buildings

MDPI

Article

A Workflow Investigating the Information behind the Time-Series Energy Consumption Condition via Data Mining

Xiaodong Liu, Shuming Zhang, Weiwen Cui, Hong Zhang *, Rui Wu, Jie Huang, Zhixin Li, Xiaohan Wang, Jianing Wu and Junqi Yang

School of Architecture, Tsinghua University, Beijing 100084, China; liuyan606431@tsinghua.edu.cn (X.L.); zhangsm20@mails.tsinghua.edu.cn (S.Z.); cuiww19@mails.tsinghua.edu.cn (W.C.); wu-r22@mails.tsinghua.edu.cn (R.W.); jie-huan18@mails.tsinghua.edu.cn (J.H.); lizhixin22@mails.tsinghua.edu.cn (Z.L.); xh-wang22@mails.tsinghua.edu.cn (X.W.); wujn21@mails.tsinghua.edu.cn (J.W.); yang-jq23@mails.tsinghua.edu.cn (J.Y.)
* Correspondence: zhanghong@tsinghua.edu.cn

Abstract: The purpose of this study is to develop a framework to understand building energy usage pattern finding using data mining algorithms. Developing advanced techniques and requirements for carbon emission reduction provides higher demands for building energy efficiency. Research conducted so far has mainly focused on total energy consumption data clusters instead of time-series curve peculiarity. This research adopts the time-series cluster algorithm k-shape and the ARM Apriori method to study the simulation database generated by the official restaurant energy model. These advanced data mining techniques can discover potential information hidden in a big database that has not been identified by people. The results show that the restaurant time-series energy consumption curve can be clustered into four type patterns: Invert U, M, Invert V, and Multiple M. Each mode has its own variation characteristics. Two aspects for the solution of intensity and peak shift are proposed, achieving energy savings and focusing on different curve modes. The conclusion shows that the combination of time-series clustering and the ARM algorithm work flow can successfully discover the building operation pattern. Some solutions focusing on restaurant energy usage issues have been proposed, and future investigations should pay more attention to building area-influenced factors.

Keywords: cluster analysis; association rule mined analysis; energy consumption; time-series; data mining

check for updates

Citation: Liu, X.; Zhang, S.; Cui, W.; Zhang, H.; Wu, R.; Huang, J.; Li, Z.; Wang, X.; Wu, J.; Yang, J. A Workflow Investigating the Information behind the Time-Series Energy Consumption Condition via Data Mining. *Buildings* **2023**, *13*, 2303. https://doi.org/10.3390/buildings13092303

Academic Editor: Francesco Nocera

Received: 25 July 2023
Revised: 16 August 2023
Accepted: 23 August 2023
Published: 10 September 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In recent years, to meet energy efficiency requirements, the building sector has contributed to energy-saving technologies and policies [1]. Although energy consumption for the building segment is the same as in 2019, carbon emissions are highest (10 $GtCO^2$) as construction activities and electricity intensity increase [2,3]. In line with the Global Status Report for the building sector, from 2020 to 2030, a reduction in carbon emissions of 6% per year is demanded with respect to the building segment [2]. Over the past decade, building benchmarking has been successfully addressed by city governments in 40+ cities in the United States to combat inefficient energy use in buildings [4]. To reach this target, diminishing highly intensive electricity use in building operations is significant.

With the development of data mining technologies, a lot of building energy efficiency research has been conducted due to the large amount of existing power data. In this studied domain, the unsupervised data mining approach presents great advantages in terms of data mining in building operation data. In this case, regular discoveries in operation data can provide an innovative energy-saving strategy for buildings. With respect to unsupervised algorithms, the cluster and association rule mining (ARM) solutions are the

most widely utilized in this domain. This paper aims to make full use of the data mining method to study building performance data. Some hidden information and issues are found by achieving the energy efficiency purpose. In line with these discovered features, some focused power-saving suggestions can be provided to run the building.

The cluster algorithm mainly aims to recognize the data variation law based on the relationship among datasets. In architectural research, the cluster method is primarily harnessed to identify building operation data, such as power load. Chicco [5] compared different cluster performances in terms of recognizing building electricity data profiles. The results showed that the k-means indicate a better recognition capability than other algorithms. Klemp et al. adopted a k-means clustered manner, predicting the building material U-value ranges successfully [6]. In addition, k-means has showed an excellent performance in terms of heating load classification [7]. Miller et al. improved the k-means cluster by identifying the daily building power variation pattern [8]. There are no limitations in terms of architectural types; the k-means algorithm has also been employed on hotel buildings [9,10]. Gao utilized the k-means approach, grouping twelve buildings in line with a performance characteristic [11]. Jaeger et al. constructed a building cluster approach by hierarchically replacing the urban energy simulation flow [12]. Andrews and Jain built a mixed dataset integrating the attributes of grid-interactive and efficient buildings [13]. K-medoids using Gower's Distance have been used to cluster these constructions, analyzing demand flexibility benchmarking. Walsh et al. put forward a performance-driven method for climate zoning that sought to address these limitations by leveraging archetypes, building performance simulation, and GIS to a great extent [14]. Coupled with the k-means cluster method, the final results proved the validity of this research.

Apart from construction energy consumption, it could also be used to recognize people's behavior. Lavin and Klabjan distinguished 1000 commercial consumer daily electricity consumption patterns harnessing the k-means solution [15]. For the indoor environment, three performance indicators, namely temperature, humidity, and light, were regarded as the data showing the indoor circumstance and were analyzed using the k-means cluster [16]. In general, the k-means algorithm needs to assign the k number in advance to achieve the algorithm function. To simplify this process, some improved algorithms have been developed, eliminating this step. For example, Kwac et al. put forward an adaptable k-means algorithm to determine residential buildings' daily energy consumption patterns [17]. In addition, focusing on non-numerical data, a fuzzy c-means clustered algorithm has also been employed to identify building patterns. With the purpose of measuring the distance computational method, which is the core of the clustered algorithm, Iglesias and Kastner compared four similarity evaluation approaches of Euclidean distance, Mahalanobis distance, Pearson correlation-based distance, and Dynamic Time Warping distance [18]. The results showed that Euclidean distance made the highest achievement in terms of daily building electricity mode identification. Qin and Zhang used the same solution on office building energy consumption data [19]. Santamouris et al. grouped the load data from 320 school buildings into five types according to the fuzzy c-means calculation principle [20]. Additionally, there were some clustered solutions without using distance calculation, such as support vector machine and decision tree. Chicco and Ilie found that the support vector machine clustering method had showed a good performance when the cluster number was at a low level [21]. Petcharat, Chungpaibulpatana, and Rakkwamsuk used a new expectation maximization clustering algorithm to investigate commercial building lighting energy consumption modes [22]. Liu et al. proposed an innovative decision tree clustering algorithm and achieved load pattern recognition for a variable refrigerant flow system [23]. Culaba et al. clustered time-series energy consumption using the k-means method to discover householders' behavior differences to prepare for power intensity forecasting with a support vector machine solution [24]. Xu et al. constructed a probability distribution model integrated k-means clustering algorithm by researching residential building electricity curves and people's behavior [25]. Nisa, Kuan,

and Lai employed the Apriori method for water-cooled chiller data to identify related parameters for constructing a prediction model [26].

Association rule mining (ARM) is an increasingly popular unsupervised algorithm in the big data research area. It can discover associated relationships among various items via accounting for the frequency of occurrence. The antecedent and consequence rule found constructs of the final output results. Initially, this type of solution was mainly used for making purchasing decisions for the retailing industry. However, it gradually began to be used in other fields such as the healthcare and financial industries with the development of big data technology [27].

In the building energy-saving research field, the ARM solution has been harnessed in terms of discovering the relation among building operation data. Yu et al. employed the ARM method to mine the domestic appliance work regulation and found that there were many relationships between different activities, such as TVs and ventilators. Based on the mined information, some suggestions for energy efficiency could be proposed for residents [28]. Yu et al. investigated critical associations for the HVAC system via the ARM algorithm [29]. In this case, plant running fault and wastage patterns were recognized successfully. D'Oca and Hong revealed that there was a correlation between windows opening and relative behaviors using the ARM algorithm [30]. Cabrera and Zareipour identified an education institution lighting wastage mode using the ARM solution [31]. The found associated rules illustrated the relationship between energy consumption and various variables such as season, time, and occupancy status. Similarly, Wang and Shao focused on the lighting system, revealing the energy wastage pattern which is impacted by significant elements. With the aim of forming the regular logic of association rule mining [32], Xiao and Fan made a framework to mine the energy wastage mode in line with ARM solutions [33]. Then, Fan et al. developed this framework further [34]. Under this frame, Li et al., 2017, discovered the ARM rules about the refrigerant flow system [35]. Wang et al. studied electricity consumption modes, focusing on residential buildings, using the ARM solution [36]. Xue et al. firstly utilized three clustered solutions of k-means, k-medoids, and hierarchical to recognize the daily operation patterns for heating in winter [37]. Meanwhile, the ARM approach of the Apriori algorithm was taking advantage of mining energy wasting patterns. Chaobo et al. used the FP-Growth ARM method to investigate typical chiller plant running issues [38]. Qiang and Xiaodong et al. focused on elementary school buildings to study the associated correlation between various subentry energies via the Apriori algorithm [39]. Similarly, Qiang, Ying et al. analyzed small hotel time-series energy using curve features [40]. Xue, Shu, and Da revealed that there was a relationship among socio-demographic features of occupant age, employment, and occupancy patterns [41].

Based on the conventional ARM method, improved association rule mining has been developed gradually with the purpose of enhancing the recognition precision and utilization scope. For example, Qiu et al. proposed a new ARM solution to integrate the weight determination mechanism, increasing algorithm judgement precision [42]. Lighting system, chiller operation, and coordinated control schemes were picked up and refined. To expand the scope of application, a quantitative ARM solution was coded for research. Fan and Xiao made a comparison between the conventional ARM and quantitative ARM solution and found that the latter method could identify numerical and categorical data information which was not limited to a specified data type [43]. Fan et al. developed an ARM solution to focus on the temporal features of a building energy system and, in a further work, the dynamic operation pattern of HVAC work was also discovered using an improved algorithm [44,45]. To increase the capable of recognition, Fan and Song et al. aimed to achieve graph identification using the ARM solution [46]. In this case, Fan, Xiao, and Song et al. put forward an image association rule mining approach successfully and raised the interpretability of mined information [47]. Gunay, Shen, and Yang picked up text data information from the work record of HVAC to discover the component fault frequency [48]. Unlike the aforementioned research, Zhang et al. studied an innovative post-

mining method and discovered the associated rules relating to anomalies by matching the found rules [49,50]. In this way, some unworthiness rules for HVAC equipment energy efficiency were filtered and removed. Liu et al. integrated cluster and association rule mining solutions to mine the operation patterns of office buildings [51]. Table 1 presents some related investigations which have occurred in recent years. It can be seen that the data mining task is the main trend within research studies currently, but fault detection could be also achieved by relative data mining algorithms.

Table 1. Detailed information of some research within 5 years.

Authors	Year	Methods	Aim	Database	Types
Peng et al. [52]	2017	k-means	Detect some inaccurate maintenance and operation data	Building information modeling	Data mining
Noh et al. [53]	2017	Association rule mining algorithm	Analyze the energy efficiency-related factors for commercial building	Official database	Data mining
Ashouri et al. [54]	2018	k-means FP-Growth	Investigate relationship between occupant behavior and energy consumption	Field measurement and questionnaire	Data mining
Akil et al. [55]	2019	k-means	Detection of abnormal data for gas, water, and energy consumption	Field measurement	Abnormal data detection
Garcia and Kamsu-Foguem [56]	2019	k-means FP-Growth	Investigate the relationship between thermal performance and different variations for prefabricated construction	Building information modeling	Data mining
Sun et al. [57]	2019	k-means FP-Growth	Improve the building energy simulation software accuracy	Field measurement and questionnaire	Data mining
Zhang et al. [58]	2020	FP-Growth	Filter some efficient association rules	Field measurement	Data mining
Liu et al. [59]	2020	k-means	Fault diagnosis of chiller system	Official database	Abnormal data detection
Khani et al. [60]	2021	k-means	Discover the energy-related behavior of the buildings' residents	Field measurement and questionnaire	Data mining
Wu and Guo [61]	2021	Apriori	Build a reasonable benchmark for assessment of the conventional building energy usage mode	Field measurement and questionnaire	Data mining
Qiu et al. [62]	2021	FP-Growth	Find related parameters in passive building design	Building simulation database	Data mining
Juan and Lee [63]	2022	FP-Growth	Mine the association between building types and green building technologies	Official database	Data mining

Table 1. *Cont.*

Authors	Year	Methods	Aim	Database	Types
Geng et al. [64]	2022	k-means	Cluster the specified indoor environment quality data	Field measurement	Data mining
Shan et al. [65]	2022	k-means	Study the different building feature under various types	Official database	Data mining
Hao et al. [66]	2022	k-means	Mine the potential capability for buildings	Official database	Data mining
Lei et al. [67]	2023	K-medoids	Propose a dynamic anomaly detection algorithm for building energy consumption data	Field measurement	Abnormal data detection

According to the above reviews of data mining research, it can be seen that, currently, many data mining investigations have been conducted from different perspectives. However, energy time-series data are still not being paid enough attention by investigators. Time-series data hide significant information, reflecting the building power usage habit, while this habit can show several energy consumption issues, providing new solutions to promote energy efficiency from particular points of view. Therefore, in line with these found features, some energy-saving approaches could be proposed, focusing on the specific type of building. The purpose of this paper is to mine the time-series energy variation regulations via clustering and ARM solutions.

The first part of this article has introduced relevant studies which have been conducted. The second section, the Methodology, has described the algorithm running principles. In the Results and Discussion section, the consequences that we find and law are illustrated, and specific reasons are explained. Finally, the Conclusion section summarizes the whole research discoveries.

2. Materials and Methods

2.1. Outline

Figure 1 presents the research outline of this investigation. The entire research mainly contains four phases: data collection, data pre-processing, data mining, and results analysis. Each phase also includes some sub-steps to accomplish corresponding functions. The first significant part is to collect the database for study. In this research, an official commercial energy consumption dataset built by the American National Renewable Energy Laboratory provides the base data, supporting the data mining study.

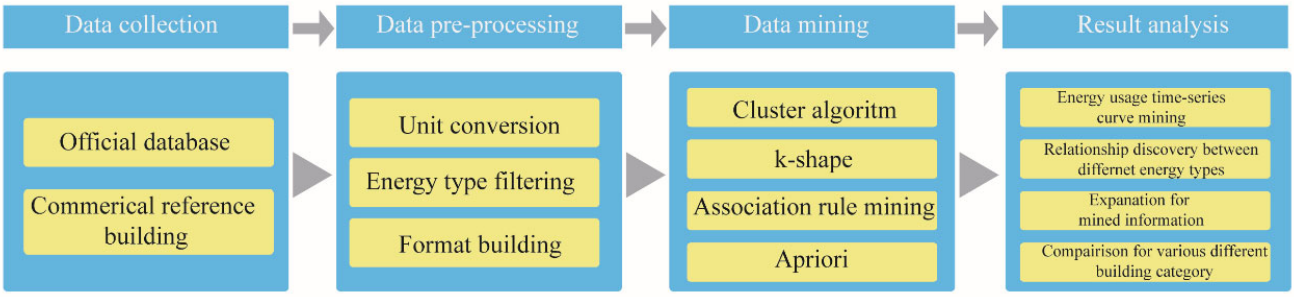


Figure 1. Outline route of this research.

The second step is to preprocess the data for computer program recognition. Three sub-sections are primarily included in this part. As the commercial energy consumption models produce simulated results in Joules, unit conversion should be conducted to con-

vert the units into kilowatt hours (kWh) format. In addition, in order to improve the efficiency of the analysis, subentry energy needs to be filtered first based on the variation feature of the energy curve. Lastly, storing all processed data in a suitable Excel document is essential for conducting post-data mining studies.

The third step primarily contains two core components: clustering and ARM algorithms. The cluster split focuses on grouping different subentry time-series energy consumption curves. K-shape is a data mining algorithm used to discover information in curve shapes, which is the main method selected in this paper. The ARM solution primarily focuses on identifying the relationship between disparate subentry energy forms. As a popular algorithm for this task, Apriori is coded in Python language and serves the ARM mission.

After completing all the steps mentioned above, the Results section presents the analysis of the studied consequences. With respect to cluster surveys, typical energy usage curves are grouped to reflect the characteristics of the building operations. Furthermore, the mined rules using the Apriori approach indicate the main influencing factor for total building energy consumption. In order to examine the reasons in detail, another office building that adopted the same algorithms is compared to the investigated result in this section. Finally, both the mined cluster curves and rules successfully describe the profiles of building energy variation.

2.2. Algorithms

2.2.1. Data Mining

Data mining refers to the progress of extracting hidden information from large datasets. It involves many application fields, such as data statistics. The purpose of data mining is to transfer information into a particular format that can be recognized by a computer system. Algorithm is the solution for achieving the corresponding data mining assignment. It is a series of task sequences coded by computer commands. In other words, data mining could also be regarded as an application system that belongs to machine learning.

Generally speaking, data mining has five common missions: outlier data detection, clustering, association rule mining, classification, and regression. Only cluster and association rule mining belong to unsupervised learning in terms of machine learning. With respect to machine learning types, there are three main categories: supervised, unsupervised, and semi-supervised. These categories are based on whether labels exist or not. Since this study does not involve labels, an unsupervised algorithm is chosen as the primary method.

2.2.2. Cluster Algorithm

Cluster analysis is a statistical technique used to classify a group of objects or subjects into homogeneous groups or clusters based on their similarities and dissimilarities. Unsupervised machine learning is a method that enables the discovery of patterns, relationships, and structures within a dataset without any predefined class labels. In this research, the core assignment is to group the time-series energy consumption data and discover potential patterns within a large dataset. Thus, the unsupervised clustering method is selected as the primary solution for clustering. It should be noted that cluster refers to a mission performed by a computer program with precision, rather than being a type of algorithm.

Table 2 presents the characteristics of various cluster algorithms. Theoretically, an autoencoder has a great power group capacity. However, the simple database logic employed in this study is not indispensable for using this complex algorithm. Among other clustering methods, k-shape has a high precision level and can handle outlier data. Therefore, it is selected as the time-series clustering method for conducting the entire research.

Table 2. Cluster algorithm feature comparison.

Algorithms	ARIMA	Shapelets	DTW	K-SHAPE	Autoencoder
Precision	Low	High	High	High	High
Efficiency	High	High	High	High	High
Complexity	Low	High	High	High	High
Database scope	Global	Local	Global	Global	Global
Outliers filter	No	Yes	No	Yes	Yes
Realization difficulty	Low	High	Low	Low	High
Memory space	Low	Low	Low	Low	High

(ARIMA: Autoregressive Integrated Moving Average Model; DTW: Dynamic Time Warping).

Generally speaking, many algorithms, such as centroid-based clustering and distribution-based clustering, can accomplish the task of clustering. However, in this study, the main method studied is the k-shape algorithm, which focuses on time-series data [68]. Attempting to group time-series curves, k-shape warps and dislocates the data graph without altering the basic shape, which is different from other conventional algorithms. In this case, the precision of recognition could be significantly improved by reducing the influence of zoom. The k-shape method consists of two primary sections: distance calculation between time-series data and picking up times-series curve centroids.

In the process of distance calculation, Equation (1) represents the computational solution, the correlation function.

$$R_k\left(\vec{x},\vec{y}\right)=\begin{cases}\sum_{l=1}^{m-k}x_{l+k}\cdot y_l\\ R_{-k}\left(\vec{y},\vec{x}\right),k<0\end{cases},k\geq 0 \tag{1}$$

where $R_k\left(\vec{x},\vec{y}\right)$ means the correlation value of two vectors of $\vec{x},\vec{y}$. M refers to the number of vectors. k is the sequence number of the correlation value. These data can ensure that the shape remains constant in the process of scaling towards a time-series curve. Equation (2) $CC_W\left(\vec{x},\vec{y}\right)$ is defined by Equation (1). In this sense, Equation (3) realizes the normalization of $NCC_q\left(\vec{x},\vec{y}\right)$ so as to extend the suitable scope. $b(NCC_b)$ is a standardized biased estimator, $u(NCC_u)$ refers to a standardized unbiased estimator, and $c(\underline{NCC_c})$ denotes the parameter standardization method.

$$CC_W\left(\vec{x},\vec{y}\right)=R_{w-m}\left(\vec{x},\vec{y}\right)=(c_1,c_2,c_3\ldots\ldots c_m),w\in\{1,2,\ldots\ldots,2m-1\} \tag{2}$$

$$NCC_q\left(\vec{x},\vec{y}\right)=\begin{cases}CC_W\left(\vec{x},\vec{y}\right)/m,\ q=b(NCC_b)\\ CC_W\left(\vec{x},\vec{y}\right)/(m-|w-m|),\ q=u(NCC_u)\\ CC_W\left(\vec{x},\vec{y}\right)/\sqrt{R_0\left(\vec{x}\cdot\vec{x}\right)\cdot R_0\left(\vec{y}\cdot\vec{y}\right)},\ q=c(NCC_c)\end{cases} \tag{3}$$

After the normalization procedure, optimization of w value can be achieved. In this case, Equation (4) outputs a similar distance SBD (shape-based distance) calculation solution for various time-series data shapes. To create a shape-based distance measure, we applied coefficient normalization to ensure that the values range from -1 to 1. Coefficient normalization involves dividing the cross-correlation sequence by the geometric mean of the autocorrelations of the individual sequences. After normalizing the sequence, it is indispensable to identify the optimum position w, where $NCC_c\left(\vec{x},\vec{y}\right)$ is maximized. The

value ranges from 0 to 2, in which 0 means that the two time-series curves have the highest similarity.

$$SBD\left(\vec{x},\vec{y}\right) = 1 - \max_{w}\left(CC_W\left(\vec{x},\vec{y}\right)/\sqrt{R_0\left(\vec{x}\cdot\vec{x}\right)\cdot R_0\left(\vec{y}\cdot\vec{y}\right)}\right) \tag{4}$$

The computation of $CC_W\left(\vec{x},\vec{y}\right)$ for all values of w requires O(m^2) time, where m is the time-series length. However, by using the convolution theorem and Fast Fourier Transform (FFT) algorithm, the time complexity can be reduced to O(mlog(m)). To further improve the performance, padding zeros is conducted to ensure that the length of the time-series is the next power of two after (2m–1). Therefore, it proves that SBD is a highly competitive method that delivers comparable outcomes to cDTW (constrained Dynamic Time Warping) and DTW (Dynamic Time Warping) but with significantly faster processing times, often differing by several orders of magnitude [68].

After enabling the distance calculation approach, it is indispensable to search the cluster center next. As for time-series clustering methodologies, homologous centroids are also a series of energy consumption data. Concentrating on the features of the time-series, k-shape solution overcomes the disadvantage of the common clustered searching procedure which computes the average value for each time. Figure 2 presents centroid search methods using a disparate cluster algorithm of common and k-shape [68].

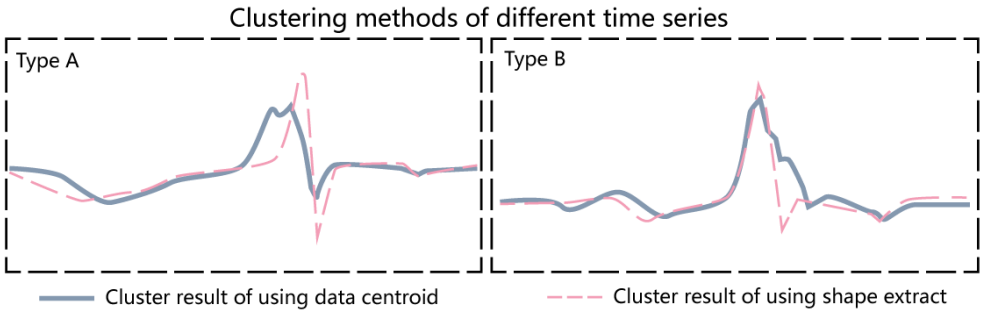


Figure 2. Cluster centroid condition under various methods.

Instead, k-shape transforms the averaged value calculation process into an optimization issue. In this case, the problem of the sum of squared distances is transferred to the eigenvector issue when the similarity is maximized. In other words, the process where time-series data close to the centroid is known as eigenvector optimization proceeding.

$$\begin{aligned}\vec{u_k}^{*} &= arg\max_{\vec{u_k}}\sum_{\vec{x_i}\in P_k} NCC_c\left(\vec{x_i},\vec{u_k}\right)^2 \\ = arg\max_{\vec{u_k}}\sum_{\vec{x_i}\in P_k}&\left(max_w CC_W\left(\vec{x_i},\vec{u_k}\right)/\sqrt{R_0\left(\vec{x_i}\cdot\vec{x_i}\right)\cdot R_0\left(\vec{u_k}\cdot\vec{u_k}\right)}\right)^2\end{aligned} \tag{5}$$

Equation (5) indicates the eigenvector computational manner. Where u_k and P_k are the eigenvector and corresponding vector space, respectively. Equation (6) shows the final eigenvector calculation form which inserts vector patterns. Where $\vec{u_k} = \vec{u_k}\cdot Q$, M = Q·T·S·Q and Q = I – (1/m)O, S = $\sum_{\vec{x_i}\in P_k}(\vec{x_i},\vec{x_i}^{T})$. Capital I represents the unit matrix, and O is the matrix when each element is 1, M = Q·T·S·Q. In the end, the optimization course transfers into a classical maximum of the Rayleigh Quotient issue.

$$\vec{u_k}^{*} = arg\max_{\vec{u_k}}\frac{\vec{u_k}^{T}\cdot Q^{T}\cdot S\cdot Q\cdot\vec{u_k}}{\vec{u_k}^{T}\cdot\vec{u_k}} = arg\max_{\vec{u_k}}\frac{\vec{u_k}^{T}\cdot M\cdot\vec{u_k}}{\vec{u_k}^{T}\cdot\vec{u_k}} \tag{6}$$

2.2.3. Association Rule Mining Algorithm

The ARM algorithm aims to investigate the relationship among various variables. Some significant related rules are discovered by counting the frequency for every itemset [69]. To achieve this purpose, several algorithms have been developed, such as Apriori and FP-Growth.

The association rule mining algorithm mainly contains two types of Apriori and FP-growth. Table 3 compares the related characteristics of two methods. It can be seen that the FP-growth approach has a higher efficiency and great process capability in terms of a large database. However, Apriori has simpler building logic. Because of the clear dataset structure used in this research, using the Apriori solution is sufficient.

Table 3. Association algorithm feature comparison.

Algorithm	Apriori	FP-Growth
Building logic	Simple	Complex
Scan times	Depend on dataset	Two
Computational manner	Iteration	Scan
Efficiency	Depend on dataset	High
Regular database	High	High
Sparse database	Low	High
Memory space	High	Low
Time	Depend on dataset	Fast

In this research, Apriori is considered to be the main solution to develop the ARM assignment. This method was proposed by R. Agrawal and R. Srikant in 1994 [70]. It computes the itemset occurrence time to judge the association relationship among each variable, which has high precision and efficiency in proceedings. Although there are many algorithms concentrating on this issue, the basic principles for the ARM mission are similar to each other.

The ARM solution is commonly defined as follows:

$$\text{Items}: \ I = \{i_1, i_2, i_3, \ldots i_n\}$$

$$\text{Transactions}: \ D = \{d_1, d_2, d_3, \ldots d_n\}$$

where i_n refers to the specific item. d_n means the transactions which constitute the database. Every transaction D includes some items in I. In this case, the final mined rule is formed as:

$$X => Y, where \ X, Y \subseteq I$$

X and Y mean the itemset in I. After mining the rules, in order to increase the explanation level, X can be called antecedent and Y is consequent. In this sense, the corresponding format for this rule can be translated as if X happens, and then Y occurs under a particular frequency.

2.3. Perform Indicator

2.3.1. Cluster Algorithm Indicators

Focusing on the cluster algorithm, the silhouette coefficient is selected as a perform indicator to show the performance of the cluster number. The number of clusters corresponding to a larger silhouette coefficient is the seeking target. The calculation process is shown as follows:

Compute the averaged distance $a(i)$ between sample i and other data in the same cluster. The mean value of distance $a(i)$ for all samples in a cluster is the similarity.

In terms of different clusters, compute the mean distance $b(ij)$ between sample i and data, which is in another cluster. $b(i) = \min\{bi1, bi2, \ldots, bik\}$ represents the distinction degree for the variable and corresponded cluster.

In this case, Equations (7) and (8) are the silhouette coefficient calculation manners.

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{7}$$

$$S(i) = \begin{cases} 1 - \frac{a(i)}{b(i)}, \ a(i) < b(i) \\ 0, a(i) = b(i) \\ \frac{a(i)}{b(i)} - 1, a(i) > b(i) \end{cases} \tag{8}$$

In line with the computational result, the cluster number could be determined. If the S(i) is close to 1, this indicates that the number of clusters is set reasonably. In this case, it is indispensable to set the cluster number first based on the silhouette coefficient value before running the algorithm.

2.3.2. Association Rule Mining Algorithm

In the ARM process, three indicators of support, confidence, and lift could reflect the final mining consequence directly. Thereinto, confidence is used the most in this paper. Confidence means the proportion of transactions containing X, also including Y. Equations (9) and (10) present the confidence generation solution.

$$\text{confidence}(X \rightarrow Y) = P(Y|X) = P(X\&Y)/P(X) \tag{9}$$

$$\text{confidence}(Y \rightarrow X) = P(X|Y) = P(X\&Y)/P(Y) \tag{10}$$

Confidence measures the reliability of mined rules. Focusing on the rules of X=>Y, with high confidence, this rule occurs with a higher possibility. This research adopts confidence to evaluate the ARM performance.

2.4. Database

The database supporting this study is the American Commercial Reference Buildings Database [71]. This database contains 16 types of commercial building energy consumption models of Energyplus. All these categories include almost all the architectural types within America. These models are validated by researchers and follow the relative standard ASHRAE 90.1-2004 [72] in order to improve their precision significantly. Meanwhile, some weight factors are also integrated into models to correct the simulation model's accuracy. All models have been set totally; in this case, final energy simulation results can be generated directly. Table 4 indicates all building types in the commercial reference building database [71].

Table 4. Building types in the commercial reference building energy database.

Building Type Name	Floor Area (ft^2)	Number of Floors
Large office	498,588	12
Medium office	53,628	3
Small office	5500	1
Warehouse	52,045	1
Stand-alone retail	24,962	1
Strip mall	22,500	1
Primary school	73,960	1
Secondary school	210,887	2
Supermarket	45,000	1

Table 4. *Cont.*

Building Type Name	Floor Area (ft^2)	Number of Floors
Quick-service restaurant	2500	1
Full-service restaurant	5500	1
Hospital	241,351	5
Outpatient health care	40,946	3
Small hotel	43,200	4
Large hotel	122,120	6
Midrise apartment	33,740	4

2.5. Building Condition

This paper selects a full-service restaurant as the energy model prototype based on its universality. Houston is adopted as the location for supporting the simulated climatic condition. The total area of this restaurant is 511 m^2 with one floor and a shed roof. Two zones were set; the kitchen and dining room constitute the entire district. In terms of area for each zone, 1.39 m^2/person controls the size of the dining area for occupancy, and 15.6 m^2/person manipulates the kitchen area. Figure 3 shows the full-service restaurant simulation model. Table 5 indicates the basic building parameters used in this research [71].

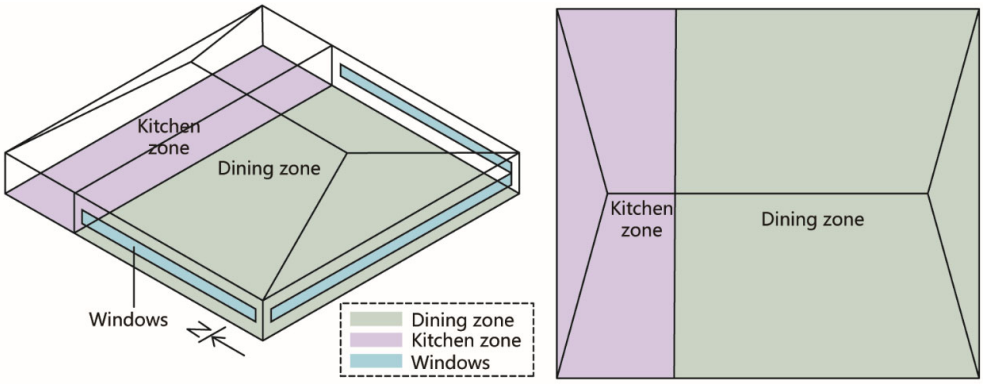


Figure 3. Simulation model basic situation.

Table 5. Full-service restaurant basic parameters.

Building Area	Length-Width Ratio	Floors	Height	Clear Height	Glazing Ratio
511 m^2	1.0	1	3.05 m	3.05 m	0.17

With respect to the requirement of airflow, in architecture, different zones correspond to different settings. Table 6 presents the settings for the air demand of disparate zones for the architecture model [71]. All the related arguments are determined in line with official building standards coupled with a field survey. In addition, the air infiltration class is set as lower than 0.4 cfm/ft^2 per 2 L/s/m^2 under 75 pa pressure according to the standard 90.1-2004.

Table 6. Configuration for building model ventilation.

Zones	Person (m^2/Person)	Requirement for Outdoor Air	Requirement for Airflow
Kitchen	18.59	7.08 L/s/person	0.38 L/s/m^2
Dining	1.39	9.44 L/s/person	6.77 L/s/m^2

In terms of the energy demand, researchers have investigated some typical buildings and related standards to ensure specific values. Table 7 indicates the setting of relative

parameters in the model to reflect the architectural energy consumption situation [71]. The energy intensity of each zone is determined by per capita format. Furthermore, in terms of the building's surroundings, the model is constructed under a type of loft integrated with a steel frame wall. The heat transfer coefficient is 0.006 (W/M^2 K) for the loft and 0.704 (W/M^2 K) for the surroundings, respectively.

Table 7. Full-service restaurant energy demand.

Zones	Area (m^2)	Volume (m^3)	Area per Person (m^2/Person)	Lighting Load (W/m^2)	Electricity Load (W/m^2)
	Natural Gas Load (W/m^2)	**Ventilation Load (L/s)**	**Exhaust Load (L/s)**	**Air Infiltration (ACH)**	**Life Water (L/h)**
Dinning	372	1133	1.39	22.6	60.3
	0.0	2667.7	0.0	0.52	0.0
Kitchen	139	425	18.58	12.91	376.6
	1197.9	60.0	1887.8	0.63	503.5
Loft	511	856	-	-	-
	-	-	-	1.0	0.0

For the building equipment running configuration, three main facilities of lighting, HVAC, and refrigerator constitute the simulation model. Equations (11) and (12) control the area factor to correct the lighting power density.

$$AF = 0.2 + 0.8(1/0.9^n) \tag{11}$$

$$n = \frac{10.21(CH - 2.5)}{\sqrt{A_\tau}} - 1 \tag{12}$$

where AF means the area factor. CH is the room height and A_τ indicates the space area. The final restaurant lighting power density is set as 23.3 W/m^2. The HVAC system primarily defines each type of data in line with actual field measurement buildings. Furnace is the heating system and packaged air-conditioning unit supplies for cool air. Single-zone constant air volume manipulates the ventilation distribution manner. Thereinto, the significant operation of the cooling system depends on the coefficient of performance (COP) argument. In Energyplus software, a COP for the compressor is required to build the cooling module. Equation (13) presents this coefficient calculation solution.

$$COP_{comp} = ((EER/3.413) + R)/(1 - R) \tag{13}$$

where EER refers to the energy efficiency ratio. R is set as 0.12, which indicates the ratio of the rated fan power to the total power [73]. With regard to the refrigerator module, Table 8 illustrates specific settings from disparate views [71]. According to the kitchen application, both the cooler and freezer need to store various food under different circumstances. In addition, a schedule is indispensable for energy consumption simulation. Each facility schedule is determined based on multiple actual building running situations. All these options guarantee the precision of the simulated energy consumption model, which clearly reflects the typical architectural operational mode.

Table 8. Refrigerator setting parameters.

Equipment	Area (m^2)	Length (m)	Load (kW)	COP	Evaporator Fan (W)	Condenser Fan (W)	Lighting (W/m^2)	Defrost (W)	Anti-Condensation (W)	Working Temperature (°C)
Cooler	9.3	3.0	2.2	3	200	330	10.8	0.00	0	1.7
Freezer	7.4	2.4	1.7	1.5	180	329	10.8	2000	0	–23.3

2.6. Workflow

Regarding the entire algorithm operation logic, the aforementioned sections build a research frame for clustering and ARM. Figure 4 presents the clustered analysis process workflow. It should be noted that all the steps are coded by python language, supported by the software pycharm, and the consequences of the final output results are listed under some of the graphs. Furthermore, the definition of the k value depends not only on the silhouette coefficient but also considers the specific operation effect.

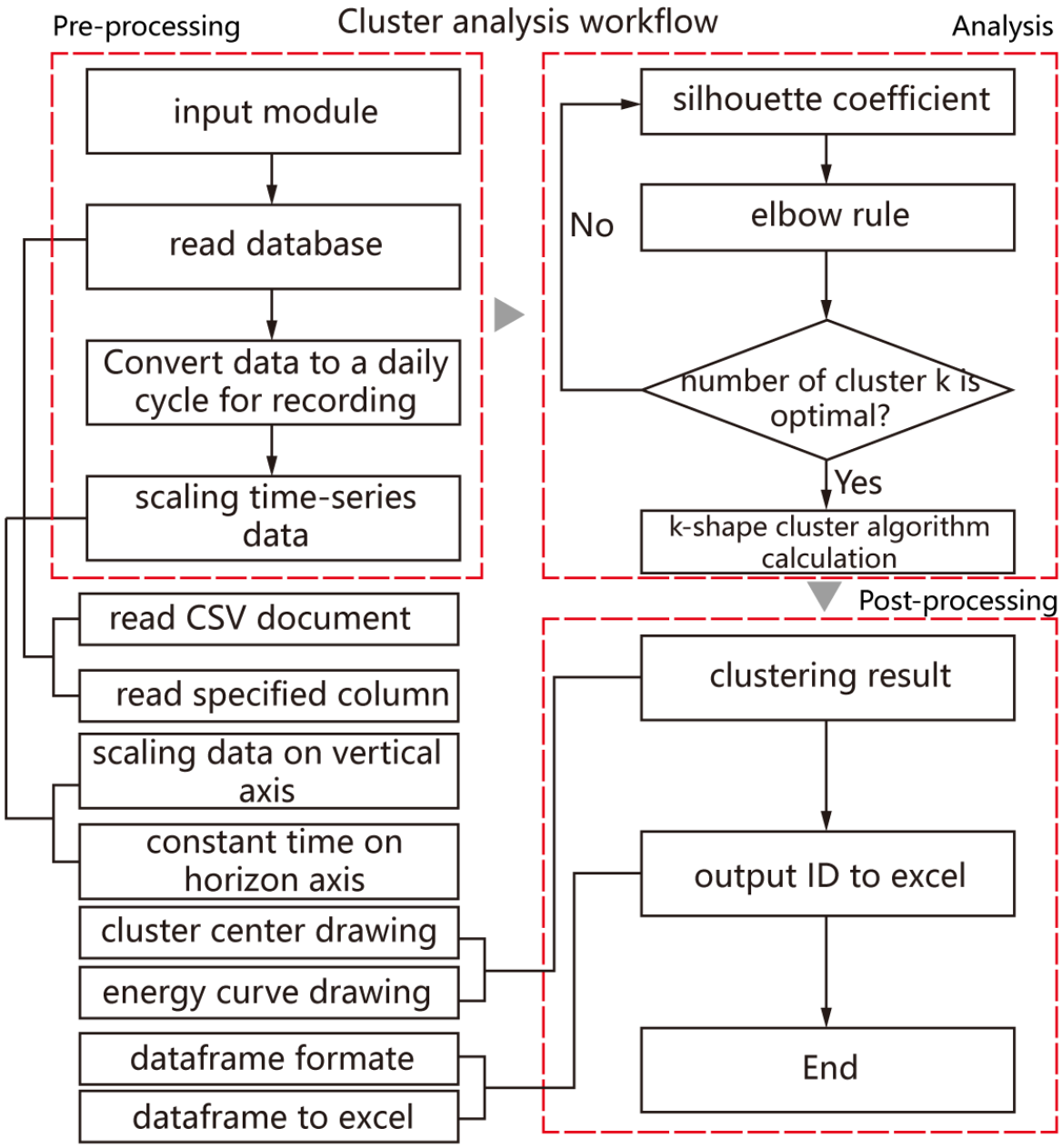


Figure 4. Running logic of cluster analysis.

Figure 5 shows the ARM research logic work flow. It is observed that the significant step is to sign a different value between adjacent moments. Based on each signature, all the associated rules could be mined by the Apriori algorithm.

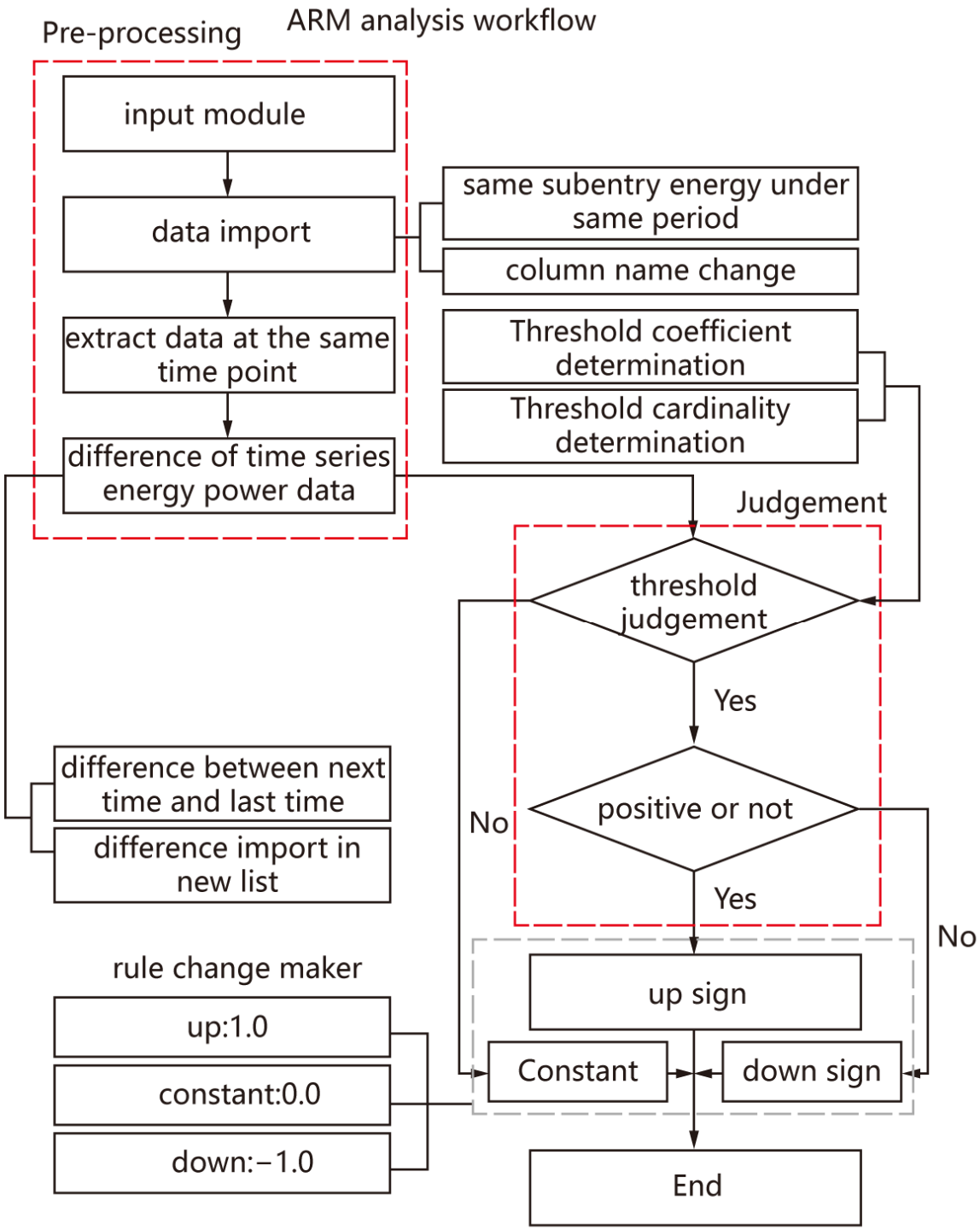


Figure 5. Operation logic of association rule mining algorithm.

3. Results

Considering the features of different subentry energy types, it is imperative to decide which type of energy should be analyzed first. This is because some load shapes only show one variation form. In this case, it is not meaningful to mine potential information from this mode. Figure 6 compares two kinds of energy variation forms. The data sources derive from the official American measured database [71].

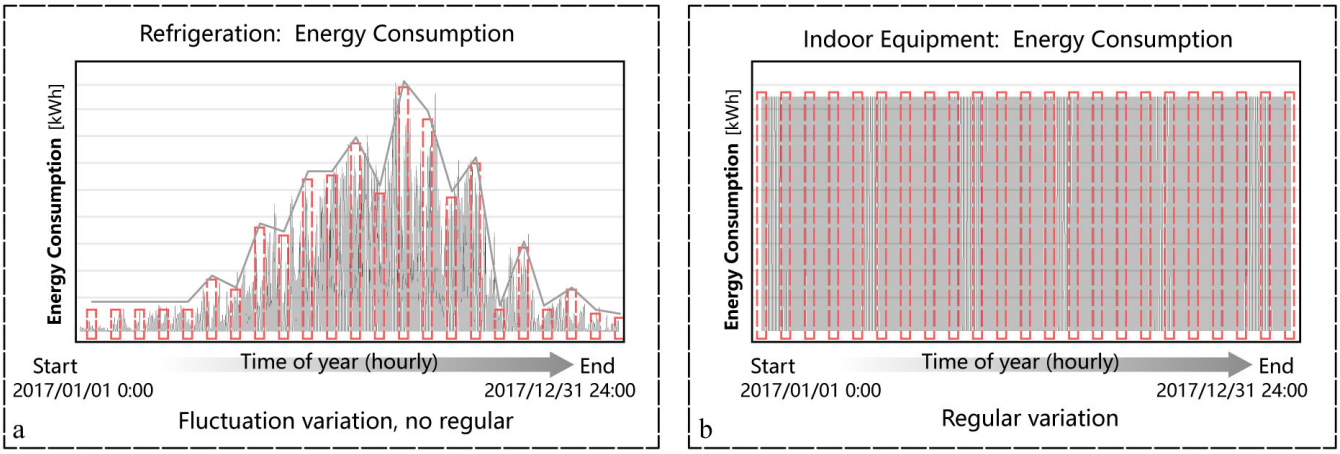


Figure 6. Two types of energy variation forms.

This Figure 6 was generated via exhibiting an annual time-series energy consumption situation. The grey color line refers to the energy curve, and the red line means the daily power variation condition. It can be seen that Figure 6a displays no significant fluctuation along the time being worth investigation potential information, while Figure 6b has the same energy changing shape each day, which means that there is no valuable pattern required to be surveyed. After primary energy filter process, four sorts of subentry power load are picked up to be studied such as cooling, heating, refrigeration power, and pump load. Additionally, total energy consumption should also be analyzed to discover the relationship between it and subentry items. Each group of clustered time-series data spans from 1:00 am to 12:00 pm within a day, while the enter database contains all the energy consumption data of the year 2017, that is 365 time-series power load curves which reflect the energy consumption condition used each day for clustering all year round.

Figure 7 displays the whole research purpose and solution. For cluster, the aim is to group the time-series energy consumption curve for each day with 24 data because there are 24 h in a day. In the last, 365 energy variation curves are classified into various categories. Regarding the ARM task, the relationship between the loading variance of subentry adjacent energy is mined within the same time. In this scenario, it is necessary to ascertain the changing energy threshold between the adjacent moment first. Equation (14) presents the judgement condition for the charge difference.

$$\begin{cases} |S| \geq E_{T+1} \cdot a \\ |S| \geq b \end{cases} \tag{14}$$

where S means the difference in the value of T and $T + 1$ time. T refers to the specific time in a day. E_{T+1} is the energy consumption value for $T + 1$ moment, and a and b denote threshold coefficients of 0.01 and 0.05, respectively. Only if both of the limitations are overcome does it mean that the power load can be regarded as changed under contiguous time. In this case, final minded rules could be shown as formatted as A↑ ⇒ B↓, for which A,B illustrate subentry energy types and the symbol ↑↓ denotes an increasing or decreasing changing status.

3.1. Cluster Analysis

3.1.1. Cooling

Cooling power consumption refers to the energy consumption used to cool the air in the building, that is, the air-conditioning cooling part of the HVAC system. Table 9 presents the silhouette coefficient value with different k values. When the k is 3, the time-series curve shows the best clustered performance.

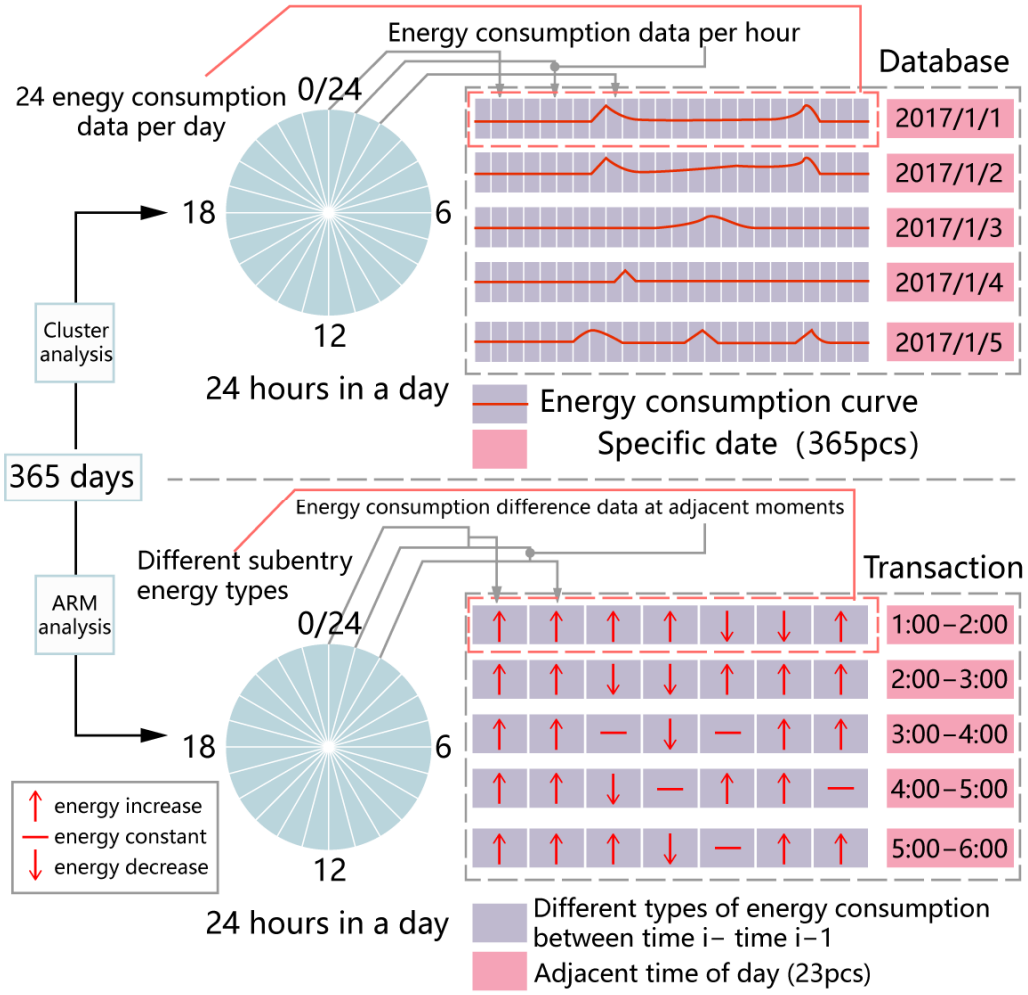


Figure 7. Cluster and ARM analysis relationship.

Table 9. Silhouette coefficient of cooling consumption.

K	2	3	4	5	6	7	8	9
Silhouette coefficient	0.31	0.32	0.11	0.13	0.10	0.11	0.13	0.10

Figure 8 indicates the clustering consequence for the cooling energy consumption curve and the corresponding date for each type. The blue dotted box denotes the energy curve adjustment method. The result shows that three forms are grouped with disparate shapes. Type A and C expresses invert U geometry, while type B appears in M form with multiple peaks. This is because canteen is required to employ the air conditioner all day due to hot weather in the summer. Moreover, in winter, the number of customers increases the load of the cooling system; thus, type B displays a high relevance with human traffic. For example, the curve peaks of type B apparently occur at 6:00–9:00 and 12:00–14:00 when it is dining time. Based on the features of group performance, to save energy expenditure, type A and C could try to diminish the cooling time, such as by stopping HVAC while importing outdoor cold air into the interior at night, especially during transitional seasons. Therefore, it is recommended that the cooling time be extended to correspond with peak moments, meal times, which could relieve the curve fluctuation degree in winter. Under this circumstance, the goal of diminishing electricity short-term pulse wastage is achieved.

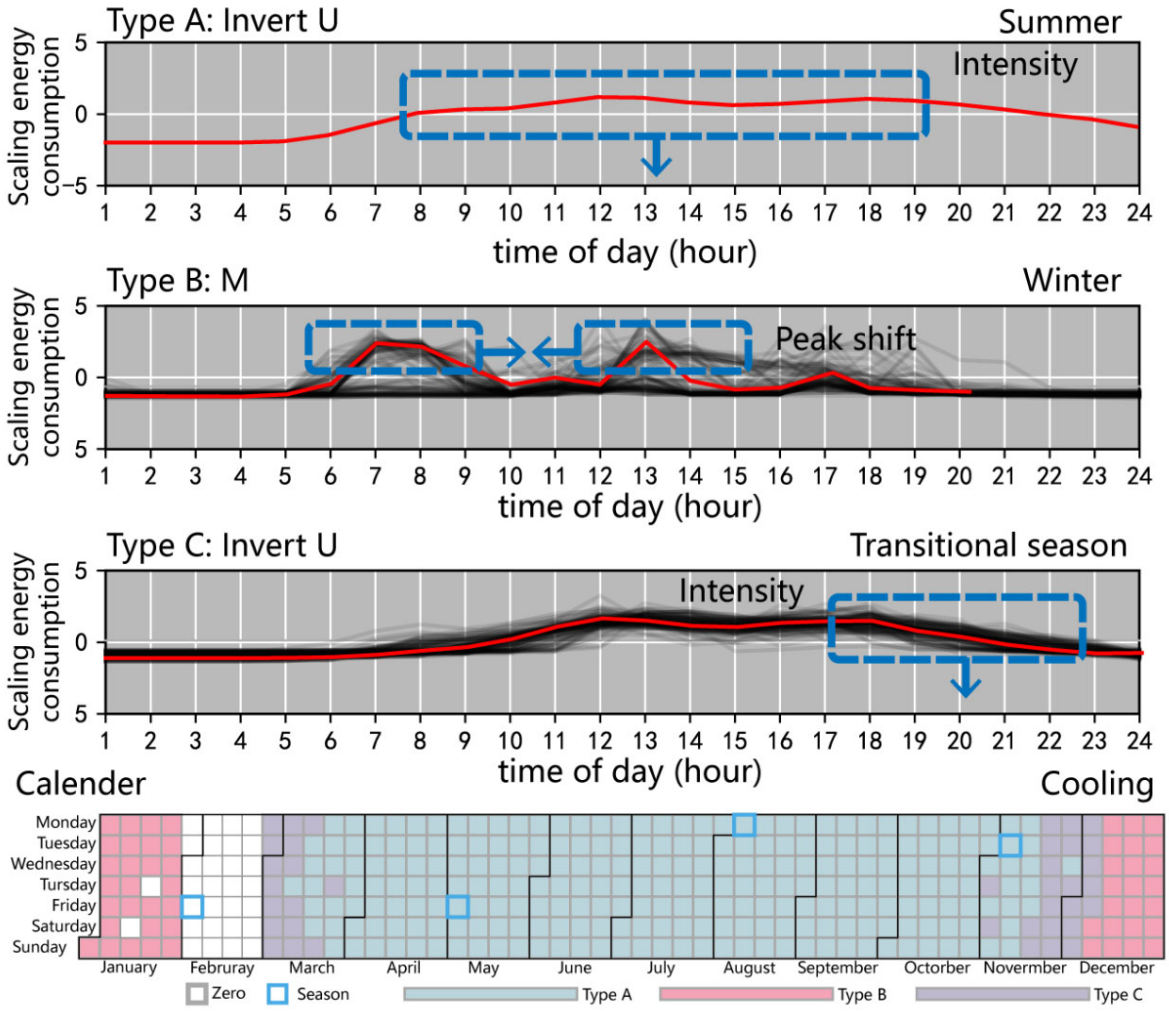


Figure 8. Clustered result of cooling energy consumption.

3.1.2. Heating

A heating facility is used in restaurants each season. Table 10 presents the k-value condition, showing that two is the best clustered number for research. Figure 9 indicates the corresponding clustering consequence. The blue dotted box denotes the energy curve adjustment method. The result shows that type A expresses invert V shape, and M form correlates the type B curve. In line with the calendar information, as shown in type A, heating used in the morning mainly applies for dehumidification in the summer and transitional seasons. In winter, peaks only occur in the morning because of the cold weather; moreover, it is harnessed at a relatively low level during the daytime due to the weather outside and many people being inside for a meal. For instance, type B impresses M from a time-series shape, illustrating high features at both ends and low ones in the middle.

Table 10. Silhouette coefficient of heating consumption.

K	2	3	4	5	6	7	8	9
Silhouette coefficient	0.56	0.32	0.31	0.4	0.29	0.51	0.33	0.29

Concentrating on this characteristic of heating utilization, some passive ventilation solutions could be employed for dehumidification, diminishing the heating load in summer and transitional seasons. In terms of winter, considering that the cooling system runs in the daytime, heating equipment should decrease to zero instead of running at a low degree during this time, saving the consumption of natural gas.

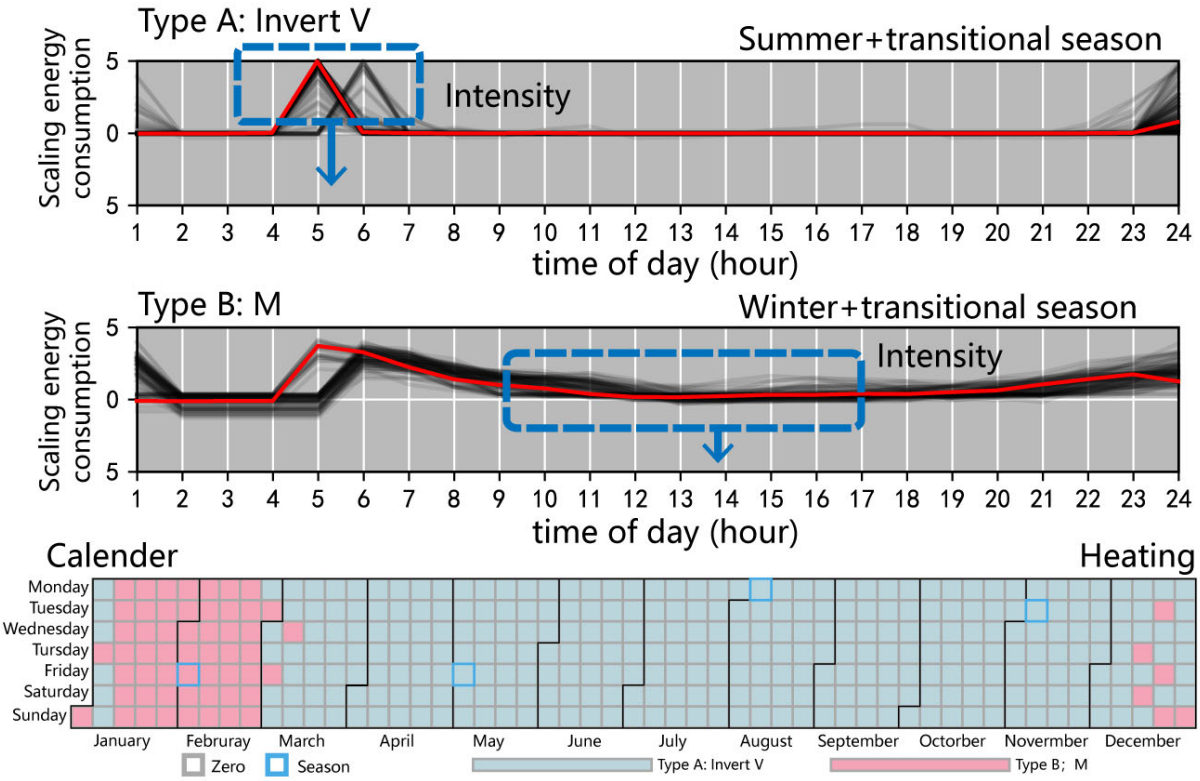


Figure 9. Clustered result of heating energy consumption.

3.1.3. Pump Load

Table 11 illustrates the silhouette coefficient condition of pump energy expenditure. The optimal number of k is three. Figure 10 suggests three sorts of curves after categorizing. Pump electricity power usage reflects the amount of life service water usage. It can be seen that work time distributes these different types rather than seasons mined in aforementioned subentry energy types. Type A approximately expresses an inverted U-shaped graph, while type B and C display M forms. The M pattern owes to the weekday when two peaks correspond to the meal time at noon and night. On the other hand, with type A, the meal time in the daytime is put off on weekends to integrate it with other dining times, shaping inverted U-shaped figures. As an example, the meal time in the morning is postponed from 7:00 on weekdays to 9:00. In addition, it should be noted that the difference between type B and C only concentrates on the distinction between summer- and wintertime. Generally speaking, at the moment of business beginning and ending, there is peak energy usage, respectively.

Table 11. Silhouette coefficient of pump energy consumption.

K	2	3	4	5	6	7	8	9
Silhouette coefficient	0.56	0.82	0.81	0.4	0.89	0.52	0.31	0.39

It can be obtained that the pump operation characteristics mainly relate to the work time instead of outside temperature, which is different with cooling and heating. Focusing on this feature, the key to the pump power energy efficiency solution is to store water in advance. Storing water at the trough and using it at the peak can effectively reduce the fluctuation degree of the energy usage curve, achieving the energy-saving goal.

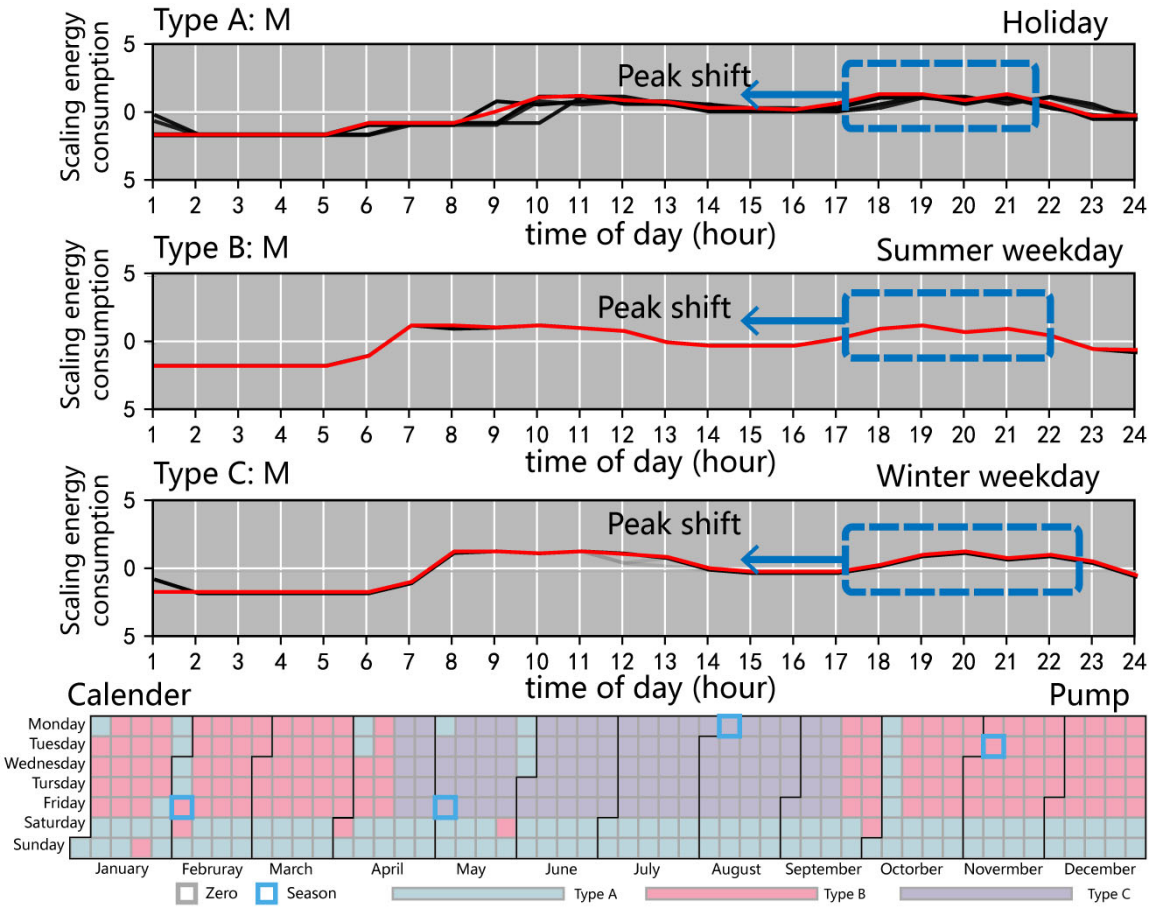


Figure 10. Clustered result of pump energy consumption.

3.1.4. Refrigeration Power

The refrigeration power load reveals the canteen workload to some extent. Table 12 shows that it is better to classify this kind of energy expenditure into two patterns. Figure 11 shows the result after clustering the k-shape algorithm. The graph illustrates that these two modes are mainly distributed on the energy peak between 3:00 and 5:00 in the morning, expressing invert V shapes. This is mainly attributed to the fact that restaurant buildings have a good exhibition process, while this procedure occurs on each hour of Tuesday and Friday, as presented by type B. Other refrigeration power loads show wave peaks concentrating on meal times for cooking requirements and rest times for food storage requirements. As an example, the peaks primarily exist at noon for lunch at 10:00–12:00 and at night for rest at 22:00–24:00.

Table 12. Silhouette coefficient of refrigeration power consumption.

K	2	3	4	5	6	7	8	9
Silhouette coefficient	0.31	0.09	0.13	0.14	0.04	−0.16	−0.42	−0.39

According to the aforementioned curve features, two solutions could be harnessed, diminishing the energy consumption. On the one hand, decreasing the frequency of opening and closing refrigerators can be an effective method to cut down the peak value. Therefore, it should be better to take out most frequently used food for cooking from the refrigerator in advance. On the other hand, postponing goods exhibition actions before dawn and combining this with the meal time at noon could reduce the fluctuation level of the curve, achieving the energy consumption objective.

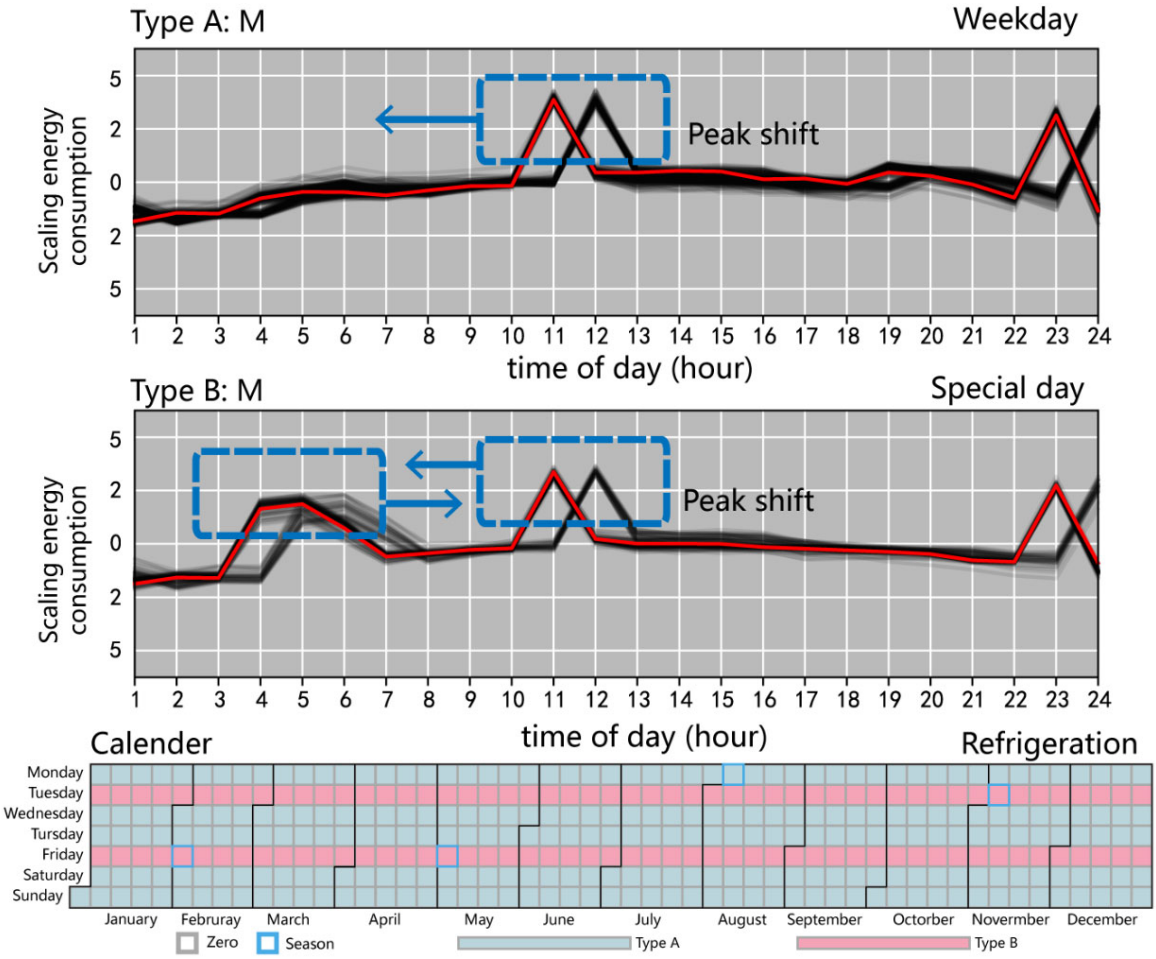


Figure 11. Clustered result of refrigeration energy consumption.

3.1.5. Total Energy Consumption

Total energy consumption refers to the sum of all subentry energies not limited to the load categories listed above. Table 13 shows the silhouette coefficient conditions of different k values. Figure 12 displays the clustering results of two, named, respectively, type A and B.

Table 13. Silhouette coefficient of total energy consumption.

K	2	3	4	5	6	7	8	9
Silhouette coefficient	0.49	0.30	0.08	0.13	0.41	0.36	0.41	0.03

The figure shows that load curves are highly related to breakfast, lunch, and dinner times in summer, resulting in a multi-peak pattern known as type A. Type B, presenting a generally similar feature to type A, has nevertheless a higher loading degree during the morning and night due to the increase in heating requirements. Two methods are introduced to decrease the energy load. One is to reduce the energy intensity at some special moment, while the other is to shift load peaks to fill up the troughs. Reducing energy intensity mainly adopts certain passive building design manners, while peak load shift needs to change people's action patterns under specific circumstances. Through analyzing different energy consumption fluctuation curves, some sequential patterns have been acquired after the clustering algorithm. These results can support building energy efficiency from a new perspective.

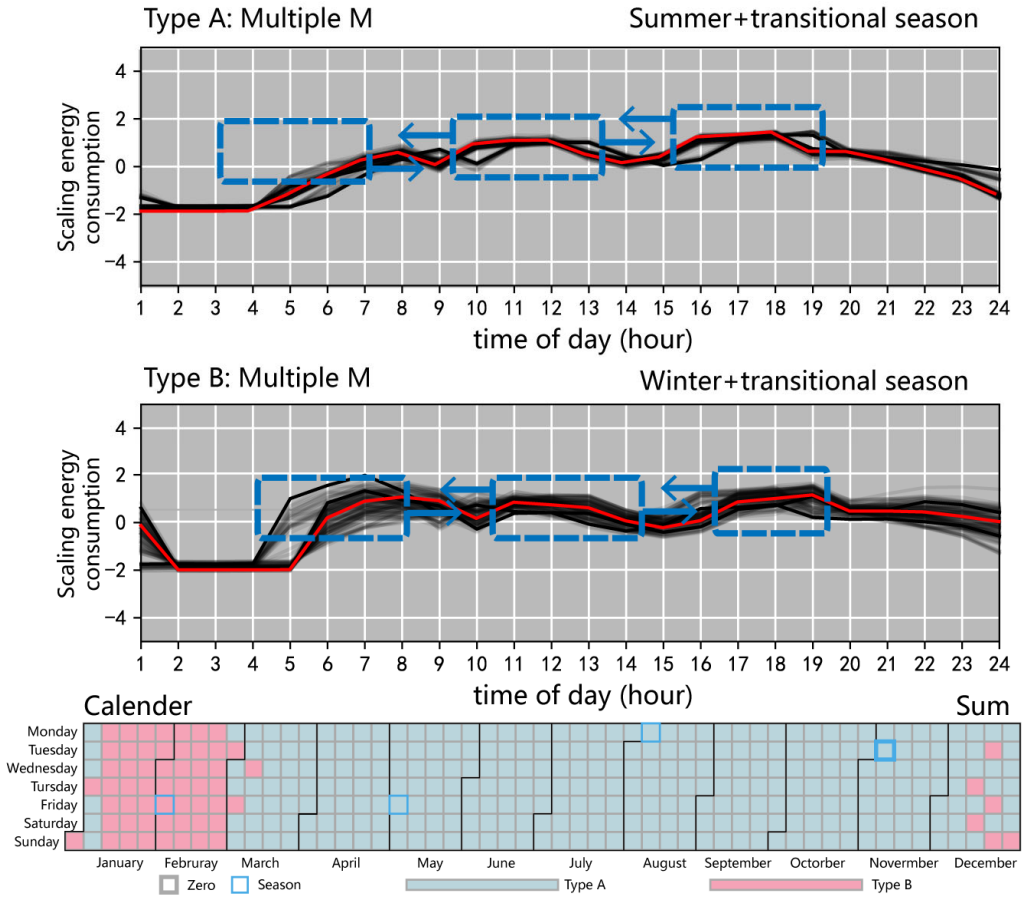


Figure 12. Clustered result of total energy consumption.

3.2. Association Rule Mining Analysis

Rules mined by association algorithms are listed hourly. Moreover, buildings chiefly run at working time instead of all hours. Thus, in this paper, efficiency rules mainly take over between 6:00 and 24:00. It should be noted that rules are filtered if their confidence thresholds stay under 60.00% due to reliability considerations. In addition, multiple ARM studies are performed in summer and winter, respectively, since all sequential curves are primarily clustered into these two seasons in line with the aforementioned clustering discovery.

3.2.1. Summer

Table 14 and Figure 13 present the ARM results. It can be seen that cooling basically shows a positive correlation with total energy fluctuation during each period. While, in terms of the pump load condition, in the period of 6:00–7:00 and 22:00–23:00, it promotes total energy fluctuation to some extent. Refrigeration energy consumption mainly has a positive influence on the cooling and total energy at lunch time and at night. It can be seen that, in summer, cooling power adoption is the most efficient method, contributing to total energy consumption. Moreover, pump load is not the major factor in energy consumption. For refrigeration, it shows a positive relationship with cooling and total energy consumption, as the stream of people who are dining can also affect the interior temperature. For example, at breakfast and dinner time, cooling shift shows the same trend as refrigeration, as is shown in Table 14. According to all the discussions above, it is recommended to focus on the energy fluctuation of cooling and refrigeration to reduce energy consumption because of their high correlations. Reducing the frequency of opening and closing refrigerators, such as taking out frequently used cooking ingredients in advance at a dining time, can also smoothen the fluctuation curve.

Table 14. Rules mined in summer.

Period	Rules	Confidence
6:00–7:00	Cooling↑=》 Total energy↑	79.16%
	Cooling↑=》 Pump load↑	79.16%
	Pump load↑=》 Total energy↑	66.66%
7:00–8:00	Cooling↑=》 Pump load↓	70.83%
	Pump load↓=》 Total energy↑	70.83%
	Cooling↑=》 Total energy↑	100.00%
8:00–9:00	-	-
9:00–10:00	-	-
10:00–11:00	Cooling↑=》 Refrigeration↑	100.00%
	Cooling↑=》 Total energy↑	100.00%
11:00–12:00	Cooling↑=》 Refrigeration↑	100.00%
	Cooling↑=》 Total energy↑	100.00%
12:00–13:00	Cooling↓=》 Refrigeration↓	79.16%
	Cooling↓=》 Total energy↓	79.16%
	Refrigeration↓=》 Total energy↓	100.00%
13:00–14:00	Cooling↓=》 Total energy↓	95.83%
14:00–15:00	-	-
15:00–16:00	-	-
16:00–17:00	Cooling↑=》 Pump load↑	75.00%
	Cooling↑=》 Total energy↑	70.83%
	Pump load↑=》 Total energy↑	87.5%
17:00–18:00	Cooling↑=》 Pump load↑	100.00%
	Cooling↑=》 Total energy↑	100.00%
	Pump load↑=》 Total energy↑	100.00%
18:00–19:00	Cooling↓=》 Total energy↓	95.83%
	Cooling↓=》 Refrigeration↑	70.83%
	Cooling↓, Refrigeration↑=》 Total energy↓	70.83%
19:00–20:00	-	-
20:00–21:00	Cooling↓=》 Total energy↓	95.83%
	Refrigeration↓=》 Total energy↓	100.00%
	Cooling↓, Refrigeration↓=》 Total energy↓	95.83%
21:00–22:00	Cooling↓=》 Total energy↓	95.83%
	Refrigeration↓=》 Total energy↓	100.00%
	Cooling↓, Refrigeration↓=》 Total energy↓	95.83%
22:00–23:00	Cooling↓=》 Total energy↓	100.00%
	Pump load↓=》 Total energy↓	100.00%
	Cooling↓, Pump load↓=》 Total energy↓	100.00%
23:00–24:00	-	-

(↑: energy increase, ↓: energy decrease, -: same with previous energy condition, =》 : lead to).

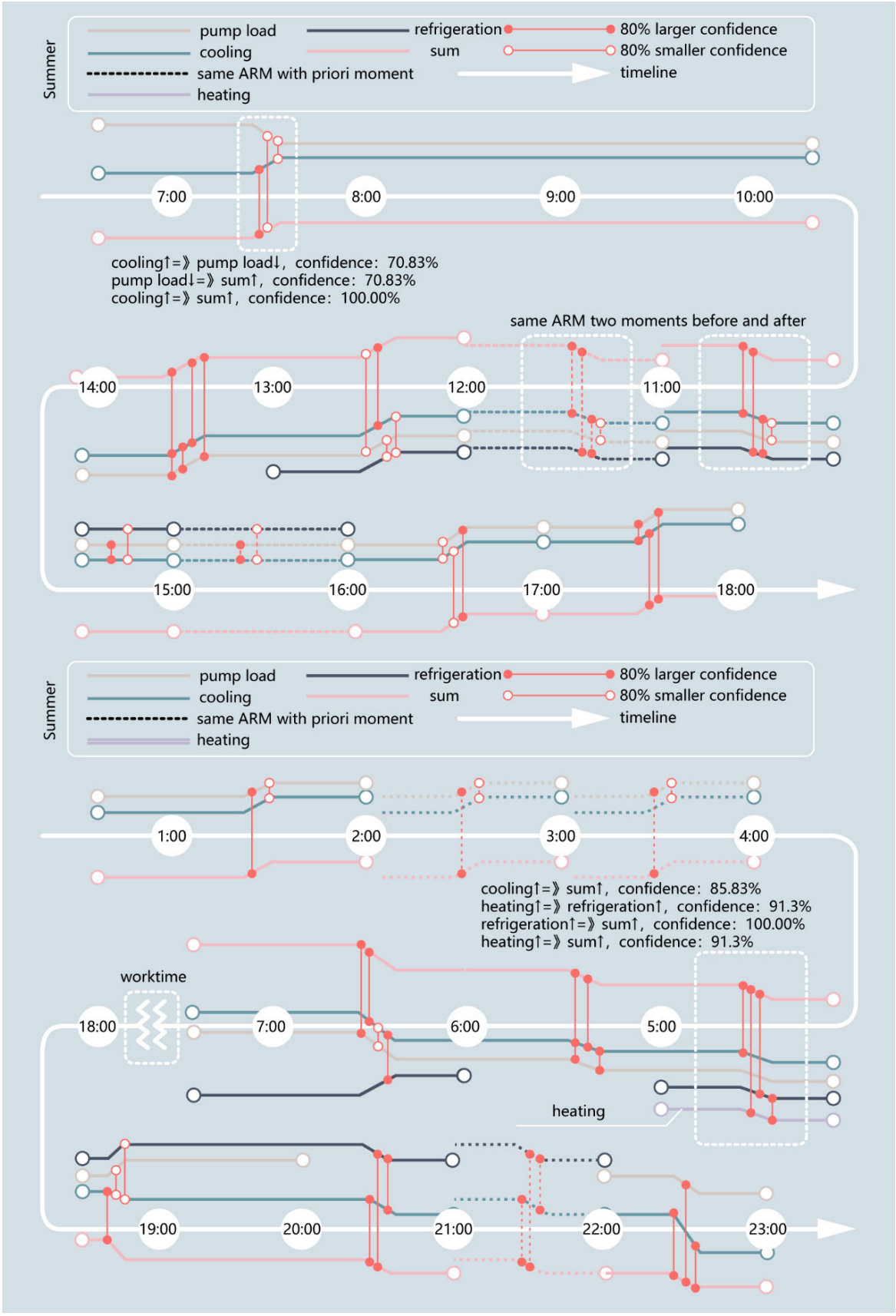


Figure 13. Visualization of ARM rules.

3.2.2. Winter

Table 15 and Figure 14 present the ARM results. It can be seen that cooling basically shows a positive correlation with total energy fluctuation during each period. While in terms of the pump load condition, in the period of 6:00–7:00 and 22:00–23:00, it promotes a fluctuation in the total energy to some extent. Refrigeration energy consumption mainly has a positive influence on the cooling and total energy at lunch time and at night. It can be seen that, in summer, cooling power adoption is the most efficient method contributing to total energy consumption. Moreover, pump load is not the major factor in energy consumption. For refrigeration, it shows a positive relationship with cooling and total energy consumption, as the stream of people dining can also affect the interior temperature. For example, at breakfast and dinner time, cooling shift has the same trend as refrigeration, as shown in Table 15. According to all the discussions above, it is recommended to focus on the energy fluctuation of cooling and refrigeration to reduce energy consumption because of their high correlations. Reducing the frequency of opening and closing refrigerators, such as taking some frequently used cooking ingredients out in advance at a dining time, can also smoothen the fluctuation curve.

Table 15. Rules mined in winter.

Period	Rules	Confidence
4:00–5:00	Heating↑=》 Refrigeration↑	91.66%
	Heating↑=》 Total energy↑	100.00%
	Refrigeration↑=》 Total energy↑	100.00%
	Heating↑, Refrigeration↑=》 Total energy↑	100.00%
5:00–6:00	Heating↓=》 Pump load↑	100.00%
	Heating↓=》 Total energy↑	100.00%
	Pump load↑=》 Total energy↑	100.00%
6:00–7:00	Cooling↑=》 Pump load↑	64.16%
	Cooling↑, Pump load↑=》 Heating-	66.66%
	Cooling↑=》 Total energy↑	64.16%
	Pump load↑=》 Total energy↑	79.16%
7:00–8:00	Pump load↑=》 Heating↓	70.83%
	Cooling↑=》 Total energy↑	100.00%
	Refrigeration-=》 Total energy↑	100.00%
8:00–9:00	Pump load↓=》 Total energy↓	66.66%
	Heating-=》 Total energy↓	100.00%
	Cooling↓=》 Total energy↓	66.66%
9:00–10:00	Cooling↓=》 Pump load↓	62.5%
	Heating-=》 Refrigeration-	100.00%
	Cooling↓, Pump load↓=》 Total energy↓	87.5%
10:00–13:00	Cooling↑=》 Pump load↑	79.16%
	Cooling↑=》 Refrigeration↑	100.00%
	Cooling↑=》 Heating-	100.00%
	Cooling↑, Pump load↑, Refrigeration↑=》 Total energy↑	100.00%

Table 15. *Cont.*

Period	Rules	Confidence
13:00–14:00	Cooling↓=》 Pump load↓	87.50%
	Cooling↓=》 Total energy↓	95.83%
	Cooling↓=》 Refrigeration-	100.00%
	Cooling↓, Pump load↓=》 Total energy↓	95.83%
14:00–16:00	Cooling-=》 Pump load-	66.66%
	Pump load-=》 Total energy-	100.00%
16:00–17:00	Cooling↑=》 Total energy↑	70.83%
	Cooling↑=》 Pump load↑	75.00%
	Pump load↑=》 Total energy↑	87.50%
	Cooling↑=》 Heating-	75.00%
17:00–18:00	Cooling↑=》 Pump load↑	100.00%
	Cooling↑=》 Total energy↑	100.00%
	Pump load↑=》 Total energy↑	100.00%
18:00–20:00	Cooling↓=》 Total energy↓	95.83%
	Cooling↓=》 Heating-	95.83%
	Cooling↓=》 Refrigeration↑	70.83%
	Cooling↓, Pump load↑, Refrigeration↑=》 Total energy↓	70.83%
20:00–22:00	Pump load↓=》 Refrigeration↓	95.83%
	Pump load↓=》 Total energy↓	100.00%
	Refrigeration↓=》 Total energy↓	100.00%
	Pump load↓, Refrigeration↓=》 Heating-	95.83%
	Pump load↓, Refrigeration↓=》 Total energy↓	100.00%
22:00–23:00	Pump load↓=》 Total energy↓	100.00%
	Heating↑=》 Pump load↓	87.5%
	Pump load↓=》 Refrigeration↑	100.00%
	Pump load↓, Refrigeration↑=》 Total energy↓	100.00%
23:00–24:00	Pump load↓=》 Refrigeration↓	79.16%
	Heating↑=》 Refrigeration↓	87.50%
	Refrigeration↓=》 Total energy↓	100.00%
	Heating↑, Refrigeration↓, Pump load↓=》 Total energy↓	87.50%

(↑: energy increase, ↓: energy decrease, -: same with previous energy condition, =》: lead to).

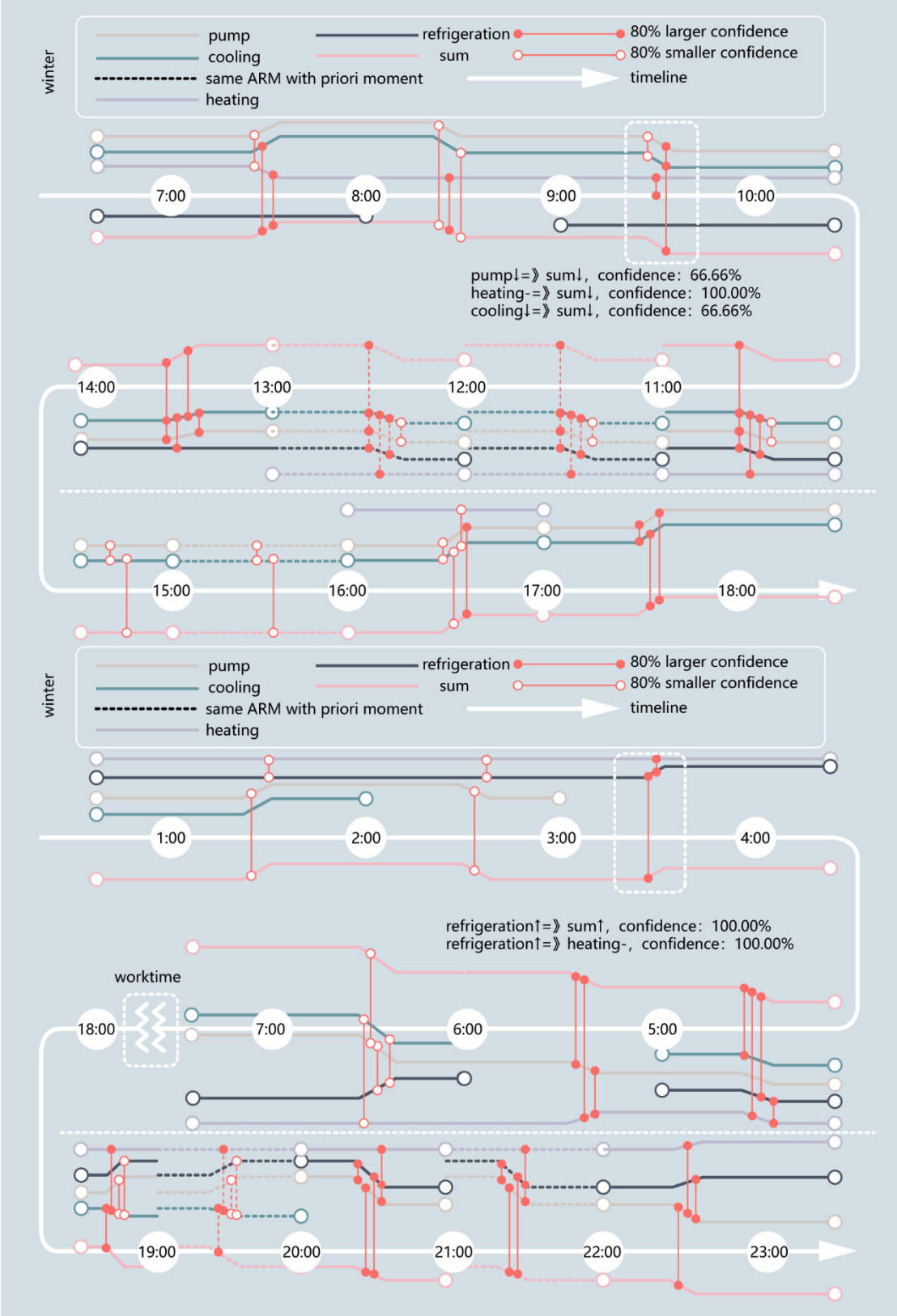


Figure 14. ARM consequences in winter.

4. Discussion

The presented solution can help pick up time-series energy consumption features. The discovered regulated characteristics illustrate the building running profile and energy

usage issue. Based on these discovered issues, some energy-saving strategies could be proposed to target more specific architecture. The studied workflow improves the energy profile recognition efficiency and fully harnesses time-series monitoring data that have not been utilized in the past. In this case, related researchers could make more scientific decisions regarding energy efficiency.

4.1. Modes Characteristic

In line with the aforementioned clustering and ARM studies, it is noticeable that several time-series curve patterns have been discovered, as shown in Table 16. In the process of research, clusters are responsible for finding patterns, while the ARM algorithm infers reasons for this special mode of curves and then provides specific solutions. In accordance with this table, four type patterns of Invert U, M, Invert V, and Multiple M are summarized. While in terms of solution, intensity refers to a decrease in the energy consumption value during a certain period. Peak shift is meant to cut the peak and fill the valley, reaching an energy balance.

Table 16. Summary of different energy consumption clustered results.

Energy	Period	Pattern	Solution
Cooling	Summer	Invert U	Intensity
	Winter	M	Peak shift
	Transition	Invert U	Intensity
Heating	Summer + transition	Invert V	Intensity
	Winter + transition	M	Intensity
Pump load	Holiday	M	Peak shift
	Summer weekday	M	Peak shift
	Winter weekday	M	Peak shift
Refrigeration	Weekday	Invert V	Peak shift
	Special day	M + Invert V	Peak shift
Total energy	Summer + transition	Multiple M	Peak shift
	Winter + transition	Multiple M	Peak shift

With respect to the found shape, inverted U form generation can be attributed to the outside steady circumstance. Due to there being no significant fluctuation situation on the curve, a major diminished energy method focuses on decreasing its intensity, such as a passive design plan. M form displays two peaks due to the local flow of people. The peak position chiefly appears during the dining period. Facing this feature, reducing the peak value and filling up the valley could improve energy efficiency. It is notable that since heating is produced by natural gas instead of electricity, a peak shift does not result in it being lessened. An inverted V shape presents a short-term pulse form induced by some special demands, such as a refrigeration machine. Two aspects of solution with intensity and peak are appropriate for this energy-saving pattern. In regard to the multiple M form, a peak shift method could be used to develop an energy-saving target based on many peak features.

4.2. Comparison

As a special building type, restaurant architecture has its own operational energy consumption pattern. It is indispensable to compare it with other building types when investigating the running characteristics. Liu et al. [51] studied the running characteristics of an office using clustering and ARM solutions. Figure 15 compares the clustered modes found from the restaurant and office building. The result shows that office buildings dis-

play more clustered categories than restaurant buildings due to the higher complexity of the architectural equipment system. Considering the building function, office construction indicates strong work time relation traits, while a restaurant correlates to meal times much more strongly. For example, whether it is a weekday and weekend has an intense influence on the operation mode of office building energy; nevertheless, restaurant power varies only a little during this period.

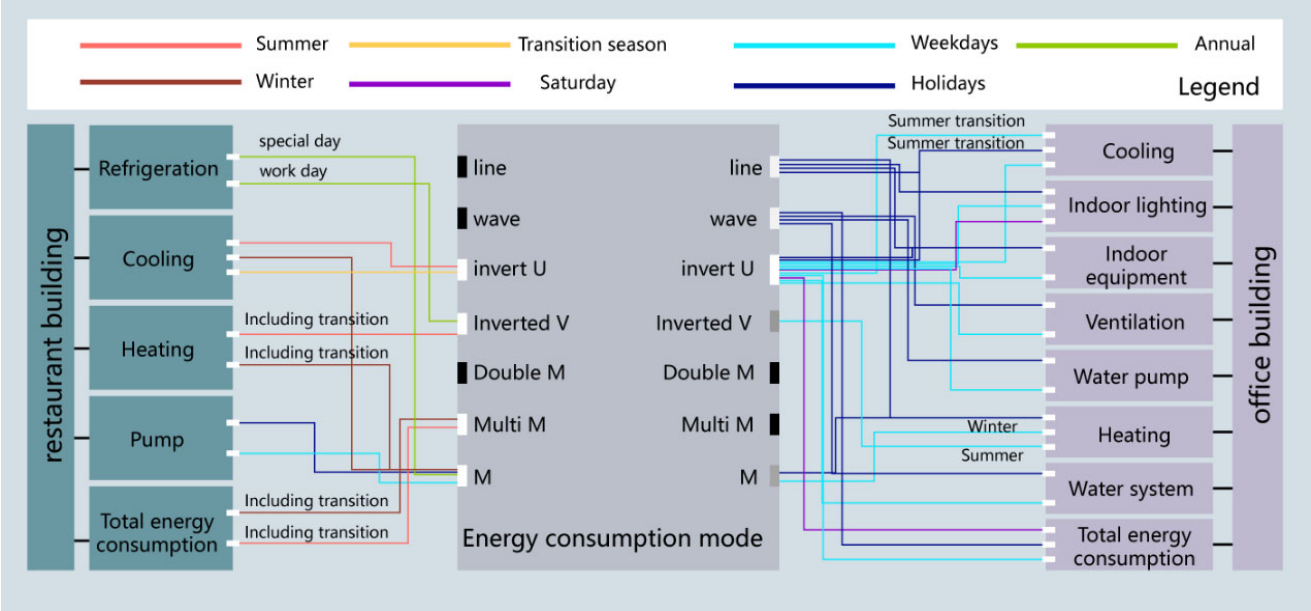


Figure 15. Comparison of restaurant and office building energy-saving patterns.

In addition, these two time-series curves seem to complement the characteristics of each other during a particular moment. This is mainly because the operating hours of the office building and the catering one are staggered in line with its architectural function. Figure 16 indicates this complementation phenomenon. The same roman numeral represents the energy condition could be complemented. It can be seen that decreasing the energy level during the lunch break and leave time of an office building can efficiently neutralize the restaurant energy peak at the same period. In this case, the energy efficiency of these two kinds of buildings can be enhanced. Therefore, canteen and office buildings are recommended to be built into an electrical power system.

Chaobo et al. found useful association rules related to component/sensor faults, control strategies, abnormal operation patterns, and normal operation patterns for chiller plants [38]. This research only focused on one building equipment with historical data. The discovered associated rules generated by the FP-growth solution resemble with this research achieved by the Apriori algorithm. But this paper investigates the abnormal condition of the building in terms of time-series, which differs from the outlier measured point data of the chiller plant. In this sense, it provides a fully innovative view of searching building energy control strategies. In addition, Xue et al. [41] analyzed household occupancy patterns and assessed the relationship between these patterns and socio-demographic characteristics using statistical methods. It adopts the common data collection approach of a questionnaire survey mode which has a lower level of objectivity in terms of science compared with measured data. The simulation or measured information can reflect people's behavior objectively. In terms of workflow, Nisa et al. [26] developed a similar target with this investigation. Nevertheless, the investigator built the simulation model via a neural network and clustered the final simulated result of total energy consumption. Focusing on the disadvantages of previous research, this paper grouped time-series building data to compensate for issues in past research papers regard-

ing the data dimension. Overall, this research makes up for the shortcomings of previous studies with respect to objective scope, study methodology, and energy usage patterns.

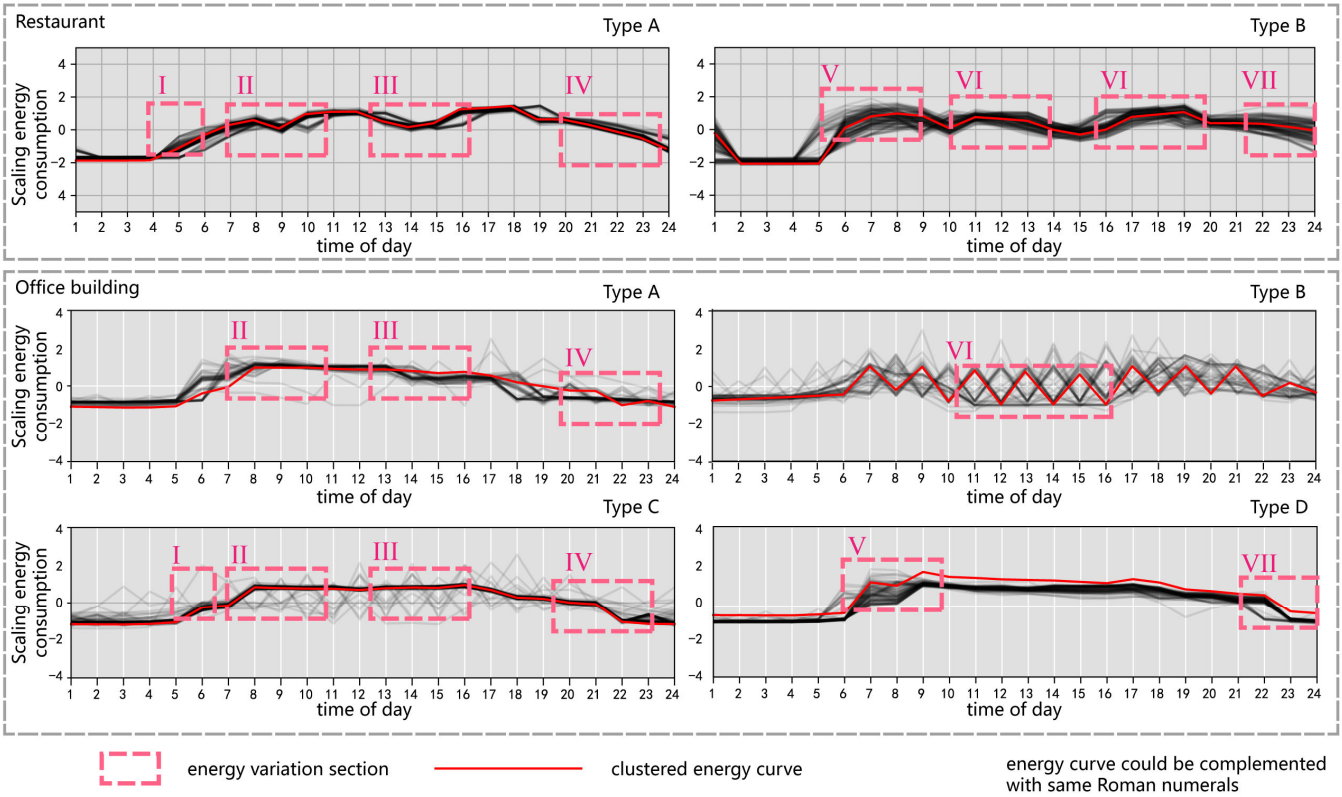


Figure 16. Complementation phenomenon for restaurant and office buildings.

4.3. Limitations

In this study, energy consumption simulation data are considered where they clearly reflect the corresponding building operation features. However, actual power load data are more random than the simulation database, thereby achieving energy diagnosis. A future study needs to be performed based on a field measurement building running dataset to discover the specific unreasonable pattern.

On the other hand, this research considers the subentry energy variation condition, neglecting the area impact. Actually, different zones in buildings also hide several associated energy expenditure relationships. In this case, clustering and ARM investigations are also required to be conducted on disparate building areas to improve energy efficiency.

In terms of time-series data resolution, this study mainly focuses on hourly accuracy. Moreover, under different time dimensions, it is possible to make multiple discoveries. Thus, future work will concentrate on comparing various energy data clustering results at different time resolutions.

5. Conclusions

This paper aims to study restaurant energy usage patterns by means of the data mining approach. Official building models provide simulation data to make up the dataset for investigation. Time-series cluster and ARM algorithms are adopted to analyze energy consumption curves. Some conclusions have been generated, as follows:

A combination of time-series clustering and ARM algorithm work flows could successfully discover the building operation pattern. Focusing on these regulations, some more scientific energy efficiency suggestions are proposed.

In the process of investigation, the cluster method mines various energy consumption patterns over time, which reflect the building's characteristics and problems, while the

association rule-mined algorithm discovers relationships between different energy types under the same moments, which deduces the reason behind the phenomenon.

Restaurant time-series energy consumption curves could be clustered into four types: Invert U, M, Invert V, and Multiple M. Each mode has its own variation characteristics. Two aspects of intensity and peak shift are proposed for achieving energy savings, focusing on different curve modes.

In terms of the subentry energy type, cooling and refrigeration are the two most influential factors for total energy. Outside circumstances and people's flow are the two significant factors influencing the energy usage pattern. With respect to the seasonal element, in summer, outside temperature primarily affects interior energy consumption, while in winter, human traffic becomes the major influential factor.

Regarding canteen architecture, the key to energy conservation should be how to fill up the power curve valley at off-peak times, such as by changing the refrigerator's mode. In addition, eliminating unnecessary energy usage is another efficient way, like closing heating equipment at peak times when there are a lot of people in the winter.

Final research results establish a new workflow for relative investigations and reveal several common problems of restaurant operation that could provide references for related energy policy determination. This study analyzes simulated energy consumption data that accurately represent building operational features, but future research should focus on field measurements to identify specific irregular patterns in actual power load data and the building area impact performance. In terms of the time issue, future work should also be conducted to compare various energy data clustering results under different time resolutions.

Author Contributions: Conceptualization, X.L. and H.Z.; methodology, X.L.; software, X.L.; validation, S.Z. and W.C.; formal analysis, R.W.; investigation, J.H.; resources, Z.L.; data curation, J.Y.; writing—original draft preparation, X.L.; writing—review and editing, X.L.; visualization, X.W., J.W.; supervision, H.Z.; project administration, H.Z.; funding acquisition, H.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China with grant number of 52078265 and the Ministry of Housing and Urban–Rural Development with grant number of R20220430.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chen, X.; Yang, H. Combined thermal and daylight analysis of a typical public rental housing development to fulfil green building guidance in Hong Kong. *Energy Build.* **2015**, *108*, 420–432. [CrossRef]
2. International Energy Agency. *2019 Global Status Report for Buildings and Construction*; United Nations Environment Program; International Energy Agency: Paris, France, 2019.
3. UN Environment Programme. *2020 Global Status Report for Buildings and Construction*; UN Environment Programme: Nairobi, Kenya, 2020.
4. Institute for Market Transformation. *Map: U.S. City, County, and State Policies for Existing Buildings: Benchmarking, Transparency and Beyond*; Institute for Market Transformation: Washington, DC, USA, 2021.
5. Chicco, G. Overview and performance assessment of the clustering methods for electrical load pattern grouping. *Energy* **2012**, *42*, 68–80. [CrossRef]
6. Klemp, S.; Abida, A.; Richter, P. A method and analysis of predicting building material U-value ranges through geometrical pattern clustering. *J. Build. Eng.* **2021**, *44*, 103243. [CrossRef]
7. Nikolaou, T.G.; Kolokotsa, D.S.; Stavrakakis, G.S.; Skias, I.D. On the Application of Clustering Techniques for Office Buildings' Energy and Thermal Comfort Classification. *IEEE Trans. Smart Grid* **2012**, *3*, 2196–2210. [CrossRef]
8. Miller, C.; Nagy, Z.; Schlueter, A. Automated daily pattern filtering of measured building performance data. *Autom. Constr.* **2015**, *49*, 1–17. [CrossRef]
9. Pieri, S.P.; Tzouvadakis, I.; Santamouris, M. Identifying energy consumption patterns in the Attica hotel sector using cluster analysis techniques with the aim of reducing hotels' CO_2 footprint. *Energy Build.* **2015**, *94*, 252–262. [CrossRef]

10. Farrou, I.; Kolokotroni, M.; Santamouris, M. A method for energy classification of hotels: A case-study of Greece. *Energy Build.* **2012**, *55*, 553–562. [CrossRef]
11. Gao, X.; Malkawi, A. A new methodology for building energy performance benchmarking: An approach based on intelligent clustering algorithm. *Energy Build.* **2014**, *84*, 607–616. [CrossRef]
12. De Jaeger, I.; Reynders, G.; Callebaut, C.; Saelens, D. A building clustering approach for urban energy simulations. *Energy Build.* **2020**, *208*, 109671. [CrossRef]
13. Andrews, A.; Jain, R.K. Beyond Energy Efficiency: A clustering approach to embed demand flexibility into building energy benchmarking. *Appl. Energy* **2022**, *327*, 119989. [CrossRef]
14. Walsh, A.; Cóstola, D.; Labaki, L.C. Performance-based climatic zoning method for building energy efficiency applications using cluster analysis. *Energy* **2022**, *255*, 124477. [CrossRef]
15. Lavin, A.; Klabjan, D. Clustering time-series energy data from smart meters. *Energy Effic.* **2015**, *8*, 681–689. [CrossRef]
16. Wu, S.; Clements-Croome, D. Understanding the indoor environment through mining sensory data—A case study. *Energy Build.* **2007**, *39*, 1183–1191. [CrossRef]
17. Kwac, J.; Flora, J.; Rajagopal, R. Household Energy Consumption Segmentation Using Hourly Data. *IEEE Trans. Smart Grid* **2014**, *5*, 420–430. [CrossRef]
18. Iglesias, F.; Kastner, W. Analysis of Similarity Measures in Times Series Clustering for the Discovery of Building Energy Patterns. *Energies* **2013**, *6*, 579–597. [CrossRef]
19. Zhang, J.Q. Sampling for building energy consumption with fuzzy theory. *Energy Build.* **2017**, *156*, 78–84. [CrossRef]
20. Santamouris, M.; Mihalakakou, G.; Patargias, P.; Gaitani, N.; Sfakianaki, K.; Papaglastra, M.; Pavlou, C.; Doukas, P.; Primikiri, E.; Geros, V.; et al. Using intelligent clustering techniques to classify the energy performance of school buildings. *Energy Build.* **2007**, *39*, 45–51. [CrossRef]
21. Chicco, G.; Ilie, I.-S. Support Vector Clustering of Electrical Load Pattern Data. *IEEE Trans. Power Syst.* **2009**, *24*, 1619–1628. [CrossRef]
22. Petcharat, S.; Chungpaibulpatana, S.; Rakkwamsuk, P. Assessment of potential energy saving using cluster analysis: A case study of lighting systems in buildings. *Energy Build.* **2012**, *52*, 145–152. [CrossRef]
23. Liu, J.; Wang, J.; Li, G.; Chen, H.; Shen, L.; Xing, L. Evaluation of the energy performance of variable refrigerant flow systems using dynamic energy benchmarks based on data mining techniques. *Appl. Energy* **2017**, *208*, 522–539. [CrossRef]
24. Culaba, A.B.; Del Rosario, A.J.R.; Ubando, A.T.; Chang, J. Machine learning-based energy consumption clustering and forecasting for mixed-use buildings. *Int. J. Energy Res.* **2020**, *44*, 9659–9673. [CrossRef]
25. Xu, J.; Kang, X.; Chen, Z.; Yan, D.; Guo, S.; Jin, Y.; Hao, T.; Jia, R. Clustering-based probability distribution model for monthly residential building electricity consumption analysis. *Build. Simul.* **2021**, *14*, 149–164. [CrossRef]
26. Nisa, E.C.; Kuan, Y.D.; Lai, C.C. Chiller Optimization Using Data Mining Based on Prediction Model, Clustering and Association Rule Mining. *Energies* **2021**, *14*, 6494. [CrossRef]
27. Han, J.W.; Kamber, M.; Pei, J. *Data Mining: Concepts and Techniques*; Morgan Kaufman Publishers: Burlington, MA, USA, 2012.
28. Yu, Z.J.; Fung, B.C.; Haghighat, F.; Yoshino, H.; Morofsky, E. A systematic procedure to study the influence of occupant behavior on building energy consumption. *Energy Build.* **2011**, *43*, 1409–1417. [CrossRef]
29. Yu, Z.J.; Haghighat, F.; Fung, B.C.; Zhou, L. A novel methodology for knowledge discovery through mining associations between building operational data. *Energy Build.* **2012**, *47*, 430–440. [CrossRef]
30. D'Oca, S.; Hong, T. A data-mining approach to discover patterns of window opening and closing behavior in offices. *Build. Environ.* **2014**, *82*, 726–739. [CrossRef]
31. Cabrera, D.F.M.; Zareipour, H. Data association mining for identifying lighting energy waste patterns in educational institutes. *Energy Build.* **2013**, *62*, 210–216. [CrossRef]
32. Wang, Y.; Shao, L. Understanding occupancy pattern and improving building energy efficiency through Wi-Fi based indoor positioning. *Build. Environ.* **2017**, *114*, 106–117. [CrossRef]
33. Xiao, F.; Fan, C. Data mining in building automation system for improving building operational performance. *Energy Build.* **2014**, *75*, 109–118. [CrossRef]
34. Fan, C.; Xiao, F.; Yan, C. A framework for knowledge discovery in massive building automation data and its application in building diagnostics. *Autom. Constr.* **2015**, *50*, 81–90. [CrossRef]
35. Li, G.; Hu, Y.; Chen, H.; Li, H.; Hu, M.; Guo, Y.; Liu, J.; Sun, S.; Sun, M. Data partitioning and association mining for identifying VRF energy consumption patterns under various part loads and refrigerant charge conditions. *Appl. Energy* **2017**, *185*, 846–861. [CrossRef]
36. Wang, F.; Li, K.; Duić, N.; Mi, Z.; Hodge, B.-M.; Shafie-Khah, M.; Catalão, J.P. Association rule mining based quantitative analysis approach of household characteristics impacts on residential electricity consumption patterns. *Energy Convers. Manag.* **2018**, *171*, 839–854. [CrossRef]
37. Xue, P.; Zhou, Z.; Fang, X.; Chen, X.; Liu, L.; Liu, Y.; Liu, J. Fault detection and operation optimization in district heating substations based on data mining techniques. *Appl. Energy* **2017**, *205*, 926–940. [CrossRef]
38. Zhang, C.; Zhao, Y.; Li, T.; Zhang, X.; Luo, J. A comprehensive investigation of knowledge discovered from historical operational data of a typical building energy system. *J. Build. Eng.* **2021**, *42*, 102502. [CrossRef]

39. Gong, Q.; Liu, X.; Zeng, Y.; Han, S. An energy efficiency solution based on time series data mining algorithm on elementary school building. *Int. J. Low-Carbon Technol.* **2022**, *17*, 356–372. [CrossRef]
40. Gong, Q.; Zeng, Y.; Adu, E.; Han, S.; Zhang, S.; Cui, W.; Sun, H.; Liu, X. An energy efficiency solution based on time series data mining algorithm-a case study of small hotel. *Int. J. Low Carbon Technol.* **2022**, *17*, 1406–1419. [CrossRef]
41. Liu, X.; Hu, S.; Yan, D. A statistical quantitative analysis of the correlations between socio-demographic characteristics and household occupancy patterns in residential buildings in China. *Energy Build.* **2023**, *284*, 112842. [CrossRef]
42. Qiu, S.; Feng, F.; Li, Z.; Yang, G.; Xu, P.; Li, Z. Data mining based framework to identify rule based operation strategies for buildings with power metering system. *Build. Simul.* **2019**, *12*, 195–205. [CrossRef]
43. Fan, C.; Xiao, F. Assessment of Building Operational Performance Using Data Mining Techniques: A Case Study. *Energy Procedia* **2017**, *111*, 1070–1078. [CrossRef]
44. Fan, C.; Xiao, F.; Madsen, H.; Wang, D. Temporal knowledge discovery in big BAS data for building energy management. *Energy Build.* **2015**, *109*, 75–89. [CrossRef]
45. Fan, C.; Sun, Y.; Shan, K.; Xiao, F.; Wang, J. Discovering gradual patterns in building operations for improving building energy efficiency. *Appl. Energy* **2018**, *224*, 116–123. [CrossRef]
46. Fan, C.; Song, M.; Xiao, F.; Xue, X. Discovering Complex Knowledge in Massive Building Operational Data Using Graph Mining for Building Energy Management. *Energy Procedia* **2019**, *158*, 2481–2487. [CrossRef]
47. Fan, C.; Xiao, F.; Song, M.; Wang, J. A graph mining-based methodology for discovering and visualizing high-level knowledge for building energy management. *Appl. Energy* **2019**, *251*, 113395. [CrossRef]
48. Gunay, H.B.; Shen, W.; Yang, C. Text-mining building maintenance work orders for component fault frequency. *Build. Res. Inf.* **2018**, *47*, 518–533. [CrossRef]
49. Zhang, C.; Zhao, Y.; Zhang, X. An improved association rule mining-based method for discovering abnormal operation patterns of HVAC systems. *Energy Procedia* **2019**, *158*, 2701–2706. [CrossRef]
50. Zhang, C.; Xue, X.; Zhao, Y.; Zhang, X.; Li, T. An improved association rule mining-based method for revealing operational problems of building heating, ventilation and air conditioning (HVAC) systems. *Appl. Energy* **2019**, *253*, 113492. [CrossRef]
51. Liu, X.; Sun, H.; Han, S.; Han, S.; Niu, S.; Qin, W.; Sun, P.; Song, D. A data mining research on office building energy pattern based on time-series energy consumption data. *Energy Build.* **2022**, *259*, 111888. [CrossRef]
52. Peng, Y.; Lin, J.-R.; Zhang, J.-P.; Hu, Z.-Z. A hybrid data mining approach on BIM-based building operation and maintenance. *Build. Environ.* **2017**, *126*, 483–495. [CrossRef]
53. Noh, B.; Son, J.; Park, H.; Chang, S. In-Depth Analysis of Energy Efficiency Related Factors in Commercial Buildings Using Data Cube and Association Rule Mining. *Sustainability* **2017**, *9*, 2119. [CrossRef]
54. Ashouri, M.; Haghighat, F.; Fung, B.C.; Lazrak, A.; Yoshino, H. Development of building energy saving advisory: A data mining approach. *Energy Build.* **2018**, *172*, 139–151. [CrossRef]
55. Akil, M.; Tittelein, P.; Defer, D.; Suard, F. Statistical indicator for the detection of anomalies in gas, electricity and water consumption: Application of smart monitoring for educational buildings. *Energy Build.* **2019**, *199*, 512–522. [CrossRef]
56. Garcia, L.C.; Kamsu-Foguem, B. BIM-oriented data mining for thermal performance of prefabricated buildings. *Ecol. Inform.* **2019**, *51*, 61–72. [CrossRef]
57. Sun, C.; Zhang, R.; Sharples, S.; Han, Y.; Zhang, H. Thermal comfort, occupant control behaviour and performance gap—A study of office buildings in north-east China using data mining. *Build. Environ.* **2019**, *149*, 305–321. [CrossRef]
58. Zhang, C.; Zhao, Y.; Li, T.; Zhang, X. A post mining method for extracting value from massive amounts of building operation data. *Energy Build.* **2020**, *223*, 110096. [CrossRef]
59. Liu, J.; Shi, D.; Li, G.; Xie, Y.; Li, K.; Liu, B.; Ru, Z. Data-driven and association rule mining-based fault diagnosis and action mechanism analysis for building chillers. *Energy Build.* **2020**, *216*, 109957. [CrossRef]
60. Khani, S.M.; Haghighat, F.; Panchabikesan, K.; Ashouri, M. Extracting energy-related knowledge from mining occupants' behavioral data in residential buildings. *J. Build. Eng.* **2021**, *39*, 102319. [CrossRef]
61. Wu, D.; Guo, M. Application of Data Mining in Traditional Benchmark Evaluation Model for Buildings Energy Consumption. *Sci. Program.* **2021**, *2021*, 8610050. [CrossRef]
62. Qiu, Z.; Wang, J.; Yu, B.; Liao, L.; Li, J. Identification of passive solar design determinants in office building envelopes in hot and humid climates using data mining techniques. *Build. Environ.* **2021**, *196*, 107566. [CrossRef]
63. Juan, Y.-K.; Lee, P.-H. Applying data mining techniques to explore technology adoptions, grades and costs of green building projects. *J. Build. Eng.* **2021**, *45*, 103669. [CrossRef]
64. Geng, Y.; Ji, W.; Xie, Y.; Lin, B.; Zhuang, W. A sub-sequence clustering method for identifying daily indoor environmental patterns from massive time-series data. *Autom. Constr.* **2022**, *139*, 104303. [CrossRef]
65. Shan, X.; Deng, Q.; Tang, Z.; Wu, Z.; Wang, W. An integrated data mining-based approach to identify key building and urban features of different energy usage levels. *Sustain. Cities Soc.* **2022**, *77*, 103576. [CrossRef]
66. Zhou, H.; Tian, X.; Yu, J.; Zhao, Y.; Lin, B.; Chang, C. Identifying buildings with rising electricity-consumption and those with high energy-saving potential for government's management by data mining approaches. *Energy Sustain. Dev.* **2022**, *66*, 54–68. [CrossRef]
67. Lei, L.; Wu, B.; Fang, X.; Chen, L.; Wu, H.; Liu, W. A dynamic anomaly detection method of building energy consumption based on data mining technology. *Energy* **2023**, *263*, 125575. [CrossRef]

68. Paparrizos, J.; Gravano, L. k-Shape: Efficient and Accurate Clustering of Time Series. In Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, Melbourne, Australia, 31 May–4 June 2015; SIGMOD Record. pp. 69–76. [CrossRef]
69. Gregory, P.-S. *Discovery, Analysis, and Presentation of Strong Rules, Knowledge Discovery in Databases*; AAAI/MIT Press: Cambridge, MA, USA, 1991.
70. Agrawal, R.; Srikant, R. Fast algorithms for mining association rules. In Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, Santiago, Chile, 12–15 September 1994; pp. 487–499.
71. Deru, M.; Field, K.; Studer, D.; Benne, K.; Griffith, B.; Torcellini, P.; Liu, B.; Halverson, M.; Winiarski, D.; Rosenberg, M.; et al. *U.S. Department of Energy Commercial Reference Building Models of the National Building Stock*; National Renewable Energy Laboratory: Golden, CO, USA, 2011. [CrossRef]
72. ASHRAE Standing Standard Project Committee 90.1. *ASHRAE Standard 90.1-2004*; Energy Standard for Buildings Except Low-Rise Residential Buildings. ASHRAE Standing Standard Project Committee 90.1: Atlanta, GA, USA, 2004.
73. Pacific Northwest National Laboratory. *Technical Support Document: Energy Efficiency Program for Commercial and Industrial Equipment: Advanced Notice of Proposed Rulemaking for Commercial Unitary Air Conditioners and Heat Pumps*; Pacific Northwest National Laboratory: Washington, DC, USA, 2004.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.