PREDICTIVE ANALYTICS FOR BUILDING POWER DEMAND: DAY-AHEAD FORECASTING AND ANOMALY PREDICTION

by

Jing Lin

B. Eng., Jiangnan University, 2018

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

THE COLLEGE OF GRADUATE STUDIES
(Electrical Engineering)

The University of British Columbia
(Okanagan)

February 2021

The following individuals certify that they have read, and recommend to the College of Graduate Studies for acceptance, the thesis entitled:

Predictive Analytics for Building Power Demand: Day-Ahead Forecasting and Anomaly Prediction

submitted by Jing Lin in partial fulfillment of the requirements for

the degree of Master of Applied Science.

Examining Committee:

Zheng Liu, School of Engineering
supervisor

Kasun Hewage, School of Engineering
Supervisory Committee Member

Chen Feng, School of Engineering
Supervisory Committee Member

Jian Liu, School of Engineering
University Examiner

Abstract

The rising energy demand of buildings contributes to global resource consumption and greenhouse gas emissions. Understanding the energy consumption habits of a building is the first and most crucial step in order to achieve energy efficiency. Forecasting and anomaly prediction of power demand play an essential role in the electric industry, as it provides the basis for making decisions in power system planning and operation. However, there are several challenges in predictive analytics to improve energy efficiency. For instance, the information in the forecast needs to be interpreted by a person with domain expertise. Moreover, automated interpretation of upcoming abnormal behaviors needs ground-truth labeling, but labels are not always available from power meter data.

To address those problems, a novel Predictive Power Demand Analytics methodology (PPDAM) is proposed in this thesis, based on long short-term memory neural networks and symbolic aggregate approximation. The main target of PPDAM is to identify upcoming normal and anomalous demand patterns, through mining historical power demand and temperature data, and day-ahead power demand forecast. The PPDAM consists of two modules: power demand forecasting and anomaly prediction. The power demand forecasting module is utilized to predict day-ahead power profiles (time sequences), as a baseline for pattern generation and anomaly prediction. In the module of anomaly prediction, the historical and predicted time-series profiles are first transformed into patterns and later retrieved from a system of labeled repositories. The patterns are classified as normal or anomalous based on their frequency of appearance in the pattern repositories.

The power demand data cover the total electrical power consumption in seven buildings at the University of British Columbia (UBC). The experimental results indicate that a power forecast could be mapped as different foreseeable demand patterns, each with a specific probability of occurrence. Classification results indicate that our method is robust and fairly accurate in predicting anomalies despite the diversities of building types and behaviors. The outcomes of this work could

provide building operators with a solution to derive latent information in power consumption data. The derived information could be used to improve the working conditions of the building's power system.

Lay Summary

The energy demand of buildings continues to rise, having a significant impact on global resource consumption and greenhouse gas emissions. Energy efficiency is one of the lowest-cost methods to reduce greenhouse gas emissions and the cleanest way to meet the increasing demand for energy. In order to improve energy efficiency in buildings, upcoming normal and anomalous power demand patterns need to be identifies, through mining historical power demand and temperature data, and day-ahead power demand forecast.

The main contribution of this study lies in a new data-driven predictive analytics methodology for the total power consumption in the non-residential building sector. The novelty of this work includes the implementation of a new solution, based on Long Short-Term Memory (LSTM) Deep Neural Network (DNN) and the Symbolic Aggregate Approximation (SAX) data mining method, which offers major advantages over traditional approaches. Aligned with the current research challenges in the field, the proposed methodology takes advantage of real-time data to predict an activity indicator, providing accurate insight regarding the power demand of the building. This is the first time that LSTM neural networks and SAX are integrated for accurate insights regarding the power needs of the building, intending to support more efficient energy management.

Table of Contents

List of Tables

List of Figures

Glossary

PPDAM Predictive Power Demand Analytics methodology

UBC University of British Columbia

LSTM Long Short-Term Memory

DNN Deep Neural Network

SAX Symbolic Aggregate Approximation

UBC University of British Columbia

EOS Earth and Ocean Science

ML Machine Learning

ARX Auto-Regressive models with eXogeneous inputs

DL Deep Learning

MEM Main Electric Meter

RWS Rooftop Weather Station

MSE Mean Squared Error

USB University Services Building

AERL Aquatic Ecosystems Research Laboratory

CKC C. K. Choi Building

ESB Earth Sciences Building

FNH Food, Nutrition and Health Building

FSC Forest Sciences Center

PSB Pharmaceutical Sciences Building

STD Standard Deviation

Acknowledgments

I deeply acknowledge my supervisor Prof. Zheng Liu, Prof. Walter Mérida and Dr. Julián A. Fernández whose advice, encouragement, and support made this work possible. Not only did they grant me a great freedom in my study, but they also delivered me lots of academic knowledge. It is my honor to study and work under their guidance.

Thanks for the financial support from Natural Resources Canada (NRCAN) [grant number EVID-1026]. Thanks also to committee members Prof. Kasun Hewage and Prof. Chen Feng for their crucial role in completing my thesis.

I would also like to thank my dear colleagues Dr. Omar E. Herrera, Ms. Rakiba Rayhana, Mr. Amirhossein Zaji, Mr. Ran Zhang for sharing their academic experiences and constructive viewpoints generously with me during our discussions. Also, I need to thank Mr. Patrick Clark, who helped polish my writing.

Finally, sincere thanks to my family and friends for their understanding, encouragement and support in all these years.

Chapter 1: Introduction

1.1 Background

During the past two decades (2008 to 2018), the world energy consumption increased by 83.7% (13181 to 24738 TWh); the world carbon dioxide (CO_2) emissions increased by 49.7% (22394 to 33513 Mt) [1]. Current predictions show that this growing trend will continue. The buildings and buildings construction sectors combined are responsible for over one-third of global final energy consumption and nearly 40% of total direct and indirect CO_2 emissions.

The rapidly growing world energy use has already raised concerns over supply difficulties, exhaustion of energy resources and heavy environmental impacts (ozone layer depletion, global warming, climate change, etc.) [2]. The emission of greenhouse gases including CO_2 in higher layers of the atmosphere is known as the main cause of global warming phenomena [3]. Energy-related CO_2 emissions from buildings have risen in recent years. Several factors have contributed to this rise, including growing energy demand for heating and cooling with rising air-conditioner ownership and extreme weather events. Enormous emissions reduction potential remains untapped due to the continued use of fossil fuel-based assets, a lack of effective energy-efficiency policies and insufficient investment in sustainable buildings.

Among the aforementioned reduction potential, energy efficiency is one of the lowest-cost methods to reduce greenhouse gas emissions and the cleanest way to meet the increasing demand for energy [4]. Improved energy efficiency helps households, businesses, and institutions save money, be more competitive, and contribute to an improved quality of life. For this reason, improving energy efficiency in buildings is today a prime objective for energy policy at regional, national, and international levels, to reduce the amount of gas emission and fossil fuel consumption [2]. One most effective approach to reducing CO_2 emission and energy consumption with regards to new buildings is to consider energy efficiency at a very early design stage. On the other

hand, efficient energy management and smart refurbishments can enhance the energy performance of the existing stock. All these solutions entail accurate predictive analytics for understanding the energy consumption habits of a building and optimal decision making, which is the first and most crucial step in order to achieve energy efficiency.

1.2 Research Motivations

Predictive analytics is an area of statistics that deals with extracting information from data and using it to predict trends and patterns [5]. Predictive analytics for building's power demand can improve the supervision of building operation, establish a baseline for energy consumption, track annual progress of intensity improvements, help increase energy efficiency, and maximize energy savings [6]. Among various data-driven predictive analytics methods for improving building energy efficiency, predicting and monitoring the power consumption with the aim of identifying abnormal behaviors is promising and cost-effective [7]. Forecasting consumption patterns is also one of the most important stages for implementing energy management systems in buildings. Methods of power demand forecasting and anomaly prediction with time-series data are reviewed and presented in detail in Chapter 2. There are several research gaps in the usage of day-ahead forecasting for anomaly detection.

The power consumption in buildings is influenced by many factors, such as ambient weather conditions (especially the temperature), building structure and characteristics, the operation of sub-level components like lighting and HVAC (Heating, Ventilating, and Air-Conditioning) systems, occupancy, and their behaviors [8]. This complex situation makes it very difficult to accurately implement the prediction of building power consumption. First, the information contained in time series predictions usually needs to be interpreted by a human with domain expertise. Daily load time series can be quantified with a set of meaningful statistics. For example, the daily average load, the minimum and maximum load, the median load and so on, all could be useful statistics in later modeling and analysis. Besides these common descriptive statistics, other meaningful statistics, such as near-base load, near-peak load, high-load duration, rise time and fall time, could

be defined to describe the load profiles, [9]. These statistics are related to load time series shape and time and are useful for building energy management. However, the information contained in load profiles needs to be further processed and interpreted by expertise to apply. Additionally, as reviewed in Section 2.1 regarding power demand forecasting, it is unclear how the performance of models is influenced by input variables and model architectures.

Anomaly prediction is the process of identifying a prediction that deviates greatly from other observations in a data set. In the field of building energy consumption, the anomaly is equal to abnormal energy consumption, which is generally caused by equipment faults or improper operation. However, automated interpretation of upcoming abnormal behaviors needs ground-truth labeling, but labels are not always available from power meter data [10]. For practical applications, anomaly detection without anomaly labels is closer to the actual requirements of building operation managers, because it is expensive and time-consuming to generate labeled data for abnormal power consumption in buildings. Therefore, a promising way to solve this dilemma is to find an algorithmic way of labeling anomalies. If unsupervised methods were to be used to create those labels (as reviewed in Section 2.2), those labels would have to be interpreted to describe anomalies in consumption behaviors.

These remaining problems are motivations of seek effective methods, capable of labeling and predicting power demand, and extracting anomalous information from those predictions. To this end, a novel Predictive Power Demand Analytics methodology (PPDAM) is presented in this thesis to address the aforementioned research gaps.

1.3 Research Objectives and Proposed Methodology

The main target of PPDAM is to identify upcoming normal and anomalous demand patterns, through mining historical power demand and temperature data, and day-ahead power demand forecast. Therefore, a new solution advancing the area of predictive analytics of time series data can be provided. The power demand data are the total electrical power consumption in the building, i.e., the summation of the HVAC system, plug loads (computers, kettles, microwaves, etc.), lighting,

etc. It is measured by the main electric meter installed on the main electrical feeder into the building. The temperature data are generated by the Earth and Ocean Science (EOS) Rooftop Weather Station at the University of British Columbia (UBC). The combination of a supervised method, i.e., LSTM networks (part of the family of deep learning models) and an unsupervised clustering method, i.e., SAX, are explored. The PPDAM consists of two modules: power demand forecasting and anomaly prediction.

The power demand forecasting module is utilized to predict day-ahead power profiles (time sequences), as a baseline for pattern generation and anomaly prediction. Power demand forecasting facilitates the estimation of the required electricity, in advance, to control the demand effectively (peak shaving and load shifting) [11]. Predicted power demand sequences could be used for implementing model-predictive control strategies for HVAC equipment [12]. Automation systems can benefit from this information in order to make decisions autonomously by following energy-saving optimization strategies.

In the module of anomaly prediction, the historical and predicted time-series profiles are first transformed into patterns and later retrieved from a system of labeled repositories. The patterns are classified as normal or anomalous based on their frequency of appearance in the pattern repositories. The aim of these data is to provide greater insight into how a building consumes energy and, therefore, what improvements are likely to be most effective in reducing consumption. Because historical consumption data containing abnormal building behavior is often ignored, information gained from the anomaly prediction module can be used to notify building managers of issues with minimal delay, helping to reduce energy costs.

1.4 Contributions and Thesis Outline

The main contribution of this study lies in a new data-driven predictive analytics methodology for the total power consumption in the building sector. The novelty of this work includes the implementation of a new solution, based on LSTM deep neural networks and the SAX data mining method, which offers major advantages over traditional approaches. Aligned with the current

research challenges in the field, the proposed methodology takes advantage of real-time data to predict an activity indicator, providing accurate insight regarding the power demand of the building.

On one hand, the day-ahead power demand forecasting module, comprehending various input combinations and LSTM model architectures, increases the adaptability to day-types and improves the resulting accuracy by including associated last-week data. On the other hand, the anomaly prediction module, based on the unsupervised SAX method, automates pattern-labeling and pattern-retrieving of time series power consumption and eliminates the difficulty in obtaining ground truth labels of anomalous consumption in the field of pattern detection. This is the first time that LSTM neural networks and SAX are integrated for accurate insights regarding the power needs of the building, intending to support more efficient energy management. Building operators could take actions on completing building maintenance tasks according to the insights obtained.

This thesis is organized into five chapters. Chapter 1 gives the background information and motivation of this research. This chapter also briefly introduces the novelty of the proposed predictive power demand analytics methodology and the contributions of this research. Chapter 2 gives a comprehensive literature review of existing methods, from the aspects of power demand forecasting and anomaly detection of time series data. Engineering, statistical and machine learning methods in the field of power demand forecasting are reviewed. Statistical and machine learning methods in the field of anomaly prediction of time series data are reviewed. In Chapter 3, the PPDAM is proposed and described in detail. The three-step application workflow of PPDAM is introduced first to show the feasibility and potential of real-world implementation. Furthermore, the modeling of power demand forecasting and anomaly prediction modules are elaborated, including theories of long short-term memory networks and symbolic aggregate approximation, models training and testing, creation and labeling of pattern repositories, and the design of classification tests. Chapter 4 presents the experimental results of the modeling of the aforementioned two modules. Section 4.1 covers the information of seven buildings and datasets involved in the experiments. The optimal model configuration (inputs and model architecture) is revealed in Section 4.2;

labeled pattern repositories are portrayed in Section 4.3. Section 4.2 also shows the anomaly classification results of the seven buildings. Summary of the research, conclusions, limitations and future research are drawn in Chapter 5.

Chapter 2: Related Work

This thesis focuses on understanding how to utilize historical and predicted data to obtain the pattern profile of power consumption in a building, and to predict normal and anomalous patterns for the next 24 hours. Therefore, power demand forecasting and anomaly detection techniques are reviewed in this chapter.

2.1 Power Demand Forecasting

In the methodology presented in this thesis, power demand forecasting is the foundation of pattern mining. The power consumption in buildings is influenced by many factors, such as ambient weather conditions (especially the temperature), building structure and characteristics, the operation of sub-level components like lighting and HVAC systems, occupancy, and their behaviors [8]. This complex situation makes it very difficult to accurately implement the prediction of building power consumption. Recently developed forecasting methods include engineering methods, Statistical methods, machine learning (ML) methods. Zhao et al. [8] and Hernandez [13] provided a comprehensive review of research work related to the modeling and prediction of building energy consumption. Different methods are compared on the accuracy to forecast the building load given available information on weather conditions, building structure, operation of sub-level components, occupancy, and building use.

2.1.1 Engineering Methods

The engineering methods use very elaborate physical functions or thermal dynamics to precisely calculate the power consumption for all components of the building with building's and environmental information, such as external climate conditions, building construction, operation and HVAC equipment, as the inputs. They have been adequately developed over the past fifty years.

Touretzky and Patil [14] developed Auto-Regressive models with eXogeneous inputs (ARX)

to forecast power demand in conjunction with existing physics-based modeling approaches. Al-Homoud [15] reviewed two simplified engineering methods for estimating small and large buildings' energy consumption, respectively. Yao and Steemers [16] developed a simple engineering method of predicting a daily energy consumption profile for one season, for the design of a renewable energy system for residential buildings. They used the average consumption from large amounts of statistical data for the modeling of each electric appliance in the building.

Although these elaborate simulation tools are effective and accurate, in practice, there are some difficulties. Since these tools are based on physical principles, to achieve an accurate simulation, they require details of building and environmental parameters as the input data. On one hand, these parameters are unavailable to many organizations. For example, the information on each classroom in a building on campus is always difficult to obtain. This lack of precise inputs will lead to a low accurate simulation. On the other hand, operating these tools normally requires tedious expert work. The calculation of energy use for space heating and cooling for a building and its components may require the knowledge of a standard developed by the ISO (International Organization for Standardization). Those expertise making engineering methods difficult to perform and cost-inefficient. For these reasons, engineering methods are not considered in this study.

2.1.2 Statistical Methods

Statistical regression models correlate the power consumption with the influencing variables. These empirical models are developed from historical performance data, which means that before training the models, it needs to collect enough historical data. In terms of the research problems it solves, statistical regression models can be divided into two categories.

The first is to predict the power usage over simplified variables such as one or several weather parameters. Ansari et al. [17] calculated the cooling load of a building by adding up the cooling load of each component of the building envelope. Each sub-level cooling load is a simple regression function of the temperature difference between inside and outside. Dhar et al. [18] modeled heating and cooling load in commercial buildings with outdoor dry-bulb temperature as the only

weather variable. The second one is to estimate important parameters of power usage, such as total heat loss coefficient, total heat capacity and gain factor, which is useful in analyzing the thermal behavior of a building or sub-level systems. Jiménez and Heras [19] used the Auto-Regressive model with eXtra inputs (ARX) to estimate the U and g values of building components. Auto-Regressive Integrated Moving Average (ARIMA) models and ARIMA with eXternal inputs (ARIMAX) models [20] have also been applied to some applications, predicting and controlling the peak electricity demand for commercial buildings and predicting the power demand of the buildings.

2.1.3 Machine Learning Methods

Sometimes machine learning is intertwined with statistical modeling, but they are different even though there is an overlapping. Machine Learning is the field of study interested in the development of black-box algorithms to transform data into intelligent action without relying on rule-based programming. In the scenario where the need for accuracy in results is higher than the interpretability of the model, machine learning models will suffice.

In order to compare the performance of ML and statistical models in terms of accuracy and computational requirements, Makridakis et al. [21] evaluated models across multiple forecasting horizons using a large subset of 1045 monthly time series used in the M3 Competition. After comparing the performance of ten ML methods with that of eight traditional statistical ones, they found that the former are dominated across both accuracy measures used and for all forecasting horizons examined. Moreover, it was observed that their computational requirements are considerably greater than those of statistical methods.

ML approaches can be divided into supervised, unsupervised, reinforcement learning, etc. Time series forecasting falls in the field of regression problem in supervised ML learning. Many regression ML algorithms have been explored in the literature, which can be categorized as shallow ML and deep learning models. Shallow ML models include K-Nearest Neighbor regression (KNN), CART regression trees (CART), Support Vector Regression (SVR), Gaussian Processes (GP), etc. Zhou in [22] compared four ML-based algorithms for cooling load forecasting of a

large-scale shopping mall, including Chaos-SVR and wavelet-SVR hybrid forecasting algorithms, back propagation neural network and SVR forecasting algorithms.

According to Cai et al. in [23], deep learning (DL) techniques for the day-ahead building load forecast show higher performance compared to traditional ML algorithms. It can be concluded deep neural networks (DNNs) are much more widely used by the scientific community and the industry due to their high efficiency, the relatively small amount of time needed to set up the system, and their performance. A comprehensive overview of deep learning methodologies can be found in [24]. Specifically, LSTM networks are especially suited for time series forecasts due to their capability for learning long-term dependencies and avoiding vanishing gradients [25]. The performance of LSTM based models to address a challenging load forecasting problem is evaluated in [26] and [27]. By exploiting the long term dependencies in the time series, LSTM is capable of forecasting complex univariate electric load time series with strong non-stationarity and non-seasonality. LSTM networks are selected due to their suitability to process time-series data as well as because of its robustness in forecasting day-ahead power demand profiles.

2.2 Anomaly Prediction of Time Series Data

Anomaly prediction in this thesis is the process of identifying a prediction that deviates greatly from other observations in a data set. In the field of building energy consumption, the anomaly is equal to abnormal energy consumption, which is generally caused by equipment faults or improper operation. In recent years, data-driven approaches are widely adopted to achieve reliable and robust anomaly detection for building energy consumption. Wang et al. [28] summarize anomalies into two levels according to the scope of the research object: whole building level, and system and component level. The whole building anomaly detection only focuses on the overall consumption of the building, while system and component anomaly detection can identify and locate exactly which component or subsystem leads to anomalous issues. This thesis focuses on whole building anomaly detection.

Despite the simplicity of the aforementioned anomaly classification, a more common concern

during anomaly detection is that whether there are explicit labels for anomalies. Therefore, existing anomaly detection methods relevant to buildings can be divided into two categories [10]: anomaly detection with anomaly labels and anomaly detection without anomaly labels. Essentially, the aim of anomaly detection with anomaly labels is to obtain accurate classification models, because the prediction needs to be classified as normal or anomalous. However, for practical applications, anomaly detection without anomaly labels is closer to the actual requirements of building operation managers, because it is expensive and time-consuming to generate labeled data for abnormal power consumption in buildings. Methods of anomaly detection without anomaly labels can be divided into statistical methods and machine learning methods.

2.2.1 Statistical Methods

When a power system is in operation, each node generates a large volume of operational data in real-time, including power consumption, voltage, current, active power and reactive power of the node. These data are random due to random disturbances from nonelectrical (e.g., the external environment (region, climate and human Factors)) and electrical factors (e.g., technical failure of the internal equipment of the power grid and connections of the distributed power sources), thereby forming random power variables (parameters). These parameters, which change with time, reflect the operating state of the power grid. Zhang [29] presented a theoretical high-dimensional random matrix model approach based on big data. Simulation of massive volumes of data based on a big data processing platform is the key to the efficiency of the proposed method. However, power system data are still considered confidential information that belongs to the operating companies. Ordinary researchers are unable to obtain the corresponding operational data for power grids, which is a bottleneck in big data research. Thus, considering the limited data size and number of parameters (only power consumption and temperature), this method is not feasible to be implemented in the data collected from one building.

Chou and Telaga [30] applied a two sigma rule to detect anomalies, and this rule simply classifies any points outside of two standard deviations from the mean as anomalous data. Generalized

extreme studentized deviate (GESD) algorithm has been implemented in anomaly detection in building energy consumption and proved to be computationally efficient in handling masses of building energy data [31]. These statistical methods have some inherent disadvantages, because the anomaly is generally defined as a point outside of several standard deviations from the mean. This kind of definition is straightforward and intuitive, but it may lead to the rough classification of high energy consumption, and the complexity of electricity consumption patterns may be ignored.

Fonseca [32] presented a SAX-based statistical method to automatically cluster typical days of energy consumption in one or several buildings. The method is based on an optimized version of the Symbolic Aggregate approXimation (SAX) method, which is adopted to tackle the challenge of high dimensionality in time series data. SAX is a data mining technique for clustering time series with applications in building fault detection and building performance assessment. The SAX process converts time-series data into a set of strings that can then be grouped together for the purposes of clustering, motif and discord analysis, and finding anomalous behavior. The number of clusters and accuracy of SAX highly depends on two highly sensitive input variables, i.e., the word size and the alphabet size. One novelty of this work is the automation of parameter selection of load profiles for different building use types. They proposed the use of the genetic algorithm NSGA-II to optimize the number of words and alphabet size of SAX subjected to three fitness objectives, i.e., maximize data accuracy and compression and minimize complexity. Potential future uses of the approach include advanced studies in fault detection and calibration of urban building performance models. Miller et al. in [33] developed a SAX-based framework for motif and discord extraction, and clustering to detect the underlying structure of building performance data. SAX allows discretization of time series data which facilitates the use of various motif and discord detection algorithms. The process breaks time series data into subsequences which are converted into an alphabetic symbol. These symbols are combined to form strings to represent the original time series enabling various mining and visualization techniques. Infrequent daily patterns are filtered and tagged for analysis of potential energy savings opportunities. They applied the approach on 474 days of example data from an international school campus in a tropical climate

and 407 days of data from an office building from a temperate European climate. Discords are filtered resulting in 17 and 22 patterns found. Selected discords are investigated and many correlate with specific failures and energy savings detected by the on-site operations staff.

2.2.2 Machine Learning Methods

Machine learning techniques for anomaly detection without labels can be classified into unsupervised learning-based methods and regression-based methods [34]. Commonly used unsupervised learning algorithms include principal component analysis (PCA), nearest-neighbor based techniques, clustering, and association rule [35]. Miller et al. in [36] reviewed research on unsupervised machine learning techniques as applied to non-residential building performance control and analysis. Goldstein et al. evaluated 19 different unsupervised anomaly detection algorithms on 10 datasets from different application domains and revealed the strengths and weaknesses of the different approaches [37].

PCA is one of the most popular algorithms in the field of building system fault detection. It transforms a group of correlated variables into a new group of variables that are uncorrelated or orthogonal to each other. The k-nearest-neighbor global unsupervised anomaly detection algorithm focuses on global anomalies and is not able to detect local anomalies. First, for every record in the dataset, the k-nearest-neighbors have to be found. Then, an anomaly score is computed using these neighbors, whereas two possibilities have been proposed [38]. The clustering algorithm divides measurements concerned into several clusters according to the geometric distances between them. Data in a cluster have similar statistical characteristics. Data in different clusters have quite different statistical characteristics. In general, the statistical characteristics of faulty data are different from those of normal data. They should belong to different clusters [39]. The Association rule algorithm is powerful in discovering relations between various measurements. The discovered relations could reveal the system operation patterns, but it needs domain experts to further analyze the patterns for fault detection [40].

Regression-based methods include deep learning models, support vector regression (SVR),

etc. Deep learning techniques can be applied to build a regression model of which the output is a continuous variable such as energy consumption, temperature and so on. Mavromatidis et al. trained an artificial neural network to predict the energy benchmarking values for abnormal detection of supermarket energy consumption [21]. SVR finds a hyperplane that maximizes the margin and minimizes the errors which are outside of the margins. Zhao et al. proposed an SVR-based method to detect system-level incipient faults in HVAC systems. The support vector regression algorithm is used to develop the reference performance index models [41].

For machine learning methods, the anomaly is generally defined as a point outside the boundaries based on the predicted value, and in this context, the position of the boundaries is not fixed. However, the above methods can only judge whether the object is an outlier, while the severity of the outlier is not evaluated. Moreover, machine learning methods rely on the training data heavily. They could obtain a very high fault detection and diagnosis accuracy and could work in the cases that some sensors are missing. They are feasible to be trained and applied automatically, but the models should be trained using a large amount of normal data for fault detection and faulty data for fault diagnosis. They cannot extrapolate beyond the range of the training data.

It can be concluded that the SAX-based statistical method is shown to be a simple but robust option for automated power profile labeling and anomaly detection. Labeling patterns can be helpful for load condition monitoring [42]. In this thesis, SAX approach is applied to identify day-ahead anomalous or normal power patterns, due to its calculation simplicity compared to machine learning methods, lower dependence on big data, and the capability to reduce the high-dimensionality. The categories of tasks undertaken by SAX include clustering, novelty detection, motif and discord detection, rule extraction, and visual analytics.

Chapter 3: Predictive Power Demand Analytics Methodology

The proposed Predictive Power Demand Analytics Methodology (PPDAM) is thoroughly explained in this chapter. Section 3.1 introduces the overall architecture of the system, including the workflow of data collection, power demand forecasting, and anomaly prediction modules. Section 3.2 introduces the design of the power demand forecasting module, including theories of LSTM and SAX, options of LSTM model architectures and inputs, and how to evaluate forecasting performance. Section 3.3 introduces how to create and label pattern repositories, and classify patterns as normal and anomalous.

3.1 The Overall Application Workflow of PPDAM

The application of PPDAM is illustrated in Figure 3.1. The overall system includes three steps: data collection, power demand forecasting, and anomaly prediction. The power demand data cover

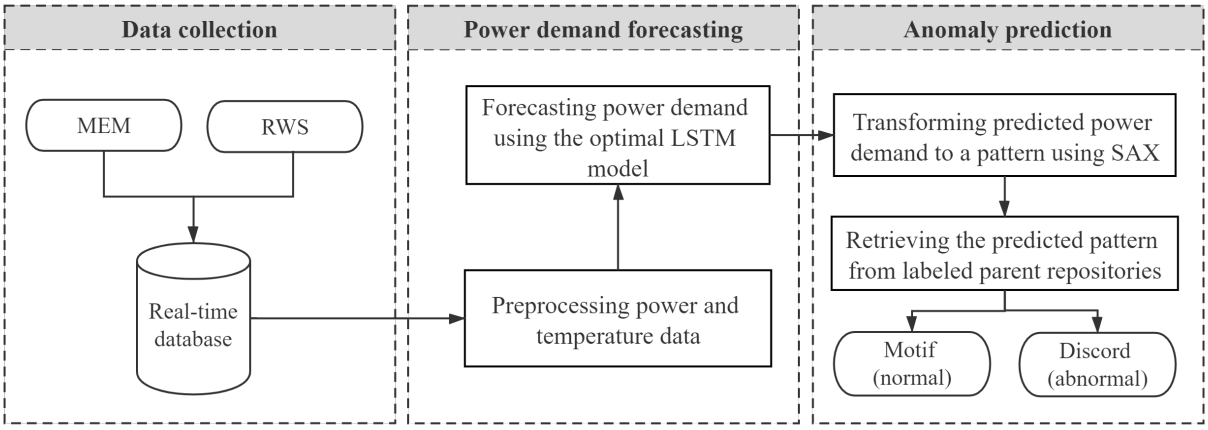

Figure 3.1: Workflow of the application of PPDAM

the total electrical power consumption in the building, measured by the main electric meter (MEM) installed on the main electrical feeder into the building. The temperature data are generated by the EOS Rooftop Weather Station (RWS) at UBC. The optimal combination of power and temperature data (introduced in Section 3.2.2 and Section 4.2.2) is extracted from the database and sent to

the data preprocessing unit, where raw inputs get cleaned and normalized. Processed data flow into the optimal LSTM model, which predicts the day-ahead time-series power demand. The predicted power demand sequence, in the form of continuous data, is converted into a discrete power demand pattern by a SAX pattern transformer. The power demand pattern is retrieved from labeled pattern repositories in which patterns are classified respectively as frequent patterns (motifs) and infrequent patterns (discord, anomalies). Thus, the motif or discord label is assigned to the predicted pattern.

3.2 Power Demand Forecasting

Figure 3.2 shows the design of the power demand forecasting module based on LSTM modeling. Historical power and temperature data are gathered and fed into the data preprocessing unit, where raw data are cleaned and standardized. Processed data are utilized to train and test LSTM models, and to generate time-series predictions of electricity power for the next 24 hours. Metrics results, e.g., Mean Squared Error (MSE), are evaluated during model testing. Power demand predictions and historical power demand are both converted to discrete patterns by SAX pattern transformers and the average distance between predicted and historical patterns is calculated. In order to select the optimal LSTM model architectures and inputs, the metrics results and SAX pattern distance need to consider and balance.

3.2.1 Long Short-Term Memory Neural Networks

LSTM neural networks are a type of recurrent neural networks (RNNs). RNNs take the form of a chain of looping modules of the neural network, which means connections between nodes form a directed graph along a temporal sequence. Thus, RNNs can save the representation of previous input(s) in memory while generating current output(s). LSTM neural networks also contain this chain structure, as shown in Figure 3.3. The input of LSTM neural network X is in three dimensions - (m, T_x, n_x), where m is the number of data samples and T_x is the number of timesteps of each sample. In Figure 3.3, $X^{\langle t \rangle}$ represents the input X at the timestep t and n_x is the number of

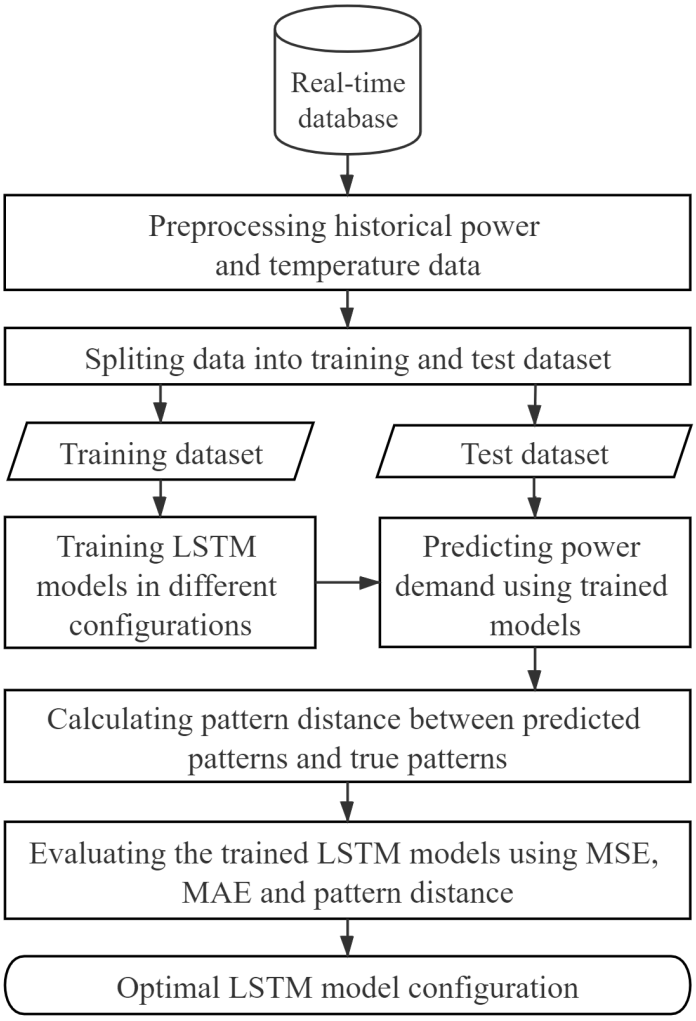

Figure 3.2: Design of the power demand forecasting module based on LSTM modeling

input variables. For example, if one thousand samples of the last 24-hour power and temperature data are taken as input, the input shape to the LSTM model would be (1000, 24, 2). The output for each timestep is represented by $a^{\langle t \rangle}$. Hidden state $c^{\langle t \rangle}$ is a representation of previous inputs, enabling the participation of inputs from previous timesteps in the estimation of current output.

The inner cell structure is illustrated in figure 3.4. For one-time forward propagation, parameters are the same for each time step of input. There are three gates in LSTM cell: input gate, forget

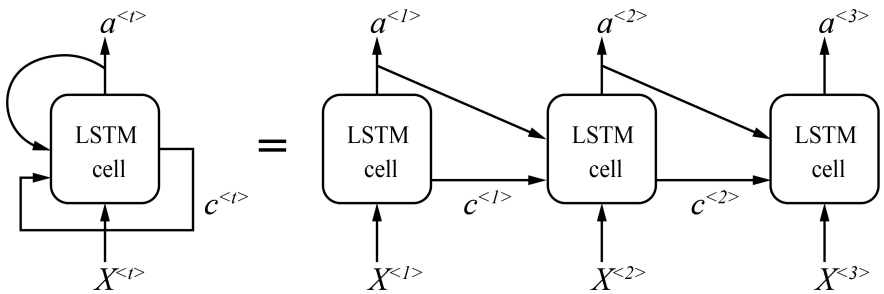

Figure 3.3: Chain structure of LSTM neural networks

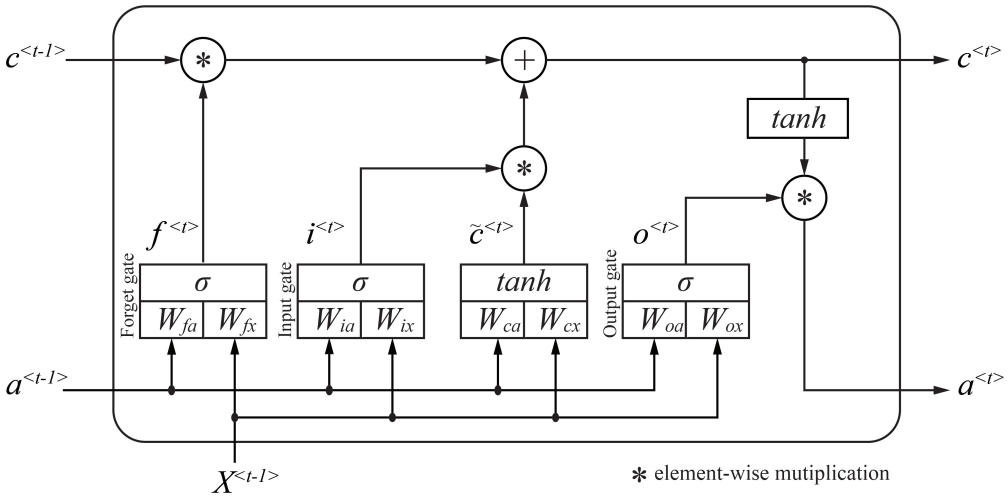

Figure 3.4: Cell structure of LSTM networks

gate and output gate. Default activation function of these gates is sigmoid function (Eq. (3.1)).

$$Sigmoid(x) = \sigma(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{1+e^x} \tag{3.1}$$

As shown in Figure 3.5, sigmoid function has domain of all real numbers with return value monotonically increasing from 0 to 1.

Inputs of these three gates are $X^{\langle t \rangle}$ and $a^{\langle t-1 \rangle}$. what happened to inputs inside gates shows in Eq. (3.2), (3.3) and (3.4). Outputs of these gates are naturally between zero and one: one represents "completely keep this" while zero represents "completely get rid of this". Therefore, three gate outputs are taken as control signal later.

$$i^{\langle t \rangle} = \sigma(W_{ia} \times a^{\langle t-1 \rangle} + W_{ix} \times X^{\langle t \rangle} + b_i) \tag{3.2}$$

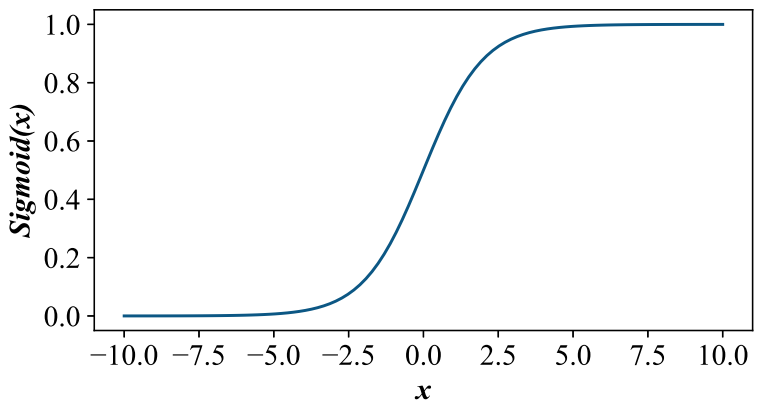

Figure 3.5: Activation function of *sigmoid*

$$f^{\langle t \rangle} = \sigma(W_{fa} \times a^{\langle t-1 \rangle} + W_{fx} \times X^{\langle t \rangle} + b_f) \tag{3.3}$$

$$o^{\langle t \rangle} = \sigma(W_{oa} \times a^{\langle t-1 \rangle} + W_{ox} \times X^{\langle t \rangle} + b_o) \tag{3.4}$$

Output of candidate box between input gate and forget gate is $\tilde{c}^{\langle t \rangle}$ which is candidate value of hidden state $c^{\langle t \rangle}$.

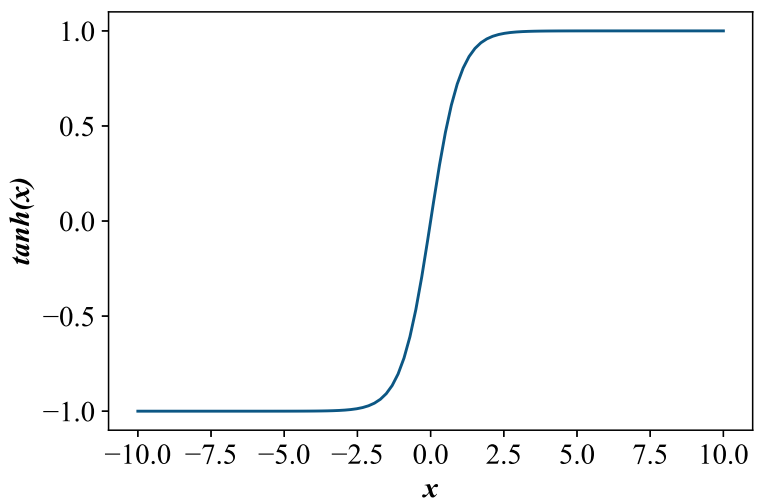

Figure 3.6: Activation function of *tanh*

$$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{3.5}$$

$$\tilde{c}^{\langle t \rangle} = tanh(W_{ca} \times a^{\langle t-1 \rangle} + W_{cx} \times X^{\langle t \rangle} + b_c) \tag{3.6}$$

After get three control signals and candidate hidden state, hidden state $c^{\langle t \rangle}$ and output for this

timestep $a^{\langle t \rangle}$ can be calculated as Eq. (3.7) and Eq. (3.8).

$$c^{\langle t \rangle} = i^{\langle t \rangle} * \tilde{c}^{\langle t \rangle} + f^{\langle t \rangle} * c^{\langle t-1 \rangle} \tag{3.7}$$

$$a^{\langle t \rangle} = o^{\langle t \rangle} * tanh(c^{\langle t \rangle}) \tag{3.8}$$

The symbol "$*$" in Figure 3.4, Eq. (3.7) and (3.8) represents the element-wise multiplication. Distinguished from the common matrix product, the element-wise multiplication takes two matrices of the same dimensions and produces another matrix of the same dimension as the operands, where each element i, j is the product of elements i, j of the original two matrices.

3.2.2 Model Inputs and Architectures Design

Four combinations of power and temperature, and three model architectures were designed to validate the performance of the LSTM models on predicting power demand. Figure 3.7 shows a

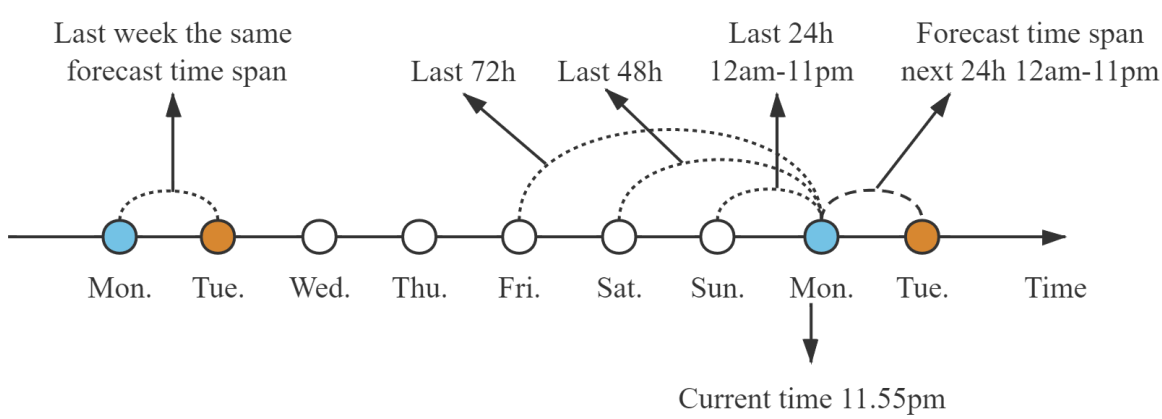

Figure 3.7: Time spans of inputs

variety of potential time spans for inputs. The sampling time is equal to one hour. In this figure, each circle represents the beginning of a day (12 am), and the dotted lines represent time spans. The last week's 24-hour realization indicates the same time span as the forecast time span but on the previous week. For example, as shown in Figure 3.7, if the present day is Sunday and the

current time is 11:55 pm, the forecast time span is 24 hours (the next day, Monday starting at 12 am), and the last week's 24h realization starts on the previous Monday at 12 am.

Four input combinations are considered for the LSTM modeling:

- Input 1: last n-hour power data samples, where $n \in \{24, 48, 72\}$.
- Input 2: last n-hour power data samples combined with last n-hour temperature samples, where $n \in \{24, 48, 72\}$.
- Input 3: last week's 24h power data samples and last n-hour power data samples, where $n \in \{24, 48, 72\}$.
- Input 4: last week's 24h power data samples combined with last week's 24h temperature samples, and last n-hour power data samples combined with last n-hour temperature samples, where $n \in \{24, 48, 72\}$.

Additionally, three LSTM model architectures are considered: standard architecture (Figure 3.8.a), single-encoder decoder architecture (Figure 3.8.b), and multi-encoder decoder architecture (Figure 3.8.c). Marino et al. in [43] compared the performance of standard architecture and single-encoder decoder architecture for load forecasting. Both were trained and tested for one-hour and one-minute timestep resolution data. The dataset [44] they used consists of four years of one-minute resolution electric power consumption data collected from a residential customer [45]. Marino et al. concluded that the standard LSTM architecture failed at one-minute resolution data while performing well on the one-hour resolution data. In addition, it was concluded that the single-encoder decoder architecture performed well in both datasets. However, to compare the effectiveness of these algorithms, the authors demonstrated that both algorithms needed to be tested on real-world datasets, which is the reason why include these two architectures as part of the experiment. Furthermore, multi-encoder decoder architecture is added to compare the performance of LSTM models given multiple inner representations of the input.

As mentioned in Section 3.2.2, the inputs for the three architectures are 3D arrays in the shape of (m, T_x, n_x). The shape of an array is the number of elements in each dimension. In the standard

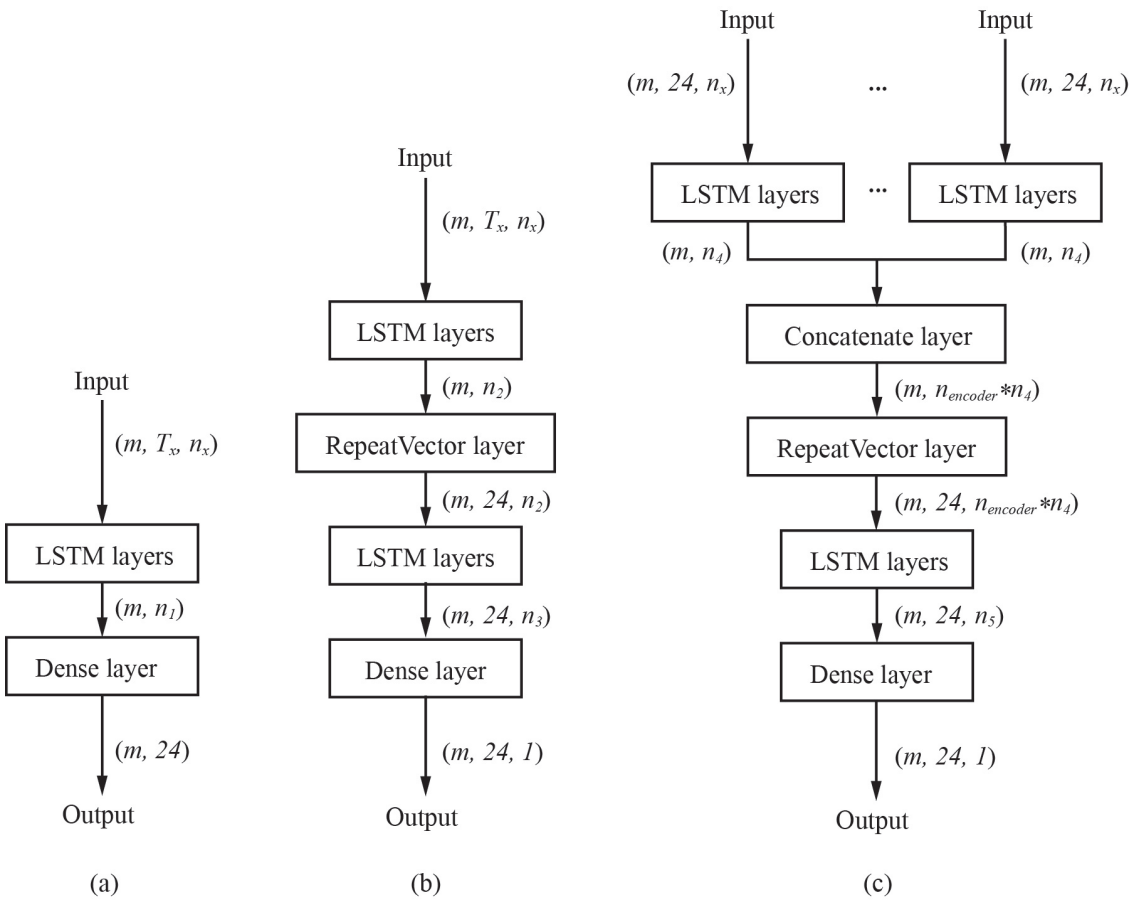

Figure 3.8: Model architectures: (a) standard architecture, (b) single-encoder decoder architecture, (c) multi-encoder decoder architecture

model architecture, the input is compressed by the LSTM layers into a 2D array in the shape of (m, n_1), where n_1 is the number of neurons in the last LSTM layer. In the three architectures, the dense layer reshapes the dimension of the data flowing out from the last LSTM layer, to guarantee that the output shape is as expected. The output of the standard model has the shape of (m, n_{dense}), where n_{dense} is the number of neurons in the dense layer and it is set equal to T_y. T_y represents the length of forecast time span, and it is 24h in the implementation.

In the single-encoder decoder architecture, the input is converted by the LSTM layers into a 2D inner representation of which the shape is (m, n_2), where n_2 is the number of neurons of the last LSTM layer. As opposed to standard architecture, this inner representation is decoded into a 3D tensor of which the shape is (m, T_y, n_3). Next, the 3D tensor passes through a dense layer, for the same reason as for the standard model. However, the output of the single-encoder decoder model

is a 3D tensor in the shape of (m, T_y, 1). The last digit represents the number of output variables, which in this research is equal to one because only power demand is being predicted.

In the multi-encoder decoder architecture, each element (the shape is (m, 24, n_x)) in an input set of size $n_{encoder}$ is assigned an individual LSTM encoder. As show in Figure 3.8.c, each LSTM encoder converts the corresponding element into a 2D representation in the shape of (m, n_4). Then, all encoded 2D inner representations are concatenated, prior to being decoded. Subsequent processing stages are similar to the single-encoder decoder architecture, as well as the output's shape.

To select the best LSTM architecture, all models are compared under the same set of inputs and value of n. Furthermore, to select the best configuration and set of inputs, the following analysis was performed, under the same LSTM architectures:

- By varying the values of n for each input, one can conclude if increasing n can improve forecasting accuracy.
- By comparing separately Input 1 and Input 2, and Input 3 and Input 4 for the same value of n, one can conclude if adding temperature data improve predicting performance.
- By comparing separately Input 1 and Input 3, and Input 2 and Input 4, one can conclude how the last week's realization improves the predictability.

3.2.3 Metrics for LSTM Models Training and Testing

There are several metrics in place to update the weight values of LSTM networks, such as the Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), etc. By minimizing the error up to a certain acceptable threshold, the most optimal model is achieved. Zhu et al. in [7] utilized MAE and MAPE as metrics to evaluate prediction models. Raza et al. in [46] present a methodology on how to use MSE to train supervised learning models. Wang et al. in [47] reviewed the statistical deterministic performance of MAE, MAPE, and RMSE for forecasting of wind speed.

Table 3.1 shows the formulas of the two figures of merit utilized in the methodology to estimate

Table 3.1: Figures of merit for LSTM models

Name	Formula	Purpose
MSE	$1/n_y \times \sum_{i=1}^{n_y}(y_i - \hat{y}_i)^2$	Loss function
MAE	$1/n_y \times \sum_{i=1}^{n_y}\lvert y_i - \hat{y}_i \rvert$	Metric

the prediction accuracy of the forecasting model. In formulas, n_y is the length of one predicted power demand sequence, which is equal to 24 in this research. MSE is used as the loss function, and MAE is used as a metric when training and testing the models. The loss function is used as the objective function to be minimized by the optimizer in Keras [48]. In Keras a metric is used to judge models' performance, which is independent of the optimization process. Keras is briefly introduced in Section 4.2.1.

3.2.4 SAX Distance Validation

In addition to MSE and MAE, the average error (Eq. (3.11)) between historical patterns (HPs) in the test set and their forecasts (PPs) is considered when selecting the best model architecture and inputs. First, to define the error between one true pattern and the prediction of it, the concept of distance between SAX representations is utilized [49]. The distance between two SAX representations, $\hat{H}$ and $\hat{P}$, requires first obtaining the distances between each pair of symbols. To calculate the distance between the r-th symbol and s-th symbol ($r,s \in \{1,2,3,4\}$) in the chosen alphabet, Eq. (3.9) is used.

$$dist(symbol_r, symbol_s) = \begin{cases} 0, & if \mid r-s \mid \leq 1 \\ \beta_{max(r,s)-1} - \beta_{min(r,s)}, & otherwise \end{cases} \tag{3.9}$$

Table 3.2 shows a lookup table of the distance between symbols in the alphabet ['a', 'b', 'c', 'd']. The distance between two SAX representations is given by Eq. (3.10) where $\hat{h}_i$ and $\hat{p}_i$ are the i-th symbols ($i \in \{1,2,...,w\}$) of $\hat{H}$ and $\hat{P}$.

Table 3.2: A distance lookup table when alphabet size is 4

	a	b	c	d
a	0	0	0.67	1.34
b	0	0	0	0.67
c	0.67	0	0	0
d	1.34	0.67	0	0

$$MINDIST(\hat{H},\hat{P}) = \sqrt{\frac{n}{w}}\sqrt{\sum_{i=1}^{w}(dist(\hat{h}_i,\hat{p}_i))^2} \tag{3.10}$$

$$DIST(HPs,PPs) = \frac{1}{m}\sum_{j=1}^{m}(MINDIST(\hat{H}_j,\hat{P}_j)) \tag{3.11}$$

As shown in Eq. (3.11), the average distance between Historical Patterns (HPs) and Predicted Patterns (PPs) is the sum of the distances between each predicted and true pattern divided by the number of predicted patterns (as well as the number of historical patterns) m. The selected optimal model minimizes $DIST(HPs,PPs)$ so that the accuracy of power demand pattern forecasting is guaranteed.

3.3 Anomaly Prediction

As described in Section 3.1, a SAX pattern can be classified as a motif or discord by retrieving this pattern from the labeled parent repositories. The labeled parent repositories are the core of the anomaly prediction process. The creation and labeling of parent pattern repositories are described in Section 3.3.2. A classification test designed to evaluate the anomaly prediction performance of the PPDAM system is introduced in Section 3.3.3.

3.3.1 Symbolic Aggregate Approximation

As with most problems in computer science, the suitable choice of representation affects the ease and efficiency of time series data mining. SAX transformation could generate concise and explicit indicators of the power systems' working conditions because symbolic representations allow the labeling of patterns [33]. This section briefly introduces the basics of SAX theory.

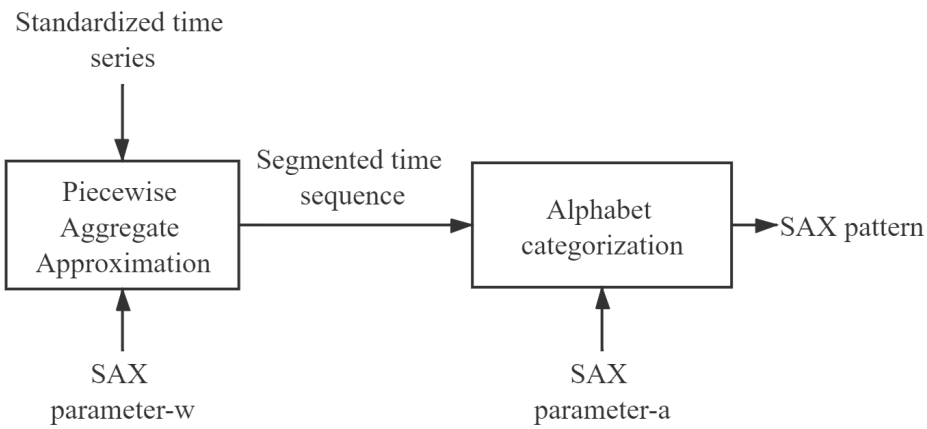

Figure 3.9: Process of SAX pattern transformation

SAX transforms a time series of an arbitrary length n to a string of an arbitrary number of characters w. This transformation allows dimension reduction, given that $n > w$. The string takes characters from an alphabet of length a, where $a > 2$. Lin et al. in [49] provide a detailed illustration of the SAX method. A brief description of the SAX method is presented in the paragraphs below.

It first requires the transformation of the standardized data into a Piecewise Aggregate Approximation (PAA) representation, as shown in Figure 3.9. The main target of PAA is to reduce the time series from n dimensions (n equals 24 for daily power profiles) to w dimensions. Thus, the data are divided into w equal sized segments and the mean value of the data falling within a segment is assigned to that segment. The solid black line in Figure 3.10 shows an example of the load shape after sampled down to w segments, where w is equal to 6 and the time span for each time window is 4 hours. Each grey line represents the mean value of the data falling within this segment.

The next stage of the SAX method is to categorize the PAA vector to obtain a discrete representation. Breakpoints are presented as a sorted list of numbers $\beta_1, \ldots, \beta_{a-1}$ such that the area under a $N(0,1)$ Gaussian curve from β_i to β_{i+1} is equal to $1/a$ (β_0 and β_a are defined as $-\infty$ and $+\infty$, respectively). As shown in Figure 3.10, the dashed blue, orange and black line represent β_1, β_2 and β_3, respectively. The distribution space of the y-axis in the figure is divided into four equiprobable

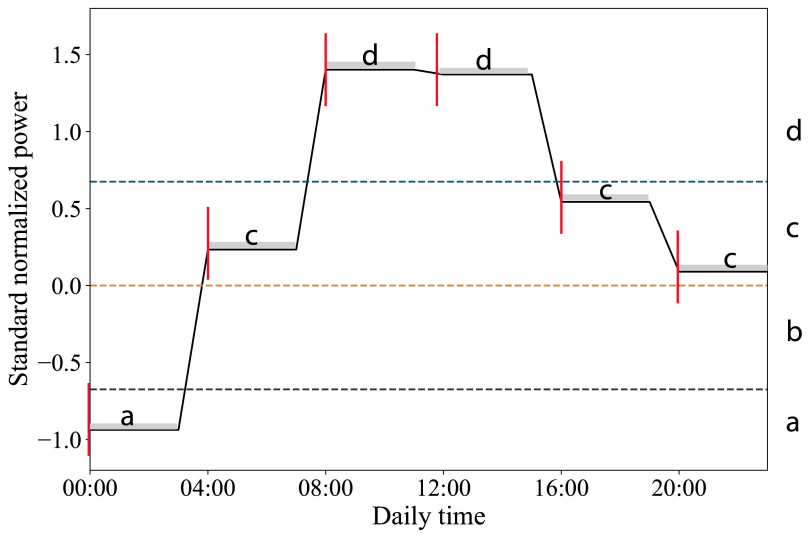

Figure 3.10: A transformed daily profile

segments (a equal to 4). Each value of the PAA vector is mapped to a symbol corresponding to the region in which it resides.

3.3.2 Creation and Labeling of SAX Pattern Repositories

Let's assume that a power demand pattern could be interpreted as a statement of the working condition of a building's power system in a day. Thus, a collection of historical patterns could represent working records of the power system for a building, in the same analogy. By creating pattern repositories, the performance of a building's power system could be tracked, predicted, and diagnosed. The process of diagnosing the power system performance using PPDAM is not covered in this paper and will be explored in future research.

There are two types of pattern repositories: parent and child pattern repositories. The structure of the repositories is illustrated in Figure 3.12. These three repositories that contain all of the true historical patterns are defined as parent repositories. Each parent repository has a certain number of child repositories. Each child repository contains all predicted patterns from one true pattern in a corresponding parent repository. Therefore, the number of child repositories in a specific parent repository is equal to the number of parent patterns in this parent repository. For instance,

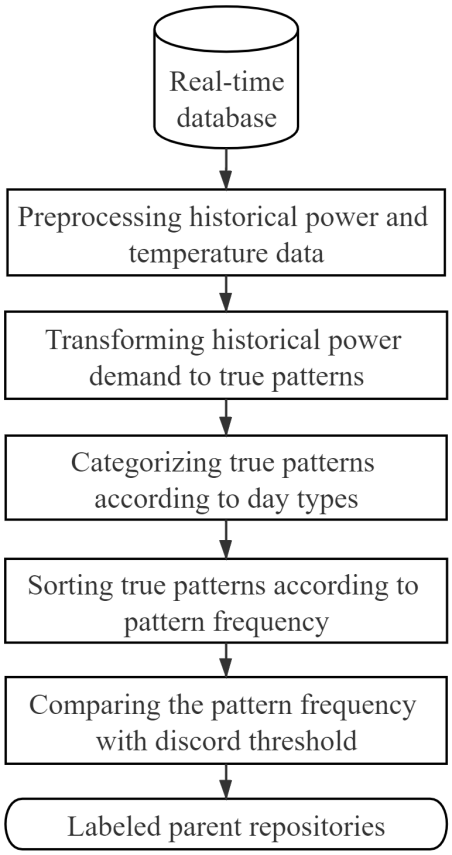

Figure 3.11: Creation and labeling of parent pattern repositories

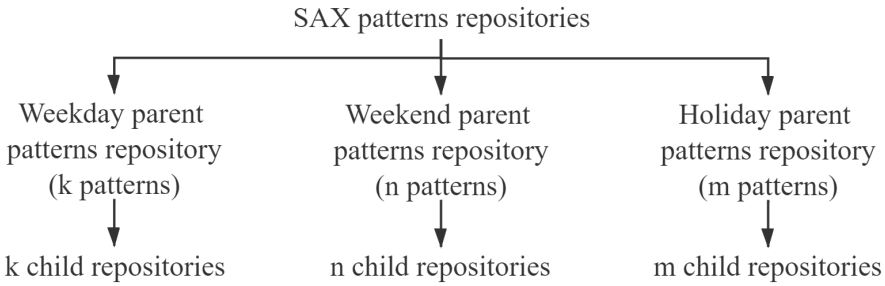

Figure 3.12: Hierarchical structure of patterns repositories

the weekday parent repository has k child repositories since there are k patterns in this weekday parent repository. Patterns in the child repositories, also called child patterns, are the result of the prediction using the LSTM model and the SAX representation.

The process of creating and labeling of parent pattern repositories is shown in Figure 3.11. A

certain period of historical power demand records of the selected building in the database are data source of the pattern repositories. Each daily power demand profile in the datasets is converted to a SAX pattern (hereafter referred to as parent pattern). All parent patterns are categorized into three groups according to three types of days (weekday, weekend, and holidays). By separating patterns according to their day type, the energy use of the building can be better differentiated and identified. As shown in Figure 3.13, each pattern in the parent repository (the table on the left) has a value (defined as parent frequency) corresponding to the frequency of occurrence of that pattern in the historical records. In each parent repository, the parent patterns are sorted based on their parent frequency. After creating parent pattern repositories, predicted patterns of all historical patterns are generated, in order to constitute child repositories. Predicted patterns generated are grouped and sorted if they are associated with the same parent pattern and the same day type.

Figure 3.13 provides an example: a weekday parent repository, the child repository of the pattern ‘aaaaaa’ in this weekday parent repository. All patterns in the table on the left are historical

Weekday parent repository

Patterns	Parent frequency	Labels
acddcc	166	Motif
acddcb	87	Motif
aaaaaa	80	Motif
...	...	...
bbbaaa	2	Discord
bbbaab	1	Discord

Weekday child repository of 'aaaaaa'

Patterns	Child frequency
aaaaaa	69
aaaaba	4
baaaaa	1
abbbbb	1
abdcbb	1
abccaa	1
aababa	1
aadcaa	1
bbbbbb	1

Figure 3.13: An example of weekday repositories

true patterns, and all patterns in the child repository of ‘aaaaaa,’ the table on the right, are predicted patterns. Each pattern in the parent repository has a value associated (defined as parent frequency),

corresponding to the frequency of occurrence in the historical records. In the child repository, each child pattern has a value associated (defined as child frequency), corresponding to the number of times the parent pattern was predicted (using the LSTM model) into that child pattern. Let's assume the parent pattern 'aaaaaa' appears 80 times in history, and it belongs to the weekday parent repository. In the child repository corresponding to 'aaaaaa,' the correctly predicted pattern appears 69 times, which means that 69 out of 80 times, the pattern 'aaaaaa' is correctly predicted. Thus, the accuracy of the LSTM model and the SAX model, in terms of each parent pattern, can be illustrated by all child repositories. This thesis focuses on anomaly detection with parent repositories and the solid functionalities of child repositories will be explored in the future study.

Motifs and discords are two key concepts to consider when applying anomaly detection to building power demand. A motif is a typical pattern that occurs regularly within a data set [50]. A discord is an unusual pattern within a data set that can be identified as infrequent behavior [51]. A threshold needs to be set in order to distinguish motifs and discords. Miller et al. in [33] defined a discord as a day-pattern with a parent frequency of less than 2% of the total count of days available. A parent pattern is labeled as a motif if its parent frequency is greater than the threshold, or it is labeled as a discord if its parent frequency is less than the threshold. For example, as shown in Figure 3.13, all parent patterns are labeled as motifs or discords.

3.3.3 Pattern Retrieval and Prediction

Figure 3.14 illustrates the classification test designed to test the efficacy of the anomaly prediction module. Classification is a process of categorizing a given set of data into classes. Thus, there are two classes in the classification test: motif and discord. Historical power and temperature data are extracted from the database to form the test set to predict the patterns of 500 days. For each day in the test set, a predicted profile is estimated. For both historical and predicted profiles, a SAX transformation is performed, considering the appropriate SAX parameters (a and w) for a desired granularity.

As mentioned in Section 3.3.2, motif and discord labels are assigned to patterns in each parent

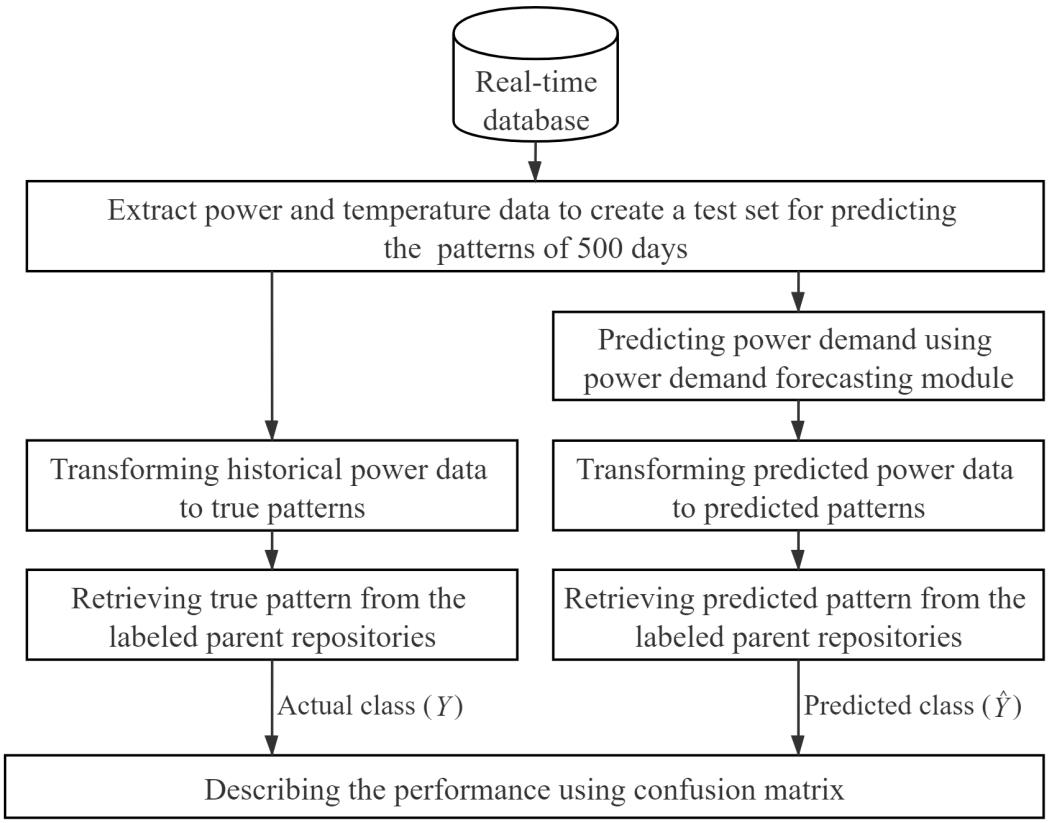

Figure 3.14: Design of the anomaly classification test

repository. A SAX pattern can be classified as a motif or a discord when retrieved from the labeled parent repositories. Therefore, the actual class (Y) for the classification test can be obtained by retrieving historical patterns in the test set from the labeled repositories (e.g. Figure 3.13). In parallel, the predicted class ($\hat{Y}$) of patterns in the test set can be obtained by retrieving predicted patterns from the labeled repositories. In this paper, the actual class (labels associated with historical patterns) is considered the ground truth.

The summary of the prediction results on a classification problem is recorded in a confusion matrix, as shown in Table 3.3. The number of correct and incorrect predictions are summarized by the count values and broken down by class. The confusion matrix gives an insight not only into the classification errors but also into the types of errors.

Explanations of terms shown in Table 3.3 are as follows. Positive (P) means that the predicted pattern is a discord. Negative (N) means that the predicted pattern is a motif. A true positive

Table 3.3: Confusion matrix

		Predicted class	
		Discord	Motif
Actual class	Discord	*TP*	*FN*
	Motif	*FP*	*TN*

(*TP*) means that the true pattern is a discord, and it has been predicted as a discord. A false negative (*FN*) means that the true pattern is a discord, but it has been predicted as a motif. A true negative (*TN*) means that the true pattern is a motif, and it has been predicted as a motif. A false positive (*FP*) means that the true pattern is a motif, but it has been predicted as a discord. Accuracy, Precision, Recall and F1 Score are used to evaluate the performance of motif and discord detection: $Accuracy = \frac{TP+TN}{TP+FN+FP+TN}$, $F1score = \frac{2 \times Recall \times Precision}{Recall+Precision}$, where $Precision = \frac{TP}{TP+FP}$, $Recall = \frac{TP}{TP+FN}$. Accuracy is the ratio of correctly predicted observations to the total number of observations. Precision is the ratio of correctly predicted positive observations to the total number of predicted positive observations. Recall is the ratio of correctly predicted positive observations to all observations with actual positive labels. F1 Score is the weighted average of Precision and Recall. Thus, it takes both FP and FN into account to strike a balance between precision and Recall [52].

Chapter 4: Experimental Results

Two sets of experiments are designed to implement and test the modules of power demand forecasting and anomaly prediction. Section 4.1 shows the information of seven buildings and all datasets involved in the experiments. The results of the two experiments are presented in Section 4.2 and Section 4.3, respectively.

4.1 Data Collection from Buildings

Seven buildings on the UBC campus are involved in the experiments. Table 4.1 shows the full names, acronyms, mean, standard deviation (STDs) and areas of seven buildings. Power data of year 2015 to 2019 are utilized to calculate the means and standard deviations of each building.

The models are built using the total electrical power consumption data measured by the main electric meter in each building. The power demand data (in the unit of kilowatt) are the summation of HVAC systems, plug loads (computers, kettles, microwaves, etc.), lighting, etc. The temperature data (in the unit of Celsius) are generated from the Rooftop Weather Station at UBC. Power and temperature data are collected from a SkySpark platform [53]. SkySpark is a platform utilized by Water and Energy Services at UBC to expose meter, BACNet signals, and weather data to promote research on building performance, all as part of the Campus as a Living Lab Initiative.

Datasets used in the experiments are introduced in this paragraph. For the experiments of

Table 4.1: Information of buildings

Building	Acronym	Area (m^2)	Mean (kW)	STD
University Services Building	USB	11598	138.52	59.28
Aquatic Ecosystems Research Laboratory	AERL	5368	54.15	14.54
C. K. Choi Building	CKC	2912	12.97	4.92
Earth Sciences Building	ESB	17755	526.57	95.14
Food, Nutrition and Health Building	FNH	5962	152.92	37.25
Forest Sciences Center	FSC	22459	549.91	111.96
Pharmaceutical Sciences Building	PSB	36138	1387.53	159.37

power demand forecasting, the dataset of the university service building is different from the ones used for the other six buildings. The time span of power data (in kW) for the university service building ranges from 12 am, October 11, 2014, to 11 pm, September 1, 2019. The raw dataset includes 42888 samples (1787 days) at a sampling time of one hour. The time span of power data (kW) for the other six buildings ranges from 12 am, January 1, 2015, to 11 pm, December 31, 2019. Thus, 43824 samples (1826 days) are collected in the raw dataset for these six building and the sampling time of the data is one hour. All datasets for power demand forecasting are split into 80% for training, and 20% for testing. In terms of anomaly prediction module, the power data ranging from 12 am, January 1, 2015, to 11 pm, December 31, 2019 are used to create pattern repositories for seven buildings. Thus, a total of 1826 days (87 holidays, 502 weekends, and 1237 weekdays) are taken into account to create the three parent repositories. The dataset used for motif and discord classification test is different from the one for power demand forecasting test. This is because all day-types and enough number of anomalous patterns need to cover in the classification test. However, patterns in the power demand forecasting test set do not provide enough true anomalous patterns. That is why a dataset for predicting the patterns of 500 days is selected as the classification test set (as described in Figure 3.14).

4.2 Experiments on Power Demand Forecasting

In order to select the optimal LSTM model for power demand forecasting in the application of PP-DAM, 34 LSTM models with various combinations of inputs and architectures were built based on the power data from the university service building. The deep learning library used are introduced in Section 4.2.1. The performance of these models is evaluated by using the MSE, MAE, and average SAX distance. The analysis of evaluation results and the optimal combination of inputs and model architecture are presented in Section 4.2.2.

4.2.1 Implementation of LSTM models

The LSTM models are implemented in Python, using the deep learning library, Keras. Keras is a high-level neural networks API, written in Python and capable of running as a frontend to TensorFlow, CNTK, or Theano [48]. The backend for the modeling is Tensorflow. One epoch is one cycle through which an entire dataset is passed once forward and backward through the neural network [54]. The weights are updated in the neural network a number of times proportional to the number of epochs, and through this process the model transforms from underfitting, to optimal, to overfitting. To ensure the model is optimal, Callback, ModelCheckpoint and EarlyStopping in Keras are configured to save the model with a minimum test loss in the training process, no matter how many epochs are set to train the model. Therefore, the 34 models analyzed in this thesis are neither underfitting nor overfitting. A detailed description of the implementation in Keras is beyond the scope of this thesis.

4.2.2 Results of Power Demand Forecasting

Table 4.2 shows MSE, MAE and SAX distance, resulting from the test of 34 LSTM configurations. In Table 4.2, there are 12 models using the standard architecture, 12 models using the single-encoder decoder architecture, and 10 models using the multi-encoder decoder architecture. Figure 4.1 plots bar charts of MSE, MAE and SAX distance, which clearly shows the trends of three metrics.

The input index represents the input configuration introduced in Section 3.2.2. The value of n refers to the last n-hour data samples. The SAX parameters a and w have been explained in Section 3.3.1. The value of a depends on how many value-categories one prefers to divide the power demand range into. The value of w depends on how many time windows within a day one wants to monitor. To meet the demands, a and w are assumed to be 4 and 6, respectively.

From Table 4.2, it is possible to conclude that the SAX distance follows a trend similar to the MSE and MAE. One can also conclude that the configurations of inputs and architecture that best minimize the losses and reduces the SAX distance are Configs 23 and 24. Config 23 is selected

Table 4.2: Configurations and prediction performance of LSTM models implemented on the dataset of the university service building

Configuration index	LSTM Architecture	Input Index	n	MSE	MAE	Distance $a = 4$ $w = 6$
1	Standard	1	24	0.4613	0.4969	0.7205
2			48	0.4394	0.4878	0.4569
3			72	0.4075	0.4564	0.4163
4		2	24	0.4457	0.4877	0.6787
5			48	0.4007	0.4548	0.3744
6			72	0.3936	0.4544	0.3368
7		3	24	0.3221	0.4144	0.2324
8			48	0.3453	0.4206	0.2853
9			72	0.3490	0.4296	0.2831
10		4	24	0.3139	0.4157	0.2355
11			48	0.3064	0.4051	0.1970
12			72	0.3205	0.4170	0.2505
13	Single-encoder decoder	1	24	0.4406	0.4790	0.7681
14			48	0.3870	0.4444	0.3185
15			72	0.3888	0.4435	0.4118
16		2	24	0.4262	0.4608	0.6074
17			48	0.3839	0.4539	0.3648
18			72	0.3793	0.4514	0.3331
19		3	24	0.3221	0.4144	0.2859
20			48	0.3242	0.3914	0.2718
21			72	0.3301	0.4078	0.2884
22		4	24	0.3006	0.3940	0.2742
23			48	**0.2762**	**0.3692**	0.1663
24			72	0.2813	0.3828	**0.1588**
25	Multi-encoder decoder	1	48	0.4295	0.4737	0.4068
26			72	0.4102	0.4476	0.4988
27		2	48	0.3996	0.4590	0.3931
28			72	0.3761	0.4377	0.4171
29		3	24	0.3233	0.3996	0.3013
30			48	0.3479	0.4171	0.2643
31			72	0.3278	0.4010	0.2900
32		4	24	0.3018	0.3936	0.3082
33			48	0.2851	0.3835	0.2798
34			72	0.2822	0.3950	0.1840

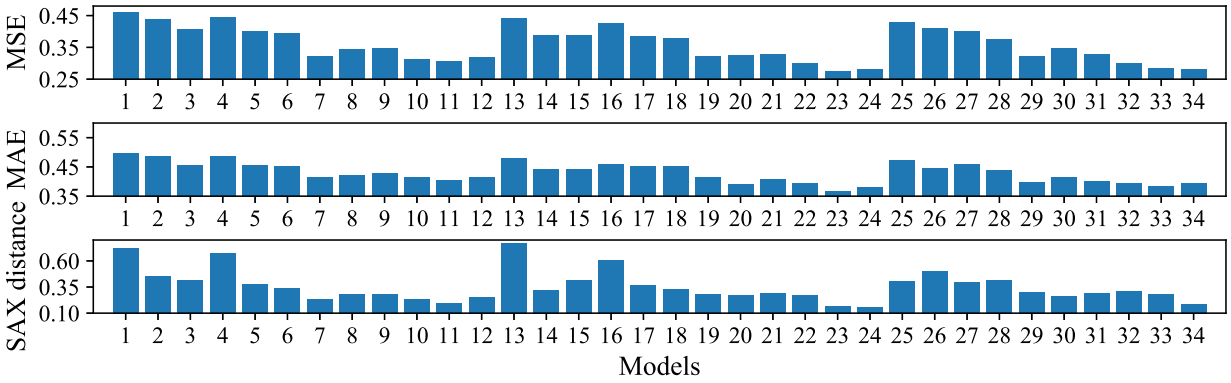

Figure 4.1: Bar plots of models' MSE, MAE and SAX distance

because it shows the lowest MSE and MAE, and the second-lowest SAX distance, behind Config 24. Additionally, compared to Config 24, Config 23 is selected because it has a lower complexity, given that it requires the data from the 48 hours before the predicting time, compared to 72 hours for Config 24, and the training time of Config 23 is 21.25% faster than Config 24.

By analyzing the results in Table 4.2, it is possible to understand how the performance of power demand forecasting is influenced by input variables and model architectures. Comparing all three LSTM architectures, one can conclude that single-encoder decoder models always produce better results (lowest MSE, MAE, distance) under the same set of inputs and value n. Comparing the standard and multi-encoder decoder architectures, the multi-encoder decoder models obtain smaller MSE and MAE, while the standard architecture performs better at minimizing the pattern distance.

Considering the value of n (comparing models with the same LSTM architecture), as the length of the last n-hour data increases (Models 1 to 6, 13 to 18, and 25 to 28), the loss (MSE) gradually decreases for models with Inputs 1 and 2 but not for Inputs 3 and 4. Therefore, it is possible to conclude that increasing the length of the historical data can improve the accuracy of the models when the previous week data is not included. For models with the same LSTM architecture and the same value of n, comparing Input 1 and Input 2, and comparing Input 3 and Input 4, one can conclude that adding temperature data improves the predicting performance. Additionally, comparing Input 1 and Input 3, and comparing Input 2 and Input 4, it can be determined how the

last week's realization improves predictability.

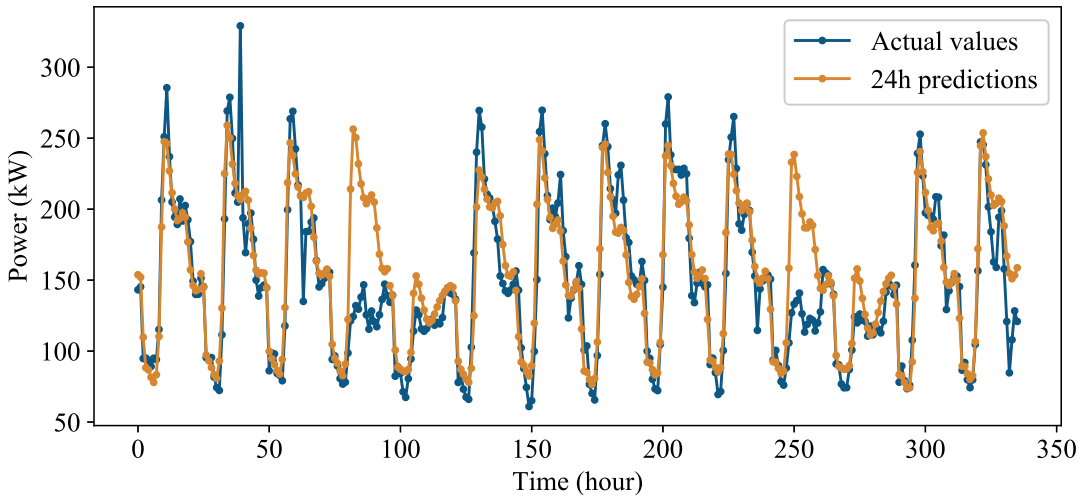

Figure 4.2: Predictions for first 336 hours in the test set of Config 17

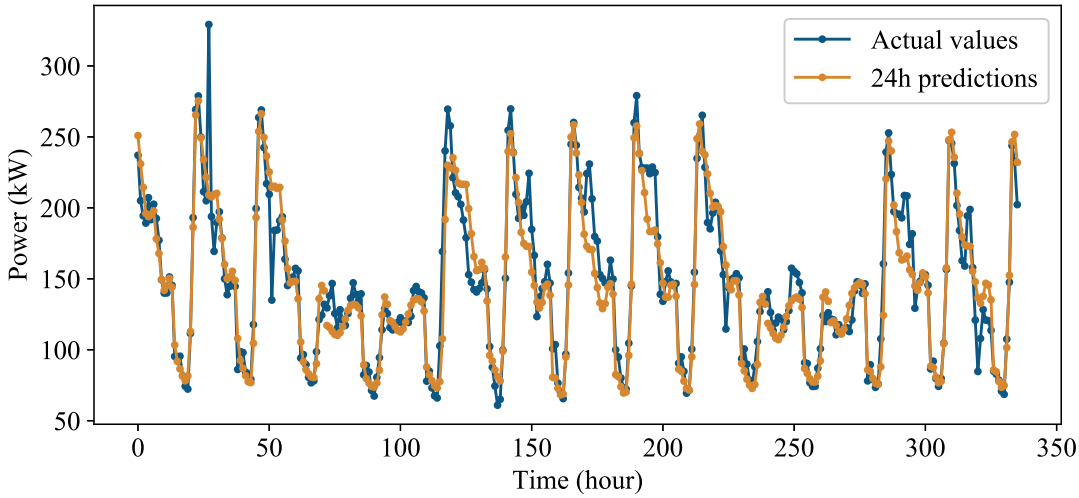

Figure 4.3: Predictions for first 336 hours in the test set of Config 23

The power demand predictions from Config 17 and Config 23, for the first 336 hours in the test set, are plotted in Figure 4.2 and Figure 4.3, respectively. Blue lines represent true power data, and orange lines indicate predicted power data. Comparing Figure 4.2 and Figure 4.3, one can conclude that Config 17 cannot predict the weekends' power demand, especially power demand on Saturday, as accurately as Config 23. This is because Config 23 uses the Input 3, which includes

Table 4.3: Prediction performance of the best models implemented on seven buildings

Configuration Index	Model Architecture	Building	MSE	MAE	Distance a = 4 w = 6
23	Single-encoder decoder	USB	0.2762	0.3692	0.1663
		AERL	0.1126	0.2345	0.0220
		CKC	0.1282	0.2446	0.0059
		ESB	0.1576	0.2972	0.0251
		FNH	0.0934	0.2145	0.0127
		FSC	0.1634	0.2645	0.0199
		PSB	0.1962	0.3288	0.0011

the last week's 24h realization. With this information, Config 23 is able to infer if the next day is a weekday or a weekend. Since the loss of models with Inputs 3 and 4 is always lower than 0.35, it can be concluded that these models perform better than models with Inputs 1 and 2 because of the last week's 24h realization data.

After select the optimal configuration of the LSTM models, i.e. Config 23, models in the optimal configuration were built for the other six building on UBC campus. The dataset for each building is split into 80% for training, and 20% for testing. As recommended by Water and Energy Services department at UBC, it might not be able to identify the changes too well on the university service building. Buildings which might work particularly well and have decent changes to point at could include PSB, FSC, and ESB. Buildings that have been pretty steady are AERL, CKC and FNH. The results (MSE, MAE, and SAX distance) of the seven building are shown in Table 4.3. The results shows that evaluation errors of the university service building is the highest.

4.3 Experiments on Anomaly Prediction

To implement the test for pattern retrieval and prediction (described in Section 3.3.3), the SAX representations of all the parent (historical) patterns corresponding to the data from seven buildings were estimated. Moreover, the SAX representations were classified and labeled as "motifs" and "discords." The creation and labeling of parent repositories is described in Section 4.3.1. The performance of pattern retrieval and prediction is evaluated in Section 4.3.2.

4.3.1 Labeled Pattern Repositories

As mentioned in Section 3.3.2, two levels of repositories are created: three parent pattern repositories (weekday, weekend, holiday), and the child repository for each parent pattern (all possible predicted patterns, for each parent pattern in each parent repository). The dataset ranging from 12 am, January 1, 2015, to 11 pm, December 31, 2019 was used to create pattern repositories for seven buildings. Thus, a total of 1826 days (87 holidays, 502 weekends, and 1237 weekdays) are taken into account to create the three parent repositories.

Table 4.4: The number of patterns in each parent repository

Building	Weekday	Weekend	Holiday
USB	114	92	**46**
AERL	76	55	35
CKC	72	59	34
ESB	**220**	**156**	39
FNH	90	80	35
FSC	96	93	30
PSB	113	90	42

By converting historical power consumption to SAX patterns, one can get the number of patterns in each parent repositories created for seven buildings, which is shown in Table 4.4. Higher the number of patterns is, more diverse the building behaviors are. For instance, Figure 4.4 shows the frequency distribution for the holiday repository (Figure 4.4.a), the weekend repository (Figure 4.4.b), and the weekday repository (Figure 4.4.c) of the university service building. 252 patterns were identified in total. The repositories contain 46, 92, and 114 patterns, respectively. The frequency distribution of a pattern differs among the three parent pattern repositories. It is important to observe that the frequency distribution of a pattern can be different among the three parent patterns repositories. For example, for pattern ‘aaaaaa,’ the frequency is 14.94% (13 out of 87) in holidays; 11.75% (59 out of 502) in weekends; and 0.57% (7 out of 1237) in weekdays. This shows that the energy consumption of a building is different depending on the type of day.

In terms of the university service building, the composition of the child repository of a pattern, e.g., ‘aaaaaa,’ is also different in each of the three parent repositories, as shown in Tables 4.5,

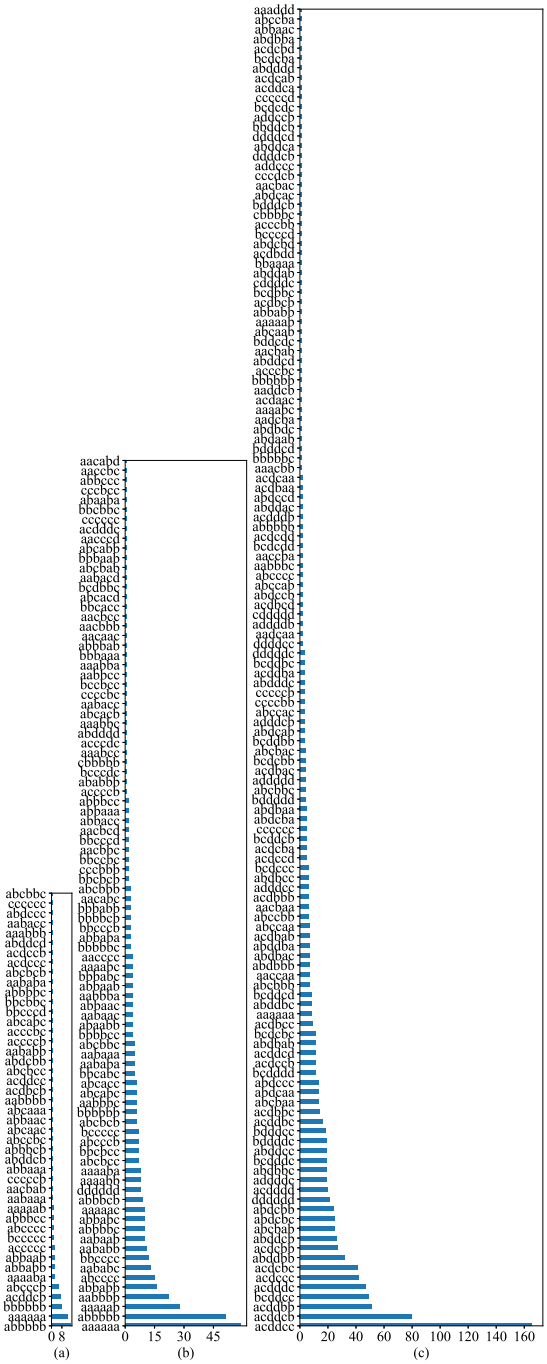

Figure 4.4: Frequency distribution of (a) holiday parent repository, (b) weekend parent repository, (c) weekday parent repository of the university service building

4.7, and 4.6. Child repositories reflect the forecast accuracy of selected Config 23 regarding each corresponding parent pattern. The forecast accuracy is 76.9% for holidays; 75% for weekdays; and 91.5% for weekends, regarding the parent pattern "aaaaaa" of the university service building.

Table 4.5: Child repository of 'aaaaaa' in holiday parent repository

Predicted pattern	Child frequency
aaaaaa	10
abcbba	1
aacbaa	1
aababa	1

Table 4.6: Child repository of 'aaaaaa' in weekend parent repository

Predicted pattern	Child frequency
aaaaaa	54
aaaaba	3
aababa	2

Table 4.7: Child repository of 'aaaaaa' in weekday parent repository

Predicted pattern	Child frequency
aaaaaa	6
aabbaa	1
bbaaaa	1

As mentioned in Section 3.3.3, Miller et al. in [33] suggest that a day-pattern with a frequency of less than 2% of the total count of days available can be considered as a discord. Table 4.8 shows the number of daily profiles in each parent repository and the threshold determined for each of the parent repositories. For instance, a SAX pattern of the university service building appeared 15 times on weekdays is labeled as discord, because its frequency, 15, is less than weekdays threshold, 25. This distinction results in discord tags for 31 days in holiday repository, 125 days in weekend repository, 304 days in weekday repository, which is 31.6%, 28.5% and 26.6% of the total count of days in the three repositories of the university service building, respectively. The number and percentage of anomalous days of all seven buildings are shown in Table 4.9. The percentages of anomalous weekdays and weekends in ESB are more than a half (55.35% and 53.98%), which are the highest among the seven buildings. It implies the highest diversity of building behaviors.

Table 4.8: Thresholds of three parent repositories

Repository	Total days	Threshold
Holidays	87	2
Weekends	502	10
Weekdays	1237	25

Table 4.9: Number and percentage of anomalous days for seven buildings

Building	Anomalous weekday		Anomalous weekend		Anomalous holiday	
	#Days	Percentage	#Days	Percentage	#Days	Percentage
USB	304	24.58%	125	24.90%	31	35.63%
AERL	293	23.69%	99	19.72%	30	34.48%
CKC	275	22.23%	128	25.50%	30	34.48%
ESB	660	**53.35%**	271	**53.98%**	37	**42.53%**
FNH	438	35.41%	137	27.29%	37	**42.53%**
FSC	450	34.86%	175	36.38%	26	29.89%
PSB	410	33.14%	179	35.65%	36	41.38%

4.3.2 Results of Anomaly Prediction

A motif and discord classification test is designed in order to validate the anomaly prediction performance of the selected model configuration (Config 23). The dataset used for motif and discord classification test is different from the one for power demand forecasting. This is because all day-types and enough number of anomalous patterns need to cover in the classification test. However, patterns in the power demand forecasting test set do not provide enough true anomalous patterns. That is why a dataset for predicting the patterns of 500 days was selected as the classification test set (as described in Figure 3.14).

Power demand data of these 500 days are extracted to generate true patterns that are taken as ground-truth labels. In this test set, true patterns of half of these 500 days are motifs, and the other half are discords. The model in Config 23 for each building generates the predicted pattern of each true pattern, and these 500 predicted patterns are classified as motifs and discords.

Results of the seven tests are summarized in Table 4.10. For instance, in terms of the university service building, there are 208 TP, 42 FN, 18 FP, and 232 TN in the results. As introduced in Section 3.3.3, accuracy, precision, recall and F1 score are used to evaluate the performance of

Table 4.10: Confusion matrix of seven buildings

			Predicted class	
			Discord	Motif
Actual class	USB	Discord	208	42
		Motif	18	232
	AERL	Discord	223	27
		Motif	19	231
	CKC	Discord	229	21
		Motif	13	237
	ESB	Discord	215	35
		Motif	23	227
	FNH	Discord	222	28
		Motif	12	238
	FSC	Discord	234	26
		Motif	14	226
	PSB	Discord	237	23
		Motif	16	224

Table 4.11: Motif and discord classification results

Building	Accuracy	Precision	Recall	F1 score
USB	0.8800	0.9200	0.8320	0.8738
AERL	0.9080	0.9215	0.8920	0.9065
CKC	**0.9320**	0.9463	**0.9160**	**0.9309**
ESB	0.8840	0.9034	0.8600	0.8812
FNH	0.9200	**0.9487**	0.8880	0.9173
FSC	0.9200	0.9435	0.9000	0.9212
PSB	0.9160	0.9368	0.9115	0.9240

motif and discord detection. Accuracy, Precision, Recall and F1 score of models implemented on seven buildings are shown in Table 4.11. The results show that the anomaly prediction modules work well on seven buildings, no matter pattern composition and diversity of building repositories.

Chapter 5: Conclusions

5.1 Summary of Contributions

The rapidly growing world energy use has already raised concerns over supply difficulties, exhaustion of energy resources and heavy environmental impacts (ozone layer depletion, global warming, climate change, etc.). Those serious issues stress the need for seeking effective energy-efficiency methods. Improving energy efficiency in buildings is today a prime objective for energy policy at regional, national, and international levels, to reduce the amount of gas emission as well as fossil fuel consumption. Among various data-driven predictive analytics methods for improving building energy efficiency, predicting and monitoring the power consumption with the aim of identifying abnormal behaviors is promising and cost-effective. However, the information contained in time series predictions usually needs to be interpreted by a human with domain expertise. For instance, meaningful statistics, such as near-base load, near-peak load, high-load duration, rise time and fall time, could be defined to describe the load profiles. These statistics are useful for building energy management but need to be further processed and interpreted by expertise to apply. Moreover, an algorithmic way of labeling anomalies needs to be explored because for practical applications, anomaly detection without anomaly labels is closer to the actual requirements of building operation managers since it is expensive and time-consuming to generate labeled data for abnormal power consumption in buildings.

Therefore, this research intends to fill up the gap by proposing a novel predictive power demand analytic methodology that is capable to identify upcoming normal and anomalous demand patterns, through mining historical power demand and temperature data, and day-ahead power demand forecast. The power demand data explored in the study are the total electrical power consumption in the building. That is the summation of heating, ventilating and air-conditioning system load, plug load, lighting, etc. It is measured by the main electric meter installed on the main electrical feeder

into the building on the campus of the University of British Columbia. The temperature data are generated by the Earth and Ocean Science Rooftop Weather Station at the University of British Columbia.

This data-driven methodology comprises two modules, i.e., the power demand forecasting module and the anomaly prediction module. The integration of two modules can derive latent information from historical and predicted power data, and provide an intuitive and effective way to identify typical (motif) and atypical (discord) behaviors by using Long Short-Term Memory neural networks and Symbolic Aggregate Approximation. This is the first time that LSTM neural networks and SAX are integrated for accurate insights regarding the power needs of the building, intending to support more efficient energy management.

The power demand forecasting module can predict time-series power demand for the next 24 hours, which is a baseline for pattern generation and anomaly prediction. In order to figure out how the performance of models is influenced by input variables and model architectures, this study explores how the forecasting performance is influenced by power and temperature inputs, and three LSTM model architectures. Two of the architectures (the standard and the single-encoder decoder architectures) have been utilized in the literature, while the multi-encoder decoder architecture is suggested for the first time. Based on the experimental results presented in the thesis, it can be concluded that single-encoder decoder models present a better performance than standard and multi-encoder decoder architectures. The optimal forecasting module is characterized by power and temperature inputs, including the last week's 24h realization and last 48h data.

In the module of anomaly prediction, the historical and predicted time-series profiles are first transformed into patterns and later are retrieved from a system of labeled repositories. The patterns are classified as normal or anomalous based on their frequency of appearance in the pattern repositories. The limitation that the information contained in time series predictions usually needs to be interpreted by a human with domain expertise is addressed since the anomaly prediction module can automatically identify the pattern from the time-series forecasts. Moreover, automated interpretation of upcoming abnormal behaviors in the anomaly prediction module avoids ground-truth

labeling, which provides an unsupervised method when labels are not available from power meter data. The performance of the anomaly prediction module is evaluated by classification tests implemented on seven buildings on the campus of the University of British Columbia. The method obtains a minimum accuracy of 88%, a maximum accuracy of 93.20%, a minimum F1 score of 87.38%, and a maximum F1 score of 93.09%.

5.2 Limitations and Future Research

This thesis shows that data-driven predictive approaches provide a promising way for power demand monitoring, which is cost-effective, accurate, robust, and scalable. However, there are also some limitations to the research. The challenges and the potential future studies for the research in this thesis are discussed as follows.

Since the approaches used in this study highly rely on the data, high-quality data are required to well implement these methods. Hence, future studies could focus on the design of more efficient and economic data collection strategies. Besides, the data used in the modeling was collected from the University of British Columbia, which may not be available in some situations and applications. Transfer learning may thus need to be further researched so that the models developed for buildings on the UBC campus could be reused as the starting point for a model on other tasks. Moreover, the condition of the building may also change due to renovations and the data collected from the building can get affected. Therefore, the sensitivity of the prediction models needs to be explored with transfer learning when the regular patterns of source data change.

The concept of child pattern repositories is introduced in the thesis, but the concrete functionalities of it are not included. The pattern distribution in the child pattern repositories indicates the historical forecast performance of the LSTM models upon each corresponding parent pattern. Thus, In building research, further work could also include monitoring strategies for building operators to obtain insights on potential faults, by analyzing the pattern distribution in child pattern repositories. Labeling SAX patterns as normal or anomalous behaviors by comparing pattern frequency with a threshold may limit the classification accuracy. In order to improve the accuracy of

anomaly classification, more classifying algorithms applied upon patterns need to explore. Given that SAX words are associated with predicted patterns, the anomaly detection strategy could be augmented by utilizing methods related to natural language processing. In order to involve the forecast results in a real-world control system, another topic to explore could be how a predicted power demand pattern could be used as a set-point to control a building management system.

To conclude, this research provides accurate insights regarding the power needs of the building, intending to support more efficient energy management. Building operators could take actions on completing building maintenance tasks according to the insights obtained. In order to improve the accuracy and feasibility of the approach, future studies could focus on the design of more efficient and economic data collection strategies, transfer learning of trained models, analyzing the pattern distribution in child pattern repositories and more efficient classifying algorithms.

Bibliography

[1] International Energy Agency, "Electricity." https://www.iea.org/fuels-and-technologies/electricity, 2020. accessed January, 2020. → page 1

[2] L. Pérez-Lombard, J. Ortiz, and C. Pout, "A review on buildings energy consumption information," *Energy and Buildings*, vol. 40, no. 3, pp. 394–398, 2008. → page 1

[3] S. Seyedzadeh, F. P. Rahimian, I. Glesk, and M. Roper, "Machine learning for estimation of building energy consumption and performance: a review," *Visualization in Engineering*, vol. 6, no. 1, p. 1, 2018. → page 1

[4] The Government of British Columbia, "Energy efficiency." https://www2.gov.bc.ca/gov/content/industry/electricity-alternative-energy/energy-efficiency-conservation, 2020. accessed January, 2020. → page 1

[5] S. Finlay, *Predictive Analytics, Data Mining and Big Data: Myths, Misconceptions and Methods*. Business in the Digital Economy, Basingstoke: Palgrave Macmillan, 1st ed., 2014. → page 2

[6] A. Crespo Márquez, A. Crespo Del Castillo, and J. F. Gómez Fernández, "Integrating artificial intelligent techniques and continuous time simulation modelling. Practical predictive analytics for energy efficiency and failure detection," *Computers in Industry*, vol. 115, p. 103164, 2020. → page 2

[7] J. Zhu, Y. Shen, Z. Song, D. Zhou, Z. Zhang, and A. Kusiak, "Data-driven building load profiling and energy management," *Sustainable Cities and Society*, vol. 49, p. 101587, 2019. → pages 2, 23

[8] H. Zhao and F. Magoulès, "A review on the prediction of building energy consumption," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 6, pp. 3586–3592, 2012. → pages 2, 7

[9] J. L. Mathieu, P. N. Price, S. Kiliccote, and M. A. Piette, "Quantifying changes in building electricity use, with application to demand response," *IEEE Transactions on Smart Grid*, vol. 2, no. 3, pp. 507–518, 2011. → page 3

[10] C. Xu and H. Chen, "A hybrid data mining approach for anomaly detection and evaluation in residential buildings energy data," *Energy and Buildings*, vol. 215, p. 109864, 2020. → pages 3, 11

[11] M. Kim, W. Choi, Y. Jeon, and L. Liu, "A hybrid neural network model for power demand forecasting," *Energies*, vol. 12, no. 5, p. 931, 2019. → page 4

[12] E. Sala-Cardoso, M. Delgado-Prieto, K. Kampouropoulos, and L. Romeral, "Activity-aware HVAC power demand forecasting," *Energy and Buildings*, vol. 170, pp. 15–24, 2018. → page 4

[13] L. Hernandez, C. Baladron, J. M. Aguiar, B. Carro, A. J. Sanchez-Esguevillas, J. Lloret, and J. Massana, "A survey on electric power demand forecasting: Future trends in smart grids, microgrids and smart buildings," *IEEE Communications Surveys and Tutorials*, vol. 16, no. 3, pp. 1460–1495, 2014. → page 7

[14] C. R. Touretzky and R. Patil, "Building-level power demand forecasting framework using building specific inputs: Development and applications," *Applied Energy*, vol. 147, pp. 466–477, 2015. → page 7

[15] M. S. Al-Homoud, "Computer-aided building energy analysis techniques," *Building and Environment*, vol. 36, no. 4, pp. 421–433, 2001. → page 8

[16] R. Yao and K. Steemers, "A method of formulating energy load profile for domestic buildings in the UK," *Energy and Buildings*, vol. 37, no. 6, pp. 663–671, 2005. → page 8

[17] F. Ansari, A. Mokhtar, K. Abbas, and N. Adam, "A Simple Approach for Building Cooling Load Estimation," *American Journal of Environmental Sciences*, vol. 1, no. 3, pp. 209–212, 2005. → page 8

[18] A. Dhar, T. A. Reddy, and D. E. Claridge, "Modeling hourly energy use in commercial buildings with fourier series functional forms," *Journal of Solar Energy Engineering, Transactions of the ASME*, vol. 120, no. 3, pp. 217–223, 1998. → page 8

[19] M. J. Jiménez and M. R. Heras, "Application of multi-output ARX models for estimation of the U and g values of building components in outdoor testing," *Solar Energy*, vol. 79, no. 3, pp. 302–310, 2005. → page 9

[20] A. J. Hoffman, "Peak demand control in commercial buildings with target peak adjustment based on load forecasting," in *Proceedings of the 1998 IEEE International Conference on Control Applications*, vol. 2, (Trieste, Italy), pp. 1292–1296, 1998. → page 9

[21] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and Machine Learning forecasting methods: Concerns and ways forward," *PLoS ONE*, vol. 13, no. 3, pp. 1–26, 2018. → pages 9, 14

[22] X. Zhou, Z. Fan, L. Liang, J. Yan, and D. Pan, "Comparison of four algorithms based on machine learning for cooling load forecasting of large-scale shopping mall," *Energy Procedia*, vol. 142, pp. 1799–1804, 2017. → page 9

[23] M. Cai, M. Pipattanasomporn, and S. Rahman, "Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques," *Applied Energy*, vol. 236, pp. 1078–1088, 2019. → page 10

[24] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015. → page 10

[25] A. Graves, A. R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proceedings of 2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, (Vancouver, BC, Canada), pp. 6645–6649, 2013. → page 10

[26] J. Zheng, C. Xu, Z. Zhang, and X. Li, "Electric load forecasting in smart grids using Long-Short-Term-Memory based Recurrent Neural Network," in *Proceedings of 2017 51st Annual Conference on Information Sciences and Systems (CISS)*, (Baltimore, MD, USA), pp. 1–6, IEEE, 2017. → page 10

[27] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 841–851, 2019. → page 10

[28] H. Wang, P. Xu, X. Lu, and D. Yuan, "Methodology of comprehensive building energy performance diagnosis for large commercial buildings at multiple levels," *Applied Energy*, vol. 169, pp. 14–27, 2016. → page 10

[29] Q. Zhang, S. Wan, B. Wang, D. W. Gao, and H. Ma, "Anomaly detection based on random matrix theory for industrial power systems," *Journal of Systems Architecture*, vol. 95, pp. 67–74, 2019. → page 11

[30] J. S. Chou and A. S. Telaga, "Real-time detection of anomalous power consumption," *Renewable and Sustainable Energy Reviews*, vol. 33, pp. 400–411, 2014. → page 11

[31] J. E. Seem, "Using intelligent data analysis to detect abnormal energy consumption in buildings," *Energy and Buildings*, vol. 39, no. 1, pp. 52–58, 2007. → page 12

[32] J. A. Fonseca, C. Miller, and A. Schlueter, "Unsupervised load shape clustering for urban building performance assessment," *Energy Procedia*, vol. 122, pp. 229–234, 2017. → page 12

[33] C. Miller, Z. Nagy, and A. Schlueter, "Automated daily pattern filtering of measured building performance data," *Automation in Construction*, vol. 49, pp. 1–17, 2015. → pages 12, 25, 30, 42

[34] Y. Zhao, T. Li, X. Zhang, and C. Zhang, "Artificial intelligence-based fault detection and diagnosis methods for building energy systems: Advantages, challenges and the future," *Renewable and Sustainable Energy Reviews*, vol. 109, pp. 85–101, 2019. → page 13

[35] Z. Yu, F. Haghighat, B. C. Fung, and L. Zhou, "A novel methodology for knowledge discovery through mining associations between building operational data," *Energy and Buildings*, vol. 47, pp. 430–440, 2012. → page 13

[36] C. Miller, Z. Nagy, and A. Schlueter, "A review of unsupervised statistical learning and visual analytics techniques applied to performance analysis of non-residential buildings," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 1365–1377, 2018. → page 13

[37] M. Goldstein and S. Uchida, "A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data," *PLoS ONE*, vol. 11, no. 4, pp. 1–31, 2016. → page 13

[38] M. Goldstein, S. Asanger, M. Reif, and A. Hutchison, "Enhancing security event management systems with unsupervised anomaly detection," in *Proceedings of the 2nd International Conference on Pattern Recognition Applications and Methods*, (Barcelona, Spain), pp. 530–538, 2013. → page 13

[39] A. Capozzoli, F. Lauro, and I. Khan, "Fault detection analysis using data mining techniques for a cluster of smart office buildings," *Expert Systems with Applications*, vol. 42, no. 9, pp. 4324–4338, 2015. → page 13

[40] N. Cotrufo and R. Zmeureanu, "PCA-based method of soft fault detection and identification for the ongoing commissioning of chillers," *Energy and Buildings*, vol. 130, pp. 443–452, 2016. → page 13

[41] Y. Zhao, S. Wang, and F. Xiao, "A system-level incipient fault-detection method for HVAC systems," *HVAC and R Research*, vol. 19, no. 5, pp. 593–601, 2013. → page 14

[42] J. A. Diaz-Acevedo, L. F. Grisales-Noreña, and A. Escobar, "A method for estimating electricity consumption patterns of buildings to implement energy management systems," *Journal of Building Engineering*, vol. 25, p. 100774, 2019. → page 14

[43] D. L. Marino, K. Amarasinghe, and M. Manic, "Building Energy Load Forecasting using Deep Neural Networks," in *Proceedings of 42nd Annual Conference of the IEEE Industrial Electronics Society*, (Florence, Italy), pp. 7046–7051, 2016. → page 21

[44] UCI Machine Learning Repository, "Individual household electric power consumption." http://archive.ics.uci.edu/ml/index.php, 2020. accessed January, 2020. → page 21

[45] E. Mocanu, P. H. Nguyen, M. Gibescu, and W. L. Kling, "Deep learning for estimating building energy consumption," *Sustainable Energy, Grids and Networks*, vol. 6, pp. 91–99, June 2016. → page 21

[46] M. Q. Raza and A. Khosravi, "A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings," *Renewable and Sustainable Energy Reviews*, vol. 50, pp. 1352–1372, 2015. → page 23

[47] K. Wang, X. Qi, and H. Liu, "A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network," *Applied Energy*, vol. 251, no. 1, p. 113315, 2019. → page 23

[48] Keras, "Keras: The python deep learning library." https://keras.io/, 2020. accessed January, 2020. → pages 24, 35

[49] J. Lin, E. Keogh, L. Wei, and S. Lonardi, "Experiencing SAX: a novel symbolic representation of time series," *Data Mining and Knowledge Discovery*, vol. 15, no. 2, pp. 107–144, 2007. → pages 24, 26

[50] P. Patel, E. Keogh, J. Lin, and S. Lonardi, "Mining motifs in massive time series databases," in *Proceedings of 2002 IEEE International Conference on Data Mining*, (Maebashi City, Japan), pp. 370–377, 2002. → page 30

[51] E. Keogh, J. Lin, and A. Fu, "HOT SAX: Efficiently finding the most unusual time series subsequence," in *Proceedings of the Fifth IEEE International Conference on Data Mining*, (Houston, TX, USA), pp. 226–233, 2005. → page 30

[52] R. Joshi, "Accuracy, precision, recall f1 score: Interpretation of performance measures." https://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures/, 2016. accessed January, 2020. → page 32

[53] Energy and Water Services UBC, "Building energy and water data." https://energy.ubc.ca/energy-and-water-data/skyspark/, 2020. accessed January, 2020. → page 33

[54] B. Hitaj, G. Ateniese, and F. Perez-Cruz, "Deep Models under the GAN: Information leakage from collaborative deep learning," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, CCS '17, (New York, NY, USA), pp. 603–618, Association for Computing Machinery, 2017. → page 35