Energy & Buildings 180 (2018) 135–145

Contents lists available at ScienceDirect

Energy & Buildings

journal homepage: www.elsevier.com/locate/enbuild

Extracting typical occupancy data of different buildings from mobile positioning data

Gu Jiefan[a], Xu Peng[a,*], Pang Zhihong[b], Chen Yongbao[a], Ji Ying[c], Chen Zhe[a]

[a] College of Mechanical and Energy Engineering, Tongji University, Shanghai 201804, China
[b] Department of Mechanical Engineering, University of Alabama, Tuscaloosa, US
[c] Beijing Key Laboratory of Green Built Environment and Energy Efficient Technology, Beijing University of Technology, Beijing 100124, China

ARTICLE INFO

Article history:
Received 28 December 2017
Revised 23 August 2018
Accepted 1 September 2018
Available online 19 September 2018

Keywords:
Typical occupancy data
Mobile device
Building energy simulation

ABSTRACT

Occupancy is one of the main factors affecting building energy consumption. The occupancy data, which refer to the occupancy number in this paper, has been widely used in the building simulation field. However, due to the stochastic nature of occupant behavior, it is hard to predict and measure how many people stay in a given building. The rapid development of mobile Internet technology provides an efficient and convenient option for occupancy detection. This paper proposes a concept of typical occupancy data (TOD), which are extracted from real-time occupancy data collected by mobile devices. *K*-means algorithm is employed to generate the TOD data through cluster analysis. An energy performance model of an office building is used as a case study to demonstrate the effectiveness of the TOD data.
© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Occupancy is one of the main factors affecting building energy consumption and it is also a main source of uncertainty in building energy simulation and prediction [1,2]. Due to the stochastic nature and privacy issue of occupancy, it is not easy to characterize occupancy in the buildings. Both the presence of occupants and their interactions with building are often simplified by some predefined schedules in current simulation programs [3], which is mainly acquired from building standards or design manuals of different countries and organizations, as shown in Fig. 1. According to the ASHRAE Standard 90.1-2013, schedules shall be determined by the designer and approved by the rating authority [4]. The user's manual provide occupancy schedules for different building types from ASHRAE standard 90.1-1989 when actual schedule is unknown [5], where occupancy factors vary on weekdays, Saturdays and Sundays. Supported by the U.S. Department of Energy (DOE) Building Energy Codes Program, Pacific Northwest National Laboratory (PNNL) made enhanced prototype building models for 16 commercial building types in 17 American climate locations with modified schedules [6]. However, actual occupancy schedules are more complicated than those in the model. Not all the schedules can be simply divided into weekdays and weekends, and not all the occupancy schedules are same for the same building type. Thus, discrepancies may exist between standard schedules and actual occupancy and then may influence the accuracy of simulation results [7,8]. How to obtain reliable occupancy data, which only refer to the count of occupancy in this paper, for building energy simulation remains as a challenged problem.

Several studies have measured actual occupancy data on sites. These measurement methods could be grouped into questionnaire method, sensor direct detection method [9–14], indirect detection method [8,15–17] and mobile device detection method [18–23]. Using questionnaires to survey real-time occupancy in the buildings saves the cost of installing sensors but needs a large work so that the data is statistic significant. Thus, questionnaires are not suitable for buildings with a large amount of occupancy and big variance, such as transportation centers. A large number of studies measured the real-time occupancy directly by installing sensors indoors, including PIR sensors, pressure sensors, imagine sensors and so on. Data collected from sensors are more accurate but require implementation of many sensors. Some sensors are expensive, such as imagine sensors, and the implementers need to be concerned with the privacy of occupants. With the indirect measure method, real-time occupancy data is obtained indirectly by detecting other environmental parameters such as CO_2 concentration, on/off lighting status, electricity consumption data, etc. These measured data are not as reliable and accurate as direct measurements. However, with the development of mobile Internet technology, it is convenient and credible to use mobile devices such as mobile phones to track human motion and detect occupancy in a specified area. Compared to the former two methods, the mobile

* Corresponding author.
E-mail address: xupengwb@gmail.com (X. Peng).

https://doi.org/10.1016/j.enbuild.2018.09.002
0378-7788/© 2018 Elsevier B.V. All rights reserved.

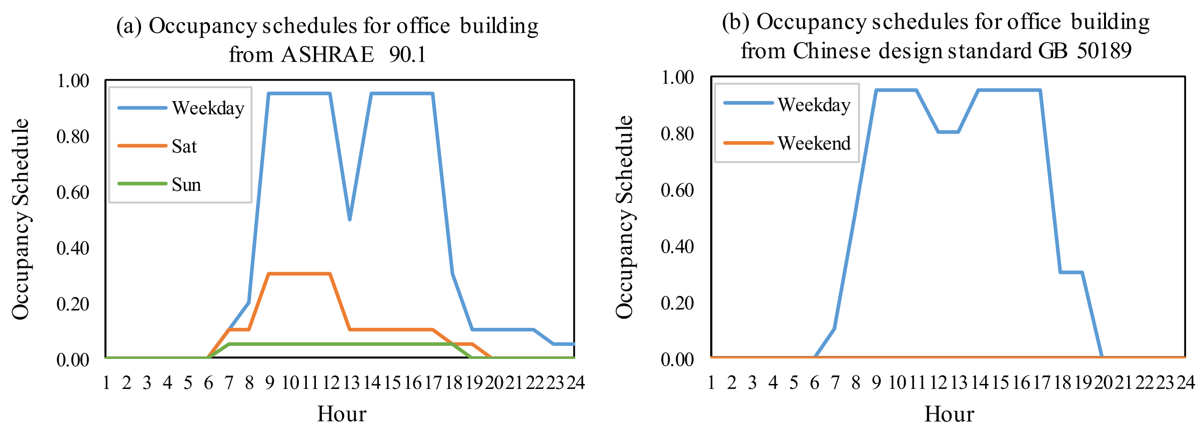


Fig. 1. Occupancy schedules for office building.

device detection method does not require any additional cost to install and maintain sensors, and the location data can be obtained in real time.

Although actual measured occupancy data has very high accuracy, it has two limitations in practical application. On one hand, it is labor intensive for installation, maintenance and computation. On the other hand, the actual measured data can be only collected from existing buildings. For buildings in the design stage, it is impossible to measure occupancy data because the building does not exist yet. However, for the majority of building types, occupancy data follow some patterns. Accurate, real-time occupancy data in building simulations and control systems may not be necessary. Instead, a series of typical occupancy data can be used to alleviate the simulation and computation workload on the premise of accuracy.

Based on the on-site measured occupancy data, several studies [24–28] used stochastic models including hidden Markov Model and Semi-Markov Model to capture the patterns of occupant presence and behaviors and make prediction. This kind of method can provide more realistic occupancy data but need a lot of work in computing, validation and calibration process. Some studies used statistical method to get occupancy patterns. Chang and Hong [29] used statistical methods to analyze lighting switch data collected from 200 cubicle offices to identify variations in occupancy patterns. Duarte et al. [7] used descriptive statistics and two-sample *t*-tests to reveal occupancy patterns in an office building. Hong and his research group proposed a series of studies to apply data-driven methods, such as clustering analysis and decision tree, to understand and predict occupancy data [30–32]. However, due to the limitation of data measurements, these studies cannot provide enough case studies and mainly focus on office buildings, where the occupants are easy to monitor.

A similar problem exists in building energy simulations, which is the treatment of weather data. Simulation software usually uses typical weather data, such as test reference year (TRY), weather year for energy calculation (WYEC) and typical meteorological year (TMY), instead of the actual weather data. TMY [33] proposed by the Sandia Laboratory and the national climatic data center (NCDC) in 1978 is the most widespread method. It constructs a year with 12 individual months of real weather data selected from historical records using Finkelstein–Schafer (FS) statistics. Considering different weather components, such as dry bulb temperature, dew point temperature, wind speed, solar radiation, etc., the FS statistics are calculated for each weather component and summed to a weighted value. TMY provides a relatively accurate and convenient method for the characterization of outdoor climatic conditions and is widely used in building design and simulation.

Referring to the TMY method, this paper proposes a concept of typical occupancy data (TOD), which is generated from real-time occupancy data collected by mobile devices. In the following section (Section 2), we introduce the data source and cluster analysis method applied to generate the typical occupancy schedule. The clustering results of different building types are demonstrated in Section 3 and a case study and application perspective are given in Section 4. The originality, limitation and future work of this study is discussed in Section 5, and the URL of the TOD database is attached at the end of this paper.

2. Method

In this paper, the building occupancy data is collected by mobile devices and stored in a database. After preprocessing, *k*-means cluster analysis is applied as the data mining method to extract the typical occupancy data (TOD). The process is illustrated in Fig. 2.

2.1. Data collection

We collected occupancy data from a mobile social network company. If users turn on location services in mobile phones, PCs and other mobile devices, when they use the applications (APPs) developed by our cooperative Internet company, the positioning software development kit (SDK) embedded in APPs would compute and store their positioning requests and form real-time positioning data. Through data analysis and feature extraction, we can obtain hourly occupancy rate and distribution of buildings. Our cooperator is one of the largest Internet companies in China, Tencent, and the services of which include social network, web portal, e-commerce, mobile games, etc. The social APPs developed by this company are involved in many basic activities of life and share approximately 60% of the market. The average daily global positioning requests have been more than 50 billion, covering more than 600 million people. The data obtained from this company are the total number and distribution of occupancy in certain buildings rather than personal information of users, so it does not involve privacy issues. To date, the hourly occupancy data of 60 buildings have been collected, and the collection work is still in process. A database needs to be built to store and manage these occupancy data.

2.2. Data classification and database construction

In existing standards and design manuals, occupancy data is related to the type, location and scale of the building. Different types of buildings have different occupancy schedules and densities. The same type of buildings in different locations follow different patterns. Buildings on different scales may differ in both

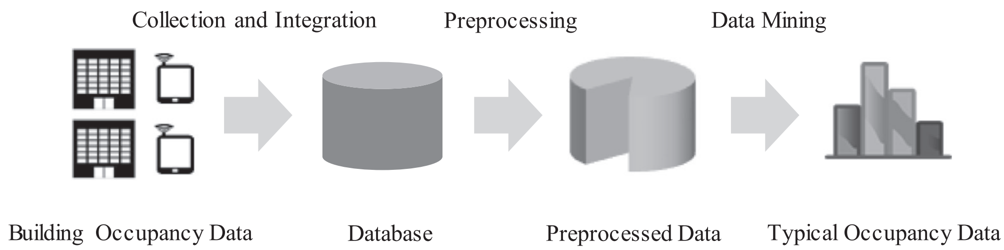
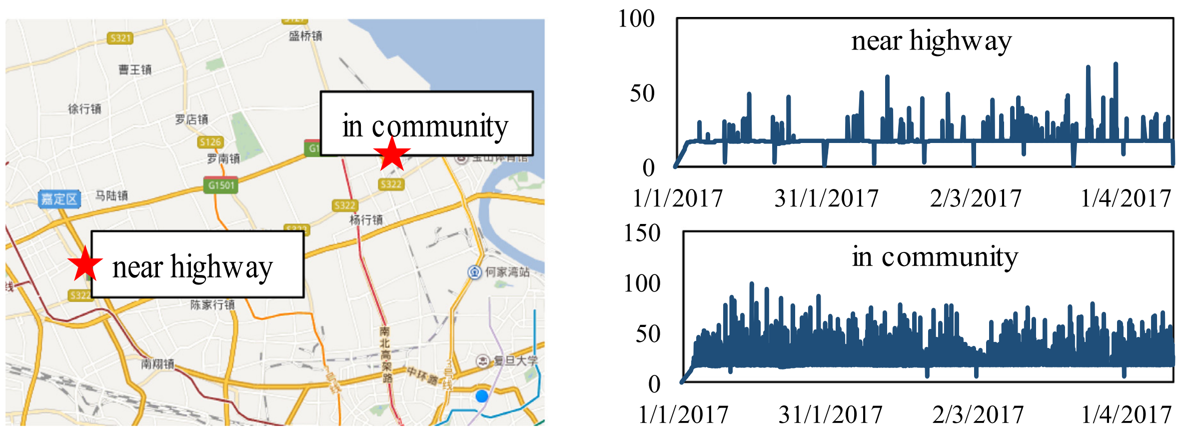


Fig. 2. Process of typical occupancy data (TOD) extraction.



(a) Location of two restaurants (b) Real-time occupancy data of two restaurants

Fig. 3. Real-time occupancy data of two quick service restaurants in different location.

count and occupancy schedule. In addition, from the collected occupancy data, we find that the age of a building has an effect on occupancy data. Therefore, in this paper, buildings descriptions are constructed with the following four attributes: building type, location, scale and age, which could be easily queried in the following preprocessing and cluster analysis.

A. Building type

Stocki et al. [34] presented a set of standardized assumptions for commercial building energy analyses. They define the following seven building types: large office, small office, retail, education, apartment, small hotel, and hospital. The 2003 Commercial Buildings Energy Consumption Survey (2003 CBECS) [35], which contains 5215 buildings separated into 29 categories and 51 subcategories, provided the most typical building types. Gowri et al. [36] added the mid-rise apartment model in a separate PNNL study. The final set of 16 building types was selected by consensus among DOE, NREL, LBNL, and PNNL, which represent approximately 70% of the commercial buildings in the U.S. [37].

Considering the existing building type categorization and the actual situation in China, the collected occupancy data of 60 buildings are classified into seven categories and 19 subcategories, which are listed in Table 1.

B. Building location

The occupancy data in buildings is also related to their geographical location. People in different cities, countries and time zones have different lifestyles and schedules. Even in the same building type in the same city, the occupancy data follow different patterns with different surrounding environment. As shown in Fig. 3 below, there are two quick service restaurants at different locations. One is near a highway and the other is in a residential community. The real-time occupancy data of them, collected from January to March, vary significantly differently. Although both restaurant occupancy peaks appear during daily meal time, in the remainder of the day, only the staff stay in the restaurants near the highway; however, in the restaurant in the community, there are several consumers all the time, so the curve is more intensive.

According to contrast result of above data, we classify the occupancy location information of buildings at four layers, including time zones, countries, cities and detail location. In detail location, we describe the surrounding environment of buildings, such as whether they are near the highway, in the community, near a school, etc.

C. Building scale

The size of a building not only affects the total number of occupancy but also the complexity of the schedule. Two shopping malls with 10 times in size difference may have entirely different schedule structures. Thus, in this paper, we do not normalize the occupancy data into density and schedule. Instead, the original occupancy data is reserved, and the building area is given to describe the scale of the building as reference.

D. Building age

The building age would also have an effect on the occupancy schedule and the number of occupancy. For instance, buildings could not be fully occupied at the beginning stage. For some public buildings such as shopping malls, the occupancy data vary with the increase of building age and the enhancement of reputation. Fig. 4 shows the real-time occupancy data of a shopping mall from January 2016 to October

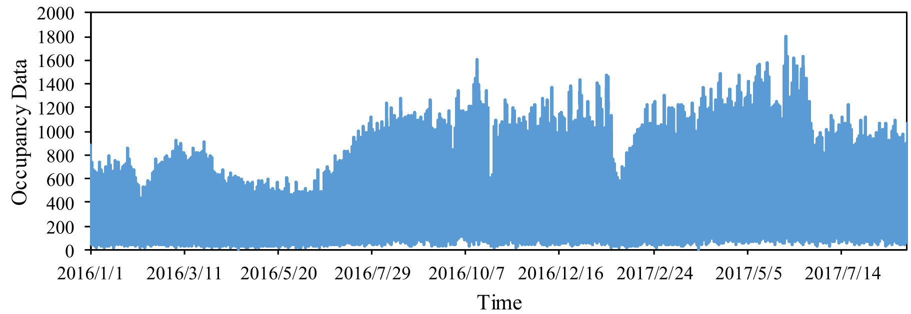


Fig. 4. Real-time occupancy data of a shopping mall.

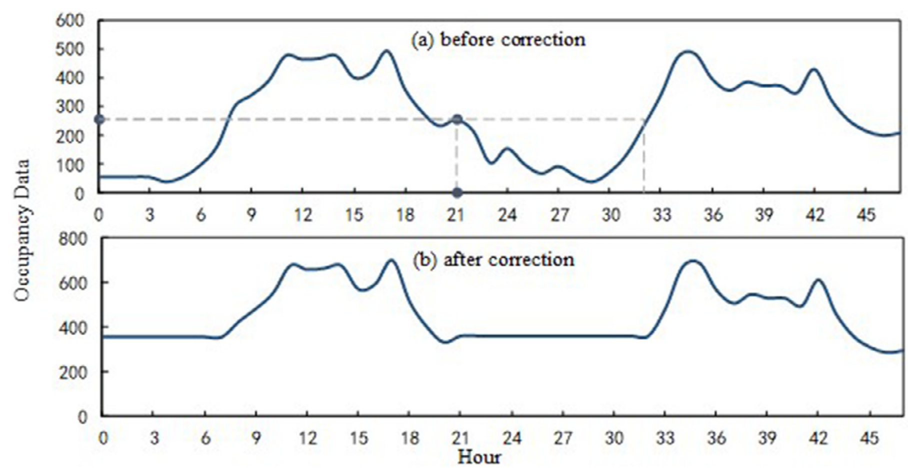


Fig. 5. Real-time occupancy data of an inpatient department of hospital for two days.

2017. Compared to 2016, the number of occupancy in 2017 increases obviously. Therefore, building age is chosen as an attribute to describe buildings in this paper.

2.3. Data preprocessing

The method used to collect occupancy data in this paper has the following three requirements: (1) using one of these popular APPs developed by our cooperative Internet company, (2) having access to network, and (3) turning on location services. If occupants do not have any one of the above, they cannot be counted and located. Thus, the following preparatory work is necessary:

(1) Correction of occupancy data due to the utilization rate of APPs

Based on the cases in public security, tourism, transportation and urban planning, the samples we collect cover approximately 70% of the total samples. Thus, we need to modify the obtained occupancy data considering that estimation by 0.7. It's noteworthy that this 0.7 estimation would vary with the popularization of APPs and needs to validation.

(1) Correction of occupancy data due to turning off APPs at night

For some building types, such as hotels, hospitals, etc., occupants will stop using APPs and turn off the Internet when sleeping, which will cause a certain deviation. From the data we collect, we find that the occupancy data of these buildings would climb to a peak during 9 p.m. to 12 p.m., which is in accordance with the bedtime. Thus, in this case, we choose the peak value of the occupancy before sleeping as the occupancy data of the entire night. Fig. 5 shows the real-time occupancy data of an inpatient department of a hospital for two days. Before correction (Fig. 5(a)), the number of occupancy climbed to a peak of roughly 260 at 9 p.m. and then dropped to 50 from 0 to 4 a.m., which differed by approximately 200 people. Therefore, we use the peak value at 9 p.m. as the occupancy data during night and multiply the data by 0.7. The corrected data is shown in Fig. 5(b).

2.4. Cluster analysis

Referring to TMY, we propose a concept of typical occupancy data (TOD). However, we do not choose the same method for TMY by separating historical data into a fixed month time span. Instead, the *k*-means algorithm is applied to extract TOD because the weather components contain dry bulb temperature, dew point temperature, wind speed, solar radiation, etc., which should be considered as the weighted sum to determine the data for the typical meteorological month. However, for occupancy data, only the count of occupancy is considered. Thus, the cluster analysis can be used to cluster the occupancy data following similar patterns, and the geometric center of each cluster can be drawn as the typical data.

Clustering is the process of grouping the data objects into subjects. Each subject is a cluster, which means an occupancy pattern in this paper. It is known as an unsupervised learning method because the class label information is unknown. This method is very useful to discover the previously unknown information within data. The *K*-means algorithm, which is one of the most widely used cluster analysis method, aims to find a partition in which the squared error between the empirical mean of a cluster and the points in the cluster is minimized [38]. Typically, Euclidean distance is used to describe the squared error (shown in Eq. (1)).

$$E = \sum_{i=1}^{k} \sum_{p \in C_i} (p - c_i)^2 \tag{1}$$

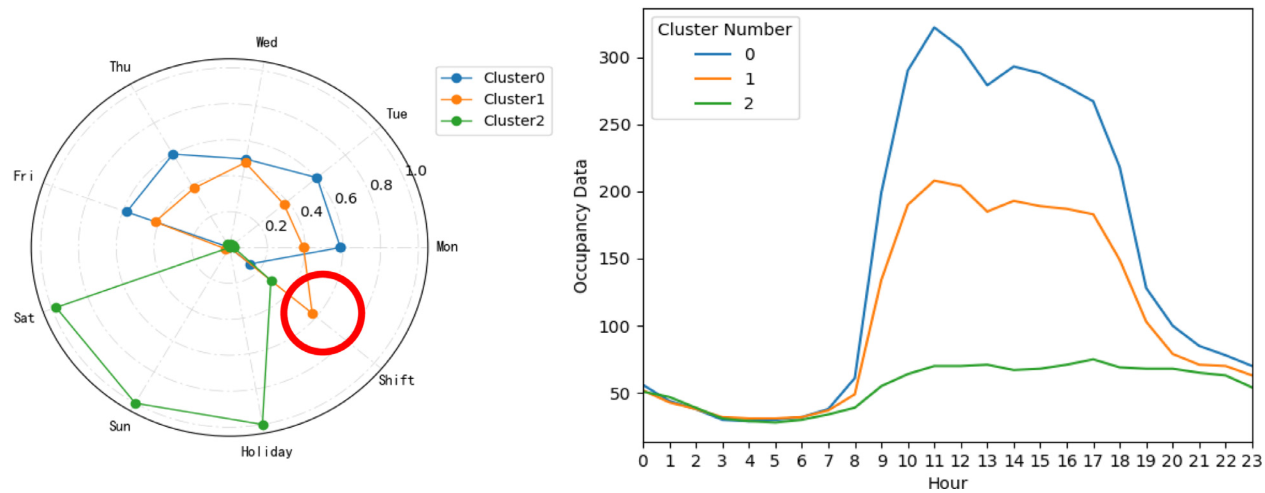


Fig. 6. Clustering results of office buildings.

Table 1
Categorization of building types in this paper.

Category	Subcategory	Category	Subcategory
Office	Government buildings	Restaurant	Full service restaurants
	Enterprise companies		Quick service restaurants
Shopping area	Shopping malls	Hospital	Outpatient department
	Supermarkets		Inpatient department
	Shops		Clinic
Hotel	Luxury hotels	Transportation	Airport
	Business hotels		Train station
	Budget Inns		Bus terminus
Education	Primary school		
	Secondary school		
	University		

where $p=(p_1,p_2,\ldots,p_n)$ are points in an Euclidean n-space and c_i is the cluster center of cluster C_i. In this paper, both p and c_i are 24-dimensional points, and p represents occupancy data of 24 h a day.

The main steps of the k-means algorithm are as follows: first select k points as initial cluster centers. The remaining points are assigned to the closest cluster center. Then, new cluster centers are calculated by repeating the previous steps until the partition is stable.

In this paper, the k-means algorithm is applied to cluster the occupancy data of each building. The optimal number of clusters k_{op} is determined by the Calinski–Harabasz (C–H) criterion [39] and is regarded as the day type number of TOD. The annual TOD is made up by the cluster centers of each day type. Considering that the internal and surrounding environment change during building lifespan and occupancy data is related to the age of the building, we update the TOD every year.

3. Results

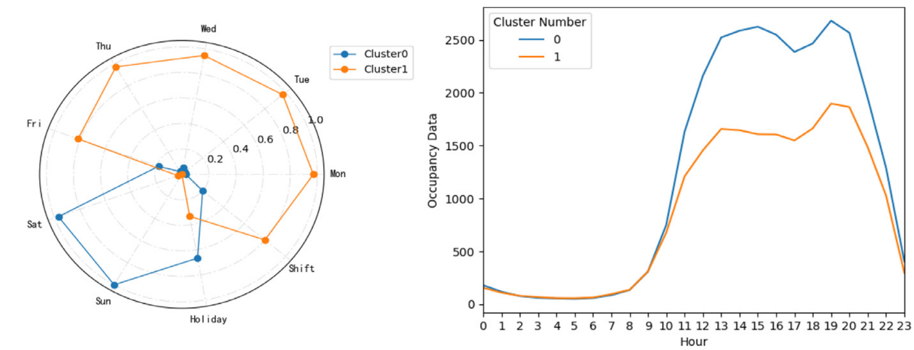
Table 2 shows the statistics of k_{op} for each building type, which stands for the number of day types of TOD. For restaurants, hotels and educational buildings, the occupancy data is classified into two types. For office buildings and hospitals, k_{op} could be 2 or 3. The clustering results of shopping area and transportation are more complicated. The optimal number of clusters varies from 2 to 4.

The following figures show the clustering results of different building types. The left figures compare the clustering results with the actual day types, while the right figures show the typical days (cluster centers). In the left figures, the actual day types are classified into nine types, including Monday to Sunday as well as holiday and shift days, which are weekends paid back for working because of legal holidays in China such as Spring Festival and National Day Holiday. The points with the same color represent that they are assigned in the same cluster. The numbers show the percentage of each actual day types assigned in this cluster. The more actual day types are in the same cluster, the points are plotted outer in left figures. For instance, the point circled in Fig. 6 represents that 60% shift days are assigned in Cluster 1 by the k-means algorithm. Some weekdays (Mondays to Fridays) are also assigned in Cluster 1, but there are still several weekdays assigned in Cluster 0. The right figure shows the cluster centers, which presents the typical occupancy data of clusters. From the clustering results, we can draw the following conclusions for different building types:

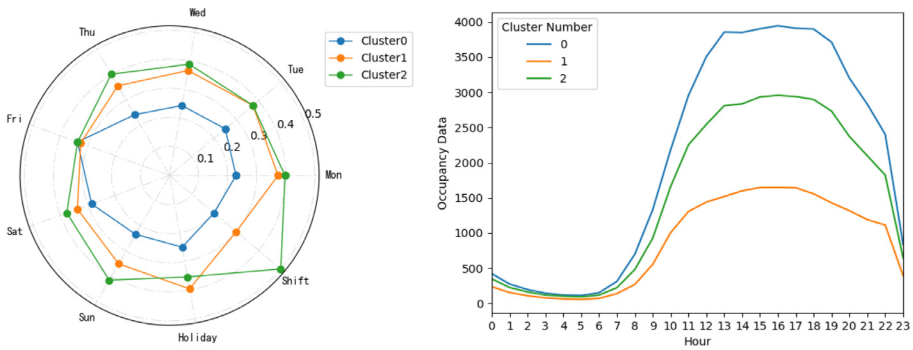
Table 2
Statistics of k_{op} for different building types.

k_{op}	2	3	4
Office	78%	22%	0%
Shopping area	55%	30%	15%
Transportation	55%	27%	18%
Hospital	57%	43%	0%
Hotel	100%	0%	0%
Education	100%	0%	0%
Restaurant	100%	0%	0%

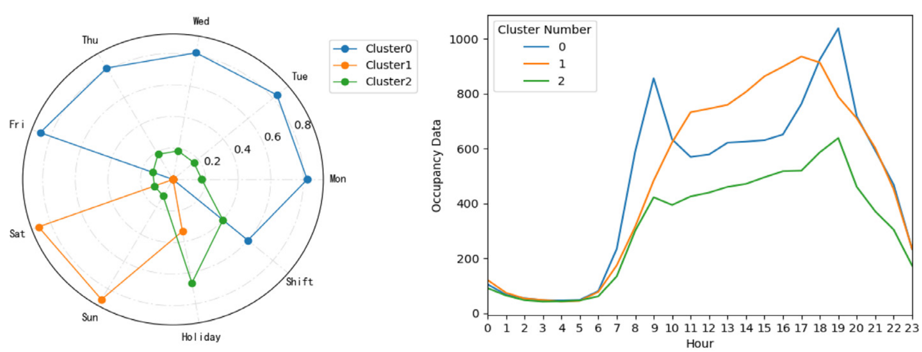
(1) For office buildings, obviously, occupancy data could be divided into weekdays and weekends. From the clustering result (shown in Fig. 6), it could be seen that when $k_{\mathrm{op}}=3$, the occupancy data of weekdays would be assigned into two clus-



(b) Shopping mall in downtown area



(c) Supermarket



(d) Shop

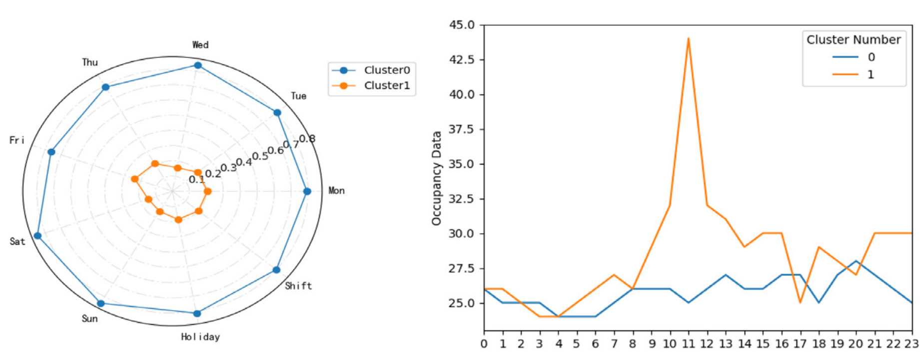


Fig. 7. Clustering results of shopping area.

ters (Cluster 0 and Cluster 1), and the typical occupancy data of Cluster 1 are fewer than Cluster 0. Looking back to the actual occupancy data, we find that the weekdays with fewer occupancy (Cluster 1) often appear before or after holidays. Most of the holidays in China, especially the holidays more than one week, spread from October to February, so the weekdays in this period tend to be assigned in Cluster 1.

(2) For the shopping area, the clustering results are more complicated. From shopping buildings at different locations and scales, the number of clusters change from 2 to 4. Fig. 7(a) shows the

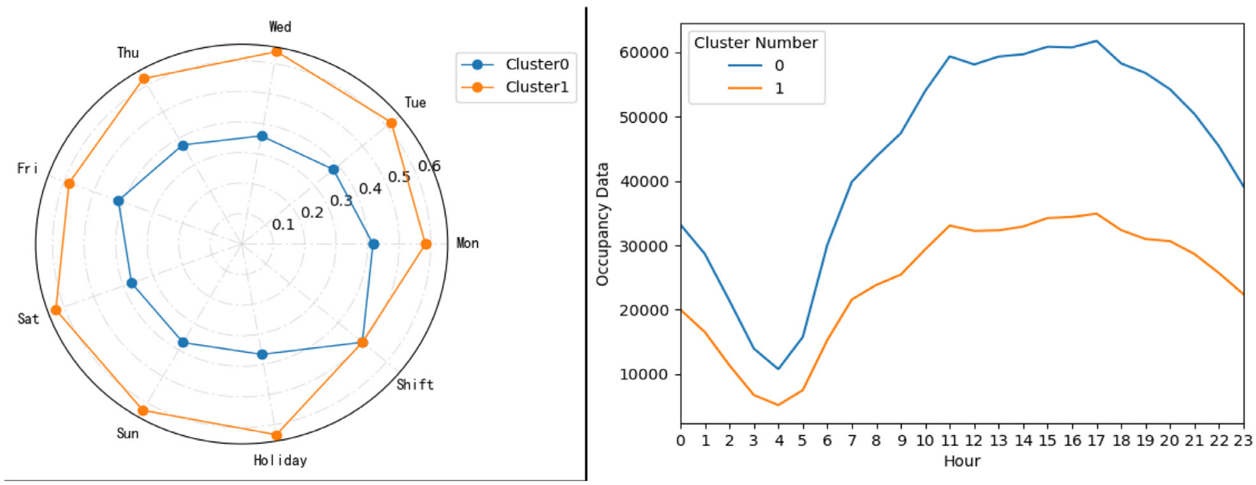


Fig. 8. Clustering results of transportation building.

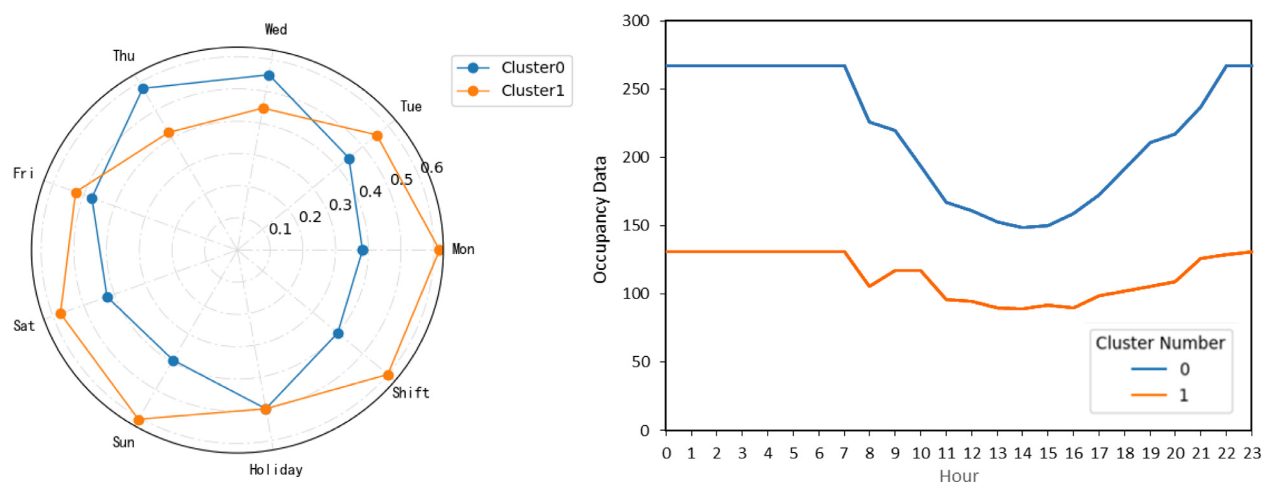


Fig. 9. Clustering results of hotel.

clustering results of a shopping mall in the suburb. The occupancy data on weekdays are fewer than weekends, but for some Fridays and shift days, people are likely to go there for shopping. Thus, parts of Fridays and shift days are also assigned in Cluster 0. Fig. 7(b) shows the clustering results of a shopping mall in downtown area. Different from the building of Fig. 7(a), there is no obvious difference of occupancy data between weekdays and weekends. Seasonal discounts are another factor that will influence the occupancy data. For the supermarket shown in Fig. 7(c), occupancy data varies differently between weekdays and weekends. On weekdays, the peak values of occupancy data appear at the morning and evening hours, while on weekends, it appears during 10 a.m. to 7 p.m. Part of holidays and shift days are assigned into Cluster 2, which often appear during the Spring Festival and New Year's Day. However, for shops, which are generally close to the community and convenient for people to buy some daily necessities, consumers will not stay too long and most of the occupants are staff in the shop. The actual day types have no effect on the number of occupancy. Occupancy data typically follow the same pattern every day. However, the occupancy data will increase suddenly for some factors, but this situation will not occur frequently and will not last for a long time. In Fig. 7(d), only 20% occupancy data are assigned in this case (Cluster 1).

(3) For transportation buildings, there is no significant difference between occupancy data of weekdays and weekends (shown in Fig. 8). The occupancy data vary due to the holidays, such as the Spring Festival, National Day Holiday, summer holidays, etc. Generally, the occupancy data of holidays and the days before or after holidays are much more than the other days. However, Spring Festival is an exception. In China, people return home to celebrate the festival. Thus, it comes to a peak value for occupancy data in transportation buildings before and after Spring Festival, while during the Spring Festival, there are fewer occupants. According to the scale and sophistication of transportation buildings, k_{op} would vary from 2 to 4.

(4) For hotels, whether luxury hotels, business hotels or budget inns, occupancy data can be divided into in-season data and off-season data that is strongly related to the location determining whether the hotel is in season or off season. Not all the hotels are in season during holidays or weekends (shown in Fig. 9).

(5) For hospitals, the occupancy data of outpatient departments differ between weekdays and weekends, while the occupancy data of inpatient departments decrease in holidays. The occupancy data in hospitals increase from April to August (see Cluster1 in Fig. 10(a and b)), which coincides with the flu season. In January and February, the occupancy data are much fewer than other times because of the Spring Festival. For clinics, the occupancy data have no obvious difference between weekdays and weekends.

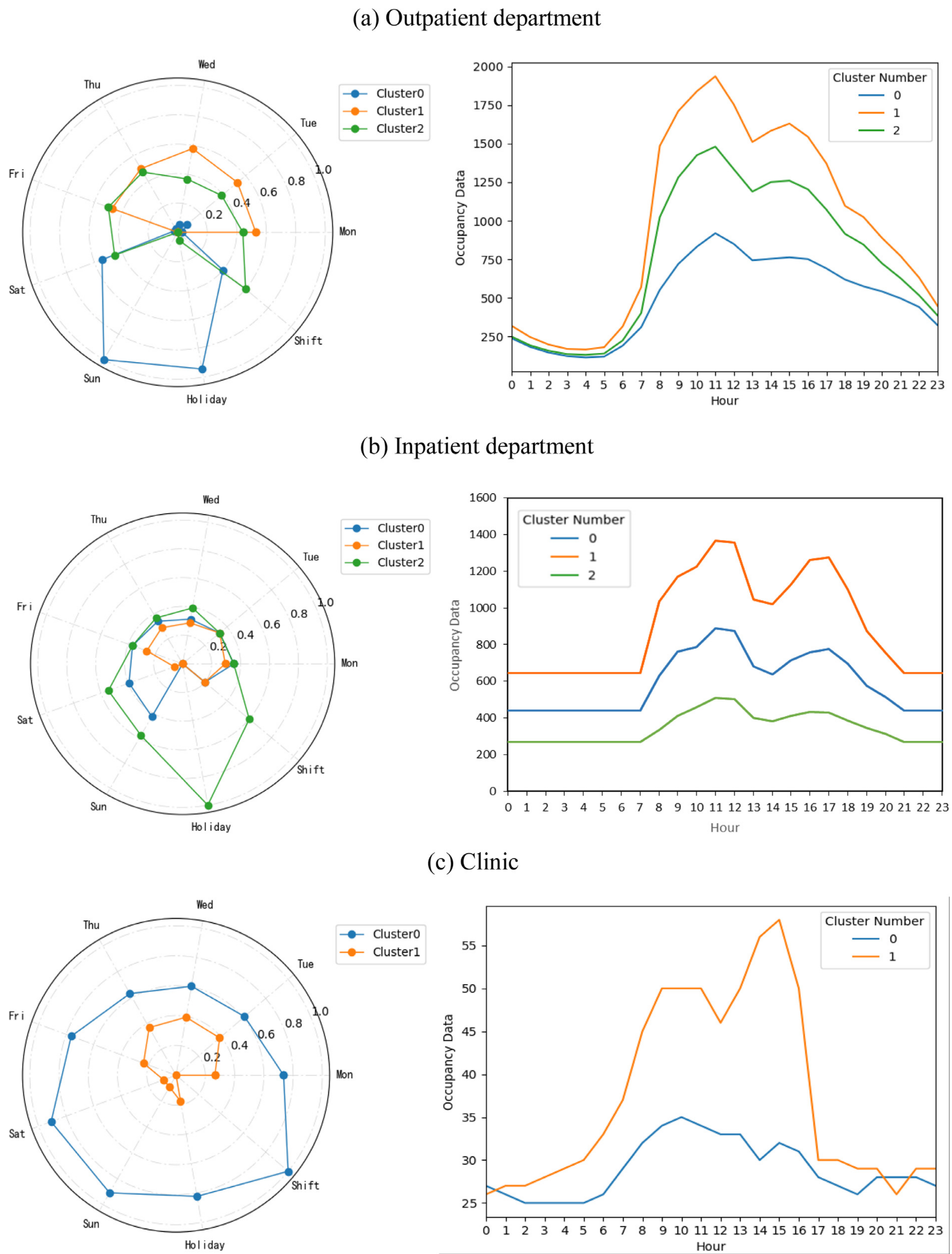


Fig. 10. Clustering results of hospitals.

(6) The clustering results of educational buildings are clear. Occupancy data for holidays including weekends, holidays in China, summer and winter holidays is Cluster 0. Non-holidays is Cluster 1 (Fig. 11).

(7) As mentioned before, the occupancy data of restaurant is related to its location, and the peak value of occupancy would appear during meal time in a day (shown in Fig. 12).

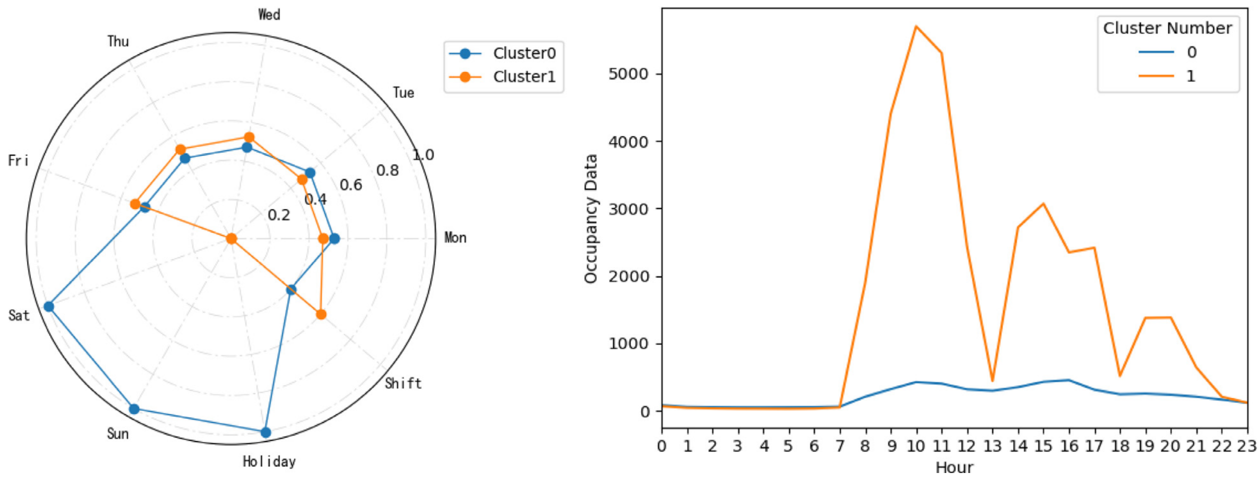


Fig. 11. Clustering results of education building.

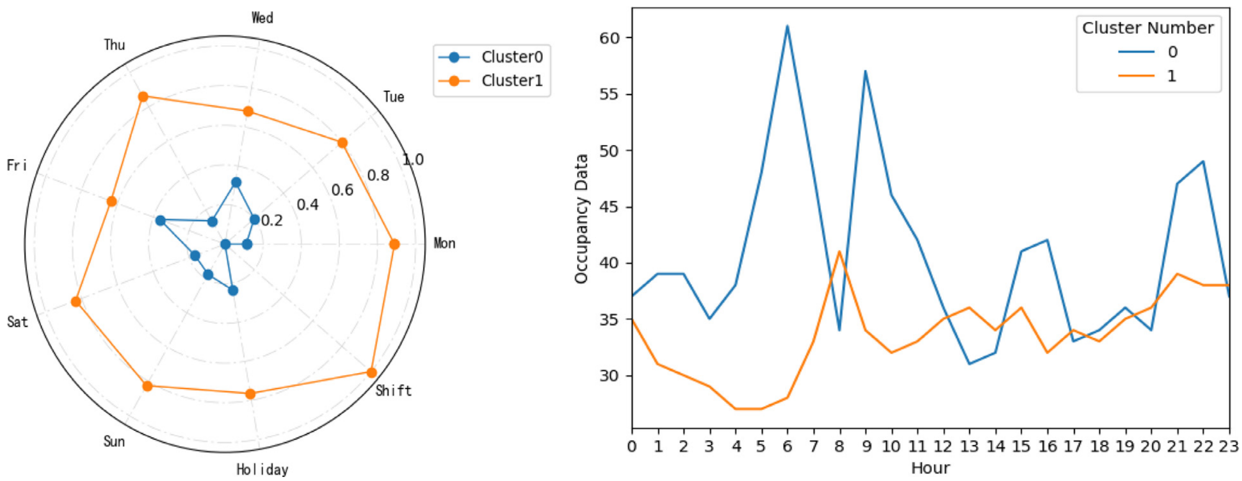


Fig. 12. Clustering results of restaurant.

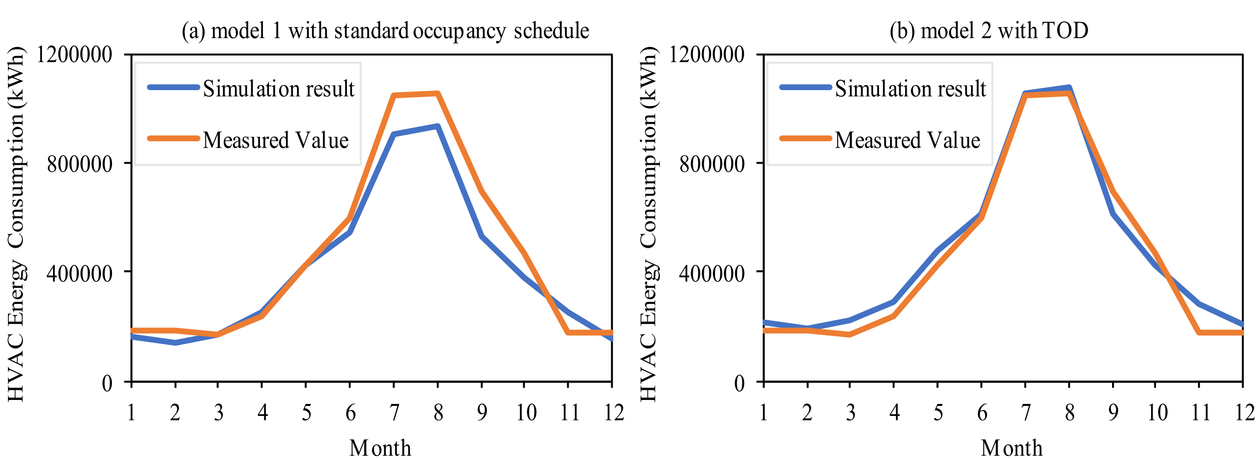


Fig. 13. Energy consumption from the simulation model vs. actual measured value.

4. Application

The TOD are described from building type, location, area and age. For application, users choose the TOD of the most similar building as reference. However, the attributes of different buildings cannot be identical. Therefore, the TOD should be modified according to the actual conditions.

A case study of an office building in Shanghai is described here to show how TOD is applied in building energy simulation and to analyze the accuracy of TOD in simulation. Two EnergyPlus models are built to compare the simulation results of using a standard occupancy schedule and TOD. For Model 1, the occupancy schedule is set according to the Chinese design standard for the energy efficiency of public buildings [40]. For Model 2, the TOD of a similar office building, which is also located in downtown Shanghai, is chosen as the input and modified about the difference of building area. The other input parameters of two models are same.

The comparison of the HVAC system energy consumption data between the actual and simulated case is shown in Fig. 13. For Model 1 (see Fig. 13(a)), the mean bias error (MBE) between the simulated and actual energy consumption is 11.69%, and the coefficient of variation of root mean square error (CV(RMSE)) is 18.81%. However, For Model 2 (see Fig. 13(b)), the MBE and CV (RMSE) drop to 4.21%, 11.59% respectively. The model using TOD has higher accuracy than that using standard schedule.

The above case study shows that using TOD in simulation can improve the accuracy of the building model and simplify the model calibration process; in particular, the proposed TOD can be used as reference for buildings without any standard occupancy schedule. TOD can also be used to develop more realistic and stochastic occupancy model through further statistical analysis. In addition, for urban planning, TODs of different buildings types in cities contribute to reasonable urban planning, including transportation planning, architectural planning, and energy planning.

5. Conclusion

Building simulation requires accurate occupancy data. Existing input methods such as fixed weekday and weekend schedules cannot meet the demand of accurate simulation and control. Precise, on-site measurements require substantial monitoring and expensive equipment. Working with one of the largest social media companies in the world and its mobile data, we propose an approach termed TOD and its extraction method to solve the problem. The simulation results demonstrated that the new approach is more accurate than the traditional one.

Compared with traditional standard occupancy schedule, TOD are collected by mobile devices, which embrace larger amounts of data and various data types. Therefore, occupancy data classification is more specific. Unlike previous schedules, in this paper, we not only provide 7 main classes and 19 subclasses of TOD of different buildings, but we also describe the TOD from the location, scale and age of buildings. Considering the change of environment during building lifespans, we will update the TOD every year. Moreover, with the increase in data collection, the existing database will continue to expand, which is very useful for further researches on occupancy.

In addition, this paper is based on the cluster analysis method to classify the occupancy data, and the clustering results are data-driven. Therefore, the results may classify issues that have not been previously considered. For example, the occupancy data in hospitals may be influenced by seasons; for example, the number of people during flu season is significantly more than the number during non-flu season and so on. These clustering results could be used in research on occupant behavior models.

Due to the limited study period, the maximum data collection time is 1.5 years, and the minimum is 0.5 years. In addition, data collection was only concentrated in Shanghai and Beijing. With the increase of acquisition time and the number of collection points, the TOD proposed in this paper will be used more widely, and the result will be more accurate. Moreover, in future research, the analysis of typical occupancy information should be considered to include gender and age ratios. These factors determine the comfort requirements of building occupants and provide the data support for further study on thermal comfort and delicate building control strategies.

Although this paper mainly introduces the TOD application in building simulation, its application scope can be extended to various industries such as transportation and urban planning, with broad application value. The TOD of buildings that were collected will be published on the Internet at http://a434.tongji.edu.cn/typical%20occupancy%20data.html and will be free accessible to researchers.

Acknowledgments

The authors wish to thank Tencent for cooperation. This work was supported by National Science & Technology Pillar Program during the thirteenth Five-year Plan Period (2017YFB0903404) and Financial Grant from Beijing Municipal Human Resources and Social Security Bureau (grant No. 2017ZZ050).

References

[1] T. Hong, et al., Advances in research and applications of energy-related occupant behavior in buildings, Energy Build. 116 (2016) 694–702.
[2] D. Yan, et al., Occupant behavior modeling for building performance simulation: current state and future challenges, Energy Build. 107 (2015) 264–278.
[3] D. Robinson, Some trends and research needs in energy and comfort prediction, Comfort Energy Use Build. (2006).
[4] ASHRAE, ASHRAE 90.1 Appendix G. Building Performance Rating Method, ASHRAE, 2013.
[5] ASHRAE, 90.1-2013 User's Manual: ANSI/ASHRAE/IES Standard 90.1, Energy Standard For Buildings Except Low-Rise Residential Buildings, ASHRAE, 2013.
[6] Goel, S., et al., Enhancements to ASHRAE standard 90.1 prototype building models. Pacific Northwest National Laboratory, PNNL-23269, 2014.
[7] C. Duarte, K. Van Den Wymelenberg, C. Rieger, Revealing occupancy patterns in an office building through the use of occupancy sensor data, Energy Build. 67 (2013) 587–595.
[8] Y. Geun Young, et al., A field survey of occupancy and air-conditioner use patterns in open plan offices, Indoor Built Environ. 20 (1) (2010) 137–147.
[9] T. Teixeira, G.D., A. Savvides, A survey of human-sensing: methods for detecting presence, count, location, track, and identity.
[10] D.B. Floyd, D.S. Parker, J.R. Sherwin, Measured field performance and energy savings of occupancy sensors: three case studies, Comp. Anal. (1996) 16.
[11] E. Naghiyev, M. Gillott, R. Wilson, Three unobtrusive domestic occupancy measurement technologies under qualitative review, Energy Build. 69 (2014) 507–514.
[12] T. Labeodan, et al., Occupancy measurement in commercial office buildings for demand-driven control applications—a survey and detection system evaluation, Energy Build. 93 (2015) 303–314.
[13] T. Harada, T. Sato, T. Mori, Human motion tracking system based on skeleton and surface integration model using pressure sensors distribution bed, in: Human Motion, 2000. Proceedings, 2000 Workshop on.
[14] Y. Benezeth, et al., Towards a sensor for detecting human presence and characterizing activity, Energy Build. 43 (2-3) (2011) 305–314.
[15] S. Wang, J. Burnett, H. Chong, Experimental validation of CO_2-based occupancy detection for demand-controlled ventilation, Indoor Built Environ. 8 (6) (1999) 377–391.
[16] B. Dong, et al., An information technology enabled sustainability test-bed (ITEST) for occupancy detection through an environmental sensing network, Energy Build. 42 (7) (2010) 1038–1046.
[17] Y.S. Kim, J. Srebric, Improvement of building energy simulation accuracy with occupancy schedules derived from hourly building electricity consumption, ASHRAE Trans. 121 (1) (2015) 353–360.
[18] R. Ahas, et al., Using mobile positioning data to model locations meaningful to users of mobile phones, J. Urban Technol., 17 (1) (2010) 3–27.
[19] S. Zhang, et al., Mobile sensing and simultaneously node localization in wireless sensor networks for human motion tracking, International Conference on Control Automation Robotics & Vision, 2011.
[20] R. Rana, et al., Novel activity classification and occupancy estimation methods for intelligent HVAC (heating, ventilation and air conditioning) systems, Energy 93 (2015) 245–255.
[21] W. Wang, J. Chen, X. Song, Modeling and predicting occupancy profile in office space with a Wi-Fi probe-based dynamic Markov Time-Window inference approach, Build. Environ. 124 (2017) 130–142.
[22] N. Mohammadi, J.E. Taylor, Urban energy flux: spatiotemporal fluctuations of building energy consumption and human mobility-driven prediction, Appl. Energy 195 (2017) 810–818.
[23] Z. Pang, et al., Application of mobile positioning occupancy data for building energy simulation: an engineering case study, Build. Environ. 141 (2018) 1–15.
[24] D. Wang, C.C. Federspiel, F. Rubinstein, Modeling occupancy in single person offices, Energy Build. 37 (2) (2005) 121–126.
[25] J. Page, et al., A generalised stochastic model for the simulation of occupant presence, Energy Build. 40 (2) (2008) 83–98.
[26] C. Wang, D. Yan, Y. Jiang, A novel approach for building occupancy simulation, Build. Simul. 4 (2) (2011) 149–167.
[27] B. Dong, K.P. Lam, A real-time model predictive control for building heating and cooling systems based on the occupancy behavior pattern detection and local weather forecasting, Build. Simul. 7 (1) (2014) 89–106.
[28] L.M. Candanedo, V. Feldheim, D. Deramaix, A methodology based on Hidden Markov Models for occupancy detection and a case study in a low energy residential building, Energy Build. 148 (2017) 327–341.

[29] W.K. Chang, T. Hong, Statistical analysis and modeling of occupancy patterns in open-plan offices using measured lighting-switch data, Build. Simul. 6 (1) (2013) 23–32.

[30] S. D'Oca, T. Hong, Occupancy schedules learning process through a data mining framework, Energy Build. 88 (2015) 395–408.

[31] X. Liang, T. Hong, G.Q. Shen, Occupancy data analytics and prediction: a case study, Build. Environ. 102 (2016) 179–192.

[32] X. Luo, et al., Performance evaluation of an agent-based occupancy simulation model, Build. Environ. 115 (2017) 42–53.

[33] J.J. Hall, et al., Generation of a typical meteorological year, in: Proceedings of Analysis for Solar Heating and Cooling Conference, San Diego, CA, USA, 1978.

[34] M. Stocki, D.C. Curcija, M.S. Bhandari, The development of standardized whole-building simulation assumptions for energy analysis for a set of commercial buildings, ASHRAE Trans. 113 (July) (2007) 422–435.

[35] EIA, Commercial Buildings Energy Consumption Survey. Energy Information Administration, 2003. Available at Last accessed April 2018, 2005 www.eia.doe.gov/emeu/cbecs/cbecs2003/introduction.html .

[36] K. Gowri, M.A. Halverson, E.E. Richman, Analysis of Energy Saving Impacts of ASHRAE 90.1-2004 For New York, Pacific Northwest National Laboratory, 2007 PNNL-16770..

[37] M. Deru, et al., U.S. Department of Energy Commercial Reference Building Models of the National Building Stock, National Renewable Energy Laboratory, 2011.

[38] Han, J. and M. Kamber, *Data Mining Concept and Techniques*. Morgan Kaufmann Publishers, 2001.

[39] T. Caliński, J. Harabasz, A dendrite method for cluster analysis, Commun. Stat. 3 (1) (1974) 1–27.

[40] ResearchCAoB, Chinese Design Standard For Energy Efficiency of Public Buildings GB50189—2015, China building industry press, Beijing, 2015.