnosis, (2) possibility to realize the function of automated fault diagnosis (3) potential reasons for the some common-seen HVAC faults (4) considerations of energy saving and comfort improvement for a fault correction, (5) further historical records and resources that can be available from local facility manager and useful for the Gates-Hillman building faults validation.

4.2 Survey development

To achieve the goals above, a tailored survey is needed.

4.2.1 Respondents

Every building has its own unique attributes, and the BAS also vary building by building. Moreover, the information required by this study is highly technical. Given this context, the best interview targets are people who have a good understanding of the both the target building and domain knowledge in building automation, alarm management, and fault detection and diagnostics.

Facility managers: Facility management is an interdisciplinary profession which focuses on delivering support services for the organization it serves. Facility managers are individuals who are responsible for making sure the building and its systems meet the needs of its occupants. Normally, they are the direct user of BAS dashboard and will be notified when alarms occur.

BAS field engineers: BAS Field engineers are usually responsible for the installation, start-up, troubleshooting, commissioning and servicing of DDC building automation systems. They also perform equipment repairs, building system calibrations, control logic designs.

4.2.2 Data collection method

Since building operation and maintenance is a highly technical task, and the information we need is complicated. It involves alarms acknowledgement sequences, fault detection and diagnostic methods, and occupant feedbacks. Thus, in person interview is an effective way of gathering that information. To better record the answers from the respondents, we decide to develop a survey.

4.2.3 Survey question design

As discussed before, each building has its unique attributes and building systems. However, their automation systems and control sequences have a lot of things in common. For example, the buildings using the BAS solution from a same vendor may have similar control sequences and nomenclature for the components. The key steps in fault detection and the parameters used for fault detection can also be similar in different buildings. Therefore, it’s necessary to get high-level information about BAS alarm management and fault detection and diagnostics. This study aims to investigate the alarms and faults in a case study and develop a framework for optimizing alarm management. General questions and specific questions are included in the survey. A sample of the survey can be found in the appendix.

4.3 Key findings

(1). **Currently, facility managers don't rely too much on the alarm management function.** There are several reasons why they don't use if often. Firstly, there are too many of the alarms. It's not possible to be notified (usually by message or email) when a single alarm occurs. Secondly, the information provided by the alarms is limited. Alarms are just a reference source for facility managers. Thirdly, they make operational decisions based on their experience, or occupants continuously complain about some issues.

(2). **Building system's normal operation has higher priority than occupant comfort and energy consumption**. Facility managers usually care most about the safety and normal functions of the building systems. When facing an alarm, facility managers rank it by three tiers. Tier 1: life safety, Tier 2; system normal operation, Tier 3: occupant comfort and energy consumption. This is similar in alarm management. Facility managers usually have their own experience about the severity of the alarms. For example, they consider the hot water's temperature abnormal more serious than the low temperature in an office room. Because low hot water temperature may be caused by faults in boilers, which has more serious impacts than the low air temperature in an office.

(3). **Building system maintenance is mainly accomplished by annual inspection**. Despite the BAS's monitoring and alarm functions, little information is provided for fault detections and diagnostics. Facility managers usually won't go check each component when they see an alarm or receive occupant complaints. The first reason is the manual cost: as long as there is no significant fault in the equipment, it's not cost effective to inspect it and fix the minor problem indicated by the alarms. For example, when a zone's air temperature is lower than the cooling setpoint in summer, BAS raises a temperature too low alarm. Facility managers typically won't go to the field and fix the problem. Since the labor cost can be higher than the energy saved.

(4). **There are a lot of fault detection and diagnostics researches, but it's hard to implement**. Each building has its own attributes. The FDD methods are just not practical in many situations.

For example, some of the FDD methods require high quality, sub-metering data, which is very rare in the buildings nowadays. In addition, the accuracy of FDD highly relies on the sensor data. However, most HVAC systems use feedback controls. The data quality can be influenced by a single sensor's failure.

(5). **The needs for future alarm management tools: simple, accurate, and powerful.** The future alarm management tools should be more user-friendly. It should provide very clear information about the alarm and avoid meaningless and lengthy descriptions. It should be accurate—the trivial alarms should appear on the dashboard. It should help building operators make better operation decisions. For example, it can embed FDD algorithms and show the root cause of the alarms, and display the alarm and faults on the floorplan. It can also provide suggestions of how to react to the alarms and what the potential impacts are.

4.4 Reflections on the interview

(1). **Improve the current alarm management function.** The interview reinforces our findings from the literature review that building system operation and maintenance is getting more and more challenging. There is a gap between what the user needs for BAS alarm management and what the tools provide. Because of the poor alarm generating rules, an overwhelming amount of alarms are generated every day. Currently, facility managers mainly rely on their experience about what to do when they see the alarms. Since the organization already invested a huge amount of money on the BAS, it's worthy to make use of the existing tools. This again justifies the significance of this study.

(2). **Consider of health and productivity savings.** From the interview, we found that facility managers are more concerned of the normal functions of buildings systems than energy consumption and occupant comfort. However, the preliminary analysis of the alarms shows that occupant-related alarms should not be ignored. Moreover, a vast amount of studies proved that there are positive relationships among good indoor air quality, good thermal quality, and the health and productivity of the occupants. It may be costly if facility sends someone to fix the problems that causes abnormal indoor temperature and high CO_2 concentration from their perspective. But we also need to look at the organizational savings that come from more comfortable indoor thermal quality and air quality. It can be very cost-effective if those factors are taken into consideration.

(3). **Fault-preventive operation.** Ignoring the alarms and simply clicking on the "acknowledge alarm" button is not the original intent of BAS designers. However, given the real situation, building operators don not have many options when they see the alarms. It will be very helpful if the alarm management tool can guide them to make informed decisions. One important function of the alarm management tool should be evaluating the impacts of the alarms and help facility managers with fault-preventive operations.

5. Framework Development

An alarm filtering and ranking tool is a key component of this study. Figure 7 shows the framework of the tool development. The framework includes the following parts:

(1). Retrieving raw BAS alarms. The plain text format raw alarm data is collected from BAS. Each entry represents an alarm which has some descriptive features. The alarms from the target buildings are extracted for next steps.

(2). Data pre-processing. The raw data has only several descriptive features, which is not interpretable for manual alarm analysis and not readable for data mining algorithms. In this step, alarm data is parsed in a tabular format

(3). Categorizing BAS alarms. After the alarms are pre-processed, they can be categorized into several categories (e.g., equipment related, occupant related, critical operation related).

(4). Prioritizing BAS alarms. The impacts of certain alarms are evaluated based on the findings from literature reviews, third-party FDD methods, or first principle calculations. The durations of the alarms can be predicted with a Decision Trees machine learning model. With the quantified impacts and durations, the alarms can be ranked per user preferences.

(5). Feedback to facility managers. Based on user preferences, the alarms with top impacts are shown to facility managers along with actionable information (e.g., measures to acknowledge the alarm, control sequences, root cause of the alarms, and maintenance recommendation.)

This chapter presents the detailed implementations of this framework with Gates Hillman Center at Carnegie Mellon University.

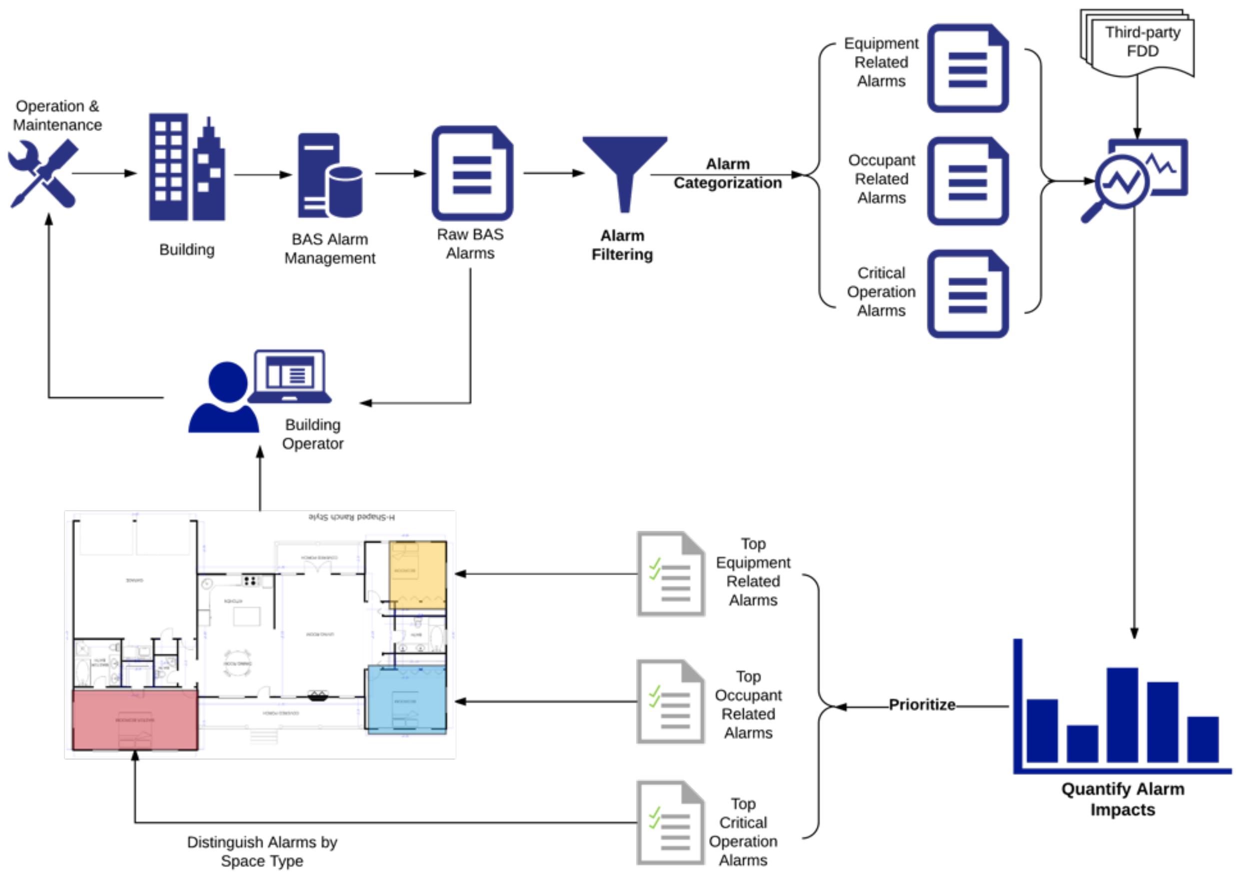


Figure 7. Framework of BAS alarm management tool

5.1 Retrieving raw BAS alarms

First, the raw alarm data is downloaded from the BAS dashboard. It contains the alarms from all buildings with AutomatedLogic®, during Feb 2010 to Feb 2016. The raw BAS alarms are stored in a plain text file. Each line of the plain text is an alarm instance.

```
Feb 19, 2010 8:39:16 AM,NORMAL,SCSC Hillman / Fourth Floor / VAV Room 4837 / Input Micro-Block Error,Input Error Alarm,Input Micro-Block Error: has returned to normal.,Feb
19, 2010 2:19:15 PM,Cory Bushman
Feb 18, 2010 7:02:03 PM,OFF NORMAL,SCSC Hillman / Fifth Floor / FCU-9 / Input Micro-Block Error,Input Error Alarm,message_prefix$ Navigate to the equipment and generate
network I/O report to view any network errors.,Feb 19, 2010 2:20:28 PM,Cory Bushman
Feb 18, 2010 7:07:00 PM,NORMAL,SCSC Hillman / Fifth Floor / FCU-9 / Input Micro-Block Error,Input Error Alarm,Input Micro-Block Error: has returned to normal.,Feb 19, 2010
2:20:28 PM,Cory Bushman
Feb 18, 2010 7:42:03 PM,OFF NORMAL,SCSC Hillman / Fifth Floor / FCU-9 / Input Micro-Block Error,Input Error Alarm,message_prefix$ Navigate to the equipment and generate
network I/O report to view any network errors.,Feb 19, 2010 2:20:28 PM,Cory Bushman
Feb 18, 2010 7:47:00 PM,NORMAL,SCSC Hillman / Fifth Floor / FCU-9 / Input Micro-Block Error,Input Error Alarm,Input Micro-Block Error: has returned to normal.,Feb 19, 2010
2:20:28 PM,Cory Bushman
Feb 18, 2010 7:57:03 PM,OFF NORMAL,SCSC Hillman / Fifth Floor / FCU-9 / Input Micro-Block Error,Input Error Alarm,message_prefix$ Navigate to the equipment and generate
network I/O report to view any network errors.,Feb 19, 2010 2:20:28 PM,Cory Bushman
Feb 18, 2010 8:02:00 PM,NORMAL,SCSC Hillman / Fifth Floor / FCU-9 / Input Micro-Block Error,Input Error Alarm,Input Micro-Block Error: has returned to normal.,Feb 19, 2010
2:20:28 PM,Cory Bushman
Feb 18, 2010 8:07:03 PM,OFF NORMAL,SCSC Hillman / Fifth Floor / FCU-9 / Input Micro-Block Error,Input Error Alarm,message_prefix$ Navigate to the equipment and generate
network I/O report to view any network errors.,Feb 19, 2010 2:20:28 PM,Cory Bushman
Feb 18, 2010 8:11:51 PM,NORMAL,SCSC Hillman / Fifth Floor / FCU-9 / Input Micro-Block Error,Input Error Alarm,Input Micro-Block Error: has returned to normal.,Feb 19, 2010
2:20:28 PM,Cory Bushman
Feb 18, 2010 8:17:03 PM,OFF NORMAL,SCSC Hillman / Fifth Floor / FCU-9 / Input Micro-Block Error,Input Error Alarm,message_prefix$ Navigate to the equipment and generate
network I/O report to view any network errors.,Feb 19, 2010 2:20:28 PM,Cory Bushman
Feb 18, 2010 8:21:51 PM,NORMAL,SCSC Hillman / Fifth Floor / FCU-9 / Input Micro-Block Error,Input Error Alarm,Input Micro-Block Error: has returned to normal.,Feb 19, 2010
2:20:28 PM,Cory Bushman
Feb 18, 2010 10:07:03 PM,OFF NORMAL,SCSC Hillman / Fifth Floor / FCU-9 / Input Micro-Block Error,Input Error Alarm,message_prefix$ Navigate to the equipment and generate
network I/O report to view any network errors.,Feb 19, 2010 2:20:28 PM,Cory Bushman
Feb 18, 2010 10:11:54 PM,NORMAL,SCSC Hillman / Fifth Floor / FCU-9 / Input Micro-Block Error,Input Error Alarm,Input Micro-Block Error: has returned to normal.,Feb 19, 2010
2:20:28 PM,Cory Bushman
Feb 18, 2010 10:37:03 PM,OFF NORMAL,SCSC Hillman / Fifth Floor / FCU-9 / Input Micro-Block Error,Input Error Alarm,message_prefix$ Navigate to the equipment and generate
network I/O report to view any network errors.,Feb 19, 2010 2:20:28 PM,Cory Bushman
Feb 18, 2010 10:41:54 PM,NORMAL,SCSC Hillman / Fifth Floor / FCU-9 / Input Micro-Block Error,Input Error Alarm,Input Micro-Block Error: has returned to normal.,Feb 19, 2010
2:20:28 PM,Cory Bushman
Feb 19, 2010 8:17:03 AM,OFF NORMAL,SCSC Hillman / Fifth Floor / FCU-9 / Input Micro-Block Error,Input Error Alarm,message_prefix$ Navigate to the equipment and generate
network I/O report to view any network errors.,Feb 19, 2010 2:20:28 PM,Cory Bushman
Feb 19, 2010 8:21:57 AM,NORMAL,SCSC Hillman / Fifth Floor / FCU-9 / Input Micro-Block Error,Input Error Alarm,Input Micro-Block Error: has returned to normal.,Feb 19, 2010
2:20:28 PM,Cory Bushman
Feb 19, 2010 8:27:03 AM,OFF NORMAL,SCSC Hillman / Fifth Floor / FCU-9 / Input Micro-Block Error,Input Error Alarm,message_prefix$ Navigate to the equipment and generate
network I/O report to view any network errors.,Feb 19, 2010 2:20:28 PM,Cory Bushman
Feb 19, 2010 8:31:57 AM,NORMAL,SCSC Hillman / Fifth Floor / FCU-9 / Input Micro-Block Error,Input Error Alarm,Input Micro-Block Error: has returned to normal.,Feb 19, 2010
2:20:28 PM,Cory Bushman
Feb 19, 2010 1:42:03 PM,OFF NORMAL,SCSC Hillman / Fifth Floor / FCU-9 / Input Micro-Block Error,Input Error Alarm,message_prefix$ Navigate to the equipment and generate
network I/O report to view any network errors.,Feb 19, 2010 2:20:28 PM,Cory Bushman
Feb 19, 2010 1:46:51 PM,NORMAL,SCSC Hillman / Fifth Floor / FCU-9 / Input Micro-Block Error,Input Error Alarm,Input Micro-Block Error: has returned to normal.,Feb 19, 2010
2:20:28 PM,Cory Bushman
```

Figure 8. Raw alarms in plain text format

As shown in Figure 8, each alarm has some original descriptions separated by "/". The descriptions can be divided into 9 parts. Table 2 shows the descriptions of raw alarm data.

Table 2. Original Description of Raw Alarms

Column	Description
1	Alarm occurring date
2	Alarm occurring year, and time of the day
3	Alarm type
4	Building name/Floor/System/Short Info
5	Alarm range
6	Alarm long description
7	Alarm acknowledged date
8	Alarm acknowledged year, time of the day
9	Facility manager

Next, alarms are extracted by building names. In this study, only alarms in Gates Hillman Center are extracted.

5.2 Data pre-processing

Data pre-processing is the prerequisite for data mining. In this study, the original alarms are stored in a text file with lengthy descriptions and noisy characters. 1Therefore, it's necessary to convert the raw alarms from text format to tabular format that can be analyzed by data mining algorithms. There are four main steps in the pre-processing.

5.2.1 Data cleansing

First, messy data should be removed. A very few amount of alarms contain messy text. Those alarms do not have meaningful descriptions and are generated by the system randomly. They are detected and removed from the data. In this case, two types of alarms are removed (1836 out of 84886). After those messy alarms are removed. The dataframe is saved to a new csv file for next steps.

Table 3. Removed Messy Data

Number	Removed alarms
1	Alarms whose type (column 3 from the raw csv file) is "FAULT"
2	Alarms whose Building name/Floor/System/Short Info is "CSCS Network"

Second, certain characters in the text need to be replaced so that the original features can be separated properly. For example, the original feature in column number four in *Table* 4 has "building", "floor", "system", and "short info" attributes. Those attributes are separated by a "/". However, some alarms have the "system" named "AHU I/O" or "Copy/Print/Work" which should not be divided into different features. So, the goal of this step is to identify which information should be divided into different features and which should not be divided. Third, after the previous steps, the alarms are now saved in a tabular format with 10 features. *Table 4* shows the feature names and their meanings.

Table 4 Raw Tabular Dataframe

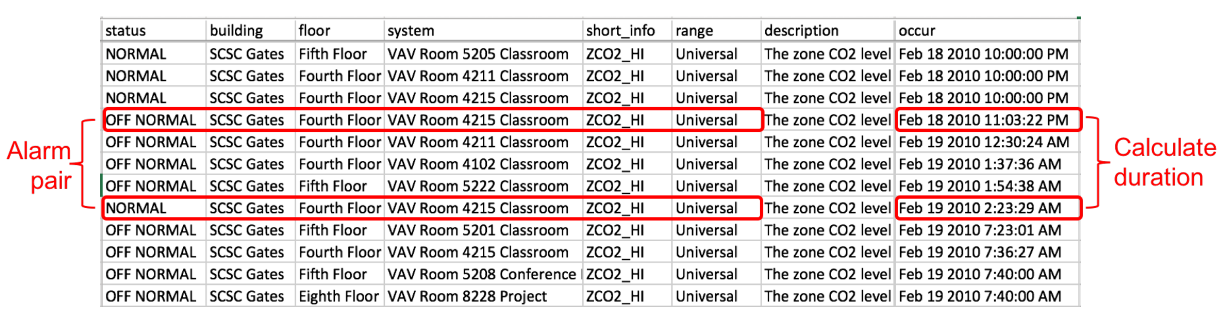
Number	Feature name	Meaning
1	status	The alarm status, "off normal", and "normal"
2	building	The building where the alarm occurs
3	floor	The floor where the alarm occurs
4	system	The system where the alarm occurs

5	short_info	Short name of the alarm from the BAS dashboard
6	range	The range of the alarm
7	description	Long description of the alarm
8	fms	The facility manager who acknowledged the
9	occur	The occurring time of the alarm
10	acknowledge	The acknowledge time of the alarm

5.2.2 Calculate alarm durations

Each alarm in the raw dataframe has a status with two possible values: "off normal" or "normal". The BAS raises an alarm when the value of a certain parameter exceeds the threshold and last for a period. When the parameter value returns to the normal range and last for some time, the BAS raises another instance with the status equals to “normal”, indicating the alarm is released.

The “normal" instances (alarms) and their corresponding “off-normal” instances have the same values in some of the features shown in *Table* 4. Those features are “building”, “floor”, “system”, and the “short info”. Each alarm also has an acknowledge time, indicating when the alarm is acknowledged (or known by the facility manager). Currently, the facility managers leave the BAS to control the systems instead for most of the times. So, the duration of the "off normal" status can be used to evaluate the impacts of the alarms. Thus, the goal of this step is to get the duration of alarms. To do this, an R script is written. The script takes the cleaned dataframe from 0. It has two loops; the outer loop checks the status of each instance. If the status of a certain instance is "off normal", an inner loop begins to look for the corresponding "normal" status instance. Once the "normal" instance is found, the program calculates the time difference between those two instances to get the duration of that alarm. The output of this script is a dataframe with all the "off normal" instances (alarm) with their durations. After this step, all the "off normal" status instances are extracted. The dataframe is saved into a new csv file.



status	building	floor	system	short_info	range	description	occur
NORMAL	SCSC Gates	Fifth Floor	VAV Room 5205 Classroom	ZCO2_HI	Universal	The zone CO2 level	Feb 18 2010 10:00:00 PM
NORMAL	SCSC Gates	Fourth Floor	VAV Room 4211 Classroom	ZCO2_HI	Universal	The zone CO2 level	Feb 18 2010 10:00:00 PM
NORMAL	SCSC Gates	Fourth Floor	VAV Room 4215 Classroom	ZCO2_HI	Universal	The zone CO2 level	Feb 18 2010 10:00:00 PM
OFF NORMAL	SCSC Gates	Fourth Floor	VAV Room 4215 Classroom	ZCO2_HI	Universal	The zone CO2 level	Feb 18 2010 11:03:22 PM
OFF NORMAL	SCSC Gates	Fourth Floor	VAV Room 4211 Classroom	ZCO2_HI	Universal	The zone CO2 level	Feb 19 2010 12:30:24 AM
OFF NORMAL	SCSC Gates	Fourth Floor	VAV Room 4102 Classroom	ZCO2_HI	Universal	The zone CO2 level	Feb 19 2010 1:37:36 AM
OFF NORMAL	SCSC Gates	Fifth Floor	VAV Room 5222 Classroom	ZCO2_HI	Universal	The zone CO2 level	Feb 19 2010 1:54:38 AM
NORMAL	SCSC Gates	Fourth Floor	VAV Room 4215 Classroom	ZCO2_HI	Universal	The zone CO2 level	Feb 19 2010 2:23:29 AM
OFF NORMAL	SCSC Gates	Fifth Floor	VAV Room 5201 Classroom	ZCO2_HI	Universal	The zone CO2 level	Feb 19 2010 7:23:01 AM
OFF NORMAL	SCSC Gates	Fourth Floor	VAV Room 4215 Classroom	ZCO2_HI	Universal	The zone CO2 level	Feb 19 2010 7:36:27 AM
OFF NORMAL	SCSC Gates	Fifth Floor	VAV Room 5208 Conference	ZCO2_HI	Universal	The zone CO2 level	Feb 19 2010 7:40:00 AM
OFF NORMAL	SCSC Gates	Eighth Floor	VAV Room 8228 Project	ZCO2_HI	Universal	The zone CO2 level	Feb 19 2010 7:40:00 AM

Figure 9. Calculate alarm durations

5.2.3 Reconfigure feature space

Currently, the dataframe has 11 features. However, many of them are some text descriptions which is hard for data mining tasks. So, the goal of this step is to reconfigure the feature space, extract short and meaningful features from current long descriptions. For example, the alarm occurring time consists of year, month, day, and time. It can be reconfigured as season, weekday/weekend, occupied hour/unoccupied hour. And the system type of the alarms can be grouped by their affected objects. Like "VAV Room Office" is occupant-related while "FCU-3" is equipment related, and "Emergency generator" is a critical condition. The occupant-related system types can be further grouped by their space types, such as "Office", "Classroom", "Conference", "Common space", etc.

First, several time-related features including season, day of week, occupied hour or not, week day or not are added. Each alarm has an "occur" feature, which consists of the occurring year, date, and time. This feature can be reconfigured into the features below:

Table 5. Reconfigured Alarm Occurring Time Features

Number	Feature	Levels	Meaning
1	occur.season	Winter -- Jan, Feb, Dec	The season when the alarm occurs
		Spring -- Mar, Apr, May	
		Summer -- Jun, Jul, Aug	
		Autumn -- Sep, Oct, Nov	
2	occur.day	Sunday ~ Saturday	The day of a week
3	occur.occupied	Yes, No	Whether the alarm occurs in occupied hour

Second, the relationship among Air Handling Units (AHUs) and end-uses can be mapped. The BAS dashboard provides two tree-views of the AHU and end-uses in its "Equipment Sources" display. *Figure 10* shows a snapshot of the tree-view. However, this information is not displayed in the alarm management tools. Since AHUs directly serve the end-uses, this mapping information is very important for detecting systematic faults in some AHUs and end-uses.

Copy to Clipboard

Heat tree

1 **(AHU-3)** - #ahu-3 (scott_hall_ahu_3)
- 1.1 **Cooridor 4S200 (VAV-4-41)** - #cooridor_4s200_vav-4-41 (scott_hall_vav_rh_fintube_2_temps)
- 1.2 **Corridor 4S300 (VAV-4-38)** - #corridor_4s300_vav-4-38 (scott_hall_vav_reheat_2_no_occ)
- 1.3 **Lobby 4S417 (VAV-4-44)** - #lobby_4s417_vav-4-44 (scott_hall_vav_rh_huh_finntube)
- 1.4 **Office 4S403, 4S405, 4S407 (VAV-4-40)** - #office_4s403_4s405_4s407_vav-4-40 (scott_hall_vav_reheat_3_temps)
- 1.5 **Dirty Tool Maintenance 4S411 (VAV-4-43)** - #dirty_tool_maintenance_4s411_vav-4-43 (scott_hall_vav_reheat_2)
- 1.6 **Tool Prototype Devel. 4S409 (VAV-4-42)** - #tool_prototype_devel_4s409_vav-4-42 (scott_hall_vav_reheat_2)
- 1.7 **Spare Parts 4S401 (VAV-4-39)** - #spare_parts_4s401_vav-4-39 (scott_hall_vav_reheat_2)

2 **RM 2251 AHU-2-2 MER** - #ahu-2-2_mer_2212 (hamburg_hall_ahu_2-2_rev1)
- 2.1 **RM 1203 Classroom 1** - #rm_1203_classroom_1 (hamburg_hall_vav_reheat_rad_co2_rev1)
- 2.2 **RM 1204 Classroom 2** - #rm_1204_classroom_2 (hamburg_hall_vav_reheat_rad_co2_rev1)
- 2.3 **RM 1205 Classroom 3** - #rm_1205_classroom_3 (hamburg_hall_vav_reheat_rad_co2_rev1)
- 2.4 **RM 1206 Presentation Classroom** - #rm_1206_presentation_classroom (hamburg_hall_vav_reheat_rad_co2_rev1)
- 2.5 **RM 2205 Office VAV2-26** - #rm_2206_office (hamburg_hall_vav_reheat_rad_rev1)
- 2.6 **RM 2203 Office VAV 2-25** - #rm_2204_office (hamburg_hall_vav_reheat_rad_rev1)
- 2.7 **RM 2202 Office VAV2-23** - #rm_2201_office (hamburg_hall_vav_reheat_rad_rev1)
- 2.8 **RM 2204 Office VAV 2-22** - #rm_2203_office (hamburg_hall_vav_reheat_rad_rev1)
- 2.9 **RM 2206 Office VAV 2-21** - #rm_220511_office (hamburg_hall_vav_reheat_rad_rev1)
- 2.10 **RM 2208 Office VAV 2-19** - #rm_2207_office (hamburg_hall_vav_reheat_rad_rev1)
- 2.11 **RM 2210 PhD Meeting VAV 2-18** - #rm_2209_phd_meeting (hamburg_hall_vav_reheat_rad_rev1)
- 2.12 **RM 2212 Office VAV 2-17** - #rm_2211_office (hamburg_hall_vav_reheat_rad_rev1)
- 2.13 **RM 2214 office VAV 2-14** - #rm_2213_office (hamburg_hall_vav_reheat_rad_rev1)
- 2.14 **RM 2216 Office VAV 2-12** - #rm_2215_office (hamburg_hall_vav_reheat_rad_rev1)
- 2.15 **RM 2218 Office VAV 2-10** - #rm_2217_office (hamburg_hall_vav_reheat_rad_rev1)
- 2.16 **RM 2220 Office VAV 2-8** - #rm_2219_office (hamburg_hall_vav_reheat_rad_rev1)
- 2.17 **RM 2222 Office VAV 2-6** - #rm_2221_office (hamburg_hall_vav_reheat_rad_rev1)
- 2.18 **RM 2224 Office VAV 2-5** - #rm_2223_office (hamburg_hall_vav_reheat_rad_rev1)

Figure 10. Tree view of AHU and end-uses from BAS dashboard

The AHUs are added to the alarms as a new feature. As the figure above shows, each AHU is responsible for many rooms and VAV boxes. Each alarm also has a “system” feature which indicates its room type and room number. Thus, the AHU serving a room can be searched and attached to the alarms that occurred in that room. This is accomplished with some R scripts which are included in the Appendices.

Third, two new features: "type", and "affected.object" are created. The value of this feature for each instance is determined by the value of its "system" feature. As discussed before, the feature "system" has too many labels. For example, there are more than a hundred office rooms in Gate-Hillman center. Each of the office room has a unique name, which makes the feature too many possible values. However, it's not practical to have such huge number of levels in data mining process. Only space type matters in this case. Thus, two new features are created. Below are the definitions of the features. The first feature is “type”, which indicates the space type of equipment type.

Table 6 Definition of “Type” Feature

Feature	Levels	Examples
type	Office	Office, Dean's Suite
	Conference	Conference, Future Use
	Classroom	Classroom, Project, Reading
	Common	Common, Cafe, Corridor, Bridge, Carrell, Lobby, Collaborate, Nursing
	Service	Kitchenette, Copy, Storage, Reception
	HVAC.equip	AHU, EF, FCU
	CRAC	CRAC
	Sensor.Meter	Sensor, Meter, Monitoring
	Critical	Emergency Generator, Garage Exhaust, Electrical Closet, Chilled Water System, Hot Water System, Rainwater System, Elevator Vent

Another new feature "affected.object" is also created. It denotes whether the affected object is occupant or equipment.

Table 7. Definition of “Affected.object” Feature

Feature	Levels	Examples
affected.object	Occupant	Office, Conference, Classroom, Common, Service
	Equipment	HVAC.equip, Sensor.Meter, CRAC
	Critical	Critical

5.2.4 Add weather data

Fourth, another important factor that could affect the system operation is the weather condition. To examine the possible relationships between weather condition and alarm patterns, several weather features are added to the dataframe. Both daily and hourly weather data are downloaded. The daily data contains weather information such as temperature, humidity, visibility, wind speed, precipitation, and other weather events (e.g. rain, snow, thunderstorm, hail). The historic data from 2010 to 2016 is manually downloaded and processed before combining to the alarm dataframe. And some R scripts are used to combine weather condition to the alarm dataframe. Shows the features after combining weather conditions.

Finally, after the feature reconfigurations, the original alarm descriptions are divided into algorithm readable features. The relationships among AHUs and end-uses are also added to the alarm dataframe. Environmental factors include daily and hourly weather conditions are added to each alarm as several new features. shows the features of the alarm dataframe.

Table 8. Reconfigured Feature Space

Number	Feature	Meaning	Type
1	building	The building where the alarm occurs	Nominal
2	floor	The floor where the alarm occurs	Nominal
3	short.info	Short description of the alarm	Nominal
4	range	The range of the alarm	Nominal
5	occur.date	The date when the alarm is raised by BAS	Date
6	occur.time	The time when the alarm is raised by BAS	Time
7	occur.season	The season.	Nominal
8	occur.day	The day of week. (Mon, Tue, etc.)	Nominal
9	occur.occupied	If the building is occupied hour when the alarm occurs	Nominal
10	AHU	The AHU that serves the end-use in alarm description	Nominal
11	type	The type of the space	Nominal
12	affected.object	The affected object. (Occupant, equipment)	Nominal
13	rain	If it is rainy that day	Nominal
14	snow	If it's snowing that day.	Nominal
15	thunderstorm	If there is thunderstorm during that day.	Nominal
16	fog	If it's foggy that day.	Nominal
17	hail	If there is hail during that day.	Nominal
18	avg.temp.day	The average temperature of that day	Numeric
19	air.temp.hour	The average temperature of that hour	Numeric
20	air.rh	the average relative humidity of that hour	Numeric
21	wind.drct	The wind direction of that hour	Numeric
22	wind.speed	The wind speed of that hour	Numeric
23	precip.hour	The total precipitation of that hour	Numeric
24	gust	The gust in that hour	Numeric
25	avg.humidity	The average relative humidity of that day	Numeric
26	avg.wind	The average wind speed of that day	Numeric
27	avg.precip	The average precipitation of that day	Numeric
28	manual.time	The time between an alarm occurs and manual acknowledgement.	Numeric
29	duration	The time between an alarm occurs and released by BAS	Numeric

5.3 Alarms Categorization

The BAS has the basic alarm categorization function which allows facility managers to view alarms in certain categories. *Figure* 11 shows the alarm categories. However, the rules that define the categories are developed by the BAS solution provider which is invisible to building operators. The user interface does not provide any explanation of the alarm root causes or actionable information.

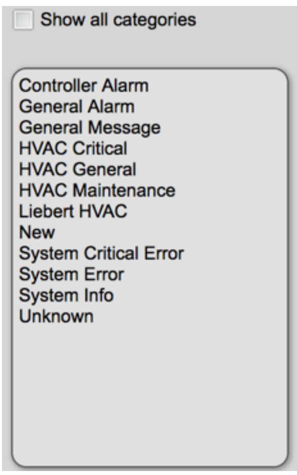


Figure 11. Original alarm categorization of the BAS

Moreover, the existing alarm categorization function does not distinguish similar alarms clearly. For example, Figure 12 and Figure 13 indicate two alarms with similar information. Both alarm have the "Zone temperature is too cool" annotations. But the first one is categorized as "HVAC Critical" and the second one is categorized as "HVAC General". None of the two alarms displays the components operation status, occupant status, and measured zone temperature. Without any additional information, it is hard to distinguish those two alarms and make operation and maintenance decisions.

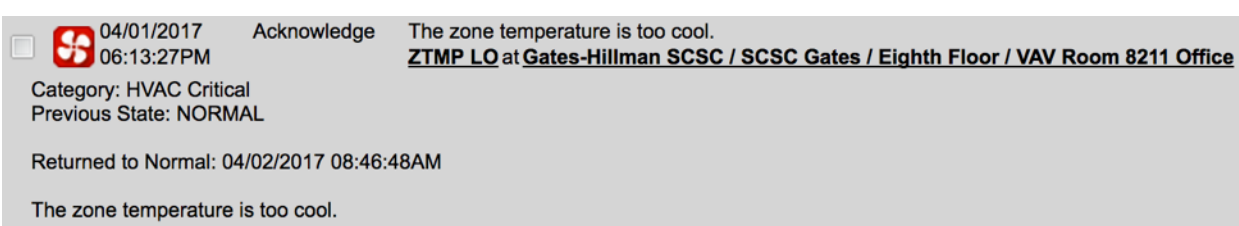


Figure 12. "Zone temperature too cool" alarm in HVAC Critical category

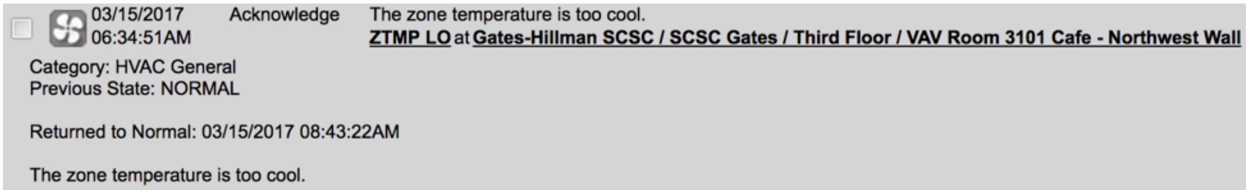


Figure 13. "Zone temperature too cool" alarm in HVAC General category

Given the existing situation, it is necessary to categorize the alarms in a meaningful way such that it allows deeper analysis in alarms' impacts. After manual exploration, the alarms are assigned into three categories: critical operation, equipment-related, and occupant-related. Alarms from those three categories have different direct impacts.

Critical operation alarms are related to safety and critical operations. Examples are hot water system failure, emergency generator down, elevator fault. Those alarms have low frequency, but can lead to serious results if ignored. Alarms in this category should not be simply ignored, but more detailed actionable information is needed.

Equipment-related alarms are from HVAC system, lighting system, sensors and meters. For example, an equipment-related alarm can be "Fan-hand alarm" which indicates the fan operation schedule may be incorrectly overwritten. It can also be "Static pressure too high in AHU", "Freezestat triggered", or "BACnet errors". Those alarms have the highest frequency, but no direct impacts on indoor environment quality and energy consumption. For instance, the "Freezestat triggered" alarm indicates the temperature of a heat exchanger is too low. It acts as a self-protection for the AHUs. But this alarm is very common in winters simply because the outdoor air temperature is low in winter. Since the BAS continuously monitors the status of freeze stats, it generates an "off-normal" alarm whenever there is an abnormal reading and another "normal" alarm whenever the reading backs to normal. Although the equipment-related alarms are the majority, most of them do not have high impacts and may inundate other important alarms. Most of the alarms in this category could be filtered out.

Occupant-related alarms are alarms that may directly affect occupant comfort. For example, it can be "Zone temperature too low/high", "Zone CO2 level too high". Although those alarms have low frequency compare to the equipment-related alarms, their potential impacts have a large variance. For example, a "Zone temperature too high" alarm in an office room is likely to have higher impacts than a “Zone CO2 level too high” alarm in a corridor. Thus, a deeper analysis is needed to prioritize alarms in this category.

In summary, the alarms could be assigned into different categories based on their affected objects. Below are some general findings from the alarm categorization. The brief understanding of the alarm distribution is helpful in quantifying their impacts in next steps.

5.4 Alarm Impact Quantification

5.4.1 Energy Impact

Due to the lack of energy metering and sub-metering in Gates Hillman Center, there is no direct way of evaluating the energy impacts of the alarms. However, since most of the spaces in the building are served by VAV terminal units. The zone air temperature setpoints, discharge air flow rate setpoints, temperature thresholds that cause an alarm can be acquired from the BAS dashboard. The real discharge air flow rate and discharge air temperature are also available on the PI Coresight system. Thus, the first principle formulas can be used to evaluate the energy impacts at the component scale. For a VAV terminal, the energy transfer between discharging air and the room air is:

Equation 1. Energy transfer between discharging air and room air

$$\dot{Q} = \dot{V} C_v (t_{discharge} - t_{room})$$

where:

$\dot{V}$ -- the volumetric flow rate.

C_v -- the volumetric specific heat capacity of air.

$t_{discharge}$ -- the VAV terminal discharge air temperature

t_{room} -- the room air temperature

Because of the abnormal condition, the VAV terminal discharge air temperature and flow rate are different from the normal situations. Therefore, the energy saving or waste for a single VAV terminal can be acquired by:

Equation 2 Energy transfer difference between normal condition and alarm condition

$$\Delta\dot{Q} = \dot{Q}_{alarm} - \dot{Q}_{normal}$$

Because the target building has more than 320 conditioned rooms with different functions, it is impossible to check the real time discharging air flow rate and temperature. A small portion of the rooms are sampled (n=37) and their discharging air flow rate and temperatures at normal and alarm conditions are recorded.

Table 9. Space Sample and population size

Room type	Sample size	Total number
Office	20	216
Classroom	5	40
Conference	3	12
Common space	5	44
Service area	3	14

The steps of sampling are: First, choose the sample spaces evenly by their type, floor and orientation. The yellow dots in Figure x to figure x indicate the spaces sampled.

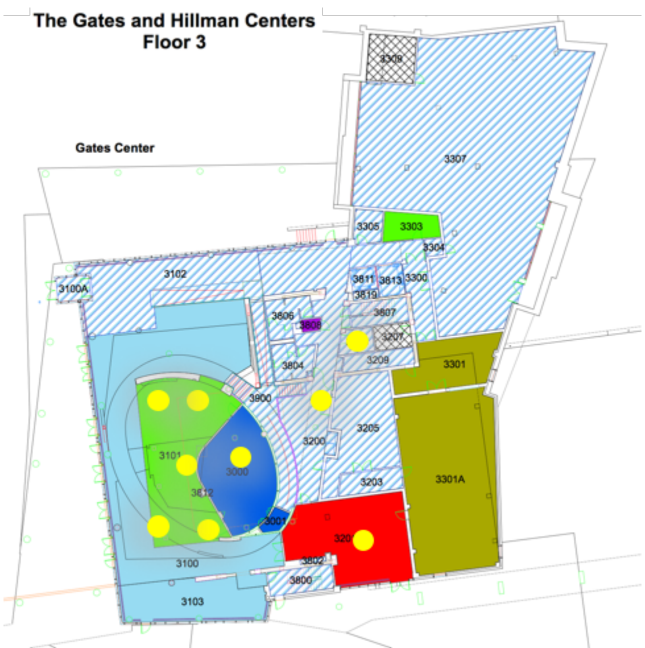


Figure 14. Sampled spaces on third floor

Figure 15. Sampled spaces on fourth floor

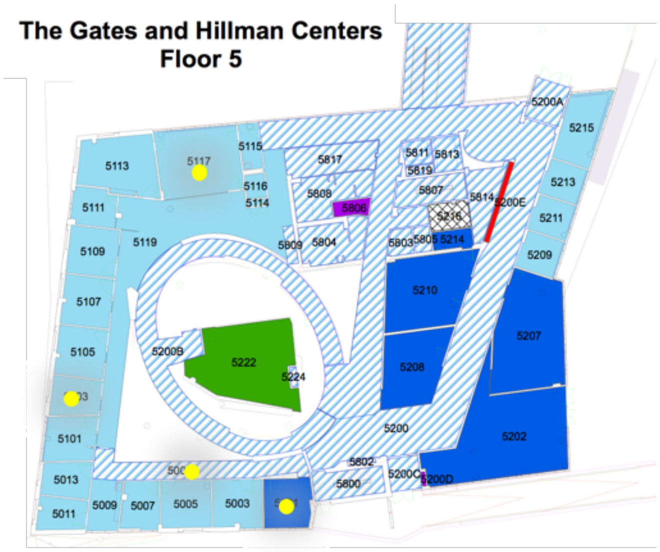


Figure 16. Sampled spaces on fifth floor

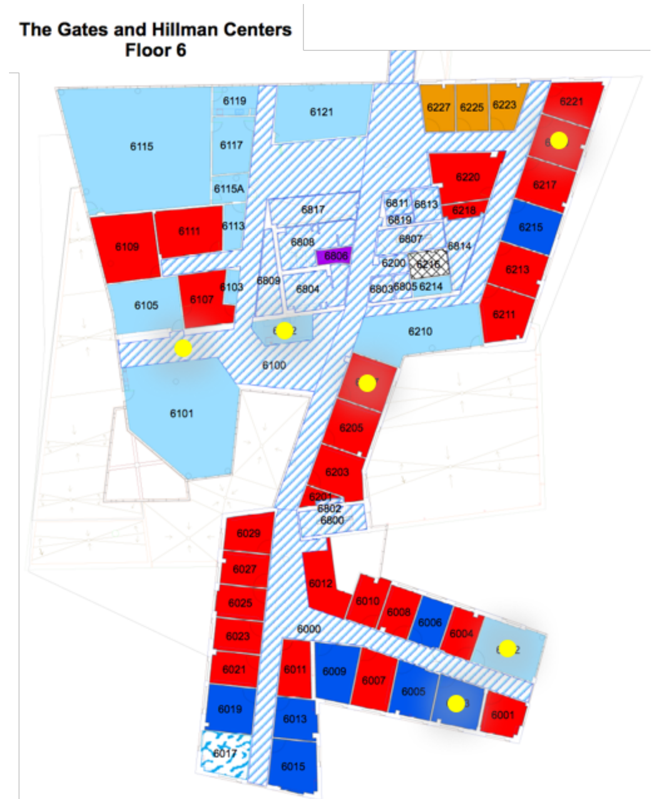


Figure 17. Sampled spaces on sixth floor

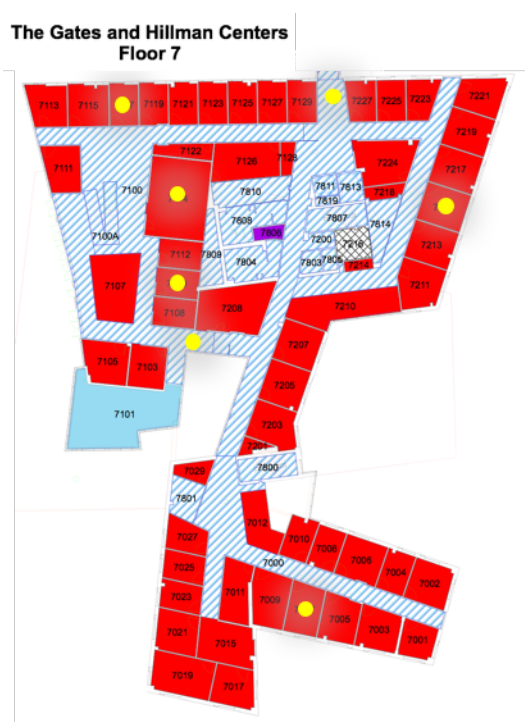


Figure 18. Sampled spaces on seventh floor

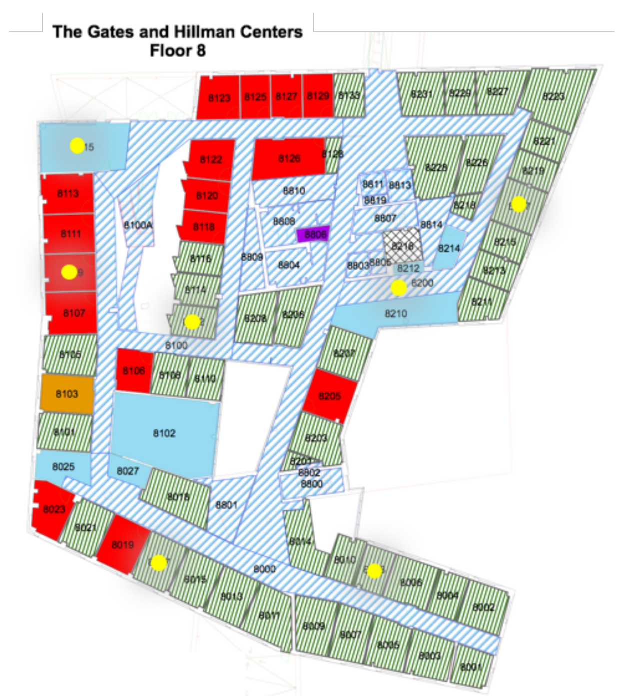


Figure 19. Sampled spaces on eighth floor

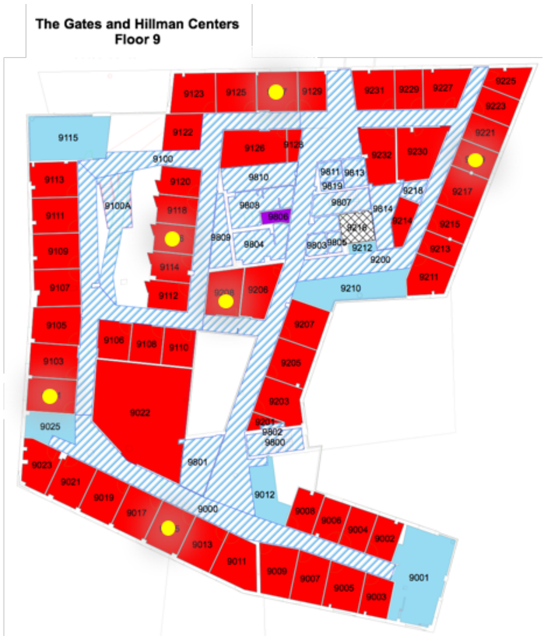


Figure 20. Sampled spaces on ninth floor

Second, look up the zone air temperature setpoints, high temperature and low temperature alarm thresholds for the sample spaces on the BAS dashboard.

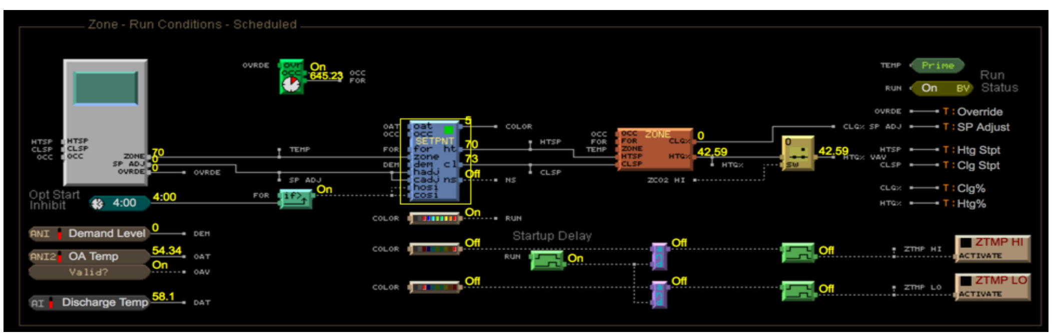


Figure 21. Zone setpoints and schedule interface on BAS dashboard

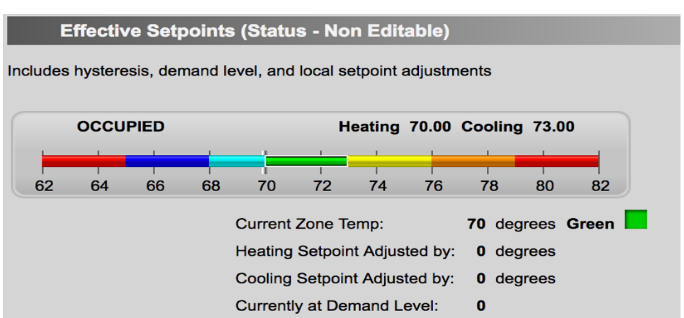


Figure 22. Setpoints and alarm thresholds of a sampled space

Third, look up the design maximum coiling and heating discharge air flow rates for the sample spaces on the BAS dashboard.

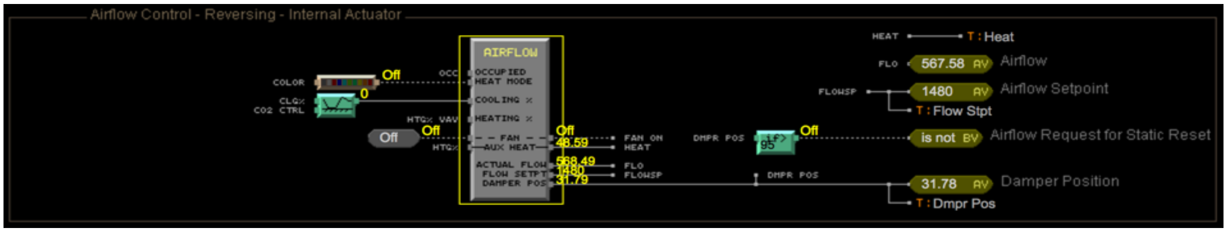


Figure 23. Zone airflow control interface on BAS dashboard

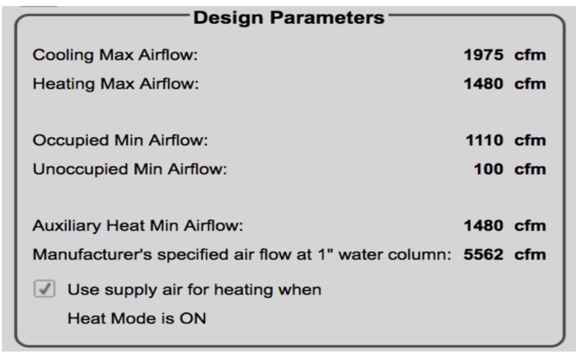


Figure 24. Design maximum cooling and heating discharge airflow rates of a sampled space

Fourth, check the historic discharge air temperatures and discharge air flow rates under normal condition and alarm condition from PI Coresight for the sample spaces.

Figure 25. Historic discharge air temperature of sampled spaces

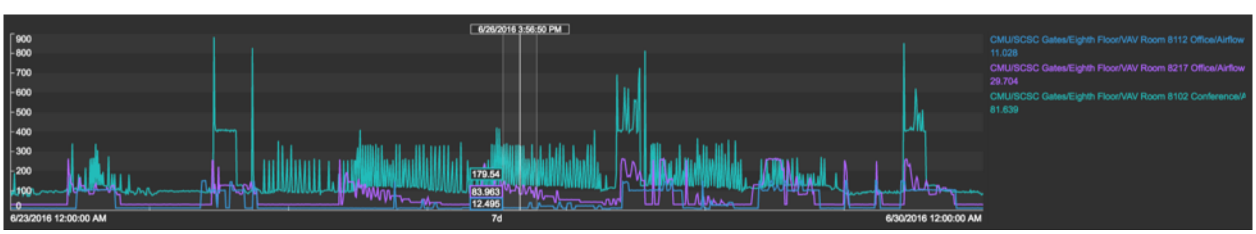


Figure 26. Historic discharge airflow rate of sampled spaces

With the data collected from the sampling, the mean values of discharging air flow rate and temperature under normal and alarm conditions can be calculated. According to the sample results, the discharge air temperature and airflow rates under each condition are shown in *Table* 10. The detailed sample results can be found in Table 27 in appendix.

Table 10. Discharge air temperature flow rates under normal and alarm conditions

Discharge Air Temperature				Discharge Air Flow Rate (% of max design rate)			
Heating		Cooling		Heating		Cooling	
normal	alarm	normal	alarm	normal	alarm	normal	alarm
87.1	109.2	60.2	55.9	92%	107%	48%	77%

With the VAV terminal unit discharge air temperatures, zone temperature setpoints, alarm temperature thresholds, and design discharge airflow rate, the energy consumption rate under normal condition and alarm condition can be calculated with Equation 1. Then, the difference between two alarm condition and normal condition can be calculated using Equation 2.

5.4.2 Thermal Comfort Impact

Thermal comfort can be quantified with Predicted Mean Vote (PMV) and Predicted Percentage Dissatisfied (PPD). According to Fanger's thermal comfort model, PMV is a function of six parameters:

Equation 3 Fanger's PMV equation

$$PMV = f(Met, clo, t_a, t_r, rh, v)$$

where:

Met – metabolic rate

clo – clothing factor

t_a – air dry bulb temperature (˚C)

t_r – mean radiant temperature (˚C)

rh – relative humidity of air (%)

v – local air velocity (m/s)

The room temperature setpoints and alarm thresholds are collected from the BAS dashboard in previous steps. Metabolic rate and clothing factor can be assumed per space type and season. Air relative humidity and local air velocity can also be assumed. Table 11 shows the input parameters for PMV calculations.

Table 11. PMV calculation input parameters

Mode	Cooling		Heating	
Space Type	Office, Classroom	Service, Common	Office, Classroom	Service, Common
Base Metabolic Rate (W)	58.15	58.15	58.15	58.15
Relative Metabolic Rate	1.1	1.2	1.1	1.2
Clothing Factor	0.5	0.5	1	1
Air Dry-bulb Temperature (˚C)	Assumed to be the average of the zone heating and cooling setpoints			
Mean Radiant Temperature (˚C)	Assumed to be the average of the zone heating and cooling setpoints			
Air Relative Humidity (%)	50%	50%	40%	40%

5.5 Alarm Duration Prediction

5.5.1 Overview

The transient energy and thermal comfort impacts of the alarms provide two interesting metrics for the alarm prioritization. However, the long-term impacts of the alarms are not embodied by them. For example, an ephemeral alarm with high transient energy and thermal comfort impacts is less severe than an alarm with low transient energy and thermal comfort impacts but a long duration.

As mentioned in 7.1.2, the duration of an alarm is the time interval between the timestamp when the alarm occurred and the timestamp when the alarm is closed by the BAS. The alarm durations range from several minutes to tens of days. Figure 27 shows the density plot of the alarms in different spaces. The durations of the most alarms are within 6 hours. But the density distributions of the alarm duration vary from space to space.

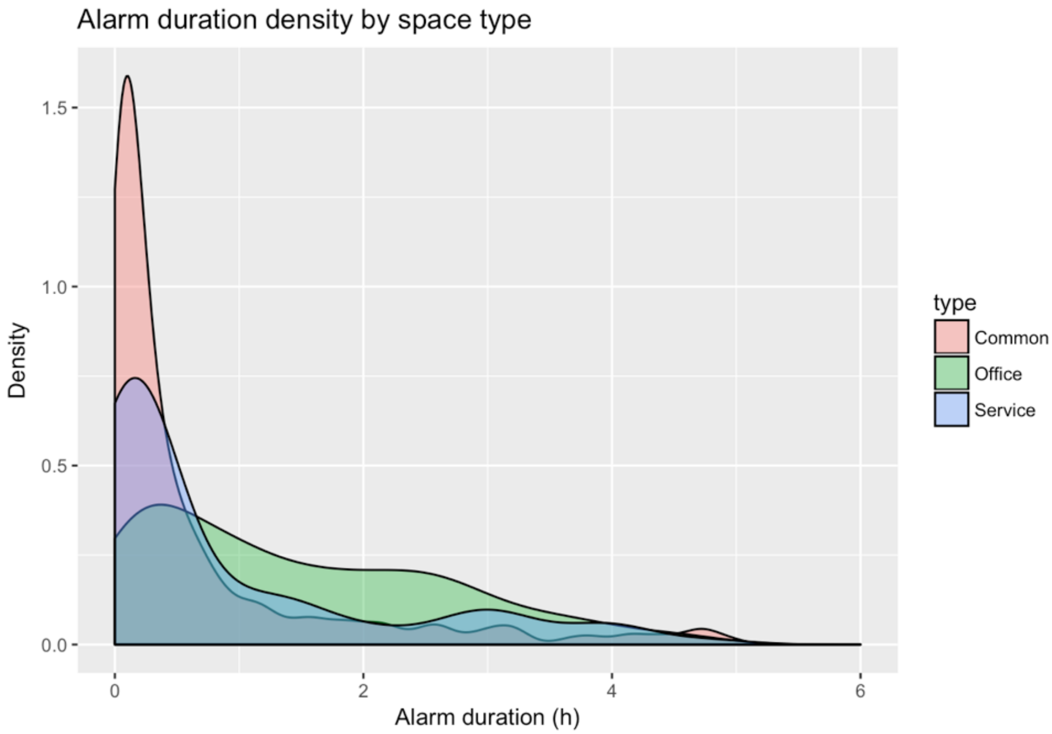


Figure 27 Alarm duration density by space type

Alarms occurred in common spaces and service areas have more concentrated density distribution than alarms occurred in offices. More than 80% of the alarms in those two types of spaces are within 2 hours. Whereas the alarms occurred in offices have a flat density distribution. This means the alarms occurred in offices have a larger variance in their duration. With the duration information, the alarms could be further prioritized. A set of rules that combines energy impacts, thermal comfort impacts and durations could be created. *Table 12* Shows an example of the rule set.

Table 12. An example of alarm prioritization rule set

Space Type	Energy Impact (kW)	PMV Impact (absolute value)	Alarm Duration	Priority
Office, Classroom	-	> 1.5	> 30 min	high
Common space	0~2	0~2	< 1d	low
Common space	4~6	0~2	> 1d	high
Service	0~2	> 1.5	2h ~ 5h	medium

5.5.2 Duration Prediction

As discussed in Chapter 5, data mining is a technique that helps to find hidden patterns behind large datasets and analyze the relationship among data features. The classification algorithms in applied machine learning are perfect tools for the alarm duration prediction. The machine learning model training process includes data preparation, data exploration, iterative model building, parameter tuning, and model evaluation. Figure 28 shows the workflow of the model training process. The model training and evaluation in this section is accomplished with an open-source data mining software named *Weka*.

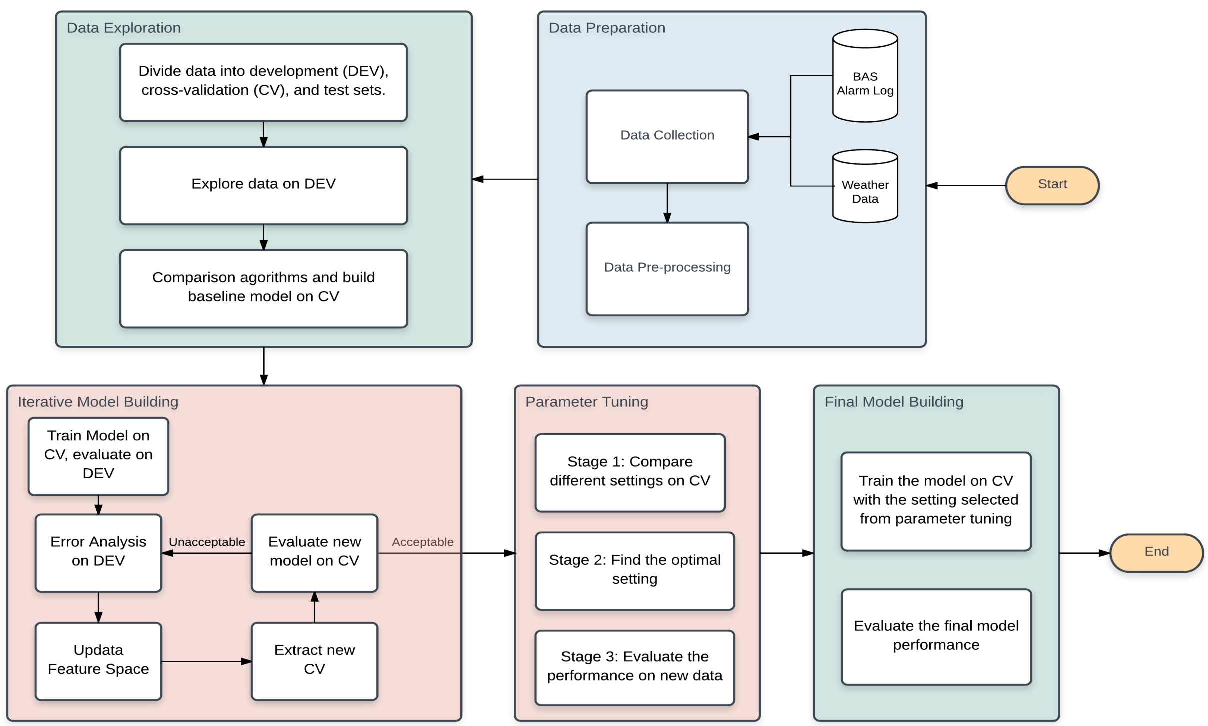


Figure 28. Workflow of machine learning model training

Since the data cleansing and feature engineering are already done in previous steps. The following sections in 7.5 focus on the data exploration, iterative model building, and parameter tuning processes. The final model evaluations are discussed in 8.2.

5.5.3 Data Exploration

Through the exploration of the whole dataset, 30834 out of 43,502 alarms are related to the Computer Room Air Conditioner (CRAC) on third floor. This specific type of alarm could seriously bias the model. Thus, they are discarded from the whole set. Similarly, 4,118 alarms with abnormal class values or missing values are discarded. Therefore, the data used for the model training is a subset with the size of 8,550. After removing the problematic instances, the new dataset is divided into three subsets—development set, cross-validation set, and final test set. The development set is used for evaluating the model trained on the cross-validation set. The cross-validation set is used train models during the iterative model building process. The final test set is used for evaluating the final model performance after iterative model building and parameter tuning process. The three subsets are extracted from the whole dataset with supervision to keep the same distribution as the original dataset. Table 13 shows the uses and sizes of the three subsets.

Table 13. Data division for machine learning model training

Dataset	Use	Size
Development set	Explore features, evaluate modeling assumptions	1425
Cross-validation set	Train models during iterative model building process	5700
Final test set	Evaluate final model performance	1425

5.5.4 Algorithms Comparison

After the basic data exploration, the next step is select an algorithm for the alarm duration prediction. Four commonly used classification algorithms are compared: Naïve Bayes, Support Vector Machine, Decision Trees, and Decision Rules. The performances of those algorithms are evaluated over the cross-validation set with 10-fold cross-validation.

Table 14 shows the performances of the algorithms.

Table 14. Algorithms comparison over cross-validation set

Algorithm	Naïve Bayes	Support Vector Machine	Decision Trees	Decision Rules
Accuracy	63.7%	71.6%	74.2%	73.6%
Kappa statistic	0.422	0.558	0.61	0.555
RMSE	0.359	0.345	0.288	0.291

The comparison indicate that Decision Trees Algorithm has the highest accuracy (74.2%), the highest Kappa statistics, and the lowest Root Mean Squared Error (RMSE). Because Decision Trees algorithm is a divide-and-conquer approach, the further down the tree goes, the less data the algorithm is paying attention to. Thus, irrelevant features may confuse the model. Therefore, the performances of the Decision Trees algorithm with and without feature selection are compared. For the feature selection, chi-square is used to evaluate the features. Top 15 features are selected for the Decision Trees model. Table 15 shows the model performances with and without feature selection.

Table 15. Decision Trees model with and without feature selection

Algorithm	Decision Trees without Feature Selection	Decision Rules with Feature Selection
Accuracy	0.744	0.747
Kappa statistic	0.6	0.61
RMSE	0.292	0.29

The comparison shows a marginal improvement with the feature selection. In addition, the original data has only 23 features. It doesn't worth picking the top 15 features at the risk of losing potential important information. So, the Decision Trees without feature selection is selected as the baseline algorithm.

5.5.5 Iterative Model Building

With the best algorithm for this task found, the next step is to optimize the feature space through the iterative model building process. In this study, three rounds of error analysis are conducted. With the feature space reconfiguration, the model performance over the cross-validation set is raised from 74.4% to 78.6%. The big picture of this process is:

(1). Train a model on the cross-validation set;

(2). Evaluate the model’s performance on development set;

(3). Conduct an error analysis and identify problematic features;

(4). Update the feature space to reduce the number of mistakes;

(5). Train a model with the cross-validation set and evaluate its performance with a 10-fold cross-validation;

(6). If the performance is acceptable, stop feature space updating process. If the performance is unacceptable, go to step (2).

In the first round, the original alarm occurring time feature is in string format. The machine learning algorithm treats it as a nominal value which doesn't provide any temporal information. It is necessary to break it down into different features. *Figure 29* shows an example of the time feature before and after reconfiguration. After breaking down the time feature, the new model's accuracy over the cross-validation set has been increased from 74.4% to 77.4%.

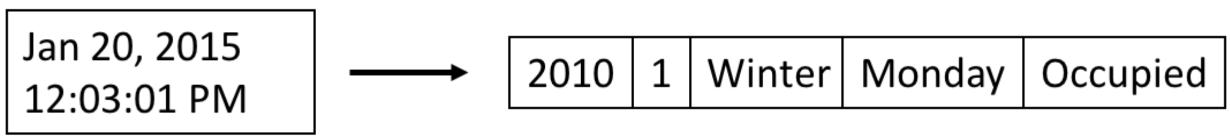


Figure 29. Example of time feature reconfiguration

In the second round, because the original "System" feature has too many different labels, it may confuse the Decision Trees algorithm. For example, it has labels such as "VAV Room 4007", "VAV Room 5017" and many similar rooms. This could lead to overfitting as there will be too many branches at the bottom of the decision tree. Moreover, there is no need to use such a detailed label for the prediction. Therefore, the original "System" feature is converted into "Type" feature, which indicate the space type where the alarm occurred. Meanwhile, an "AHU" feature is added to provide the corresponding AHU information. The evolution of the new feature space shows an accuracy of 77.8%. Although there is only a slight improvement from the first round, the new feature space could help to avoid overfitting.

In the third round, the weather features are inspected. During the error analysis, it was found that the qualitative features "Rain", "Snow", "Fog", "Thunderstorm", and "Hail" do not have high weights on the prediction. There values appear to have random relationships with the alarm durations. On the contrary, the quantitative features contribute more on the prediction. Therefore, the qualitative weather features are removed. And a further analysis has shown that for the outdoor air relative humidity, wind speed, and precipitation, hourly average values have more weights on the prediction than daily average values. Thus, for those two features, the daily average value features are removed. After the feature removal, the evaluation of the new model shows an accuracy of 78.6%.

5.5.6 Model Parameter Tuning

After the feature space being optimized, the next step is to optimize the model setting so that it performs the best on the feature space. For the Decision Trees algorithm, there are two parameters that can be tuned: (1). *confidenceFactor* C—the upper confidence limit of making an error, smaller C incurs more pruning of the decision tree; (2). *minNumObj* M—the minimum number of instances per leaf. The *CVParameter* wrapper in *Weka* is used to for the parameter tuning process. For *confidenceFactor,* the tuning is set from 0.1 to 0.9, with 9 steps. For *minNumObj*, the tuning is set from 1 to 10, with 10 steps. The wrapper first looks for the optimal setting for *confidenceFactor*, and then looks for the best setting for *minNumObj.* After the search, it evaluates the best combination of those two parameters over the cross-validation set through a 10-fold cross-validation.

The parameter tuning process indicates that the optimal setting is when *confidenceFactor=0.1,* and *minNumObj=2.* After parameter tuning, the model built on the cross-validation set has an accuracy of 79.2%. Since the accuracy of the model is significantly improved, it is worth doing the optimization. The final model will be built on the development and cross-validation combined set with the optimal parameter setting. The evaluation of the final model is discussed in 6.3.

6. Case Study & Results

6.1 Alarm Categorization Result

As discussed in Chapter 7, the alarms could be categorized based on their affected objects. *Table* 16 shows the categorization rules. The alarms are firstly divided into different space type groups. Then they are categorized into Occupant-related, Equipment-related, and Critical operation groups.

Table 16. Alarm categorization rules

Naming Convention of Alarms	Space Type	Category
VAV Room ####, VAV Room Office, Dean's Suite	Office	Occupant-related
Classroom, Project, Reading	Classroom	
VAV Room #### Conference, Future Use	Conference	
Café, Corridor, Bridge, Study Carrell, Lobby, Collaboration Space, Collaborative Common, Nursing	Common	
Kitchenette, Work/Copy/Print, Storage, Reception/Mail	Service	
AHU, FCU, CHW, HWS	HVAC Equipment	Equipment-related
Energy Meter, Monitoring, Sensor	Meters	
Emergency Generator, Garage Exhaust, Electrical Closet, CRAC, Chilled Water System, Hot Water System, Rainwater System	Critical	Critical Operation

With the alarms categorized, the next step is to investigate the patterns behind the alarms and quantify the impacts. The preliminary analysis from the alarms shows that over 87% of the total alarms are equipment-related alarms, and only 7.7% of them are occupant-related alarm.

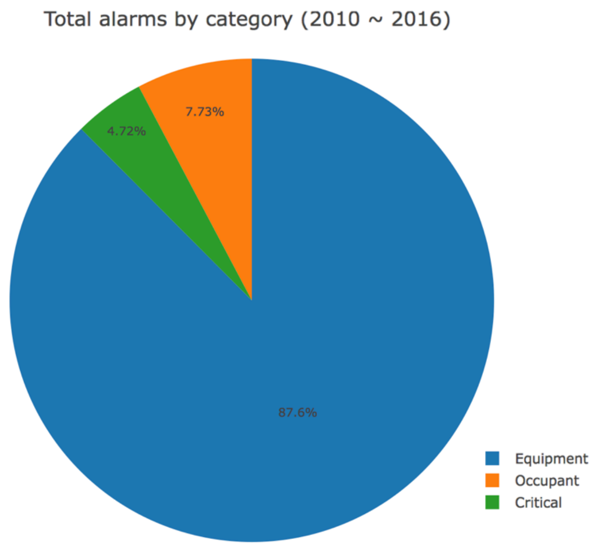


Figure 30. Alarm count by category

In addition to the total numbers, the trends of the alarms are also interesting. Figure 31 shows the trends of the alarms by category. The number of equipment-related alarms has been decreased dramatically over the years, which is the result of fine-tuning of the building systems. However, the number of occupant-related alarms and critical operation alarms remain stable. It means the fine tuning of the building only fixed some of the problems in relation to HVAC equipment.

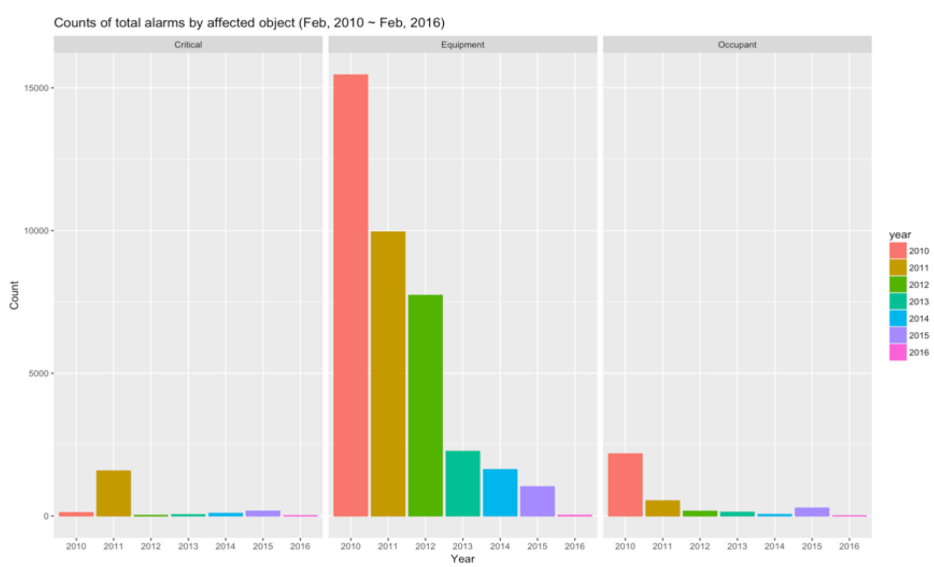


Figure 31. Alarm trends by category

As for the occupant-related alarms, Figure 32 shows the trends.

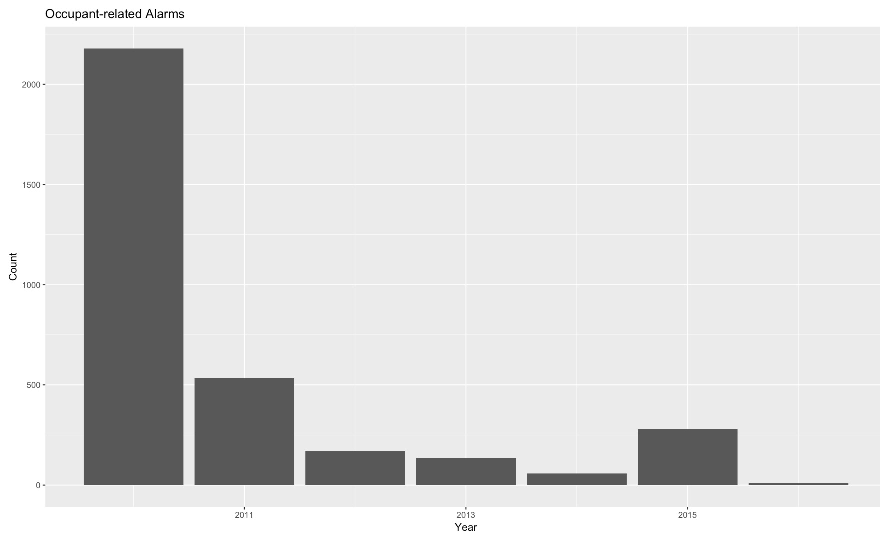


Figure 32. Occupant-related alarm trends

The number of the occupant-related alarms is decreased significantly from 2010 to 2011. After 2010, the average annual number is 240. In 2015, the number even increased. Compared to equipment-related alarms, occupant-related alarms have a larger uncertainty. This could be explained by the large diversity of alarm sources. Figure 33 shows the number of alarm sources by category on each floor.

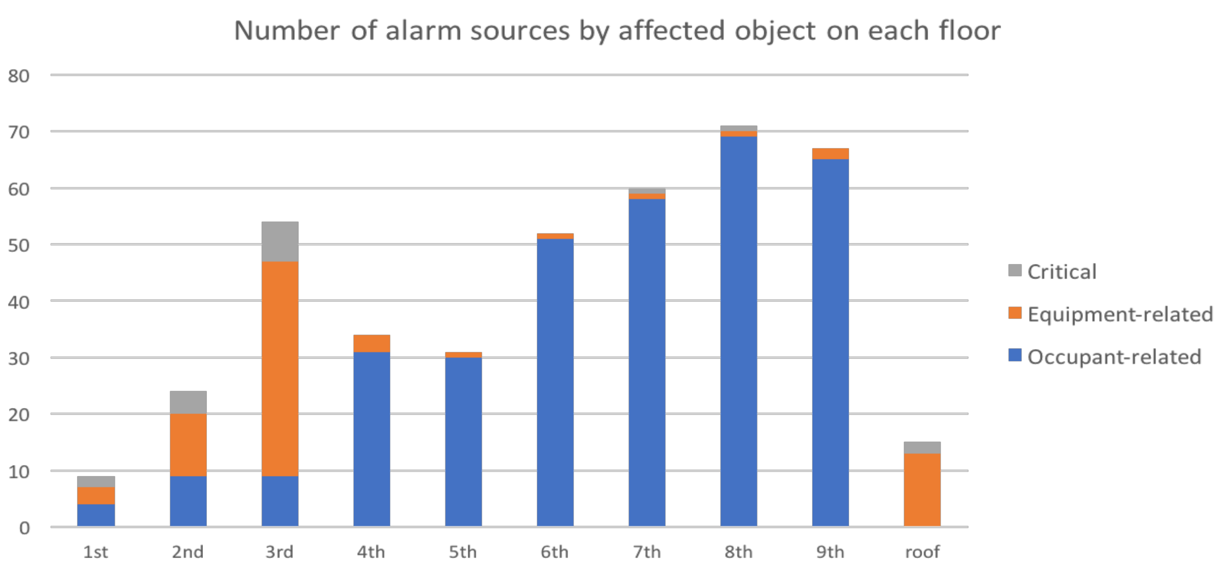


Figure 33. Number of alarm sources by category

Except for floor 2, 3 and roof, occupant-related alarms have the largest number of sources. The large number of alarm sources makes it more difficult to identify and fix the problems behind the alarms. Because equipment-related alarms have a small number of sources, fixing a few problems in the system could effectively reduce the number of the alarms. For example, there were over 14000 alarms occurred in the computer room air conditioner on third floor in 2010. After the problem was solved, the number of this alarm is decreased by 99.3% in 5 years. In 2015, the total number was only 95. However, given the large alarm source, it is challenging to eliminate occupant-related alarms without deep investigations.

6.2 Alarm Impact Quantification Result

6.2.1 Overview

There are 1043 alarms with direct impact on VAV terminal units' component energy consumption and indoor thermal quality during 2010 to 2016. The overview of those alarms' impact on energy consumption and thermal comfort are:

- More than half of the alarms (n=569, 53.6%) lead to higher energy consumption compared to normal operation, while others save energy.
- There were more alarms occurred in cooling mode (58.7%) than heating mode.
- There were more "Zone temperature too high" alarms (77.0%) than "Zone temperature too low" alarms.
- The average energy consumption rate increase is 2.2 kW with the range of -0.7 kW to 6.3 kW.
- The average thermal comfort impact (PMV) is -1.28 with the range of -3.12 to 1.26.

6.2.2 Analysis by Impact Category

Based on the relationship between the discharge air temperature and room air temperature, a VAV terminal unit can have higher or lower energy consumption rate under alarm condition compared to normal condition. The table below shows those different situations.

Table 17 Possible impacts of alarms under different situations

Alarm	Mode	Impact	
		Energy	Thermal
Zone temperature too high	Heating	Save	Hot
Zone temperature too high	Cooling	Waste	Hot
Zone temperature too low	Heating	Waste	Cold
Zone temperature too low	Cooling	Save	Cold

To visualize the impacts on energy and thermal comfort, the alarms are displayed in scatter plots where x-axis is the energy consumption rate compared to normal situation and y-axis is the PMV. Figure 34 below shows the alarms occurred cooling and heating modes. The blue dots stand for the alarms occurred in cooling season, and the red dots stand for the alarms occurred in heating season. Figure 35 shows the "Zone air temperature too high" alarms and "zone air temperature too low" alarms. The orange dots stand for the high zone air temperature alarms, and the cyan dots stand for low air temperature. The sizes of the dots indicate the duration of the alarms.

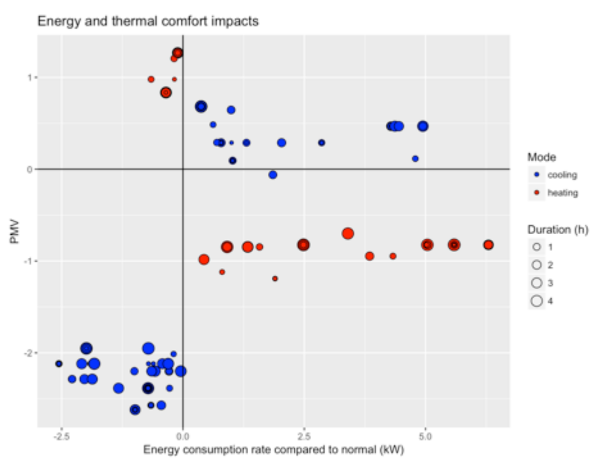


Figure 34. Alarm impacts distribution by space condition mode

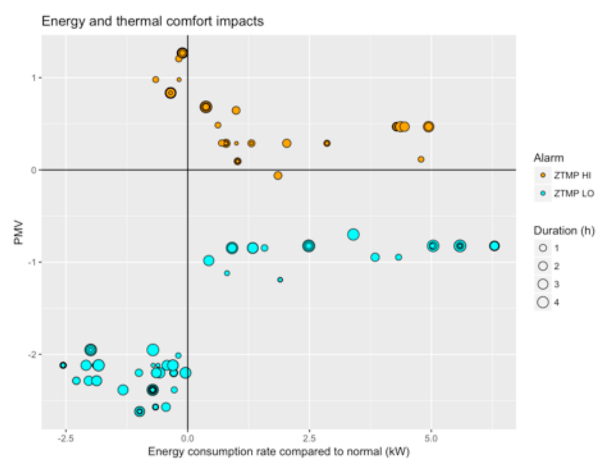


Figure 35. Alarm impacts distribution by alarm type

There are 431 alarms occurred in heating mode and 612 alarms occurred in cooling mode. The visualization results are consistent with the intuition. There are some alarms in both cooling and heating modes save energy, while others increase energy consumption compared to normal situation. 30.4% of the alarms occurred in cooling mode and 88.2% of the alarms occurred in heating mode consume more energy compared to normal situation. As the figure on the left indicate, the alarms with positive energy impacts (consume more energy) have lower impacts on thermal comfort. And the alarms with negative energy impacts have different impacts on thermal comfort. The dots on the left bottom of the plot are alarms occurred in cooling season and have the “zone temperature too low” alert. Although they have negative impact on energy consumption, their impacts on thermal comfort are very high. The dots on the right side of the plot have low impacts on thermal comfort, but have high impacts on energy consumption. To prioritize the alarms, they could be ranked by the overall impacts on thermal comfort and energy consumption. Figure 36 and Figure 37show the ranking of alarms by their thermal comfort impacts, and energy consumption impacts, respectively. The alarms are ranked with a scale of 0 to 4. The grey dots have the lowest impacts, while the red dots have the highest impacts.

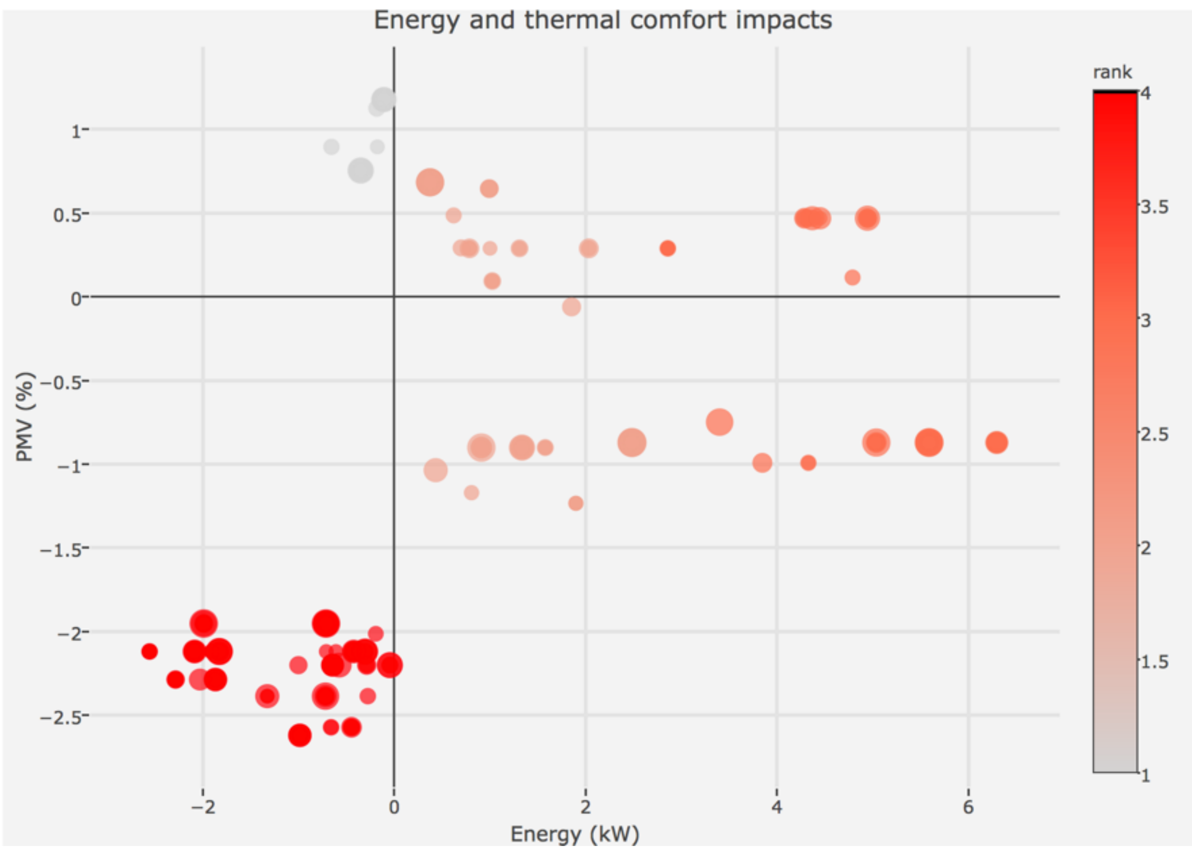


Figure 36. Alarm ranking by thermal comfort impacts

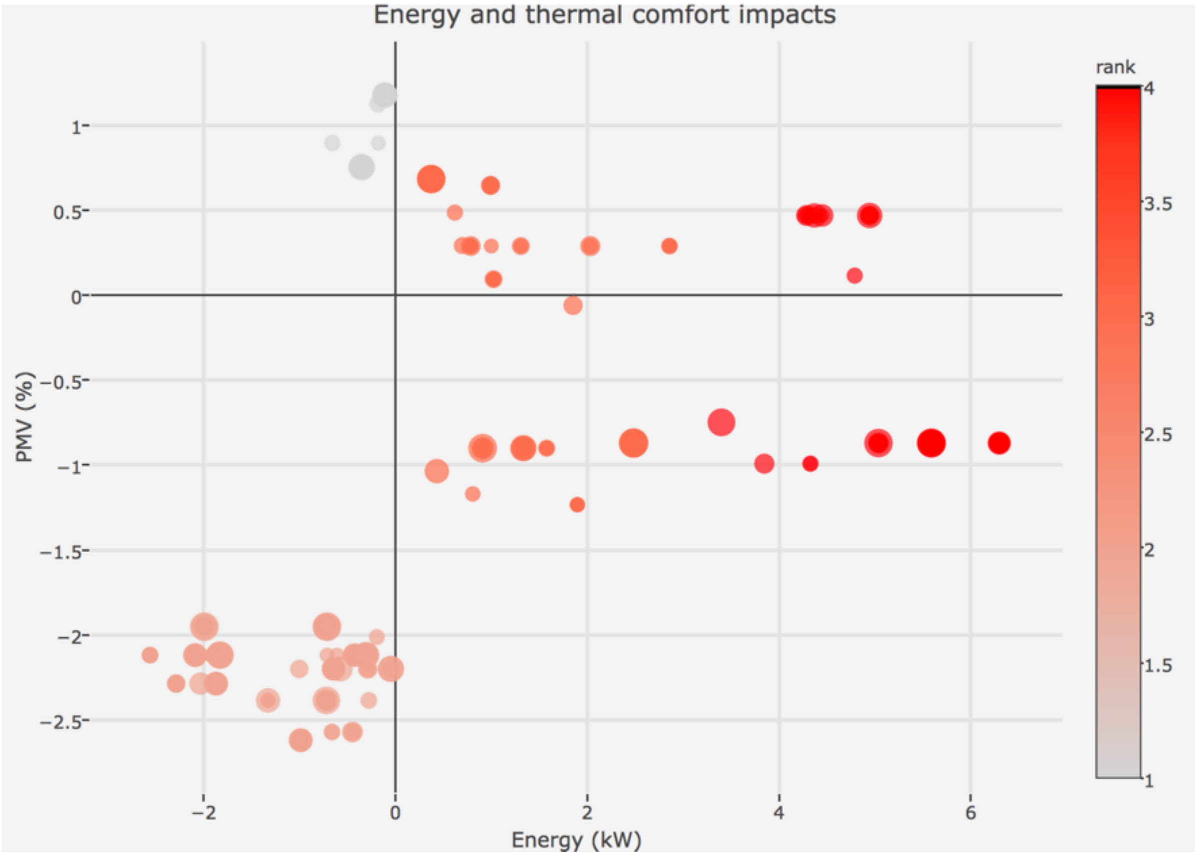


Figure 37. Alarm ranking by energy impacts

To analyze the causes of the difference between the alarms impacts, the alarms have been divided into different groups. First the alarms are divided into four different groups based on their thermal comfort impact and energy impact. The thresholds that distinguish the alarms with high and low impacts are the median values. For thermal comfort impact analysis, alarms with absolute value of PMV higher than 1 are considered as the "high" group, and the rest are considered as the "low" group. Similarly, the alarms with energy impacts greater than 1.8 kW are considered as the "high" group, and the rest are considered as the "low" group. Then the distributions of the parameters such as heating and cooling setpoints, the alarm thresholds temperatures, and the design discharge airflow rates are compared among different groups. Figure 37 shows the comparison between alarms with high absolute PMV and low absolute PMV. Figure 38 shows the comparison between the alarms was high energy impact and low energy impact.

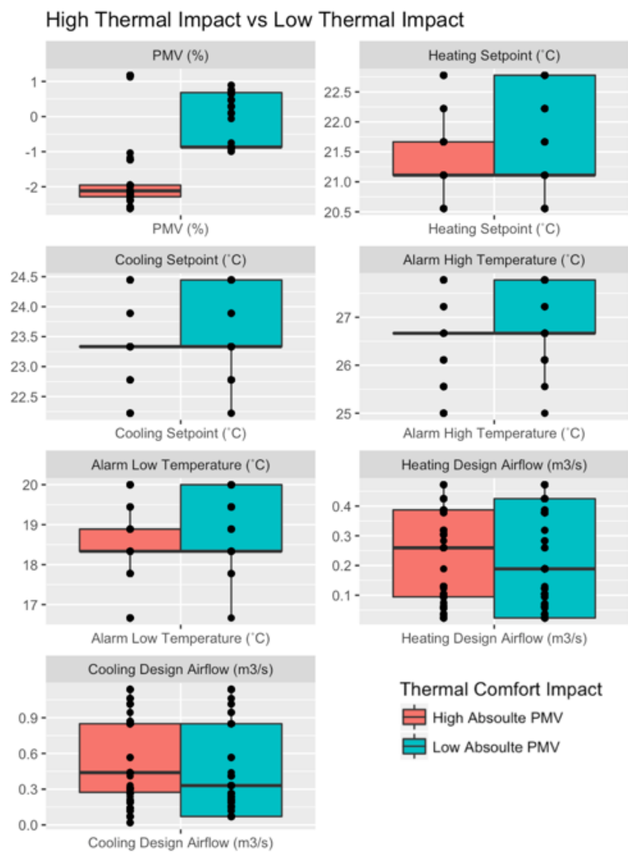


Figure 38 Key parameters comparison for alarms with high and low thermal comfort impacts

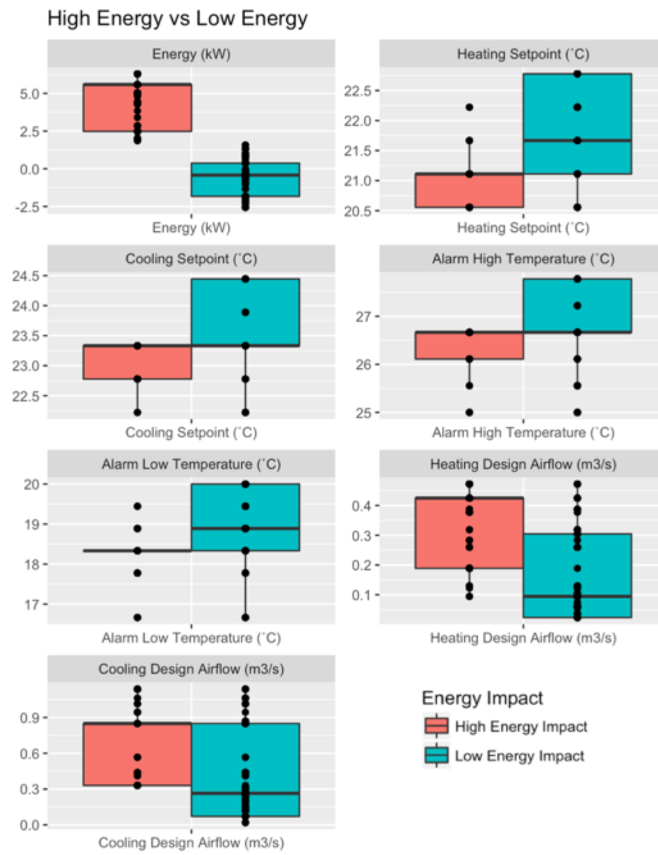


Figure 39. Key parameters comparison for alarms with high and low energy impacts

It can be seen from the figure that alarms with high thermal impacts have lower heating and cooling setpoints, lower alarms temperature thresholds, and higher discharge air flow rate. The alarms with high energy impacts have low heating and cooling setpoints, lower alarm temperature thresholds, higher heating and cooling discharge airflow rate.

6.2.3 Analysis by Space Type

It can be seen from Figure 34 that the alarms in cooling and heating modes have large variances. The scatterplot shows that the distributions of alarms' thermal comfort impacts are different in cooling modes and heating modes. The alarms have a large range in cooling mode, but a small range in heating modes. The differences in the distributions maybe caused by the space type differences. Therefore, a separate analysis of the alarms by their space type is needed. Figure 40 shows the impacts of the alarms by their space types.

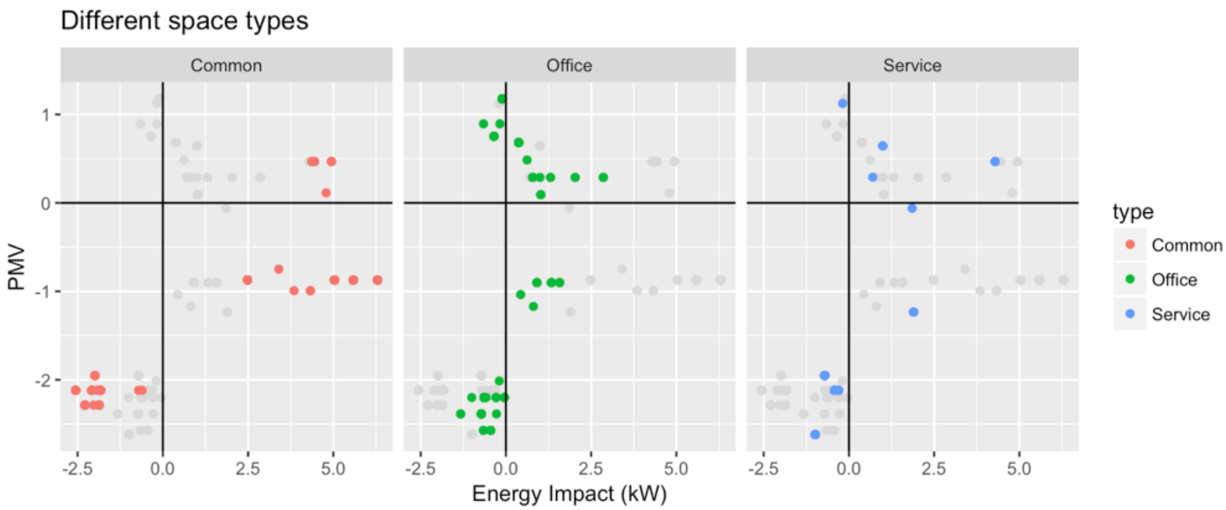


Figure 40. Alarm impacts distribution in different spaces

It can be seen from the scatterplots that the alarm occurred in different space types have different distributions on the impact plane. In terms of energy consumption, alarms occurred in common spaces have the largest impacts. Whenever there is an alarm in both heating and cooling modes, the VAV terminals in common spaces consume more than 2.5 kW higher than normal condition. In contrary, alarms that occurred in office and service areas have moderate impacts on energy consumption. In terms of thermal comfort, "Zone temperature too high" alarms from all space types have moderate impacts because most of the PMV values are within 0 and 1. It means when there is a "Zone temperature too high" alarm, the zone is slightly warm. However, when there is a "Zone temperature too low" alarm in cooling mode, the PMV values are below -2.5. That means during cooling mode, it will be very cold if there is a "Zone temperature too low" alarm.

The differences of the alarm impact distributions are caused by the zone stepoints and design discharge airflow rates, and alarm thresholds. Figure 40 shows that some alarms occurred in common spaces have the highest energy impacts. This can be explained by the highest discharge airflow rate of common spaces showed in Figure 41. As for thermal comfort, all "Zone air temperature too low" alarms occurred in cooling mode lead to low PMVs. This is because the zone temperature setpoints in cooling seasons are too low and the spaces are overcooled. The detailed analysis indicates that alarms are closely related to the zone design parameters. Potential measurements that help reduce the energy and thermal comfort impacts are proposed in Chapter 8.

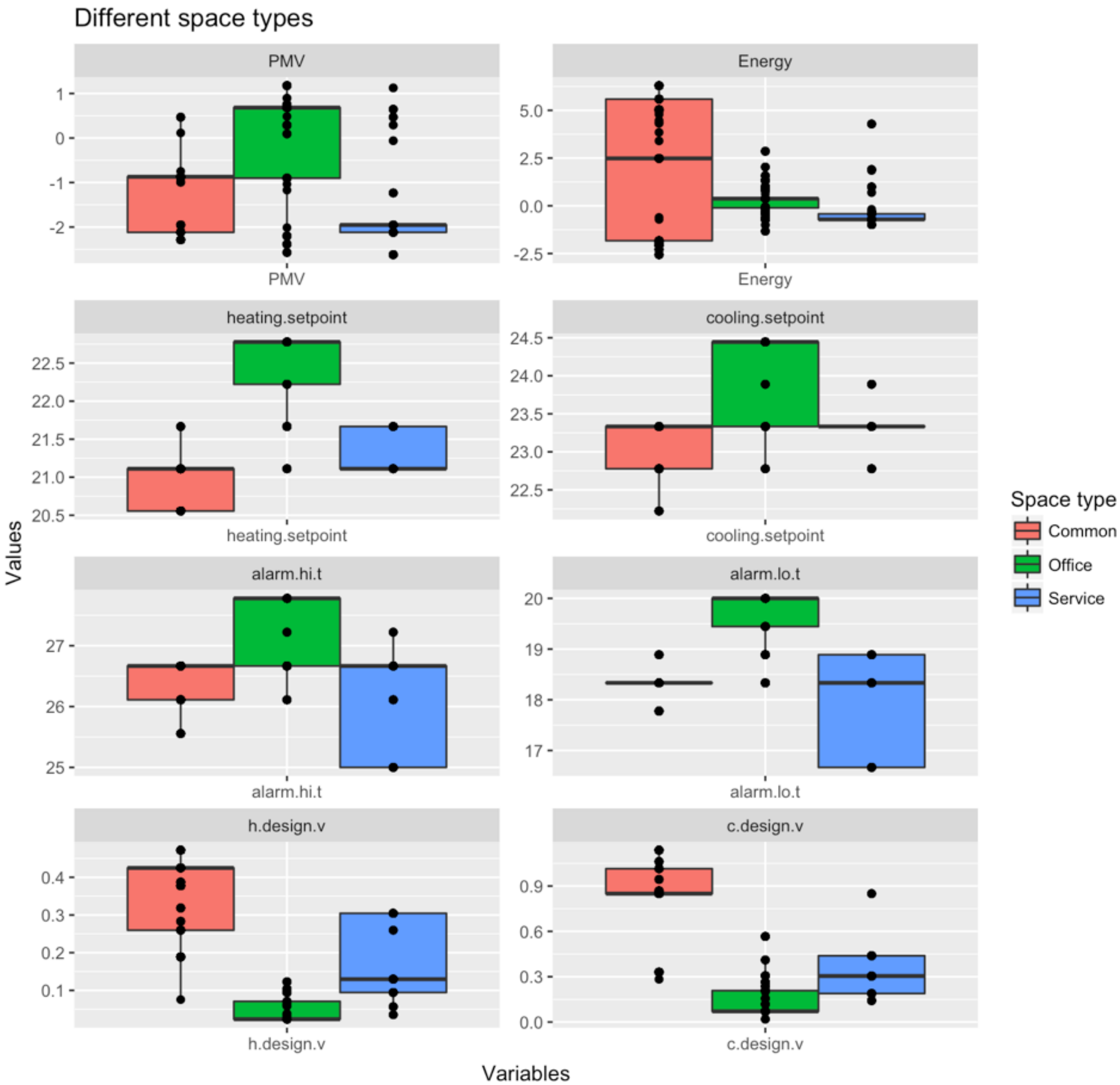


Figure 41. Design parameters of different space types

The figure below shows the average alarms impacts on energy consumption and thermal comforts in difference space types in cooling and heating modes.

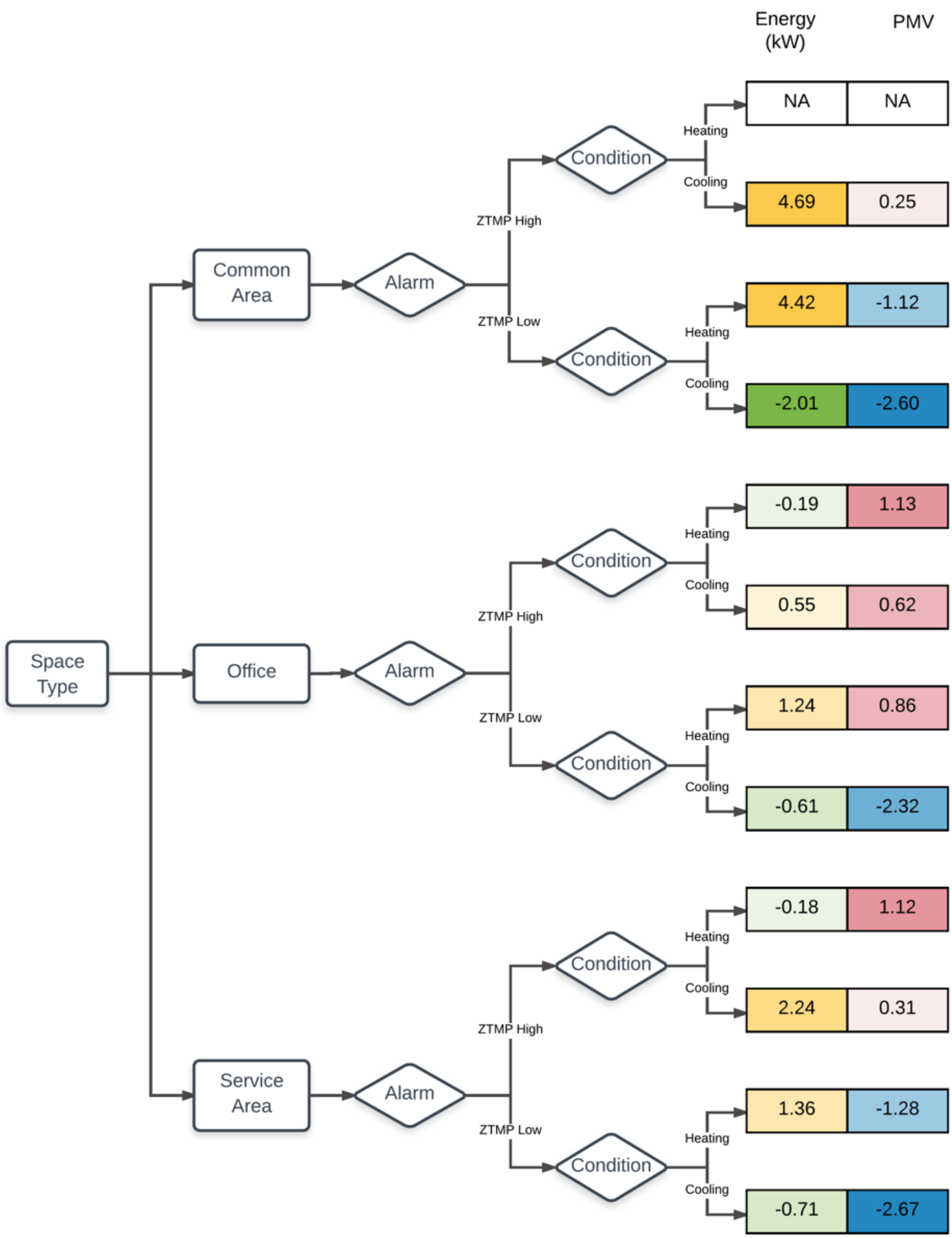


Figure 42. Typical alarms in different spaces and their average impacts on energy consumption and thermal comfort

The colors of the cells in the tree diagram indicate the severity of the alarms' impacts. For energy consumption impacts, green stands for less energy consumption compare to normal condition, while yellow stands for more energy consumption than normal condition. For thermal comfort impacts, blue stands for cold and brand stands for hot. Deep color means the alarm's impact is severe. For example, a "Zone temperature too low" alarm that are occurs in the common space in cooling mode has the negative energy consumption impact and a high thermal comfort impact. With the scatter plots and the tree diagram, it is very easy to identify the alarms with high thermal comfort impacts or energy impacts.

6.2.4 Typical Alarm Analysis

The impacts of alarms and their ranking are discussed in previous sections. In this section, typical alarms with the different energy and thermal comfort impacts are analyzed. The alarm description, location, space type, condition, and their energy and PMV impact are listed in tables. The implications of those impacts are also explained.

1). High energy impact and low thermal comfort impact alarms.

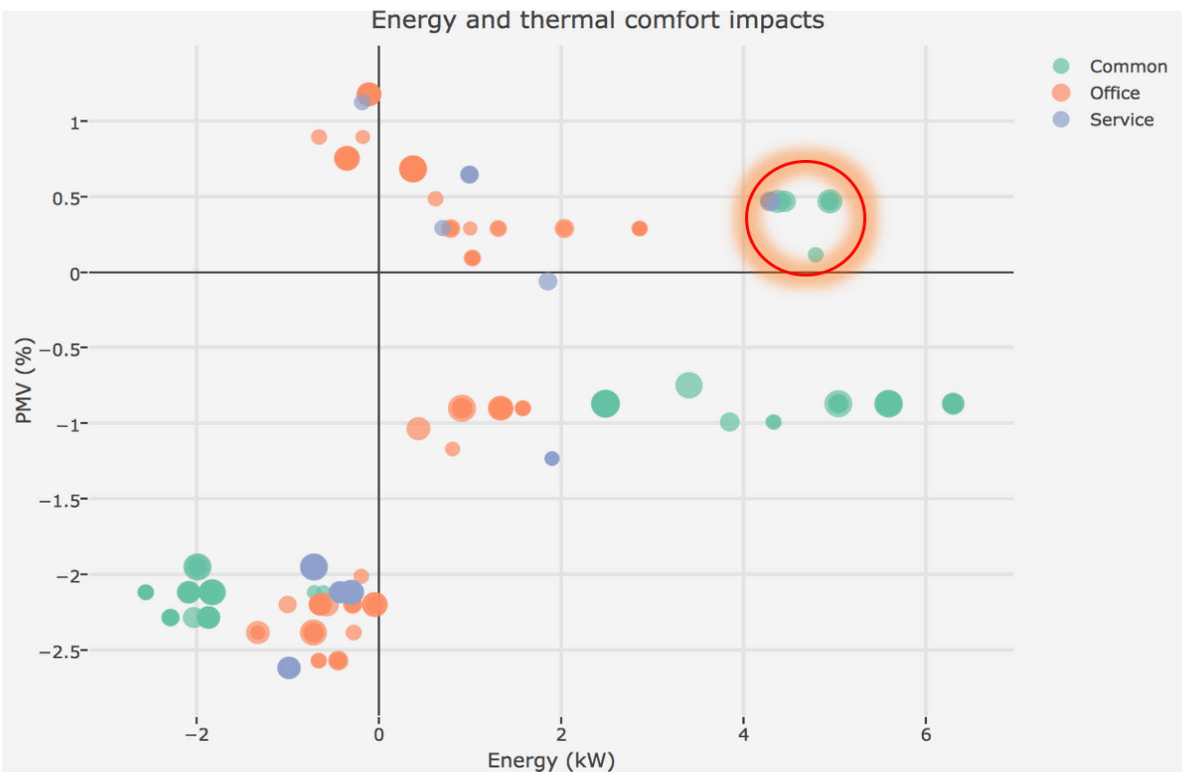


Figure 43. High energy impact and low thermal comfort impact alarms

Table 18. High energy impact and low thermal comfort impact alarms

Alarm	Zone temperature too high	Whole building average
Location	VAV Corridor 5300 North	
Space type	Common space	
Condition	Cooling	
Design cooling discharge airflow (cfm)	2000	990
Design heating discharge airflow (cfm)	600	434
Low temperature alarm threshold (˚F)	64	65.3
High temperature alarm threshold (˚F)	80	79.5
Energy Impact (kW)	4.94	-
PMV Impact	0.29	-

Implications: although the alarm shows "Zone temperature too high", the calculated PMV indicates the zone is only slightly warm under the alarm. This is because the high temperature alarm threshold for this space is 80 ˚F, which is not too high. However, the energy impact is significant. That can be explained by the high design discharge airflow rates. As shown in the table above, the design cooling discharge airflow rate of this corridor it's more than two times of the average value across the building.

2). High energy impact and medium thermal comfort impact

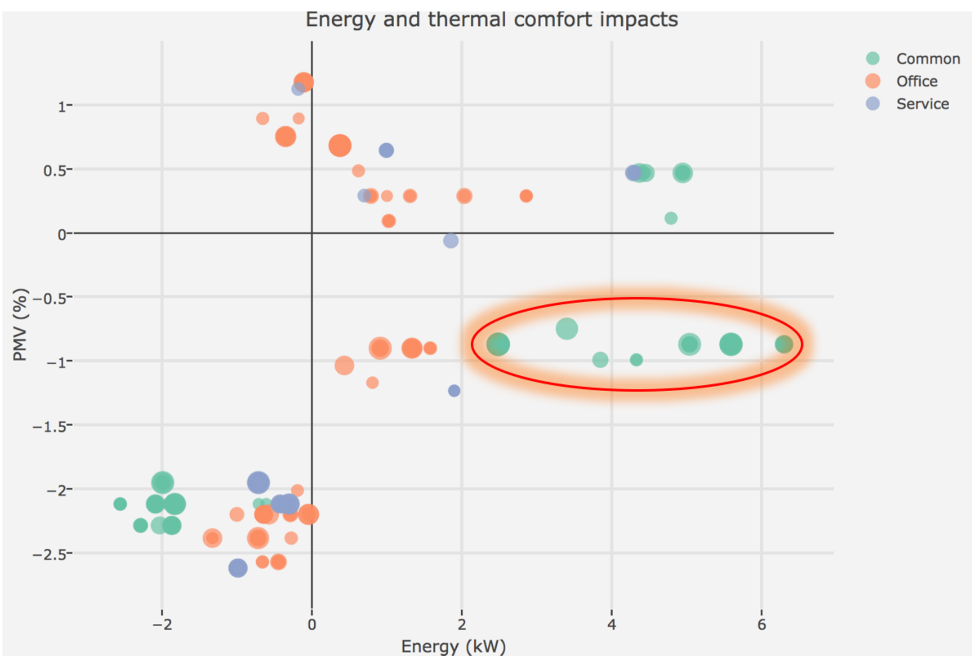


Figure 44. High energy impact and medium thermal comfort impact alarms

Table 19. High energy impact and medium thermal comfort impact alarms

Alarm	Zone temperature too low	Whole building average
Location	VAV Room 3101 Café—North wall	
Space type	Common space	
Condition	Heating	
Design cooling discharge airflow (cfm)	2410	990
Design heating discharge airflow (cfm)	1000	434
Low temperature alarm threshold (˚F)	65	65.3
High temperature alarm threshold (˚F)	80	79.5
Energy Impact (kW)	6.30	-
PMV Impact	-1.11	-

Implications: as indicated in the figure and the table above, those "Zone temperature too low" alarms occurred in heating seasons have very high energy impact and medium thermal comfort and impact. Similar with the first type of alarms, these are alarms occurred in spaces with large design discharge air flow rates.

3). Low energy impact and low thermal comfort impact

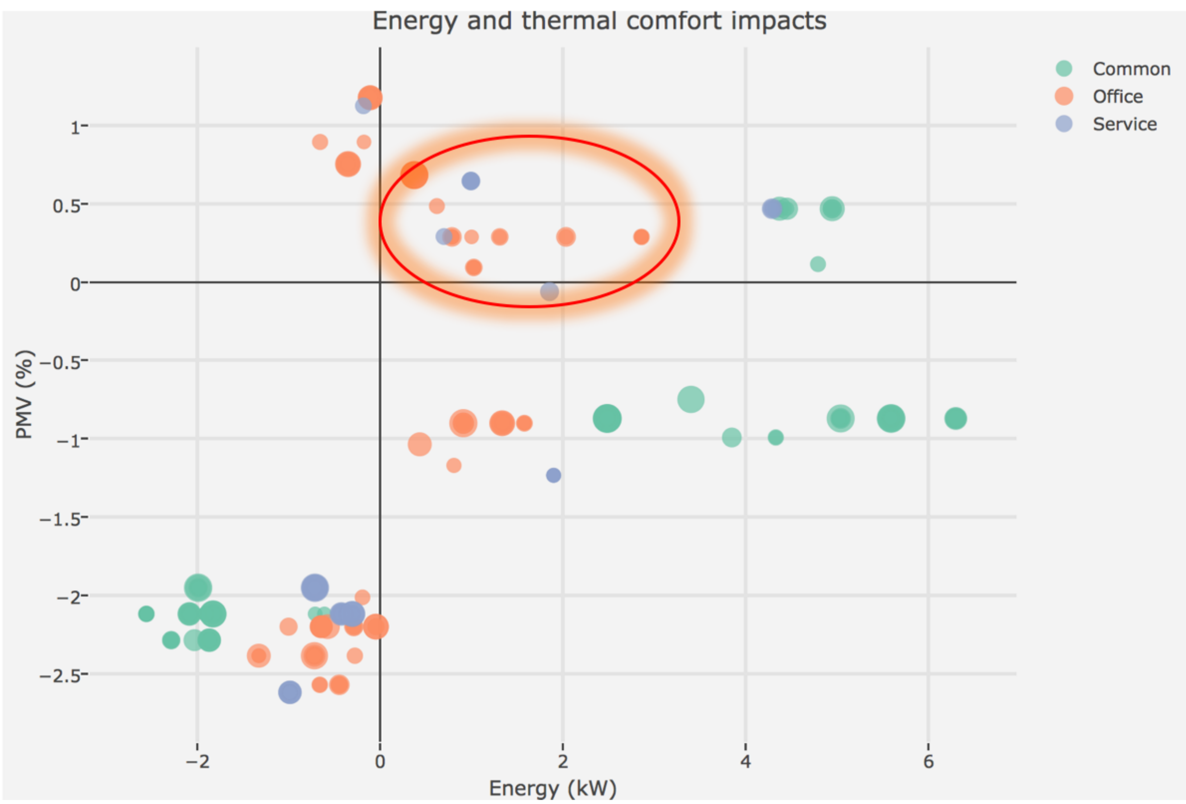


Figure 45. Low energy impact and low thermal comfort impact alarms

Table 20. Low energy impact and low thermal comfort impact alarms

Alarm	Zone temperature too high	Whole building average
Location	VAV Room 8110 Office	
Space type	Office	
Condition	Cooling	
Design cooling discharge airflow (cfm)	250	990
Design heating discharge airflow (cfm)	80	434
Low temperature alarm threshold (˚F)	67	65.3
High temperature alarm threshold (˚F)	81	79.5
Energy Impact (kW)	0.62	-
PMV Impact	0.49	-

Implications: this type of alarms have low energy impact and low thermal comfort impact. The reason that they have low thermal comfort impact is that the high temperature alarm threshold is only 81 ˚F, which is not too high according to PMV calculation result. The low energy impact is because the space has low design discharge air flow rates.

4). Low energy impact and medium thermal comfort impact

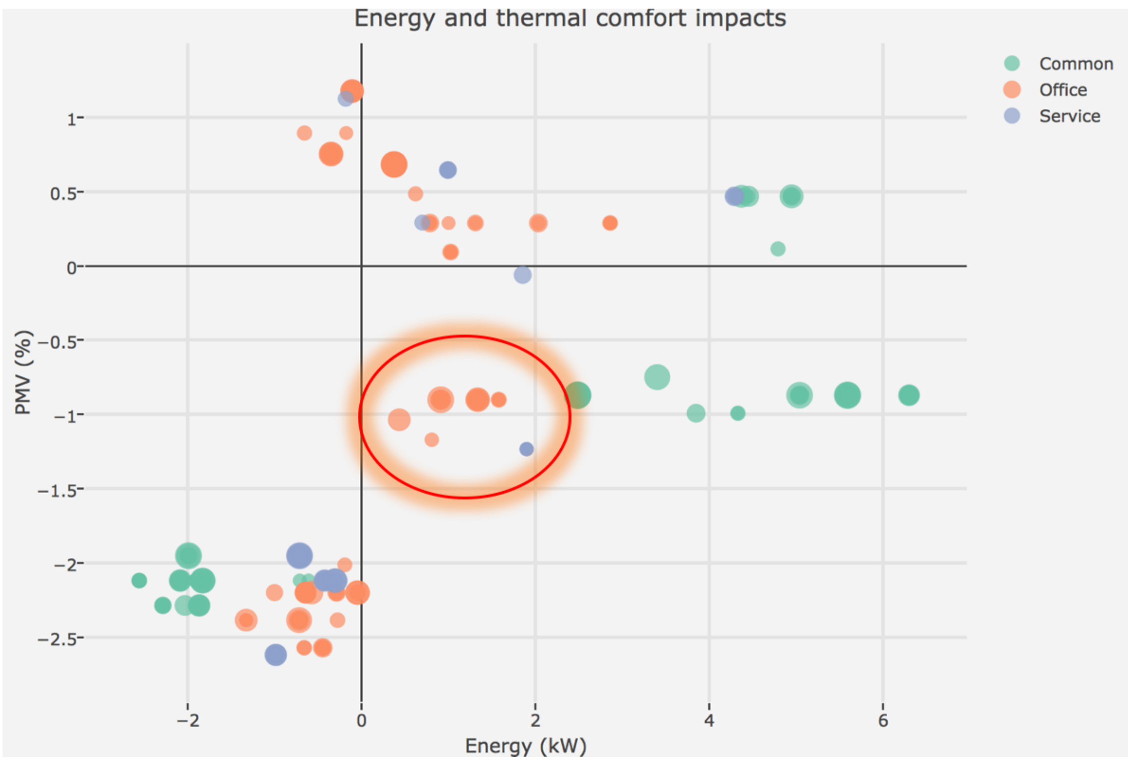


Figure 46. Low energy impact and medium thermal comfort impact alarms

Table 21. Low energy impact and medium thermal comfort impact alarms

Alarm	Zone temperature too low	Whole building average
Location	VAV Room 4007 Office	
Space type	Office	
Condition	Heating	
Design cooling discharge airflow (cfm)	440	990
Design heating discharge airflow (cfm)	130	434
Low temperature alarm threshold (˚F)	65	65.3
High temperature alarm threshold (˚F)	79	79.5
Energy Impact (kW)	0.81	-
PMV Impact	-1.12	-

Implications: this type of alarms have low energy impact and medium thermal comfort impact. The low energy impact is because the space has low design discharge air flow rates. In addition, the zone high temperature alarm threshold is only 79 ˚F in this zone. This may lead to unnecessary energy consumption and increasing amount of alarms.

5). Negative energy impact and low thermal comfort impact

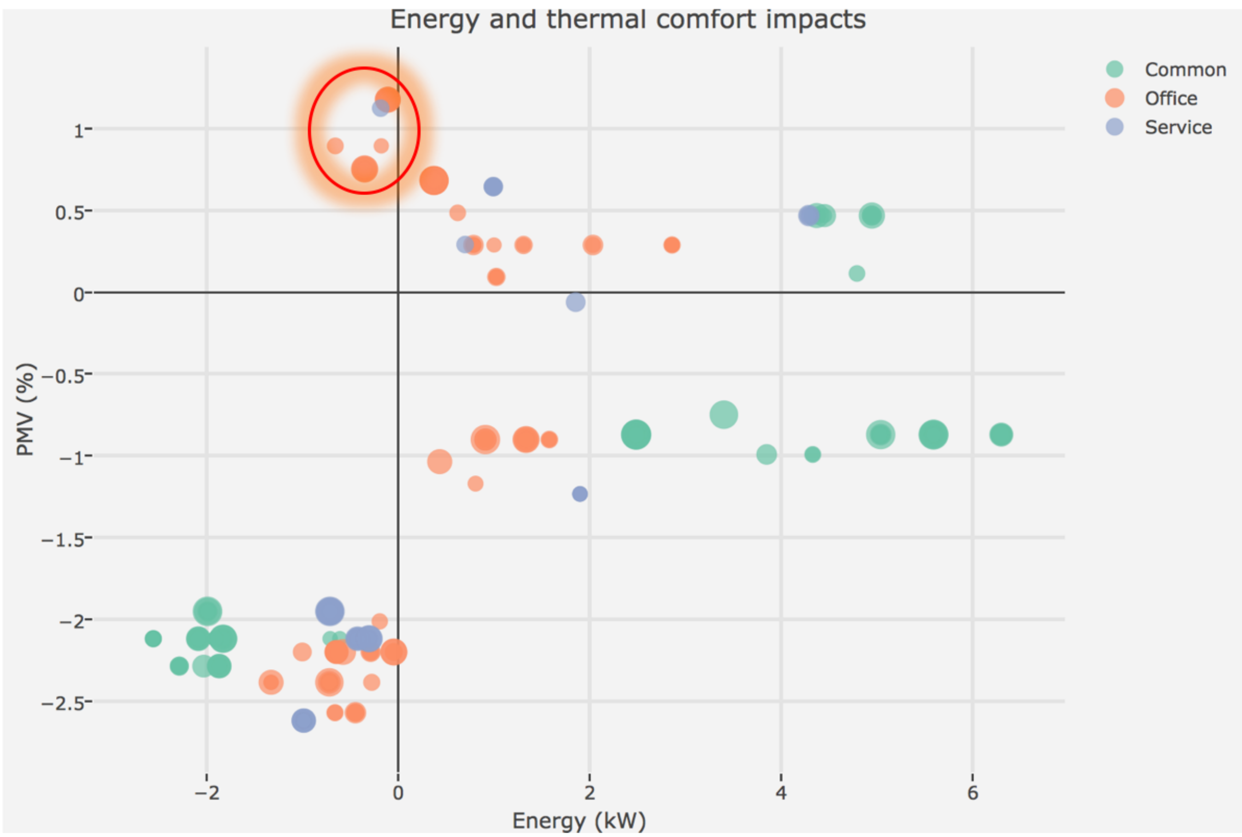


Figure 47. Negative energy impact and low thermal comfort impact alarms

Table 22. Negative energy impact and low thermal comfort impact alarms

Alarm	Zone temperature too high	Whole building average
Location	VAV Room 5003 Office	
Space type	Office	
Condition	Heating	
Design cooling discharge airflow (cfm)	250	990
Design heating discharge airflow (cfm)	70	434
Low temperature alarm threshold (˚F)	66	65.3
High temperature alarm threshold (˚F)	80	79.5
Energy Impact (kW)	-0.18	-
PMV Impact	0.98	-

Implications: this type of alarms have negative energy impact and low thermal comfort impact. When there is a "Zone temperature too high alarm" in heating mode, the discharge air flow rate would be reduced, and the discharge air temperature would be decreased. So, the VAV terminal consumes less energy than normal operation conditions. Moreover, because the high temperature alarm threshold is 80 ˚F, the zone would only be slightly warm. This type of alarms has the least impacts on both thermal comfort and energy consumption.

6). Negative energy impact and high thermal comfort impact

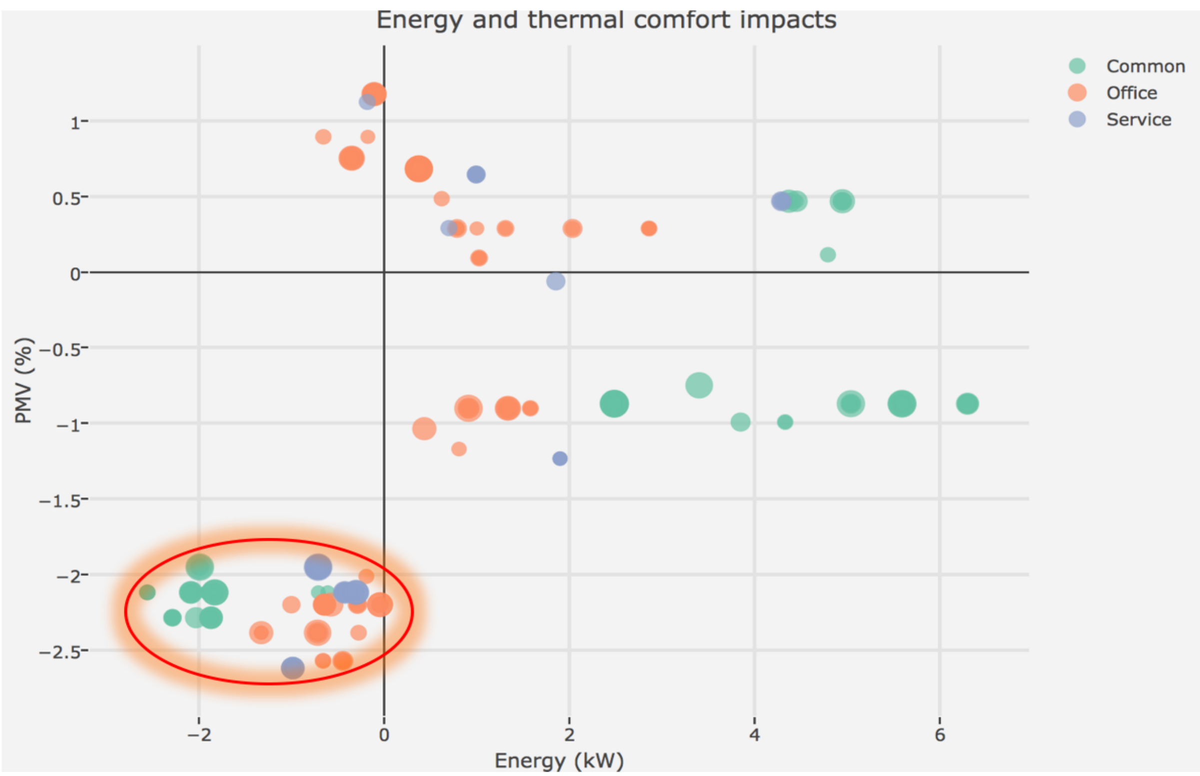


Figure 48. Negative energy impact and high thermal comfort impact alarms

Table 23. Negative energy impact and high thermal comfort impact alarms

Alarm	Zone temperature too low	Whole building average
Location	VAV Room 4001 Office	
Space type	Office	
Condition	Cooling	
Design cooling discharge airflow (cfm)	1200	990
Design heating discharge airflow (cfm)	200	434
Low temperature alarm threshold (˚F)	66	65.3
High temperature alarm threshold (˚F)	80	79.5
Energy Impact (kW)	-1.33	-
PMV Impact	-2.4	-

Implications: this type of alarms have negative energy impact and high thermal comfort impact. Because the "Zone temperature too low" alarms occurred in cooling mode, occupants have a low clothing factor. The result is the low PMV, which means the zones are very cold. But since the discharge air flow rates would be decreased and the discharge air temperature would be raised, the energy consumption rate is decreased under alarm conditions. The reason for this type of alarms may be that the spaces are overcooled before the alarms are generated. And the overcooling of the space lead to unnecessary energy consumptions. Thus, this type of alarms has a high priority.

6.3 Alarm Duration Prediction Results

The alarm dataset's feature space is optimized through the iterative model building process discussed in 5.5.5. The optimal model setting for Decision Trees algorithm is found through the parameter tuning process discussed in 5.5.6. This section presents the final model performance and its comparison with the baseline model.

Table 24. Comparison of baseline model and final model

Performance		Baseline Model	Final Model
Overall Accuracy		74.1%	77.8%
True Positive Rate	< 30 min	87.7%	89.4%
	30 min ~ 2h	58.8%	60.6%
	2h ~ 5h	25%	20.2%
	5h ~ 24h	68.1%	83.7%
	> 24h	66.4%	62.6%
Kappa Statistics		0.5992	0.6533
Root Mean Squared Error		0.2933	0.2579

Note: the models are trained on development and cross-validation combined set, and evaluated on final test set.

From the comparison, the final model has a significant performance improvement over the baseline model. All the three-performance metrics-- overall accuracy, Kappa statistics, and root mean squared error of the new model indicate the improvement over baseline model. The true positive rates of all the class labels have been increased except for the alarms with duration between 2 hours and 5 hours. Where the true positive rate of this class has dropped from 25.0% to 20.2%. But given the fact that alarms with durations between 2 hours and 5 hours have the lowest percentage (n=505, 5.9%) of the total alarms, this new model is acceptable.

6.4 Alarm Prioritization

With the instantaneous energy consumption rate impact, thermal comfort impact, and alarm durations, the cumulative impacts can be calculated. Then the occupant-related alarms could be prioritized based on the cumulative energy consumption impacts and thermal comfort impacts.

6.4.1 Cumulative Energy Consumption Impact

The cumulative energy consumption impact is the potential impact of certain alarms in a period. Equation 4 shows the calculation:

Equation 4 Cumulative Energy Consumption Impact

$$CECI = ECRI \times Duration$$

where:

$CECI$-- the cumulative energy consumption impact of the alarm (kWh)

$ECRI$ -- the energy consumption rate impact of the alarm (kW)

$Duration$ -- the duration of the alarm (h)

6.4.2 Cumulative Thermal Impact

Similar with energy impacts, cumulative thermal comfort impacts reflect the potential impact of certain alarms in a period. shows the calculation:

Equation 5 Cumulative Thermal Comfort Impact

$$CTCI = TCI \times Duration$$

where:

$CTCI$-- the cumulative thermal comfort impact of the alarm (PMV*h)

TCI -- the thermal comfort impact of the alarm (kW)

$Duration$ -- the duration of the alarm (h)

6.4.3 Alarm Prioritization

The prioritization of alarms is based on their comprehensive impacts on energy consumption and thermal comfort. Firstly, the energy impacts of the alarms could be divided into high, medium, and low groups as shown in Figure 49. Since energy consumption impacts are equally important in different spaces, the alarms from different spaces have the same grouping thresholds.

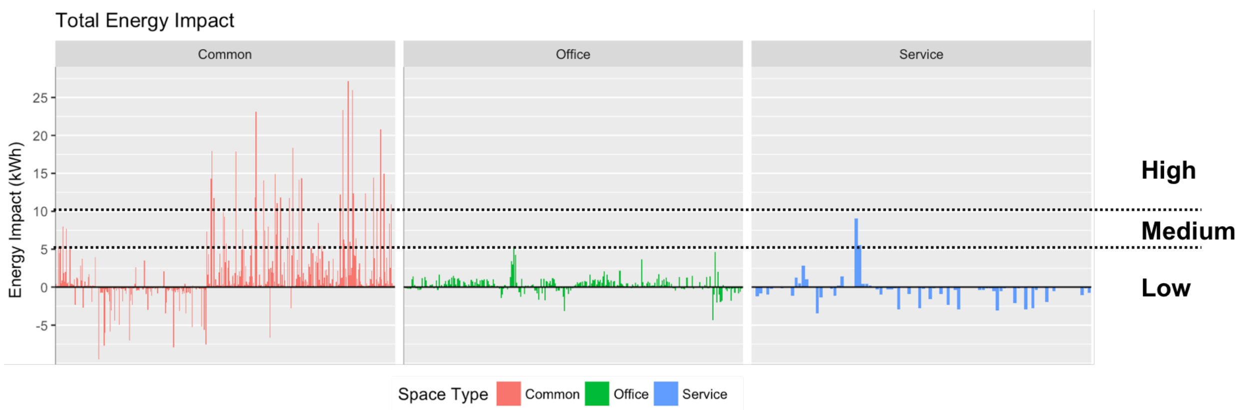


Figure 49. Cumulative enegry impacts groups

It can be seen from the figure, all the alarms with the “High energy consumption” are from common spaces. All the alarms from offices are in the “Low energy consumption” group. This finding is consistent with the analysis in 6.2.3.

For cumulative thermal comfort impacts, the alarms are also divided into “high”, “medium”, and “low” groups. However, because the alarms are not equally important in different space types, there should be different grouping thresholds. For example, occupants in common spaces and service areas have a high mobility. Since they don't stay in those areas for a very long time, it is fine that the cumulative thermal comfort impacts are higher than office rooms. *Figure 50* Shows an example of the alarms grouping.

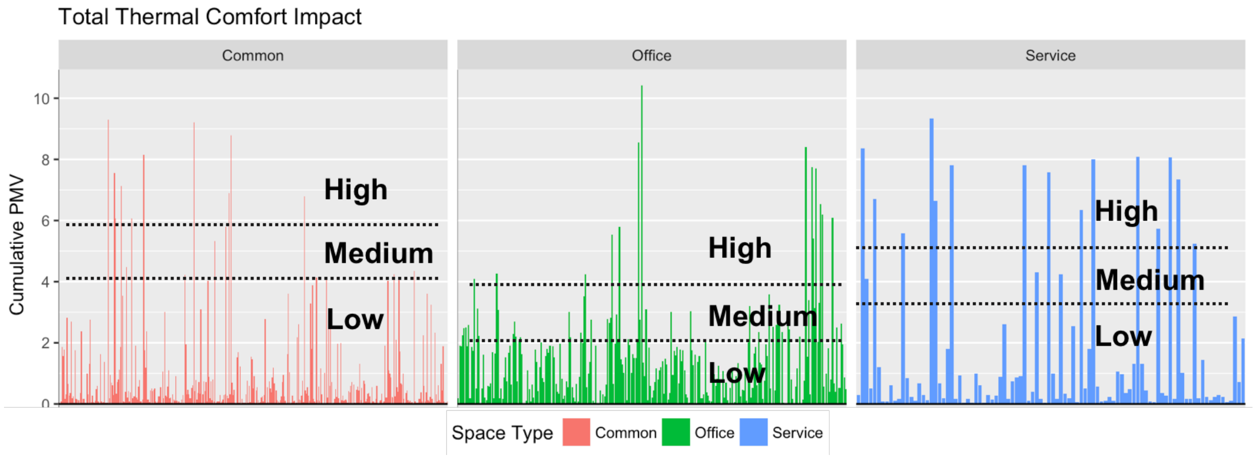


Figure 50. Cumulative thermal comfort impacts groups

With the alarms grouped by their impacts, the total impacts can be calculated with the equation below:

Equation 6 Cumulative Thermal Comfort Impact

$$Total\ Impact = A \times CECI + B \times CTCI$$

where:

A-- the weight of energy consumption impact

$CECI$-- the cumulative energy consumption impact of the alarm (kWh)

B-- the weight of thermal comfort impact

$CTCI$-- the cumulative thermal comfort impact of the alarm (PMV*h)

The framework aims to provide building operators a way to evaluate the impacts of BAS alarms. For the alarm prioritization, they can weight energy consumption and thermal comfort based on their preferences. Figure 51 shows the alarm prioritization with the same weight on energy consumption and thermal comfort. Figure 52 Shows the alarm prioritization with a higher weight on thermal comfort.

Thermal \ Energy	Low (1)	Medium (2)	High (3)
Low (1)	2	3	4
Medium (2)	3	4	5
High (3)	4	5	6

A=1, B=1

Figure 51. Alarm prioritization with same weight

Thermal \ Energy	Low (1)	Medium (2)	High (3)
Low (1)	3	4	5
Medium (2)	5	6	7
High (3)	7	8	9

A=1, B=2

Figure 52. Alarm prioritization with higher weight on thermal comfort

Extending from the combined impact metrics, alarms can be prioritized. A 3D visualization tool built on the alarm data could help us better understand what's behind the alarms. This visualization tool is made with an open source *JavaScrip* graphing library called *Plotly* in *R* programming environment. The code scripts could be found in the appendices. *Figure 53* Shows a screenshot of the 3D visualization of alarm prioritization.

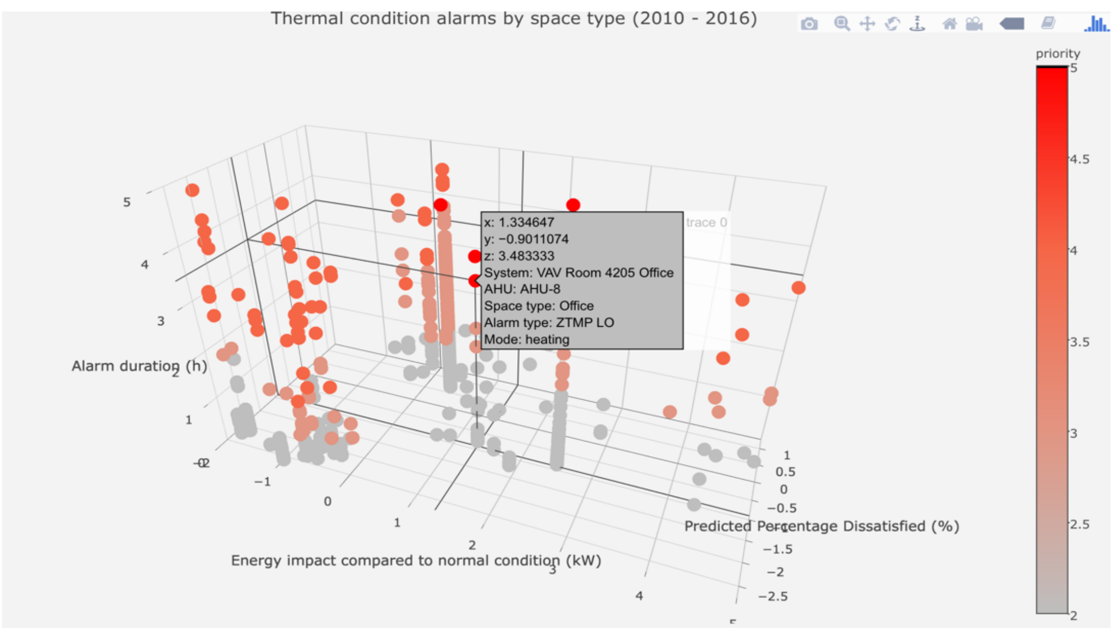


Figure 53. 3D Visualization of the alarm prioritization

In the example, the dots in red have high priority and the dots in grey have low priority. The alarm highlighted in the screenshot indicates that it occurred in heating model, with an instantaneous energy consumption rate impact of 1.33 kW, and a PMV of -0.9. It occurred in office room 4205 and this room is covered by AHU 8. The visualization tool provides not only alarm prioritization, but also necessary qualitative and quantitative information.

7. Conclusion

7.1 Summary of Key Findings

This thesis has investigated challenges in using BAS's alarm management functions to support building system operation and maintenance through literature reviews, and manual inspections of a BAS solution from a university building on CMU campus. The current problems and future expectations of BAS alarm management functions are investigated through interviews with facility managers and engineers. Then a data mining framework was built to optimize the current BAS alarm management functions. The framework is implemented with Gates-Hillman Center building on CMU campus as a case study. The key findings of this thesis can be summarized as:

- BAS is becoming normal in large buildings. The complexity of building systems makes operation and maintenance very challenging. One of the operational challenges is the huge amount of alarms generated by BAS. Because the lack of actionable information, building operators usually don't know what to do with those alarms. Moreover, the literature reviews indicate there is little research in BAS alarm management area.
- Two interviews have been done in this thesis work to investigate current problems and future expectations for BAS alarm management function. The interviews results show that facility managements in most buildings are reactive. One of the reasons for that is the lack of actionable information provided by BAS. Building system operation and maintenance requires strong domain knowledge even with BAS installed. Although BAS can convert sensor measurements into information such as alarms, that kind of information can hardly be converted into knowledge. Thus, a simple, accurate and powerful BAS alarm management function that can retrieve knowledge from information, will be helpful in operation and maintenance decision making.
- With the data mining framework, BAS alarms are categorized as occupant-related alarms, equipment-related alarms, and critical operation alarms based on their affected objects. In the case study, it was found that equipment-related alarms have the largest number. But the number of equipment-related alarms has been dramatically reduced during the past six years. However, the number of occupant-related alarms was unpredictable. The

examination of the alarms sources show that occupant-related alarms have the most sources, while equipment-related alarms have the least sources. That means occupants-related alarms are more difficult to handle given their large diversity.

- The energy and thermal comfort impacts of occupant-related alarms can be quantified. The quantifications are achieved by calculating the transient energy consumption rates (kW) and thermal comfort (PMV) impacts and multiplying them by alarm durations. The transient energy consumption rate impact is calculated with a thermodynamics equation. The transient thermal comfort impact is calculated with Fanger's PMV equation.

7.2 Contributions

As discussed before, there is little research on the BAS alarm management topic. This thesis has proposed a data mining framework that helps to pre-process the raw alarm, categorize the alarms based on their affected objects and prioritize the alarms based on potential energy consumption thermal comfort impacts. The implementation of the framework is demonstrated by a case study of Gates-Hillman Center building on CMU campus. The main contributions of this framework are:

- It outlines the process of retrieving and processing raw alarm data from BAS dashboard. Detailed data cleansing and feature engineering steps are explained. Those steps include removing useless features, handling missing values, attaching system information and weather information. Corresponding R scripts are developed to automate the process in the future.
- A machine learning model is developed as part of the data mining framework. The machine learning model is built on decision-tree model to predict alarm durations in the future. Through iterative model building, error analysis, and model parameter tuning processes, the final model could predict the alarms' durations accurately (80% overall accuracy).
- The framework quantifies BAS alarms impacts with transient impacts and durations. It then allows users to prioritize the alarms based on their cumulative impacts. Different impact metrics such as energy consumption and thermal comfort could be weighted differently. The quantification results could be easily used to prioritize and visualize the

alarms and their detailed information. Figure 54 shows an example of alarm log from BAS dashboard, which provides little useful information. Figure 55 shows an example of alarm prioritization and visualization. With this tool, facility managers can navigate to the top alarms with minimum effort and obtain key relevant information behind the scene.

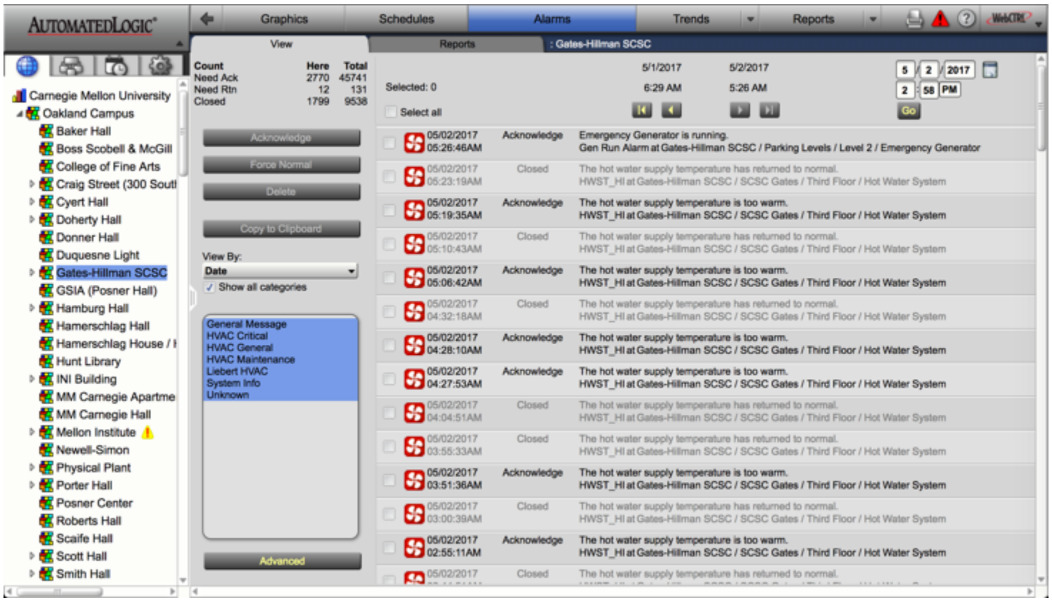


Figure 54. Original alarm log from BAS dashboard

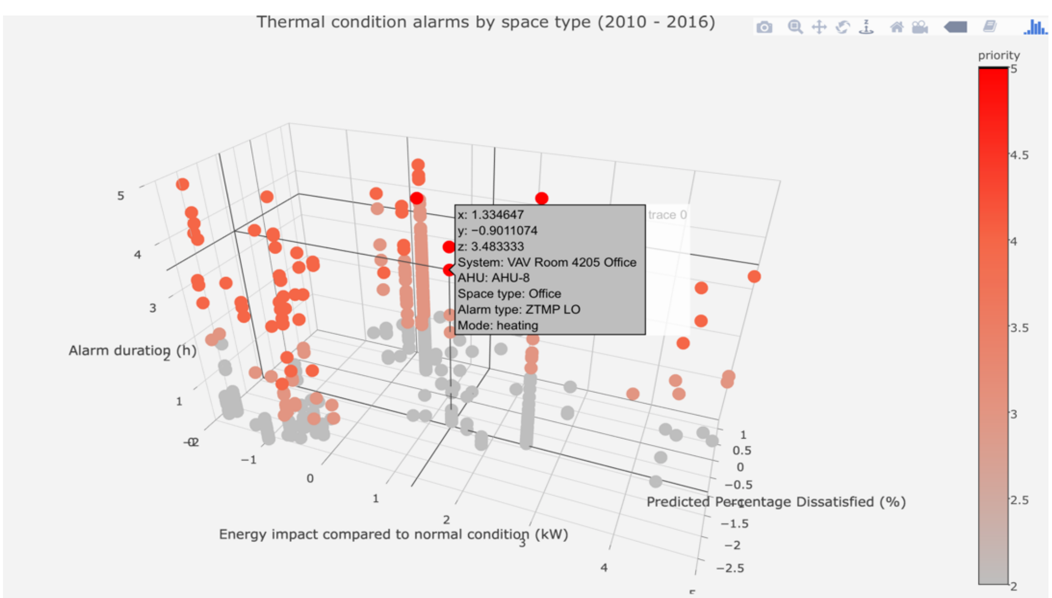


Figure 55. Alarm prioritization and visualization example

7.3 Limitations

The framework developed in this study is only a prototype which needs further researches and examinations. The case study has shown the framework's potential of converting information from BAS alarms into knowledge that can help building system operation and maintenance decision making. But it still has some limitations:

- Because there is no zone level sub-metering of the HVAC system, the energy consumption impact quantification is achieved by fist principle calculations. The alarms' impacts on the whole building energy efficiency are not captured. More specifically, energy consumption quantification is focusing on terminal units like VAV box instead of focusing on system or building level. In addition, the impacts before or after the alarms' durations are not quantified.
- For thermal comfort impacts, the framework uses Fanger's PMV equation. Because there is insufficient sensor measurement, many assumptions have been made. The assumptions include radiant air temperature, relative humidity, air velocity and clothing factor.
- During the sensor measurements collection process, around 1/8 of the conditioned spaces in the building are sampled. Although the sampling tries to keep similar space type and orientation distributions as the population, it can still be inaccurate due to sampling errors.
- Only occupant-related alarms are considered in the case study. The impacts of equipment-related alarms are not considered because they have a very small number of alarm sources, which means they are more predictable and easier to be fixed than occupant-related alarms. However, it may not be the same case in other buildings. Moreover, only energy consumption and thermal comfort impacts are quantified in this case study. Impact metrics such as air quality, cost of operation and repairs are not quantified.

7.4 Future Work

The future work of this study involves improving the data mining framework by adding more impact categories, using third party fault detection and diagnostics techniques to analyze the root causes of the alarms, and developing a better user interface.

- Currently, only energy consumption and thermal comfort impacts of the alarms are evaluated. But different stakeholders' interests may differ from each other. In the future, more impact metrics such as indoor CO_2 level, operation cost, equipment life could be added to the framework to satisfy different user preferences.
- During the manual inspection process in the case study, many defects in current rule-based alarm generation mechanism have been found. Those defects have weakened the alarm management function. In current stage, the data mining framework filters and categorizes alarms with expertise knowledge. In the future, third party fault detection and diagnostics techniques could be integrated into the framework to provide automated alarm filtering and categorizing, and to help analyze the root causes of the alarms.
- The framework can be integrated into Computerized Maintenance Management System (CMMS). CMMS could receive real-time alarm log, sensor measurements, and weather data. It then feeds those data into the data mining framework. That way, the alarms would be evaluated dynamically. Moreover, with CMMS, top alarms could be visualized with a better user interface. Operation and maintenance workflows under certain alarm conditions can be standardized.

Appendices

Appendix A. Survey and Interviews

A.1 Structured Survey

Thank you for agreeing to take part in this survey which measures your opinions of existing Building Automation System. The data we collected in this survey is used for academic purposes. The survey should take 15~20 min to complete. Your responses will be anonymous.

1. What are the main steps that you take to acknowledge an alarm from BAS dashboard? How do you rank order/prioritize the alarms based and their importance?

____

____

____

____

2. What is the difference between an 'HVAC general' alarm and an 'HVAC critical' alarm?

____

____

____

____

3. Why are some alarms 'closed' immediately while others remain 'unacknowledged'? Is there a rule that decide those two types of alarms?

____

____

____

____

4. How would you classify the following alarms? (Please select the closest category per your opinion.)

4.1

01/24/2017 04:47:10PM Closed The zone CO2 level has returned to normal.
ZCO2_HI at Gates-Hillman SCSC / SCSC Gates / Fourth Floor / VAV Room 4211 Classroom

Category: HVAC Critical
Previous State: OFFNORMAL

Returned to Normal: 01/24/2017 04:47:10PM

The zone CO2 level has returned to normal.

☐ Energy Consumption

☐ Thermal Comfort

☐ Equipment Maintenance

4.2

01/24/2017 04:46:40PM Closed DRV HAND ALM has returned to normal.
DRV HAND ALM at Oakland Campus / Scott Hall / Process & Recirculation AHU CHWS / Proces

Category: HVAC General
Previous State: OFFNORMAL

Returned to Normal: 01/24/2017 04:46:40PM

DRV HAND ALM has returned to normal.

☐ Energy Consumption

☐ Thermal Comfort

☐ Equipment Maintenance

4.3

01/24/2017 04:41:53PM Acknowledge Freezestat LT-2 indicates an alarm condition.
Freezestat LT-2 at Oakland Campus / Mellon Institute / Third Floor / Collab/Mailroom BCU-1

Category: HVAC General
Previous State: NORMAL

Freezestat LT-2 indicates an alarm condition. Click on event source for details.

☐ Energy Consumption

☐ Thermal Comfort

☐ Equipment Maintenance

4.4

01/24/2017 03:57:19PM Acknowledge The zone temperature is too cool.
ZTMP_LO at **Oakland Campus / Hamburg Hall / A-Level / A001E VRF 5-2**

Category: HVAC Critical
Previous State: NORMAL

The zone temperature is too cool.

☐ Energy Consumption

☐ Thermal Comfort

☐ Equipment Maintenance

4.5

09/20/2015 02:45:02AM Waiting for Normal The supply fan has exceeded its runtime limit.
SF RUNTIME at **Oakland Campus / Wean Hall / Fourth Floor / AHU-1**

Category: HVAC Maintenance
Previous State: NORMAL

Acknowledged by: Marty Altschul at 09/26/2016 10:43:34AM

The supply fan has exceeded its runtime limit.

☐ Energy Consumption

☐ Thermal Comfort

☐ Equipment Maintenance

5. Is there any alarm schedule? If yes, what's the logic behind them?

____

____

____

____

6. Please indicate the type of feedback and its frequency from building occupants.

	Never	Once or twice	Sometimes	A lot
Too warm	☐	☐	☐	☐
Too cold	☐	☐	☐	☐
Stuffy air	☐	☐	☐	☐
Draft	☐	☐	☐	☐

7. What features do you think would be most helpful in a BAS dashboard?

☐ View alarms on floorplan

☐ Filter out trivial alarms

☐ Rank alarms by impact categories

☐ Diagnose system faults

☐ Others:____

8. Do you estimate the potential impact of the faults/alarms on energy efficiency? Is it important? How do you estimate the potential energy saving if a fault is corrected?

____

____

____

____

9. Do you have a categorized historical fault list for Gates building that we can refer to?

☐ Yes

☐ No

Han Li

10. What are the typical faults in the system?

☐ Fan speed

☐ Damper stuck

☐ Simultaneously heating and cooling

☐ Heating and cooling cycling (alternating frequently)

☐ Heating when cooling is needed

☐ Cooling when heating is needed

☐ Others ____

11. How do you diagnose the reason for a fault?

____

____

____

____

12. Complaint calls vs. BAS system? What do you respond first? The alarms from the BAS is only used for equipment maintenance only? Occupant comfort complaints are only from phone calls/web inputs.

____

____

____

11. If you have the "perfect" BAS system, what would it be like? What does it have to do to be helpful for you to do your job?

____

____

____

A.2 Survey Response

Thank you for agreeing to take part in this survey which measures your opinions of existing Building Automation System. The data we collected in this survey is used for academic purposes. The survey should take 15~20 min to complete. Your responses will be anonymous.

1. What are the main steps that you take to acknowledge an alarm from BAS dashboard?

> Alarm logic is written by people, which may not be complete. Hence, alarms are only a reference for the facility team to get more info if a complaint is received from occupants and fatal problems occur. In real situations, we need to consider many factors. Firstly, it depends on the type of alarms. Then we need to consider:
> 1). If the alarm still exists after a while,
> 2). If anyone is continuously complaining about a certain issue,
> 3). Whether there is significant failure in the system.
> Besides that, some alarms like freezestat tripped may be caused by the self-protection function of AHU when outside air temperature is too low. For this kind of alarm, there is no need to acknowledge.
> Sometimes, we need to evaluate the cost of the faults. For example, hiring a technician to do manual inspection can cost $100/hour. However, the energy waste due to high/cold zone temperature may be less than this cost. So, we have yearly inspection/commissioning to detect faults.

2. What is the difference between an 'HVAC general' alarm and an 'HVAC critical' alarm?

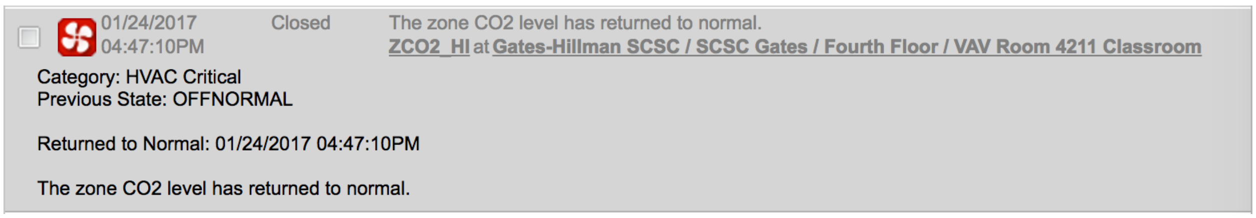
> This is just the rule written by product developer. In Gates building, the rules are reasonable for most cases.

3. Why are some alarms 'closed' immediately while others remain 'unacknowledged'? Is there a rule that decide those two types of alarms?

> Again, it depends on the rule.

4. Which of the 3 categories below do the following alarms belong to? (Please select the closest category per your opinion.)

4.1



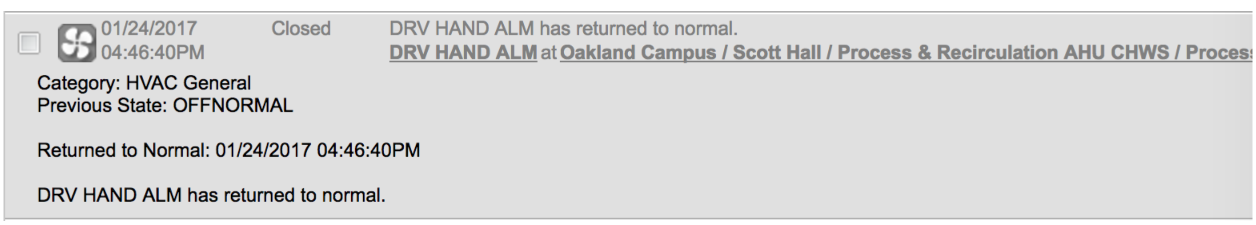
☐ Energy Consumption

☐ Thermal Comfort

☐ Equipment Maintenance

☒ **CO2 comfort**

4.2



☐ Energy Consumption

☐ Thermal Comfort

☒ Equipment Maintenance

4.3

01/24/2017 04:41:53PM Acknowledge Freezestat LT-2 indicates an alarm condition.
Freezestat LT-2 at **Oakland Campus / Mellon Institute / Third Floor / Collab/Mailroom BCU-1**
Category: HVAC General
Previous State: NORMAL

Freezestat LT-2 indicates an alarm condition. Click on event source for details.

☐ Energy Consumption

☐ Thermal Comfort

☒ Equipment Maintenance

4.4

01/24/2017 03:57:19PM Acknowledge The zone temperature is too cool.
ZTMP_LO at **Oakland Campus / Hamburg Hall / A-Level / A001E VRF 5-2**
Category: HVAC Critical
Previous State: NORMAL

The zone temperature is too cool.

☐ Energy Consumption

☒ **Thermal Comfort**

☐ Equipment Maintenance

4.5

09/20/2015 02:45:02AM Waiting for Normal The supply fan has exceeded its runtime limit.
SF RUNTIME at **Oakland Campus / Wean Hall / Fourth Floor / AHU-1**
Category: HVAC Maintenance
Previous State: NORMAL

Acknowledged by: Marty Altschul at 09/26/2016 10:43:34AM

The supply fan has exceeded its runtime limit.

☐ Energy Consumption

☐ Thermal Comfort

☒ Equipment Maintenance

5. Is there any alarm schedule? If yes, what’s the logic behind them?

There may be different setpoints and alarm thresholds in different seasons. Also, the alarms thresholds in different spaces, different days of week can also be different.

6. Please indicate the type of feedback you have heard from building occupants.

	Never	Once or twice	Sometimes	A lot	N.A.
Too warm	☐	☐	☐	☒	☐
Too cold	☐	☐	☐	☒	☐
Stuffy air	☐	☐	☐	☐	☒
Draft	☐	☐	☐	☐	☒

7. What features do you think would be most helpful in a BAS dashboard?

☐ View alarms on floorplan

☐ Filter out trivial alarms

☐ Rank alarms by impact categories

☒ Diagnose system faults (However, it's unrealistic nowadays)

☐ Others:____

8. How do you estimate the potential energy saving if a fault is corrected?

1). First to consider is what kinds of fault are related to energy. For example, if the fan is shut down. This fault is saving energy rather than wasting.

2). Too hot/too cold: may not have a large impact on energy

3). No sub metering. It is hard to estimate.

9. Do you have a categorized historical fault list for Gates building that we can refer to?

☐ Yes

☒ **No**

Han Li

10. What are the typical faults in the system?

☒ **Fan speed**

- ☒ **Damper stuck**
- ☒ Simultaneously heating and cooling
- ☒ Heating and cooling cycling (alternating frequently)
- ☒ Heating when cooling is needed
- ☒ Cooling when heating is needed
- ☐ Others ____

11. How do you diagnose the reason for a fault?

It depends on different faults. For example, zone temperature cannot reach setpoint since cooling or heating demand is beyond HVAC capacity. Main way is to inspect data. If the faults still exist, then see the current sensor data. If the faults are past, we should look at the historical data. Sometimes, it is sensor fault. Then we need to detect compare different sensor data. Since most current HVAC system is using feedback control loop, any sensor will impact entire system. It is a hard process to diagnose the system.

Appendix B. Room Sampling

B.1 Alarm Sources

Table 25. Alarm sources by space/equipment type

Floor	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	roof
Office	2	7	1	18	13	31	42	53	49	0
Classroom	0	0	2	7	3	6	8	6	8	0
Conference	0	0	0	0	2	3	1	3	3	0
Common	2	2	6	6	10	5	5	4	4	0
Service	0	0	0	0	2	6	2	3	1	0
HVAC Equipment	3	9	30	3	1	1	1	1	2	12
Meters	0	2	8	0	0	0	0	0	0	1
Critical	2	4	7	0	0	0	1	1	0	2
Total	9	24	54	34	31	52	60	71	67	15

Table 26. Alarm sources by category

Floor	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	roof
Occupant-related	4	9	9	31	30	51	58	69	65	0
Equipment-related	3	11	38	3	1	1	1	1	2	13
Critical	2	4	7	0	0	0	1	1	0	2
Total	9	24	54	34	31	52	60	71	67	15

B.2 Sampled Rooms

Table 27. Key parameters of sampled spaces

Room	type	Location	Discharge Air Temperature				Discharge Air Flow Rate (% of max design rate)				$\Delta T_{discharge}$ (Heating)	$\Delta T_{discharge}$ (Cooling)	V_{alarm}/V_{normal} (Heating)	V_{alarm}/V_{normal} (Cooling)
			Heating		Cooling		Heating		Cooling					
			normal	alarm	normal	alarm	normal	alarm	normal	alarm				
4115	office	External	89	101	60	58	90%	94%	65%	82%	12	2	1.04	1.26
4205	office	External	84	123	56	55	107%	110%	37%	75%	39	1	1.03	2.03
4124	office	Internal	NA	NA	59	54	NA	NA	71%	105%	NA	5	NA	1.48
5103	office	External	97	143	61	57	99%	101%	40%	83%	46	4	1.02	2.08
5001	office	External	69	71	61	58	47%	107%	36%	98%	2	3	2.28	2.72
6219	office	External	118	137	56	53	106%	108%	91%	130%	19	3	1.02	1.43
6207	office	External	112	139	59	56	78%	79%	124%	153%	27	3	1.01	1.23
6003	office	External	101	139	63	58	112%	114%	24%	42%	38	5	1.02	1.75
7117	office	External	77	89	61	57	101%	103%	71%	102%	12	4	1.02	1.44
7110	office	Internal	NA	NA	60	56	NA	NA	33%	42%	NA	4	NA	1.27
7007	office	External	81	106	62	56	106%	110%	27%	102%	25	6	1.04	3.78
7215	office	External	82	110	55	54	75%	83%	48%	75%	28	1	1.11	1.56
8017	office	External	106	123	57	53	93%	103%	46%	99%	17	4	1.11	2.15
8109	office	External	86	113	57	55	76%	105%	67%	102%	27	2	1.38	1.52
8112	office	Internal	NA	NA	57	52	NA	NA	35%	60%	NA	5	NA	1.71
8217	office	External	73	75	61	54	40%	102%	38%	100%	2	7	2.55	2.63
9127	office	External	90	120	63	56	75%	105%	46%	97%	30	7	1.40	2.11
9101	office	External	84	114	58	55	56%	113%	25%	64%	30	3	2.02	2.56
9219	office	External	87	118	60	56	103%	112%	43%	100%	31	4	1.09	2.33

9015	office	External	83	110	59	55	136%	169%	14%	44%	27	4	1.24	3.14
9116	office	Internal	NA	NA	58	53	NA	NA	24%	37%	NA	5	NA	1.54
4215	classroom	Internal	NA	NA	59	58	NA	NA	60%	80%	NA	1	NA	1.33
6002	classroom	External	83	100	58	56	101%	106%	34%	85%	17	2	1.05	2.50
4101	classroom	External	79	114	63	59	126%	137%	63%	63%	35	4	1.09	1.00
7114	classroom	Internal	NA	NA	63	58	NA	NA	51%	58%	NA	5	NA	1.14
9208	classroom	Internal	NA	NA	73	61	NA	NA	63%	69%	NA	12	NA	1.10
5117	conference	External	71	73	66	57	108%	109%	39%	45%	2	9	1.01	1.15
6115	conference	External	83	96	60	57	103%	109%	28%	76%	13	3	1.06	2.71
8115	conference	External	82	97	60	58	109%	116%	35%	66%	15	2	1.06	1.89
3101	common	Internal	NA	NA	64	54	NA	NA	67%	95%	NA	10	NA	1.42
4300	common	External	76	108	59	56	106%	112%	15%	32%	32	3	1.06	2.13
5000	common	External	83	107	55	53	87%	100%	63%	101%	24	2	1.15	1.60
6100	common	Internal	NA	NA	59	55	NA	NA	90%	103%	NA	4	NA	1.14
8200	common	Internal	NA	NA	65	56	NA	NA	57%	67%	NA	9	NA	1.18
6102	service	Internal	NA	NA	62	55	NA	NA	11%	13%	NA	7	NA	1.18
7200	service	External	101	103	61	60	62%	66%	5%	16%	2	1	1.06	3.20
7300	service	Internal	NA	NA	58	56	NA	NA	80%	103%	NA	2	NA	1.29
Average			87.1	109.2	60.2	55.9	92%	107%	48%	77%	22.1	4.3	1.16	1.62

Appendix C. Key Steps in Data Processing

This section includes key steps of the data processing. The complete code can be found in the miscellaneous folder.

C.1 Read and Extract Raw Alarms for Gates-Hillman Center

```
# Author: Han Li
# Date: Dec, 2016
# This is the script for extracting alarms from raw text data by building names.

# Utility functions
import.csv <- function(filename) {
  return(read.csv(filename, sep = ",", header = TRUE))
}

write.csv <- function(ob, filename) {
  write.table(ob, filename, quote = FALSE, sep = ",", row.names = FALSE)
}

# ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
# Import raw txt data
raw_all_vector <- readLines("CMU_BAS.txt") ;
# ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
# Parse raw data into different building.
Gates_Hillman_vector <- (grep('SCSC', raw_all_vector, value=TRUE));
# ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
# Output building-wise alarm data into .csv files
write(Gates_Hillman_vector, "Gates_Hillman_step1.csv")
```

C.2 Data Cleansing

```
# Import libraries
library(tidyr)
library(dplyr)
# Read the csv file created in script 1.
gates_hillman.cleaned.a <- import.csv("Gates_Hillman_step1.csv")
# Remove empty columns
gates_hillman.cleaned.a <- gates_hillman.cleaned.a[,1:9]
# Set feature names
cnames <- c("occurring_date","occurring_year_time","status","location","range",
        "description","acknowledge_date","acknowledge_year_time","fms")
colnames(gates_hillman.cleaned.a) <- cnames
# Remove alarms with undesired text from Gates_Hillman.csv
gates_hillman.cleaned.a <- clean.a(gates_hillman.cleaned.a, "FAULT", 3)
gates_hillman.cleaned.a <- clean.a(gates_hillman.cleaned.a, "SCSC Network", 4)
# Find and replace special characters.
gates_hillman.cleaned.b <- gates_hillman.cleaned.a
gates_hillman.cleaned.b$location <- gsub("I/O","I-O",gates_hillman.cleaned.a$location)
gates_hillman.cleaned.b$location <- gsub("Reception/Mail","Reception-
Mail",gates_hillman.cleaned.b$location)
gates_hillman.cleaned.b$location <- gsub("Work/Copy/Print","Work-Copy-
Print",gates_hillman.cleaned.b$location)
gates_hillman.cleaned.b$location <- gsub("Gates-Hillman SCSC /
","",gates_hillman.cleaned.b$location)
gates_hillman.cleaned.b$location <- gsub("Horn/Strobe","Horn-
Strobe",gates_hillman.cleaned.b$location)
# Split column "Building name/Floor/System/Short Info" into 4 feature columns.
gates_hillman.cleaned.c <- gates_hillman.cleaned.b %>%
  separate(location, c("building", "floor", "system", "short_info"), " / ")
# Concatenate date, year, and times.
gates_hillman.cleaned.c$occur <- paste(gates_hillman.cleaned.c$occurring_date,
                    gates_hillman.cleaned.c$occurring_year_time, sep = "")
gates_hillman.cleaned.c$acknowledge <- paste(gates_hillman.cleaned.c$acknowledge_date,
                   gates_hillman.cleaned.c$acknowledge_year_time, sep = "")
# Drop original time features.
gates_hillman.cleaned.c <- gates_hillman.cleaned.c[c(-1,-2,-10,-11)]
View(gates_hillman.cleaned.c)
class(gates_hillman.cleaned.c$occur[1])
grep("Feb-29",gates_hillman.cleaned.c$occur)
# Write the cleaned dataframe to a new csv file.
write.csv(gates_hillman.cleaned.c,"Gates_Hillman_step2.csv")
```

C.3 Calculate Alarm Durations

```
 # Pull out all the off normal alarms and calculate the time_diffs
for(i in 1:n) {
  print(i)
  row1 <- df.before[i,]
  if (row1$status == "OFF NORMAL") {
    for(j in i+1:n) {
      row2 <- df.before[j,]
      if (j == n+1){
        print("This alarm never go back to normal!")
        time_diff <- 999999  # The 'off normal instance' never go back to 'normal'.
        duration <- c(duration,time_diff)
        break # Break inner loop
      }
      else if (row2$status == "NORMAL" && row2$floor == row1$floor &&
             row2$building == row1$building && row2$system == row1$system &&
             row2$short_info == row1$short_info){
        time_diff <- row2$occur - row1$occur
        # Convert the units to hour
        if (units(time_diff) == "days") {
          time_diff <- time_diff * 24
        } else if(units(time_diff) == "mins") {
          time_diff <- time_diff / 60
        } else if(units(time_diff) == "secs") {
          time_diff <- time_diff  / 3600
        }
        duration <- c(duration,time_diff)
        break
      }
    }
  }
}
```

C.4 Add Air Handling Unit Information

```
# Author: Han Li
# Date: March 5, 2017
# This is the script for attaching AHU to occupant-related alarms
# ~~~~~~~~~~~~~~~~~~~~~~~~~~ Get AHU labels ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
# Get Gates-Hillman AHU and end-uses
AHU_tree_vector <- readLines("AHU.txt")
# Parse raw data into different building.
AHU_GH <- (grep('scsc', AHU_tree_vector, value=TRUE))
write(AHU_GH, "HVAC_t.csv")
# ~~~~~~~~~~~~~~~~~~~~~~~~~~ Import dataframes ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
################################# Start from here ################################
df.ahu <- import.csv("AHU.csv")
df.getAHU <- import.csv("Gates_Hillman_step4-0.csv")
# View(df.getAHU)

# ~~~~~~~~~~~~~~~~~~~~~~~~~~ Attach AHU to end-use ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
# Clean
df.ahu$enduse <- gsub("I/O","I-O",df.ahu$enduse)
df.ahu$enduse <- gsub("Reception/Mail","Reception-Mail",df.ahu$enduse)
df.ahu$enduse <- gsub("Work/Copy/Print","Work-Copy-Print",df.ahu$enduse)
df.ahu$enduse <- gsub("Gates-Hillman SCSC / ","",df.ahu$enduse)
df.ahu$enduse <- gsub("Horn/Strobe","Horn-Strobe",df.ahu$enduse)
df.ahu$enduse <- trimws(df.ahu$enduse)
# Create a new environment
ahu_env <- new.env()
# Store weather data into an environment (map), the key is the date
for(i in 1:nrow(df.ahu)) {
  ahu_env[[df.ahu$enduse[i]]] <- df.ahu[i,]$ahu
}
df.getAHU$AHU <- NA
# Add AHU to alarms
for(i in 1:nrow(df.getAHU)) {
  key <- as.character(df.getAHU$system[i])
  if (!is.null(ahu_env[[key]])) {
    print(ahu_env[[key]])
    df.getAHU$AHU[i] <- as.character(ahu_env[[key]])
  }
}
# Write AHU-labeled dataframe
write.csv(df.getAHU,"Gates_Hillman_step4-1.csv")
```

C.5 Add Weather Condition

```
# Author: Han Li
# Date: March 5, 2017
# This is the script for attaching weather conditions to alarms.

# ~~~~~~~~~~~~~~~~~~~~~~~~~~ Import data ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
df.addWeather <- import.csv("Gates_Hillman_step4-2.csv")
df.weather.daily <- import.csv("weather_2010-2016_daily.csv")
df.weather.hourly <- import.csv("weather_2010-2016_hourly.csv")
# ~~~~~~~~~~~~~~~~~~~~~~~~ Configure Features ~~~~~~~~~~~~~~~~~~~~~~~~
# Convert original occur.time feature to POSIXct
temp.a <- as.POSIXct(df.addWeather$occur.time, format="%I:%M:%OS %p")
# Get hour values
temp.b <- as.POSIXlt(temp.a)$hour
# Create key value for attaching hourly weather data
df.addWeather$occur.hour <- paste(df.addWeather$occur.date,temp.b,sep = "-")
df.weather.hourly$occur.hour <- paste(df.weather.hourly$date,df.weather.hourly$hour,sep = "-")
# ~~~~~~~~~~~~~~~~~~~~~~~~ Create Environment ~~~~~~~~~~~~~~~~~~~~~~~~
# Now we create two environment (hashtable) for attaching weather data.
# Then the rest roww vector is set to be the corresponding value.
# Key must be character
df.weather.daily$key <- as.character(df.weather.daily$date) # For daily data, the key is date
df.weather.hourly$key <- as.character(df.weather.hourly$occur.hour) # For hourly data, the key is date-hour
# Create new environments
env.weather.daily <- new.env()
env.weather.hourly <- new.env()

# For daily date, store weather data into the environment (map), the key is the date
for(i in 1:nrow(df.weather.daily)) {
  env.weather.daily[[df.weather.daily$key[i]]] <- df.weather.daily[i,2:6]
}
# For hourly date, store weather data into the environment (map), the key is the date-hour
for(i in 1:nrow(df.weather.hourly)) {
  env.weather.hourly[[df.weather.hourly$key[i]]] <- df.weather.hourly[i,3:19]
}
View(df.weather.hourly)
# Create a temperory weather dataframe. It will be attached to alarm_df later.
df.weather.daily.temp <- df.weather.daily[0,]
df.weather.hourly.temp <- df.weather.hourly[0,]
View(df.weather.hourly.temp)


# ~~~~~~~~~~~~~~~~~~~~~~~~ Attach Weather  ~~~~~~~~~~~~~~~~~~~~~~~~
# Search corresponding weather in the environment and attach it to the alarm instance
for(i in 1:nrow(df.addWeather)) {
  print(i)
  key <- as.character(df.addWeather$occur.date[i])
  df.weather.daily.temp <- rbind(df.weather.daily.temp,env.weather.daily[[key]])
}
# Attach hourly data in similar way
for(i in 1:nrow(df.addWeather)) {
  print(i)
  print(key)
  key <- as.character(df.addWeather$occur.hour[i])
  value <- env.weather.hourly[[key]]
  # If the there is no weather data for the hour, use previous hour's data
  if(is.null(value)){
    temp.row <- nrow(df.weather.hourly.temp)
    value <- df.weather.hourly.temp[temp.row,]
  }
  df.weather.hourly.temp <- rbind(df.weather.hourly.temp,value)
}

# Combine weather data to alarm dataframe.
df.addWeather <- cbind(df.addWeather,df.weather.daily.temp)
df.addWeather <- cbind(df.addWeather,df.weather.hourly.temp)
```

C.6 Alarm Priority Plot

```
# Generate 3D scatter plots by space type
p <- plot_ly(df.plot5, x = ~Energy, y = ~PMV, z = ~duration, color= ~priority,
             colors = c("grey", "red"),
            marker = list(symbol = 'circle', sizemode = 'diameter'),
            text = ~paste('System:',system,'<br>AHU:',AHU,'<br>Space type:',type,
                         '<br>Alarm type:',short.info,'<br>Mode:',condition))%>%
  layout(title = 'Thermal condition alarms by space type (2010 - 2016)',
        scene = list(xaxis = list(title = 'Energy impact compared to normal condition (kW)',
                              range = c(-2, 5),
                              zerolinewidth = 1,
                              ticklen = 5,
                              gridwidth = 2),
                  yaxis = list(title = 'Predicted Percentage Dissatisfied (%)',
                              # range = c(36.12621671352166, 91.72921793264332),
                              zerolinewidth = 1,
                              ticklen = 5,
                              gridwith = 2),
                  zaxis = list(title = 'Alarm duration (h)',
                              range = c(0, 5),  # alarms with duration less than 5 hours
                              zerolinewidth = 1,
                              ticklen = 5,
                              gridwith = 2)),
        paper_bgcolor = 'rgb(243, 243, 243)',
        plot_bgcolor = 'rgb(243, 243, 243)')
p
```

Bibliography

ACEEE. (2016). Retrieved December 8, 2016, from http://aceee.org/sector/commercial

Ahmed, A., Korres, N. E., Ploennigs, J., Elhadi, H., & Menzel, K. (2010). Mining building performance data for energy-efficient operation. *Advanced Engineering Informatics*, 341-354.

Ahmed, A., Ploennigs, J., Menzel, K., & Cahill, B. (2010, July). Multi-dimensional building performance data management for continuous commissioning. *Advanced Engineering Informatics*, 466-475.

Allen, W. H., Rubaai, A., & Chawla, R. (2016). Fuzzy Neural Network-Based Health Monitoring for HVAC System Variable-Air-Volume Unit. *IEEE Transactions on Industry Applications, 52*(3), 2513 - 2524.

American Society of Heating, Refrigerating and Air-Conditioning Engineers. (n.d.).

ASHRAE. (2013). *Energy Standard for Buildings Except Low-Rise Residential Buildings.* Atlanta: ASHRAE.

ASHRAE. (2016). *ASHRAE Handbook 2016: HVAC Systems and Equipment: SI Edition.* Atlanta.

ASHRAE. (2016). *ASHRAE Handbook 2016: HVAC Systems and Equipment: SI Edition.* Atlanta: ASHRAE.

Beghi, A., Brignoli , R., Cecchinato, L., Menegazzo, G., Rampazzo, M., & Simmini, F. (2016). Data-driven Fault Detection and Diagnosis for HVAC water chillers. *Control Engineering Practice, 53*, 79-91.

Beghi, A., Brignoli, R., Cecchinato, L., Menegazzo, G., Rampazzo, M., & Simmini, F. (2016). Data-driven Fault Detection and Diagnosis for HVAC water chillers. *Control Engineering Practice*, 79-91.

Bobker, M., Joseph, M., Ascazubi, M., Brown, T., Thorpe, K., & Essman, J. (2013). OPERATIONAL EFFECTIVENESS IN USE OF BAS. *13th International Conference for Enhanced Building Operations.* Montreal.

Burton, K., Raftery, P., Kennedy, B., Keane, M. M., & O'Sullivan, D. (2013). Review of automated fault detection and diagnostic tools in air handling units. *Energy Efficiency*, 335–351.

Domingues, P., Carreira, P., Vieira, R., & Kastner, W. (2015). Building automation systems: Concepts and technology review. *Computer Standards & Interfaces*, 1-12.

Evan, M. (2009). *Building Commissioning: A Golden Opportunity for Reducing Energy Costs and Greenhouse-gas Emissions.* Lawrence Berkeley National Laboratory. eScholarship.

Fan, C., Xiao, F., Madsen, H., & Wang, D. (2015, September). Temporal knowledge discovery in big BAS data for building energy management. *Energy and Buildings*, 75-79.

Fernbach, A., Granzer, W., & Kastner, W. (2011). Interoperability at the management level of building automation systems: A case study for BACnet and OPC UA. *Emerging Technologies & Factory Automation (ETFA), 2011 IEEE 16th Conference.* IEEE.

Ghosh, S., Roy, S., & Bandyopadhyay, S. K. (2012). A tutorial review on Text Mining Algorithms. *International Journal of Advanced Research in Computer and Communication Engineering*, 223-233.

Guo, Y., Wall, J., Li, J., & West, S. (2015). Intelligent Online Sensor Monitoring and Fault Alarm System in Heating Ventilation and Air Conditioning Systems. *2015 9th International Conference on Sensing Technology (ICST)* (pp. 789-794). IEEE.

Jadhav, N. Y. (2016). Building Management and Automation Technologies. In N. Y. Jadhav, *Green and Smart Buildings* (pp. 95-107). Springer.

Katipamula, S., & Brambley, M. R. (2005, January). Methods for Fault Detection, Diagnostics, and Prognostics for Building Systems— A Review, Part I. *International Journal of HVAC&R Research, 11*(1), 3-25.

Li, S., & Wen, J. (2014). A model-based fault detection and diagnostic methodology based on PCA method and wavelet transform. *Energy and Buildings*, 63-71.

Lord, S.-F., Noye, S., Ure, J., Tennant, M. G., & Fisk, D. J. (2016). Comparative review of building commissioning regulation: a quality perspective. *Building Research & Information, 44*, 630-643.

McQuiston, F. C., Parker, J. D., & Spitler, J. D. (2005). *Heating, ventilating, and air conditioning: analysis and design.* John Wiley & Sons.

Munasinghe, N. (2016). *Common Problems of our Industry*. Retrieved December 8, 2016, from http://branches.automatedlogic.com/branch/houston/release/common-problems-of-our-industry/

Namburu, S. M., Azam, M. S., & Luo, J. (2007). Data-Driven Modeling, Fault Diagnosis and Optimal Sensor Selection for HVAC Chillers. *IEEE Transactions on Automation Science and Engineering, 4*(3), 469-473.

Narayanaswamy, B., Balaji, B., Gupta, R., & Agarwal, Y. (2014). Data Driven Investigation of Faults in HVAC Systems with Model, Cluster and Compare (MCC). *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*, (pp. 50-59). New York, NY, USA.

Piette, M. A., Kinney, S. K., & Haves, P. (2001). Analysis of an information monitoring and diagnostic system to improve building operations. *Energy and Buildings*, 783-791.

Sallans, B., Bruckner, D., & Rus, G. (2006). Statistical Detection of Alarm Conditions in Building Automation Systems. *2006 4th IEEE International Conference on Industrial Informatics* (p. 6). IEEE.

SEKER, S. E., MERT, C., Al-NAAMI, K., OZALP, N., & AYAN, U. (2014). TIME SERIES ANALYSIS ON STOCK MARKET FOR TEXT MINING CORRELATION OF ECONOMY NEWS. *INTERNATIONAL JOURNAL OF SOCIAL SCIENCES AND HUMANITY STUDIES, 6*(1), 69-91.

Shukla, N. (2009). *HVAC presentation for beginers, Slide 12 Various AC systems*. Retrieved September 2016, from http://www.slideshare.net/guestf11b52/hvac-presentation-for-beginers

Simmonds, D., & Bhattacherjee, A. (2015). Smart Systems, Smarter Living: An Empirical Study of the Building Automation System in Organizations. *SIG GREEN Workshop at AIS Electronic Library* (pp. 1-13). AISeL.

Sustar, J., & Goldschmidt, I. (2007, September). Saving Energy with a Building Automation System. Retrieved from http://www.buildingautomationsystem.org/Benefits-Of-Building-Automation-Systems.html

U.S. DOE. (2013). *EnergyPlus Energy Simulation Software*. Retrieved December 16, 2013, from http://apps1.eere.energy.gov/buildings/energyplus/

U.S. DOE. (2016). *Building Energy Data Exchange Specification (BEDES)*. Retrieved September 2016, 2016, from http://energy.gov/eere/buildings/building-energy-data-exchange-specification-bedes

U.S. EIA. (2016, March 18). (U. E. Administration, Producer) Retrieved December 8, 2016, from https://www.eia.gov/consumption/commercial/reports/2012/energyusage/

U.S. Energy Information Administration. (2017, 1 22). *2012 CBECS Preliminary Results*. Retrieved from U.S.EIA: https://www.eia.gov/consumption/commercial/reports/2012/preliminary/

Urban, P., & Landryová , L. (2016). Identification and evaluation of alarm logs from the alarm management system. *Carpathian Control Conference (ICCC), 2016 17th International* (pp. 769-774). IEEE.

Wikipedia. (2017, 3 10). *Facility management*. Retrieved from https://en.wikipedia.org/wiki/Facility_management

Xiao, F., & Fan, C. (2014). Data mining in building automation system for improving building operational performance. *Energy and Buildings*, 109-118.

Yan, R., Ma, Z., & Kokogiannakisa, G. (2016). A decision tree based data-driven diagnostic strategy for air handling units. *Energy and Buildings, 133*, 37–45.

Yu, Y., Woradechjumroen, D., & Yu, D. (2014). A review of fault detection and diagnosis methodologies on air-handling units. *Energy and Buildings*, 550-562.

Yu, Z., Haghighat, F., & Fung, B. C. (2015). Advances and challenges in building engineering and data mining applications for energy-efficient communities. *Sustainable Cities and Society*, 33–38.

Han Li